Inference for a mean-reverting stochastic process with multiple change points

Fuqi Chen

Department of Statistical and Actuarial Sciences,
University of Western Ontario, London, Ontario, Canada
e-mail: fchen223@uwo.ca

Rogemar Mamon[*]

Department of Statistical and Actuarial Sciences,
Department of Applied Mathematics,
University of Western Ontario, London, Ontario, Canada
e-mail: rmamon@stats.uwo.ca

Division of Physical Sciences and Mathematics,
University of the Philippines Visayas, Iloilo, Philippines

and

Matt Davison

Department of Statistical and Actuarial Sciences,
Department of Applied Mathematics,
University of Western Ontario, London, Ontario, Canada
e-mail: mdavison@uwo.ca

Abstract: The use of an Ornstein-Uhlenbeck (OU) process is ubiquitous in business, economics and finance to capture various price processes and evolution of economic indicators exhibiting mean-reverting properties. The time at which structural transition representing drastic changes in the economic dynamics occur are of particular interest to policy makers, investors and financial product providers. This paper addresses the change-point problem under a generalised OU model and investigates the associated statistical inference. We propose two estimation methods to locate multiple change points and show the asymptotic properties of the estimators. An informational approach is employed in detecting the change points, and the consistency of our methods is also theoretically demonstrated. Estimation is considered under the setting where both the number and location of change points are unknown. Three computing algorithms are further developed for implementation. The practical applicability of our methods is illustrated using simulated and observed financial market data.

MSC 2010 subject classifications: Primary 60G20, 62P05; secondary 62M99.

---
[*]Corresponding author. Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada, N6A 5B7. Email: rmamon@stats.uwo.ca.

1. Introduction

We examine the change-point detection problem on the drift parameters of a generalised version of the Ornstein-Uhlenbeck (OU) process introduced in Dehling et al. (2010); see also Dehling et al. (2014) and Nkurunziza and Zhang (2016), which is a compact version of Zhang (2015). Such a process X_t is a solution to the stochastic differential equation (SDE)

$$dX_t = S(\theta, t, X_t)dt + \sigma dW_t, \quad 0 < t \leq T, \tag{1.1}$$

where $S(\theta, t, X_t) = L(t) - aX_t = \sum_{i=1}^{p} \mu_i \varphi_k(t) - aX_t,\ i = 1, \ldots, p,\ \theta = (\mu_1, \ldots, \mu_p, -a)^\top$ and $\top$ denotes the transpose of a matrix. Here, W_t is a one-dimensional standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$. In particular, if $L(t) = \mu$ then (1.1) is the SDE of the classical OU process, which is commonly used to model the stochastic dynamics of various financial variables.

The OU process was initially introduced for problems in the physical sciences (e.g., Lansky and Sacerdote (2001)). Further, many economic indicators, prices in the financial market as well as processes in the natural and engineering are captured sufficiently by the OU model. The classical work of Vasiček (1977) employs an OU model for bond valuation. The importance of this stochastic process is also demonstrated by its ubiquity in many fields. For instance, the OU process is used in mathematical models of the electricity market (e.g., Erlwein et al. (2010)), commodity futures market (e.g., Date et al. (2013)), weather derivatives (e.g., Elias et al. (2014)), central-bank rate setting policy (e.g., Elliott and Wilson (2007)), spreads between pairs of securities (e.g., Elliott et al. (2005)), stochastic control-driven insurance problems (e.g., Liang et al. (2011)), spot freight rates in the shipping industry (e.g., Benth et al. (2015)), risk management (e.g., Date and Bustreo (2016)), and power generation (e.g., Howell et al. (2011)). In the OU-modelling context, Tenyakov et al. (2016) proposed a signal processing-based approach to determine presence of market liquidity regimes. Various applications of the OU process are also highlighted in biology (e.g., Rohlfs et al. (2010)), neurology (e.g., Shinomoto et al. (1999)), survival analysis (e.g., Aalen and Gjessing (2004)), and chemistry (e.g., Lu (2003, 2004)).

We note that the mean-reverting level of an OU process is constant, which can be a notable weakness for many financial datasets. This may be rectified by introducing a generalised OU process where a time-dependent function describes its level of mean reversion. Such a generalised version incorporates time-inhomogeneity and seasonality of mean reversion simultaneously. Dehling et al. (2014) developed the framework to study a change-point phenomenon under the generalised OU process. This allows the model to capture drastic changes

at certain time points (e.g., dramatic changes to interest rates due to the outbreaks of financial crisis or war). In practice, many data series are characterised by some potential changes in their evolution structure, i.e., a sudden change in mean or variance and other model parameters. It is then of interest to determine the (i) existence and (ii) location of the change point. This implies segregating the data series into different segments and analysing them in a less efficient but more accurate way. Thus, our research contributions support and complement the objective of papers employing regime-switching OU-process as we provide a methodology to verify the switching phenomenon in the data. We go further by precisely estimating where the switch occurred and how many switches are possible given a data set. An instance of this support and complementarity are depicted in Subsection 3.1 of Tenyakov et al. (2016), where a simple statistical testing of regime-switching in the data was performed.

Pioneering contributions to this field of change-point detection were spearheaded by Page (1954) and Shiryaev (1963). Advances in recent years have tackled the (i) estimation of change points and coefficients of linear regression models with multiple change points (Bai and Perron (1998); Perron and Qu (2006); Lu and Lund (2007), Gombay (2010), and Chen and Nkurunziza (2015)); (ii) change-point testing for the drift parameters of a periodic mean-reverting process (cf. Dehling et al. (2014)); (iii) applications in finance (cf. Spokoiny (2009)); (iv) detection of malware within software (Yan et al. (2008)); (v) climatology (Reeves et al. (2007), Robbins et al. (2011), Gallagher et al. (2012)); and epidemiology (Yu et al. (2013)). The analysis of change points could be described more generally as a hypothesis-testing problem for the existence of change points in various locations. This could be viewed, from another perspective, as a model selection problem where the change points are additional unknown parameters to be estimated.

The change-point problems are typically examined depending on two alternatives: (i) the number of change points is known but their exact locations are unknown (Perron and Qu (2006) and Chen and Nkurunziza (2015)) and (ii) both the number and the exact locations of the change points are unknown. The estimation methods under the first scenario only require the identification of the exact locations of the change points. Clearly, the first alternative is more tractable but less realistic than the second. Closed-form solutions for the direct calculation of the change point are usually not available. Current change-point estimation approaches are normally constructed to perform a search at every possible location of unknown candidate change points via some efficient computational algorithms subject to some constraint or criteria. Examples of well-known algorithms for change point detection include: (i) the binary segmentation type algorithm (Scott and Knott (1974); Sen and Srivastava (1975)), (ii) the segment-neighbourhood type algorithm (Auger and Lawrence (1989); Bai and Perron (1998)) with adaption to the restricted regression model (Perron and Qu (2006)); and (iii) the optimal partitioning type algorithm (Jackson et al. (2002)) and its pruned version, PELT method by Killick et al. (2012). Further details of these algorithms can be found in Killick et al. (2012) and Maidstone et al. (2014).

The intents of our work are motivated by two major research results. The first motivation is from Dehling et al. (2010) that derives a maximum likelihood estimator (MLE) for the drift parameters of the diffusion process and establishes its asymptotic properties. This was extended in Dehling et al. (2014), where there is one unknown change point and a likelihood-ratio test statistic was constructed to determine such change point. The second motivation is from Nkurunziza and Zhang (2016) that establishes the asymptotic properties of both the unrestricted and restricted MLE for the drift parameters of the generalised OU process with a single change point. A James-Stein-type shrinkage estimator for the drift parameters is proposed in Nkurunziza and Zhang (2016) as an improvement and it is also shown that the previously established asymptotic properties also hold for any consistent estimator for the rate of the change point.

Neither Dehling et al. (2014) nor Nkurunziza and Zhang (2016) offer a specific methodology to identify the change point, although Nkurunziza and Zhang (2016) suggest using a method in Chen and Nkurunziza (2015). The Chen and Nkurunziza (2015) method is, however, designed to find change points in multiple regression models, and is not particularly designed for the detection of change points in OU processes. This hole in the literature led us to develop the three main contributions of the current paper. First, we extend the single-change point framework to the multiple-change point setting and present two consistent methods to estimate the unknown locations of change points. Second, we prove the asymptotic normality of the drift parameters' MLE. Third, we employ information-based statistics to resolve the issue of estimating the unknown number of change points and then created three algorithms to implement the calculations. We validate the performance of our estimation techniques using simulated and real market data.

This paper is structured as follows. In Section 2, we present the formulation of the multiple change-point problem. Section 3 summarises the results of Dehling et al. (2014) and Nkurunziza and Zhang (2016) on MLE and the related asymptotic properties, which provide an impetus on the asymptotic performance of our proposed methods. Two estimation methods are put forward to determine the unknown locations of change points in section 4 along with the discussion of the asymptotic properties of the estimators; we find that the asymptotic properties obtained in Nkurunziza and Zhang (2016) also hold in our proposed techniques. Section 5 deals with the problem of both the existence issue and location of the change points using an information approach. We develop computing algorithms in section 6 in order to implement the proposed methods. In section 7, we assess the applicability of our methods through numerical examples on simulated and observed financial market data. Finally, section 8 provides some concluding remarks.

2. Problem description in determining change points

We study the generalised version of the OU process with SDE representation given in (1.1). It is assumed that there exist m ($m \geq 1$) unknown change points

$\tau_j = s_j T$, where $j = 1, \ldots, m$ and $0 < s_1 < \ldots < s_m < 1$. To simplify the notation, we let $\tau_0 = 0$ and $\tau_{m+1} = T$. In our setup $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)\top}, \ldots, \boldsymbol{\theta}^{(m+1)\top})^\top$ with $\boldsymbol{\theta}^{(j)} = (\mu_1^{(j)}, \ldots, \mu_p^{(j)}, -a^{(j)})^\top$ for $\tau_{j-1} < t \le \tau_j$ and

$$S(\boldsymbol{\theta}, t, X_t) = \sum_{j=1}^{m+1} \left(\sum_{k=1}^{p} \mu_k^{(j)} \varphi_k(t) - a^{(j)} X_t \right) \mathbb{1}_{(\tau_{j-1} < t < \tau_j)} \tag{2.1}$$

with $\mathbb{1}_{(.)}$ as the indicator function. Note that $\boldsymbol{\theta}^{(j)}$ may also be a vector.

We start by assuming that the number of change points m is known, but the exact value of each change point denoted by $\tau_1^0, \ldots, \tau_m^0$ (and correspondingly the exact rates s_j^0, $j = 1, \ldots, m$) are unknown. Furthermore, considering that we have multiple change points in the model, we posit that these change points are asymptotically distinct. We further impose the following assumptions.

Assumption 1. $\tau_j^0 = T s_j^0$, $0 < s_1^0 < \cdots < s_m^0 < 1$. *We call* $s_j = \dfrac{\tau_j^0}{T}$ *the change points' arrival rate, and if we have* $\widehat{s}_j$, *the value* $\widehat{\tau}_j^0$ *is immediate.*

Assumption 1 implies that the length of each regime $[\tau_{j-1}^0, \tau_j^0]$ is proportional to T. The structure of the model in each regime is similar to that of the no-change point process studied in Dehling et al. (2010); see also Nkurunziza and Zhang (2016) for the case of a single change point. This means that the results established in the existing literature could also be adapted to the case of multiple change points.

MLEs for the drift parameters and their asymptotic properties were shown in Dehling et al. (2010) for the case of no change point and in Nkurunziza and Zhang (2016) for the case of one change point. Certainly, Nkurunziza and Zhang (2016) is a special case of our study with $m = 1$. The next section reviews previous results and extends them to the multiple change points problem.

3. Prior MLE-based results and our extension

The asymptotic normality for the MLE estimator of the drift parameters in Nkurunziza and Zhang (2016) assumes that the estimator is already consistent. In our case, we shall prove (rather than simply assume) that such an estimator of the change point is consistent.

In the subsequent discussion, we write "$\xrightarrow[T\to\infty]{p}$", "$\xrightarrow[T\to\infty]{D}$", and "$\xrightarrow[T\to\infty]{a.s.}$" to mean convergence in probability, convergence in distribution, and convergence almost surely, respectively. The notation $||.||$ denotes the Frobenius norm for matrices. We use bold, unitalicised English or Greek letters in lowercase for vectors; and bold, unitalicised English or Greek letters in upper case for matrices.

The "$O(\cdot)$" denotes the Landau symbol, also known as the "Big O" notation, which is used to describe the asymptotic behaviour of functions. So, for a set of random variables U_n and a corresponding set of constants a_n, $U_n = O_p(a_n)$ means U_n/a_n is stochastically bounded. Formally, this means $\forall \epsilon > 0$, $\exists \, M > 0$,

$\ni P(|U_n/a_n| > M) < \epsilon$, $\forall n$. On the other hand, the symbol involving "small o", i.e., $U_n = o_p(a_n)$ means U_n/a_n converges in probability to zero as n approaches an appropriate limit. So, since $U_n = o_p(a_n)$ is equivalent to $U_n/a_n = o_p(1)$, convergence in probability to zero is here defined as $\lim_{n\to\infty} (P(|U_n/a_n|) \geq \epsilon) = 0$.

3.1. Log-likelihood function

The following assumption from Dehling et al. (2010) is also retained here.

Assumption 2. $P\left(\int_0^T S^2(\theta, t, X_t) < \infty\right) = 1$, *for all* $0 < T < \infty$ *and elements* $\theta^{(j)}$ *of* θ *involved in* $S(\theta, t, X_t)$ *given by equation (2.1).*

Under Assumptions 1–2 and Theorem 7.6 of Lipster and Shiryaev (2001), the corresponding likelihood function in our modelling framework is

$$\ell^*(\theta, X_t) = \exp\left(\frac{1}{\sigma^2}\int_0^T S(\theta, t, X_t)\,dX_t - \frac{1}{2\sigma^2}\int_0^T S^2(\theta, t, X_t)\,dt\right).$$

The log-likelihood function is therefore

$$\log \ell(\theta, X_t) = \frac{1}{\sigma^2}\int_0^T S(\theta, t, X_t)dX_t - \frac{1}{2\sigma^2}\int_0^T S^2(\theta, t, X_t)dt$$

$$= \frac{1}{\sigma^2}\sum_{j=1}^{m+1}\theta^{(j)\top}\tilde{\mathbf{r}}_{(\tau_{j-1}^0, \tau_j^0)} - \frac{1}{2\sigma^2}\sum_{j=1}^{m+1}\theta^{(j)\top}\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}\theta^{(j)},$$

where $\tilde{\mathbf{r}}_{(\tau_{j-1}^0, \tau_j^0)} = \left(\int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_1(t)dX_t, \ldots, \int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_p(t)dX_t, -\int_{\tau_{j-1}^0}^{\tau_j^0}X_t dX_t\right)^\top$ and

$$\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)} = \begin{bmatrix} \int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_1^2(t)dt & \cdots & \int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_1(t)\varphi_p(t)dt & -\int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_1(t)X_t dt \\ \vdots & \vdots & \vdots & \vdots \\ -\int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_1(t)X_t dt & \cdots & -\int_{\tau_{j-1}^0}^{\tau_j^0}\varphi_p(t)X_t dt & \int_{\tau_{j-1}^0}^{\tau_j^0}X_t^2 dt \end{bmatrix}.$$

Note that the (i_1, i_2) entry (i.e., the element in the i_1-th row and i_2-th column) of $\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$, $i_1, i_2 = 1, \ldots, p+1$ is of the form $\int_{\tau_{j-1}^0}^{\tau_j^0}\zeta_{i_1}(t)\zeta_{i_2}(t)dt$, where $\zeta_i(t) = \varphi_i(t)$ for $i = 1, \ldots, p$ and $\zeta_{p+1}(t) = -X_t$.

3.2. Maximum likelihood estimators for the drift parameters

By setting the first partial derivatives with respect to each of the parameters of $\ell(\theta, X_t)$ to 0, we obtain the MLE of the drift parameters, provided $\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$ is invertible for each $j = 1, \ldots, m+1$. When $\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$ is invertible, Remark 3 of Dehling et al. (2010) shows the existence of $\lim_{T\to\infty}(s_j^0 - s_{j-1}^0)T\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}^{-1}$ and

examines its asymptotic properties (note that the term '$\frac{1}{T}\mathbf{Q}^{-1}$' in the remark may be a typo and the corrected form, as one can confirm from Proposition 4.5 in Dehling et al. (2010), should be $T\mathbf{Q}^{-1}$). Further, Nkurunziza and Zhang (2016) revisits this topic and employs the incompleteness condition of basis functions (see Assumption 3) to guarantee that the denominator of expression (14) in Dehling et al. (2010) does not equal to 0, which is an essential condition for $\lim_{T\to\infty}\frac{1}{(s_j^0-s_{j-1}^0)T}\mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)}$ to be positive definite and thus invertible. Nkurunziza and Zhang (2016) also extends the asymptotic results established in Dehling et al. (2010) to the case of single change point. The results in Nkurunziza and Zhang (2016) can be further adapted to give the asymptotic properties of $\frac{1}{(s_j^0-s_{j-1}^0)T}\mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)}$ (as well as its inverse) under the following assumption.

Assumption 3 (Assumption 2 in Nkurunziza and Zhang (2016)). *For any $T > 0$, the basis functions $\{\varphi_k(t), k = 1, \ldots, p\}$ are Riemann-integrable on $[0, T]$ and satisfy three properties.*

(1) Periodicity. That is, $\varphi_k(t+v) = \varphi_k(t) \ \forall \ k = 1, \ldots, p$ and v is the period in the data.
(2) Orthogonality. That is, $\forall \ k_1, k_2 = 1, \ldots, p$, $\int_0^v \varphi_{k_1}(t)\varphi_{k_2}(t)dt$ is equal to v if $k_1 = k_2$ and 0 otherwise.

Remark 1. *The two points under Assumption 3 correspond to similar assumptions in Dehling et al. (2014). Also, for T large, the family of basis functions $\{\varphi_k(t) : k = 1, \ldots, p\}$ is incomplete, and this is imposed in Nkurunziza and Zhang (2016) to establish the positive definiteness of $\frac{1}{T}\mathbf{Q}_{(0,T)}$. Note that the link between the incomplete basis functions and positive definiteness of $\frac{1}{T}\mathbf{Q}_{(0,T)}$ is discussed in Zhang (2015), and Nkurunziza and Zhang (2016) applied this result directly. The set of basis functions is incomplete in the sense that there are smooth functions which can not be represented as their linear combination (to fix ideas, consider the set $\{1, \cos(\frac{2\pi jx}{T})\}_{j=1}^p$ on the interval $[0, 2\pi]$).*

Another condition used in Nkurunziza and Zhang (2016) in proving the asymptotic properties of the proposed estimators is the boundedness of the basis function (see the proof of Theorem 6.1 in Nkurunziza and Zhang (2016) for more details). Following the methodologies in Nkurunziza and Zhang (2016), we require the incomplete basis function $\{\varphi_k(t), k = 1, \ldots, p\}$ to be sufficiently well behaved so that the asymptotic results similar to those in Nkurunziza and Zhang (2016) can be established. A restructure was to guarantee this which still admits practical choices for basis functions, for example, Fourier basis, to require $\varphi_k(t)$, $k = 1, \ldots, p$, $t \in [0, T]$ to be bounded on $\mathbb{R}^+$ (i.e. $\varphi_k(t) \le K_\varphi$ for some $0 < K_\varphi < \infty$) as for every k the basis function $\varphi_k(t)$ is bounded on $[0, T]$ and v-periodic. Moreover, for the rest of this paper, we assume without loss of generality that the sample size T is an integral multiple of the period length v, i.e., $T = Nv$ for some integer N. Without loss of generality, we let $v = 1$, which implies that $\varphi_k(t+1) = \varphi_k(t)$.

Using the results in Dehling et al. (2010) and Nkurunziza and Zhang (2016), the MLE of the drift parameters based on the log likelihood function provided

above are given by $\hat{\theta} = (\hat{\theta}^{(1)\top}, \ldots, \hat{\theta}^{(m+1)\top})^{\top}$ with

$$\hat{\theta}^{(j)} = \mathbf{Q}^{-1}_{(\tau_j^0, \tau_{j-1}^0)} \tilde{\mathbf{r}}_{(\tau_j^0, \tau_{j-1}^0)}, \quad j = 1, \ldots, m+1. \tag{3.1}$$

Substituting (1.1) into (3.1) and going through some algebraic computations will lead to

$$\hat{\theta}^{(j)} = \theta^{(j)} + \sigma T \mathbf{Q}^{-1}_{(\tau_j^0, \tau_{j-1}^0)} \frac{1}{T} \mathbf{r}_{(\tau_j^0, \tau_{j-1}^0)}, \quad j = 1, \ldots, m+1, \tag{3.2}$$

where $\mathbf{r}_{(a,b)} = \left(\int_a^b \varphi_1(t) dW_t, \ldots, \int_a^b \varphi_p(t) dW_t, - \int_a^b X_t dW_t \right)^{\top}$ for $0 \leq a < b \leq T$.

3.3. Asymptotic properties of the MLE

To study the asymptotic proprieties of the MLE in the next section, equation (3.1) to be precise, we review the established asymptotic results in Dehling et al. (2010) for the case where there is no change point ($m = 0$) and also the results in Nkurunziza and Zhang (2016) when there exists one change point ($m = 1$).

If there is no change point ($m = 0$ and $\theta = \theta^{(1)}$ for $0 < t \leq T$), the solution of the SDE (1.1) has the explicit representation

$$X_t = e^{-a^{(1)}t} X_0 + h^{(1)}(t) + \mathbf{z}_t^{(1)}, \quad 0 < t \leq T, \tag{3.3}$$

where $h^{(j)}(t) = e^{-a^{(j)}t} \sum_{i=1}^{p} \mu_i^{(j)} \int_0^t e^{a^{(j)}s} \varphi_i(s) ds$ and $\mathbf{z}_t^{(j)} = \sigma e^{-a^{(j)}t} \int_0^t e^{a^{(j)}s} dW_s$.

Note that as the process $\{X_t, t \geq 0\}$ is not stationary in the ordinary sense, it is impossible to apply the ergodic theorem directly. To circumvent this, a stationary solution, for $t \in \mathbb{R}$ instead of $t \geq 0$, was introduced in Dehling et al. (2010). Consider

$$\tilde{X}_t = \tilde{h}^{(1)}(t) + \tilde{Z}_t^{(1)}, \quad 0 < t \leq T, \tag{3.4}$$

where $\tilde{h}^{(j)}(t) = e^{-a^{(j)}t} \sum_{i=1}^{p} \mu_i^{(j)} \int_{-\infty}^{t} e^{a^{(j)}s} \varphi_k(s) ds$, $\tilde{Z}_t = \sigma e^{-a^{(j)}t} \int_0^t e^{a^{(j)}s} d\tilde{B}_s$, and $(\tilde{B}_s)_{s \in \mathbb{R}}$ denotes a bilateral Brownian motion, i.e.,

$$\tilde{B}_s = B_s \mathbb{1}_{\mathbb{R}^+}(s) + \bar{B}_{-s} \mathbb{1}_{\mathbb{R}^-}(s),$$

where $(B_s)_{s \geq 0}$ and $(\bar{B}_s)_{s \geq 0}$ are two independent standard Brownian motions. Then, from Lemma 4.3 in Dehling et al. (2010), the sequence of $\mathcal{C}[0,1]$-valued random variables $W_k(s) = \tilde{X}_{k-1+s}$, $0 \leq s \leq 1$, $k \in \mathbb{N}$ is stationary and ergodic. In this case, by Proposition 4.5 of Dehling et al. (2010),

$$\begin{aligned} \frac{1}{T} \int_0^T \tilde{X}_t \varphi_k(t) dt &\xrightarrow[T \to \infty]{a.s.} \int_0^1 (\tilde{h}^{(1)}(t)) \varphi_k(t) dt, \\ \frac{1}{T} \int_0^T \tilde{X}_t^2 dt &\xrightarrow[T \to \infty]{a.s.} \int_0^1 (\tilde{h}^{(1)}(t))^2 dt + \frac{\sigma^2}{2a^{(1)}}. \end{aligned} \tag{3.5}$$

Moreover, under Assumptions 1–3, Lemma 4.4 in Dehling et al. (2010),

$$|\tilde{X}_t - X_t| \xrightarrow[t \to \infty]{a.s.} 0. \tag{3.6}$$

Using (3.6), we have the following properties:

$$\frac{1}{T} \int_0^T \tilde{X}_t \varphi_k(t) dt - \frac{1}{T} \int_0^T X_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} 0,$$

$$\frac{1}{T} \int_0^T \tilde{X}_t^2 dt - \frac{1}{T} \int_0^T X_t^2 dt \xrightarrow[T \to \infty]{a.s.} 0.$$

It follows from (3.5) that

$$\frac{1}{T} \int_0^T X_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} \int_0^1 \tilde{h}^{(1)}(t) \varphi_k(t) dt,$$

$$\frac{1}{T} \int_0^T X_t^2 dt \xrightarrow[T \to \infty]{a.s.} \int_0^1 \tilde{h}^{(1)}(t)^2 dt + \frac{\sigma^2}{2a^{(1)}}.$$

Hence,

$$T \mathbf{Q}_{(0,T)}^{-1} \xrightarrow[T \to \infty]{a.s.} \mathbf{\Sigma}_0^{-1}, \tag{3.7}$$

where

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \mathbf{I}_p & \mathbf{\Lambda}_{(0,T)} \\ \mathbf{\Lambda}_{(0,T)}^\top & w \end{bmatrix}$$

with $\mathbf{\Lambda}_{(0,T)} = \left(\int_0^1 \tilde{h}^{(1)}(t) \varphi_1(t) dt, \ldots, \int_0^1 \tilde{h}^{(1)}(t) \varphi_p(t) dt \right)^\top$ and
$w = \int_0^1 (\tilde{h}^{(1)})^2(t) dt + \frac{\sigma^2}{2a^{(1)}}$.

Furthermore, under Assumptions 1–3, the following properties for $\mathbf{r}_{(0,T)}$ hold.

1. $\{\mathbf{r}_{(0,T)}, T > 0\}$ is a martingale.
2. $\frac{1}{T} \mathbf{r}_{(0,T)} \xrightarrow[T \to \infty]{a.s.} 0$.
3. $\frac{1}{\sqrt{T}} \mathbf{r}_{(0,T)} \xrightarrow[T \to \infty]{D} \mathbf{r} \sim \mathcal{N}_{p+2}(0, \mathbf{\Sigma}_0)$.

For more details about the above properties, readers may refer to Dehling et al. (2010) or Nkurunziza and Zhang (2016).

These properties, together with Slutsky's theorem, yield

$$\frac{1}{\sqrt{T}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[T \to \infty]{D} \mathbf{r} \sim \mathcal{N}_{p+1} \left(0, \mathbf{\Sigma}_0^{-1} \right).$$

Nkurunziza and Zhang (2016) extend the above asymptotic properties to the case of a single change point. Using similar arguments, we extend these results in the context of multiple change points. We first present a result covering the coefficients of SDE (1.1).

Proposition 3.1. *Under Assumptions 1–3, the coefficients in SDE (1.1) for $m \geq 1$ satisfy both the space-variable Lipschitz and the spatial growth conditions.*

Proof. See Appendix A. □

Using Proposition 3.1 and the mean-reverting property of OU process, together with the similar arguments employed in the proof of Theorem 3.1 (as well as Lemma 3.3) in Nkurunziza and Zhang (2016), it may be verified that SDE (1.1) admits a solution that is uniformly bounded in L^2 and

$$\sup_{t \geq 0} \mathrm{E}\left(X_t^2\right) \leq K_1 \tag{3.8}$$

for some $0 < K_1 < \infty$.

Employing (3.8), we (3.3) and (3.4) to their representations in context of multiple change points ($m \geq 1$) given by

$$X_t = e^{-a^{(j)}t} X_0^{(j)} + h^{(j)}(t) + \mathbf{z}_t^{(j)}, \quad \tau_{j-1}^0 < t \leq \tau_j^0, \quad j = 1, \ldots, m+1, \tag{3.9}$$

where $X_0^{(j)} = X_{\tau_{j-1}}$ and

$$\tilde{X}_t = \tilde{h}^{(j)}(t) + \tilde{Z}_t^{(j)}, \quad \tau_{j-1}^0 < t \leq \tau_j^0, \quad j = 1, \ldots, m+1. \tag{3.10}$$

Then

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} (s_j^0 - s_{j-1}^0) \int_0^1 \tilde{h}^{(j)}(t) \varphi_k(t) dt \tag{3.11}$$

and

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t^2 dt \xrightarrow[T \to \infty]{a.s.} (s_j^0 - s_{j-1}^0) \left(\int_0^1 (\tilde{h}^{(j)}(t))^2 dt + \frac{\sigma^2}{2a^{(j)}} \right). \tag{3.12}$$

Using (3.6) and (3.8) and similar arguments in the proof of Theorem 3.1 in Nkurunziza and Zhang (2016) with the estimated change point replaced by the exact value of the change points, the following properties hold:

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt - \frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} X_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} 0,$$

and

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t^2 dt - \frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} X_t^2 dt \xrightarrow[T \to \infty]{a.s.} 0.$$

Thus,

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} X_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} (s_j^0 - s_{j-1}^0) \int_0^1 \tilde{h}^{(j)}(t) \varphi_k(t) dt$$

and

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} X_t^2 dt \xrightarrow[T \to \infty]{a.s.} (s_j^0 - s_{j-1}^0) \left(\int_0^1 (\tilde{h}^{(j)}(t))^2 dt + \frac{\sigma^2}{2a^{(j)}} \right).$$

So,

$$\frac{1}{T}\mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)} \xrightarrow[T\to\infty]{a.s.} (s_j^0 - s_{j-1}^0)\boldsymbol{\Sigma}_j, \tag{3.13}$$

where

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \mathbf{I}_p & \boldsymbol{\Lambda}_j \\ \boldsymbol{\Lambda}_j' & w_j \end{bmatrix} \tag{3.14}$$

with $\boldsymbol{\Lambda}_j = \left(\int_0^1 \tilde{h}^{(j)}(t)\varphi_1(t)dt, \ldots, \int_0^1 \tilde{h}^{(j)}(t)\varphi_p(t)dt \right)^\top$ and $w_j = \int_0^1 (\tilde{h}^{(j)})^2(t)dt + \frac{\sigma^2}{2a^{(j)}}$, $j = 1, \ldots, m+1$. Further, by the Continuous Mapping Theorem,

$$T\mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)}^{-1} \xrightarrow[T\to\infty]{a.s.} \frac{1}{(s_j^0 - s_{j-1}^0)}\boldsymbol{\Sigma}_j^{-1}. \tag{3.15}$$

So long as Assumptions 1–3 hold and with the aid of similar argument used in the proof of Proposition 2.1 in Nkurunziza and Zhang (2016), it may be shown that $\boldsymbol{\Sigma}_j$ is positive definite.

Note that (3.13) and (3.15) are key elements in analysing the asymptotic properties of $\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}$ and its inverse, where $\hat{\tau}_j$ is the estimator of τ_j^0.

Invoking the boundedness property of $\varphi_k(t)$, we have

$$\mathrm{E}\left(\int_0^T \left(\frac{1}{\sqrt{T}}\varphi_k(t)\mathbb{1}(\tau_{j-1}^0 < t \le \tau_j^0) \right)^2 dt \right) = \int_0^T \frac{1}{T}\varphi_k^2(t)\mathbb{1}(\tau_{j-1}^0 < t \le \tau_j^0)dt$$
$$\le K_\varphi^2(s_j^0 - s_{j-1}^0) < \infty \tag{3.16}$$

for $k = 1, \ldots, p$ and $j = 1, \ldots, m$. Similarly, using (3.8),

$$\mathrm{E}\left(\int_0^T \left(\frac{1}{\sqrt{T}}X_t\mathbb{1}(\tau_{j-1}^0 < t \le \tau_j^0) \right)^2 dt \right) = \int_0^T \frac{1}{T}\mathrm{E}(X_t^2)\mathbb{1}(\tau_{j-1}^0 < t \le \tau_j^0)dt$$
$$\le K_1^2(s_j^0 - s_{j-1}^0) < \infty. \tag{3.17}$$

From Proposition 1.21 in Kutoyants (2004), we get

$$\left(\frac{1}{\sqrt{T}}\mathbf{r}_{(0,\tau_1^0)}, \ldots, \frac{1}{\sqrt{T}}\mathbf{r}_{(\tau_m^0,T)} \right) \xrightarrow[T\to\infty]{D} \mathcal{N}_{(m+1)(p+1)}\left(\mathbf{0}, \tilde{\boldsymbol{\Sigma}} \right), \tag{3.18}$$

where $\mathbf{0}$ is a vector of zeros and $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}(s_1^0\boldsymbol{\Sigma}_1, (s_2^0 - s_1^0)\boldsymbol{\Sigma}_2, \ldots, (1 - s_m^0)\boldsymbol{\Sigma}_{m+1})$. Combining (3.13), (3.15) and (3.18), along with the Slutsky's Theorem and some algebraic computations (see Proposition 6.3 in Nkurunziza and Zhang (2016) for more details),

$$\frac{1}{\sqrt{T}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow[T\to\infty]{D} \mathcal{N}_{(m+1)(p+1)}(\mathbf{0}, \sigma^2\tilde{\boldsymbol{\Sigma}}^{-1}), \tag{3.19}$$

where $\tilde{\boldsymbol{\Sigma}}^{-1} = \mathrm{diag}\left(\frac{1}{s_1^0}\boldsymbol{\Sigma}_1^{-1}, \frac{1}{s_2^0 - s_1^0}\boldsymbol{\Sigma}_2^{-1}, \ldots, \frac{1}{1 - s_m^0}\boldsymbol{\Sigma}_{m+1}^{-1} \right)$.

The above asymptotic properties are established based on the exact values of the locations of the change points τ_j^0, $j = 1, \ldots, m$. However, in practice, τ_j^0 are often unknown. Hence, we shall devise methods to estimate the unknown τ_j^0 and investigate whether the above asymptotic normality still hold for the estimated change points.

4. Estimation of change points and pertinent asymptotic properties

In Subsections 4.1 and 4.2, we develop two techniques to estimate the unknown locations of change points. The asymptotic normality of $\hat{\theta}$ based on the estimated change points is discussed in Subsection 4.3.

4.1. Least sum squared error method

We introduce the least sum of squared errors (LSSE) method then investigate the consistency of our proposed estimator. Consider a partition $0 = t_0 < \cdots < t_n = T$ on a time period $[0, T]$ with constant increment $\Delta_t = t_{i+1} - t_i$. Also, let $Y_i = X_{t_{i+1}} - X_{t_i}$ and $\mathbf{z}_i = (\varphi_1(t_i), \ldots, \varphi_p(t_i), -X_{t_i})\Delta_t$.

Due to the uncertain locations of estimated change points, the exact value of the drift parameters θ and the MLE may have different indices. For example, if $\hat{\tau}_j > \tau_j^0$, then for all $t_i \in (\tau_j^0, \hat{\tau}_j]$, the associated exact value of the drift parameters is $\theta^{(j+1)}$ but the MLE is $\hat{\theta}^{(j)}$. So here, θ_i and $\hat{\theta}_i$ refer to the exact value and MLE of the drift parameters at time point t_i, respectively. In this case, $\theta_i = \sum_{j=1}^{m+1} \theta^{(j)} \mathbb{1}(\tau_{j-1}^0 \le t_i \le \tau_j^0)$ and $\hat{\theta}_i = \sum_{j=1}^{m+1} \hat{\theta}^{(j)} \mathbb{1}(\hat{\tau}_{j-1} \le t_i \le \hat{\tau}_j)$ with $\hat{\theta}^{(j)} = \mathbf{Q}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)}^{-1} \tilde{\mathbf{r}}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)}$ for $j = 1, \ldots, m+1$. Then by the Euler-Maruyama discretisation method,

$$Y_i = \mathbf{z}_i \theta_i + \epsilon_i, \quad i = 1, \ldots, T, \tag{4.1}$$

where ϵ_i is the error term $\sigma\sqrt{\Delta_t}\omega_i$, and ω_i is the ith independent draw from a standard normal variable. Therefore, we could now use the LSSE method to estimate the change points.

From (4.1), the estimates for the multiple change points $\tau^0 = (\tau_1^0, \ldots, \tau_m^0)$ are given by

$$\hat{\tau} = \arg\min_\tau SSE([0, T], \tau, \hat{\theta}(\tau)), \tag{4.2}$$

where

$$SSE([0, T], \tau, \hat{\theta}(\tau)) = \sum_{t_i \in [0, T]} (Y_i - \mathbf{z}_i \hat{\theta}_i)^\top (Y_i - \mathbf{z}_i \hat{\theta}_i). \tag{4.3}$$

Consistency of the proposed estimator

Under Assumptions 1–3, $\sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i$ for $j = 1, \ldots, m$, the discretised versions of $\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$ are both positive definite with probability 1 provided that the basis functions $\{\varphi_k(t), i = 1, \ldots, p\}$ are incomplete. Moreover, using (3.13) (see also Proposition 2.1 in Nkurunziza and Zhang (2016)), we have that $\frac{1}{(s_j^0 - s_{j-1}^0)T} \mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$ converges in probability to some positive definite matrices for large T, as do their respective discretised versions. Hence, for large T, it is reasonable to impose a useful assumption in proving the consistency of the estimators of the change points.

Assumption 4. *For every $j = 1, \ldots, m$, there exists an $L_0 > 0$ such that for all $L > L_0$ the minimum eigenvalues of $\frac{1}{\ell} \sum\limits_{t_i \in (\tau_j^0, \tau_j^0 + L]} \mathbf{z}_i' \mathbf{z}_i$ and of $\frac{1}{\ell} \sum\limits_{t_i \in (\tau_j^0 - L, \tau_j^0]} \mathbf{z}_i' \mathbf{z}_i$, as well as their respective continuous-time versions $\frac{1}{\ell} \mathbf{Q}_{(\tau_j^0, \tau_j^0 + L)}$ and $\frac{1}{\ell} \mathbf{Q}_{(\tau_j^0 - L, \tau_j^0]}$, are all bounded away from 0.*

For the motivation of the above assumption, see Perron and Qu (2006) and Chen and Nkurunziza (2015). The next two propositions provide results characterising consistency.

Proposition 4.1. *Suppose that $\theta^{(1)} - \theta^{(2)}$, the shift in the drift parameters, is of fixed non-zero magnitude independent of T. Then, under Assumption 1–4, and $\hat{s}_j - s_j^0 \xrightarrow[T \to \infty]{P} 0$, $j = 1, \ldots, m$.*

Proof. See Appendix B. □

Proposition 4.2. *Suppose the conditions in Proposition 4.1 hold. Then, for every $\epsilon > 0$, there exists a $C > 0$ such that for large T, $P(T|\hat{s}_j - s_j| > C) < \epsilon$, for every $j = 1, \ldots, m$.*

Proof. See Appendix B. □

Proposition 4.1 shows that the estimated rate $\hat{s}_j = \frac{\hat{\tau}_j}{T}$ is consistent for s_j^0, for $j = 1, \ldots, m$. Proposition 4.2 gives the convergence rate T of $\hat{\tau}_j$, $j = 1, \ldots, m$. In reality we may encounter the case where the shift is time-dependent, and, in particular as T tends to infinity, the shift may shrink towards 0 at rate v_T, i.e., $\theta^{(j)} - \theta^{(j-1)} = \mathbf{M} v_T$, where $\mathbf{M}$ is independent of T and $v_T \xrightarrow[T \to \infty]{} 0$. In this case, the validity of Propositions 4.1 and 4.2 depends on the speed v_T. In fact, using similar arguments as in the proofs of Propositions 4.1 and 4.2, we have the following corollary.

Corollary 4.1. *Suppose that $\theta^{(j)} - \theta^{(j-1)} = \mathbf{M} v_T$, where $j = 1, \ldots, m + 1$, $\mathbf{M}$ is independent of T and $v_T \xrightarrow[T \to \infty]{} 0$ but $T^{1/2 - r^*} v_T \xrightarrow[T \to \infty]{} \infty$ for some $0 < r^* < 1/2$. Then under Assumptions 1–4, we have (i) $\hat{s} - s^0 \xrightarrow[T \to \infty]{P} 0$ and (ii) for every $\epsilon > 0$, there exists a $C > 0$ such that for large T, $P(T v_T^2 |\hat{s} - s| > C) < \epsilon$.*

Proof. See Appendix B. □

4.2. Maximum log-likelihood method

We introduce an alternative method to estimate the location of the change points based on the maximum of log-likelihood function. Recall that the log-likelihood function for (1.1) with the exact change points $\tau_1^0, \ldots \tau_m^0$ is given by

$$
\begin{aligned}
\log \ell(\tau_1^0, \ldots, \tau_m^0, \theta) &= \frac{1}{\sigma^2} \int_0^T S(\theta, t, X_t) dX_t - \frac{1}{2\sigma^2} \int_0^T S^2(\theta, t, X_t) dt \\
&= \frac{1}{\sigma^2} \sum_{j=1}^{m+1} \int_{\tau_{j-1}^0}^{\tau_j^0} S(\theta^{(j)}, t, X_t) dX_t \\
&\quad - \frac{1}{2\sigma^2} \sum_{j=1}^{m+1} \int_{\tau_{j-1}^0}^{\tau_j^0} S^2(\theta^{(j)}, t, X_t) dt.
\end{aligned}
\tag{4.4}
$$

From (4.4), the estimator of the location of the change points is given

$$
\hat{\tau} = \arg \max_{\tau} \log \ell(\tau, \hat{\theta}(\tau)),
\tag{4.5}
$$

where $\hat{\theta}(\tau)$ is the MLE of θ based on the given change points $\tau = (\tau_1, \ldots, \tau_m)$.

In practice, the calculation of $\log \ell(\tau, \hat{\theta}(\tau))$ in (4.5) relies on numerical approximation methods (see Auger and Lawrence (1989)) to compute the integrals inside $\log \ell(\tau, \hat{\theta}(\tau))$. For example, by approximating the Riemann sum based on a partition $0 = t_0^* < \ldots < t_n^* = T$ with $\Delta_t^* = t_{i+1}^* - t_i^*$, (4.5) is calculated as

$$
\begin{aligned}
\log \ell^*([0, T], \tau, \hat{\theta}(\tau)) &= \frac{1}{\sigma^2} \sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \hat{\theta}^{(j)\prime} V(t)'(X_{t_{i+1}^*} - X_{t_i^*}) \\
&\quad - \frac{1}{2\sigma^2} \sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \left(\hat{\theta}^{(j)\prime} V(t)' \right)^2 \Delta_t^*.
\end{aligned}
\tag{4.6}
$$

The approximated version of (4.5) is then

$$
\hat{\tau} = \arg \max_{\tau = (\tau_1, \ldots, \tau_m)} \log \ell^*([0, T], \tau, \hat{\theta}(\tau)).
\tag{4.7}
$$

Consistency of the proposed estimator

We now link results (4.7) and (4.2), which are the respective results from the LSSE- and MLE-based methods.

Proposition 4.3. *Consider the observed process X_t, $t \in [0, T]$. If the increment Δ_t^* is equal to Δ_t defined in Section 4.1 then under Assumptions 1–4, the asymptotic results given in Propositions 4.1 and 4.2 as well as in Corollary 4.1 also hold for (4.7).*

Proof. See Appendix B.   □

Using the consistency properties, we can establish the asymptotic normality for the MLE of drift parameters based on the estimated change points.

4.3. Asymptotic normality of $\hat{\theta}$ based on the estimated change points

Previously, we established the T-rate consistency of the estimated change points. Based on these asymptotic consistency results, we extend some of the asymptotic normality results in Nkurunziza and Zhang (2016) that are related to our study to the case of multiple change points.

Proposition 4.4. *Let* $\hat{\boldsymbol{\tau}} = \{\hat{\tau}_1, \ldots, \hat{\tau}_m\}$ *be the estimated change points using (4.2) or (4.5). Then, under Assumptions 1–4, we have that for* $i = 1, \ldots, p$,

$$\frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t \varphi_k(t) dt \xrightarrow[T \to \infty]{p} (s_j - s_{j-1}) \int_0^1 \tilde{h}(t) \varphi_k(t) dt. \tag{4.8}$$

Similarly,

$$\frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t^2 dt \xrightarrow[T \to \infty]{p} (s_j - s_{j-1}) \left(\int_0^1 \tilde{h}^2(t) dt + \frac{\sigma^2}{2a^{(j)}} \right). \tag{4.9}$$

Proof. See Appendix C. $\qquad \square$

Proposition 4.5. *Under the same conditions as in Proposition 4.4, we have that for* $i = 1, \ldots, p$,

$$P \left(\int_0^T \left(\frac{1}{\sqrt{T}} \varphi_k(t) \mathbb{1}_{\hat{\tau}_{i-1} < t \le \hat{\tau}_i} \right)^2 < \infty \right) = 1, P \left(\int_0^T \left(\frac{-1}{\sqrt{T}} X_t \mathbb{1}_{\hat{\tau}_{i-1} < t \le \hat{\tau}_i} \right)^2 < \infty \right) = 1. \tag{4.10}$$

Proof. This follows from (3.8) and similar arguments in the proofs of Lemmas 3.1 and 3.2 in Nkurunziza and Zhang (2016). $\qquad \square$

Employing Propositions 4.4 and 4.5, we have the following results.

Proposition 4.6.

$$T \mathbf{Q}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)}^{-1} \xrightarrow[T \to \infty]{P} \frac{1}{s_j^0 - s_{j-1}^0} \boldsymbol{\Sigma}_j^{-1}, j = 1, \ldots, m, \tag{4.11}$$

where $\boldsymbol{\Sigma}_j$ *is defined in (3.14).*

Proof. See Appendix C (or Proposition 6.2 in Nkurunziza and Zhang (2016) for the single change-point case). $\qquad \square$

Proposition 4.7. *Let* $\hat{\boldsymbol{\tau}} = \{\hat{\tau}_1, \ldots, \hat{\tau}_m\}$ *be the estimated change points from (4.2), and suppose that Assumptions 1–4 hold. Then,*

$$\frac{1}{\sqrt{T}} \mathbf{r}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)} - \frac{1}{\sqrt{T}} \mathbf{r}_{(\tau_{j-1}^0, \tau_j^0)} \xrightarrow[T \to \infty]{P} \mathbf{0}, \tag{4.12}$$

where

$$\mathbf{r}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)} = \left(\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \varphi_1(t) dW_t, \ldots, \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \varphi_p(t) dW_t, - \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t dW_t \right)^{\top}.$$

Proof. See Appendix C (or Proposition 6.2 in Nkurunziza and Zhang (2016) for the case of only a single change point). □

The next result following from Propositions 4.6 and 4.7 plays an essential role in proving the asymptotic normality of MLE of the drift parameter based on the proposed estimated change points.

Proposition 4.8. *Suppose that Assumptions 1–4 hold. Then*

$$\left(\frac{1}{\sqrt{T}} \mathbf{r}_{(0,\tau_1^0)}, \ldots, \frac{1}{\sqrt{T}} \mathbf{r}_{(\tau_m^0, T)} \right) \xrightarrow[T \to \infty]{D} \mathbf{r} \sim N_{(m+1)(p+1)} \left(0, \tilde{\mathbf{\Sigma}} \right), \qquad (4.13)$$

where $\tilde{\mathbf{\Sigma}} = \mathrm{diag} \left(s_1^0 \mathbf{\Sigma}_1, (s_2^0 - s_1^0) \mathbf{\Sigma}_2, \ldots, (1 - s_m^0) \mathbf{\Sigma}_{m+1} \right)$.

Proof. This proposition follows from Proposition 3.1 in Nkurunziza and Zhang (2016). □

From Proposition 4.8 together with Slutsky's Theorem, the following corollary establishes the asymptotic normality for $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\theta}}^{(1)}, \ldots, \hat{\boldsymbol{\theta}}^{(m)} \right)$.

Corollary 4.2. *Let* $\hat{\boldsymbol{\tau}} = \{\hat{\tau}_1, \ldots, \hat{\tau}_m\}$ *be the estimated change points from (4.2), and supposes that Assumption 1–4 hold.*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[T \to \infty]{D} \rho \sim N_{((m+1)(p+1))} \left(0, \sigma^2 \tilde{\mathbf{\Sigma}}^{-1} \right), \qquad (4.14)$$

where $\tilde{\mathbf{\Sigma}}^{-1} = \mathrm{diag} \left(\frac{1}{s_1^0} \mathbf{\Sigma}_1^{-1}, \frac{1}{s_2^0 - s_1^0} \mathbf{\Sigma}_2^{-1}, \ldots, \frac{1}{1 - s_m^0} \mathbf{\Sigma}_{m+1}^{-1} \right)$.

Proof. See Appendix C. □

5. Estimating the number of change points

In the last section, we developed two consistent estimation methods for the case when the number of change points is known. In this section, we extend our examination of the change-point problem when the number of change points is also unknown. Hence, we are interested in knowing the number of change points as well as their exact locations.

One popular methodology for detecting the unknown number of change points is to treat this issue as a model-selection problem. For instance, adding one change point into (1.1) brings $p+1$ extra drift parameters into the model. Thus, detecting the number of change points can be considered as selecting the most suitable statistical model from a series of candidate models with different number of change points, and this can be solved using an informational approach. Such approach deems the most appropriate model as the one which minimises the log-likelihood-based information criterion

$$\mathcal{IC}(m) = -2 \log \ell(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}}) + (m+1)h(p)\phi(T). \qquad (5.1)$$

In (5.1), $\log \ell(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}})$ is defined in (4.4); $\hat{\boldsymbol{\tau}}$ is obtained via (4.5) corresponding to each m; $h(p) = p + 1$ if there is no change in σ (or $p + 2$ if there is a change in

σ); $\phi(T)$ is a non-decreasing function of T, the length of the data set; and m is the potential number of change points to be determined.

Based on the asymptotic results for the Riemann sum approximation of the log-likelihood function $\log \ell(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}})$, we use the criterion

$$\mathcal{IC}(m) = -2 \log \ell^*([0, T], \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau})) + (m + 1)(p + 1)\phi(T), \qquad (5.2)$$

where $\log \ell^*([0, T], \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau}))$ is given in (4.6).

Note that, if the number of change points is known, the term $(m+1)h(p)\phi(T)$ is fixed and the approach covering (5.2) is equivalent to the maximum log-likelihood method introduced in the previous section. The efficiency of information criterion depends on the choice of the penalty criterion $\phi(T)$. For example, if $\phi(T) = 2$, then (5.2) reduces to the well-known Akaike information criterion (AIC) Akaike (1973). However, in practice, a model selected by minimising the AIC may not be asymptotically consistent in terms of the model order; see Schwarz (1978). Modified information criteria were, thus, proposed to overcome this problem. One example is the Schwarz information criterion (SIC) Schwarz (1978), which sets $\phi(T)$ as the logarithm of the sample size. The SIC has been successfully applied to the change-point analysis in the literature. As it gives an asymptotically consistent estimate of the order of the true model, we also adapt the SIC for our theoretical development.

Note that the penalty term $(m+1)(p+1)\phi(T)$ in SIC increases as the sample size increases. Hence, for large sample size, SIC tends to ignore the relatively small changes in the process. This feature makes it useful for those who are mainly interested in studying only the major changes within certain time period. Further, based on the SIC, we have the following asymptotic results for (5.2).

Proposition 5.1. *Under Assumptions 1–4, we have that for large T, (i) $\mathcal{IC}(m = m^0) < \mathcal{IC}(m < m^0)$ with probability 1 and (ii) $\mathcal{IC}(m = m^0) < \mathcal{IC}(m > m^0)$ with probability 1.*

Proof. See Appendix D. □

Proposition 5.1 tells us that, for large T, $\mathcal{IC}(m)$ reaches its minimum value when $m = m^0$ and this allows us to detect the exact value of m^0.

6. Computational algorithms

In this section, we put forward algorithms for computing (4.2) and (4.7) when m^0 is known. We also provide algorithms for computing (5.2) when m^0 is unknown. Based on these algorithms, a simulation study to examine the efficiency of the proposed methods for different time periods T is presented in Section 7.1. Our numerical results will show that, with our sample parameter set, the proposed methods perform well for values of T as small as $T = 5$.

6.1. Algorithm for ([4.2](#)) and ([4.7](#)) (with known m^0)

In estimating the unknown locations of change points, a standard searching method is to compute the criteria, i.e., least squared errors for ([4.2](#)) or maximum log likelihoods for ([4.7](#)), through all possible locations of change points and search for the one that returns the optimal value. However, for m change points the associated costs for the above searching procedure are of order $O((T/\Delta_t)^m)$. Thus, for large T and small Δ_t, the computations can be time consuming. To overcome this problem, we adopt to our two proposed LSSE and MLE methods a dynamic programming algorithm due to Bai and Perron ([1998](#)), see also Perron and Qu ([2006](#)), to reduce the computational cost to $O(m(T/\Delta_t)^2)$ for m change points. This algorithm is very efficient when $m \geq 2$.

Algorithm 1. Let $H_1(r, T_r)$ be either $H_1(r, T_r) = \min_\tau SSE([0, T_r], \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau}))$, the least sum squared error for ([4.2](#)) or $H_1(r, T_r) = \max_\tau \log \ell^*([0, T_r], \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau}))$, the maximum Riemann sum approximation of log likelihood for ([4.7](#)) computed based on the optimal partition of time interval $[0, T_r]$ that contains r change points. Also, let $H_2(a, b)$ be the SSE for ([4.2](#)) or Riemann sum approximation of log likelihood for ([4.7](#)) computed based on a time regime $(a, b]$. Further, we assume that Assumption [1](#) holds and let $h = \epsilon T$ be the minimal permissible length of a time regime. Then ([4.2](#)) or ([4.7](#)) with known $m = m^0$ can be computed as follows.

Step 1: Compute and save $H_2(a, b)$ for all time periods $(a, b]$ that satisfy $b - a \geq h$.

Step 2: Compute and save $H_1(1, T_1)$ for all $T_1 \in [2h, T - (m-1)h]$ by solving the optimisation problem

$$H_1(1, T_1) = \begin{cases} \min_{a \in [h, T_1-h]}[H_2(0, a) + H_2(a, T_1)] & \text{for } (4.2) \\ \max_{a \in [h, T_1-h]}[H_2(0, a) + H_2(a, T_1)] & \text{for } (4.7). \end{cases}$$

Step 3: Sequentially compute and save

$$H_1(r, T_r) = \begin{cases} \min_{a \in [rh, T_r-h]}[H_1(r-1, a) + H_2(a, T_r)] & \text{for } (4.2) \\ \max_{a \in [rh, T_r-h]}[H_1(r-1, a) + H_2(a, T_r)] & \text{for } (4.7) \end{cases}$$

for $r = 2, \ldots m - 1$, and $T_r \in [(r+1)h, T - (m-r)h]$.

Step 4: Finally, the estimated change points are obtained by solving

$$H_1(m, T) = \begin{cases} \min_{a \in [mh, T-h]}[H_1(m-1, a) + H_2(a, T)] & \text{for } (4.2) \\ \max_{a \in [mh, T-h]}[H_1(m-1, a) + H_2(a, T)] & \text{for } (4.7) \end{cases}$$

and $H_1(m-1, a) = H_2(0, a)$ if $m = 1$.

The steps in Algorithm [1](#) can be viewed as a combination of two components. Step 1 computes all possible choices of H_2, and the computations in this step are at most of order $O((T/\Delta_t)^2)$ as there are at most $(T/\Delta_t)^2$ different time periods

$(a, b]$ in the dataset. This step is useful, since in the succeeding steps some pairs of (a, b) will be visited more than once during the optimisation process, so using previously saved results for $H_2(a, b)$ will be helpful to reduce computations. Steps 2-4 can be treated as an application of the Segment Neighbourhood Search (SNS) method introduced by Auger and Lawrence (1989). The goal of Steps 2–4 is to search for the global optimal locations of the m change points and the total computation costs in these steps are also of $O(m(T/\Delta_t)^2)$. Note that when $m = m^0 = 1$ (a single change point), only the last step is needed to search for the optimal location of a change point, and the related computations costs are of $O(T/\Delta_t)$.

When $m^0 > 1$ and T is large, Algorithm 1 can be extremely time-consuming because of the $O(m(T/\Delta_t)^2)$ computations, we aim to decrease the computational costs in this case. Apparently, some computations in Step 1 may be redundant. For example, in Step 2, the domain for the optimisation problem is $a \in [h, T_1 - h]$ for each $T_1 \in [2h, T - (m^0 - 1)h]$; so the calculations of $H_2(0, b^*)$ for all $b^* > T - m^0 h$ in Step 1 become unnecessary as these results will not be used. Thus, the computations in Step 1 could be moved into Steps 2- 4 so that only the necessary H_2 are computed and stored. This means one could begin the algorithm from Step 2 by computing and storing $H_2(0, b)$ for $b \in [h, T - mh]$ and $H_2(a, b)$ for $a \in [h, T - mh]$, $b \in [2h, T - (m-1)h]$, then solve for $H_1(1, T_1)$. In Step 3, for $r = 2, \ldots, m - 1$, one only needs to compute and save $H_2(a, b)$ for $(a \in [T - (m - r + 2)h, T - (m - r + 1)h], b \in [(r + 1)h, T - (m - r)h]) \bigcup (a \in [rh, T - (m - r + 2)h], b \in [T - (m - r + 1)h, T - (m - r)h])$ before solving for $H_1(1, T_1)$. Finally, in Step 4, we compute and store $H_2(a, T)$ for $a \in [mh, T - h]$ before solving for $H_1(m, T)$.

6.2. Algorithm for (5.2) (with unknown m^0)

When m^0 is unknown, one may compute and compare the m values of (5.2) and m varies from 0 up to $m_{\max}$ for some $0 \leq m_{\max} \leq \lceil [T/h] \rceil$. The upper bound $m_{\max}$ can also be predetermined from the descriptive analysis of the observed processes. For each m, one can first apply Algorithm 1 to obtain the estimated change points and compute (5.2) accordingly. After the $m_{\max}$ computations, the desired $\hat{m}$ is the one that returns the minimum value of (5.2). By Proposition 5.1, $\hat{m}$ is consistent when T is large, provided $m^0 \in [0, m_{\max}]$.

If we directly apply Algorithm 1 to $m = 1, \ldots, m_{\max}$, the total computations will be of order $O((1 + 2 + \ldots + m_{\max})(T/\Delta_t)^2)$. To further simplify the computations, we study the behaviour of Step 2–3 in Algorithm 1 when m increases from m^* to $m^* + 1$. In this case, the ranges of T_r reduces from $T_r \in [(r+1)h, T - (m^* - 1)h]$ to $T_r \in [(r+1)h, T - (m^*)h]$ for each $r = 1, \ldots, m-1$. This implies that the stored optimisation results of Steps 2 and 3 in Algorithm 1 at the previous step ($m = m^*$) can also be used in the current step ($m = m^* + 1$). Therefore, with the previously stored results, the only step that needs to be updated for each m is when $r = m - 1$ and $r = m$, and the associated computations are of order $O((T/\Delta_t)^2)$ at $r = m - 1$ and $O(T/\Delta_t)$ at $r = m$. Based on these

considerations, we tailor the SNS algorithm for (5.2) and the related computations for $m = 0 \ldots, m_{\max}$ are of order $O(m_{\max}(T/\Delta_t)^2)$.

Algorithm 2 (SNS method). Step 1: Follows all steps in Algorithm 1 to search for the optimal locations of the m estimated change points then store the computed value of (5.2) for $m = 0, 1, 2$. Note that the results of $H_2(a, b)$ for all (a, b) such that $a - b \geq h$ as well as the optimisation results of $H_1(r, T_r)$ for all $r = 1, \ldots, m$ and $T_r \in [(r+1)h, T - (m-r)h]$ need to be stored for future use.

Step 2: For $m = 3, \ldots, m_{\max}$, first let $r = m-1$ and $T_r \in [(r+1)h, T-(m-r)h]$ then compute and store $H_1(r, T_r)$. Next let $r = m$ and the estimated change points are obtained by solving $H_1(m, T)$, where $H_1(r, T_r)$ and $H_1(m, T)$ are defined in Algorithm 1. Finally, based on the estimated m change points, compute and store $\mathcal{IC}(m)$.

Step 3: $\hat{m}$ is obtained from $m = 1, \ldots, m_{\max}$ that returns the smallest value of (5.2).

The advantage of SNS method is that it returns the optimal locations of change points for every $m = 1, \ldots, m_{\max}$. Hence, it is useful if one interested in investigating the relationships between the locations of change points and m. However, for large T and $m_{\max}$, the $O(m_{\max}(T/\Delta_t)^2)$ computational costs in the SNS method may be high.

In addition to SNS method, another dynamic programming algorithm for finding the unknown number of change points is called the Optimal Partitioning (OP) algorithm introduced by Jackson et al. (2002). The related computational costs are of order $O(T/\Delta_t)^2)$ for any m; hence, it is more efficient than SNS when $m_{\max}$ is large. Based on the OP algorithm, Killick et al. (2012) introduced the Pruned Exact Linear Time (PELT) method. Although the maximum computational costs for the PELT algorithm is still up to $O(n^2)$ for a data set with size n, the computations in the PELT method, which involved pruning of the solution space under some conditions can be much less than those required in the OP.

However, the PELT method introduced in Killick et al. (2012) may not satisfy Assumption 1, which is essential for most of the theoretical properties developed in this paper. In fact, under Assumption 1: (i) there is no change point in time period $[0, h)$, where h is defined in Algorithm 1; (ii) if there is a change point $\tau^* \in [h, 2h)$, then there is no non-zero change point prior to τ^*; and (iii) for any potential change point $\tau^* \in [2h, T]$, the minimal distance between it and the most recent change point prior to this change point is at least h. Based on these considerations, we use the following modified version of the PELT algorithm.

Algorithm 3 (Modified PELT method). Let $n = T/\Delta_t$ be the length of the data set based on the partition $0 = t_0 < t_1 < \ldots < t_n = T$ of time period $[0, T]$ with increment Δ_t, and let $t_h = h/\Delta_t$. Set $SS_{2t_h} = \{0\}$ and $F(t_i) = 0$ for $i = 1, \ldots, t_h - 1$. Then, for $i = t_h, \ldots, n$, compute and store the values obtained from the following steps.

Step 1: For $i = t_h, \ldots, 2t_h - 1$, compute and store $F(t_i) = -2 * \log \ell^*([0, t_i], \hat{\theta}) + (p+1) \log(n)$, where $\log \ell^*([a, b], \hat{\theta})$ is the Riemann sum approximation (midpoints) of the log likelihood (no change point) evaluated at time period $[a, b]$ with $\theta = \hat{\theta}$.

Step 2: For $i = 2t_h, \ldots, n$, compute and store: (i) $F(t_i) = \min_{t \in SS_{t_i}} F(t) - 2 * \log \ell^*((t, t_i], \hat{\theta}) + (p+1) \log(n)$; (2). $\tau_{t_i}^* = \arg\min_{t \in SS_{t_i}} F(t) - 2 * \log \ell^*((t, t_i], \hat{\theta}) + (p+1) \log(n)$; (3). $SS_{t_{i+1}} = \{0\} \cup \{t \in SS_{t_i} \cup \{t_i - t_h + 1\} : F(t) - 2 * \log \ell^*((t, t_i], \hat{\theta}) \le F(t_i)\}$.

Step 3: Denote $cp(0) = \emptyset$. Then, the optimal change points can be obtained by solving $cp(t^*) = (cp(t^* = \tau_{t^*}^*), \tau_t^*)$ with t^* starts from T and iterates recursively until $\tau_{t^*}^* = 0$.

7. Numerical demonstrations

The Monte-Carlo simulation technique will be used in Subsection 7.1 (i) to evaluate the comparative performance of the two estimation methods, viz. LSSE in (4.2) and MLL in (4.7) to determine the unknown location of m^0 change points assumed to already exist; and (ii) to test the method in (5.2) for detecting the unknown number of change points. In Subsection 7.2, we illustrate the various implementation details of our proposed methods on some observed financial market data.

7.1. Monte-Carlo simulation study

Our simulation considers two different scenarios (or cases). In the first case, we study the performance of the proposed methods under the classical OU process defined by

$$dX_t = (\mu^{(j)} - \alpha^{(j)} X_t)dt + \sigma dW_t, \quad \text{if } s_{j-1}^0 T < t < s_j^0 T, \quad j = 1, \ldots, m. \quad (7.1)$$

In the second case, the performance evaluation of the proposed methods is tested assuming a periodic mean-reverting OU process, with 2-dimensional periodic incomplete set of basis functions $\left\{1, \sqrt{2} \cos\left(\frac{\pi t}{2\Delta_t}\right)\right\}$ (which are orthogonal on $[0,T]$ with weight fixed to 1), given by

$$dX_t = \left[\mu_1^{(j)} + \mu_2^{(j)} \sqrt{2} \cos\left(\frac{\pi t}{2\Delta_t}\right) - \alpha^{(j)} X_t\right] dt + \sigma dW_t, \quad (7.2)$$

where $j = 1, \ldots, m$; $\Delta_t = t_{i+1} - t_i$ is the increment for $[0, T]$; and $s_{j-1}^0 T < t < s_j^0 T$.

Each case consists of 500 iterations. Although an exact solution is available, we choose to use the Euler-Maruyama discretisation scheme to be consistent with the results of Nkurunziza and Zhang (2016). In each iteration, we first generate a desired simulated process based on a given period T with pre-assigned

"true" parameters such as the number and location of change points and the model coefficients. To evaluate the performance of (4.2) and (4.7), we specify the number of change points to be known but the rate is unknown. Then, we estimate and record the change points' arrival rates by applying (4.2) and (4.7) on the simulated process. The detailed simulation setup and results are reported in Subsection 7.1.1.

In Subsection 7.1.2, we also use the Monte-Carlo simulation method to investigate the performance of (5.2). That is, we assume that m^0 is unknown and apply (5.2) with m ranging from 0 to $m^0 + 3$. Then, the m that returns the minimum value of (5.2) is chosen as the estimated value for the number of change points. After 500 iterations, we analyse the performance of the proposed methods based on the recorded results.

7.1.1. Estimating the rate s_j of change points

We first study the performance of (4.2) and (4.7) in estimating the rates of the change points with known m^0. For the simulation setup, we consider the case where $m^0 = 2$ and 3 with different time periods $T = 5, 10, 20$, respectively. The pre-assigned values of the coefficients are provided in Table 1.

For each case, after 500 iterations we record the mean of the estimates based on (4.2) and (4.7), together with the 95% empirical confidence interval (i.e., locating the 2.5 and 97.5 percentiles) and also the mean-squared error. The results are reported in Tables 2–5. In this section, we also report the following

TABLE 1
Pre-assigned coefficients (with known number of change points)

		2 change points $(0.35T, 0.7T)$			3 change points $(0.25T, 0.5T, 0.75T)$			
Case	Coefficient	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 4$
1	$\mu^{(j)}$	0.08	2.50	0.08	0.08	2.50	0.08	2.50
	$\alpha^{(j)}$	0.10	1.00	0.50	0.10	1.00	0.50	1.00
2	$\mu_1^{(j)}$	0.08	2.50	0.08	0.08	2.50	0.08	2.50
	$\mu_2^{(j)}$	0.02	1.20	0.02	0.02	1.20	0.02	1.20
	$\alpha^{(j)}$	0.10	1.00	0.50	0.10	1.00	0.50	1.00

TABLE 2
Simulation results of $\hat{s}_j$, $j = 1, \ldots, m^0$ for case 1, (7.1) with $m^0 = 2$, true values of parameters as in Table 1

	$T=5$			$T=10$		
$\hat{s}_j$	Mean	95% C.I.	MSE	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.348	(0.313, 0.371)	1.75×10^{-4}	0.349	(0.333, 0.363)	9.35×10^{-5}
$\hat{s}_1$, (MLL)	0.348	(0.313, 0.371)	1.76×10^{-4}	0.349	(0.333, 0.363)	9.35×10^{-5}
$\hat{s}_2$, (LSSE)	0.701	(0.638, 0.742)	4.61×10^{-4}	0.702	(0.676, 0.736)	1.76×10^{-4}
$\hat{s}_2$, (MLL)	0.701	(0.638,0.742)	4.61×10^{-4}	0.702	(0.676, 0.736)	1.76×10^{-4}

	$T=20$		
$\hat{s}_j$	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.350	(0.341, 0.356)	3.47×10^{-5}
$\hat{s}_1$, (MLL)	0.350	(0.341, 0.356)	3.75×10^{-5}
$\hat{s}_2$, (LSSE)	0.700	(0.682, 0.716)	5.62×10^{-5}
$\hat{s}_2$, (MLL)	0.700	(0.682, 0.716)	5.19×10^{-5}

TABLE 3

Simulation results of $\hat{s}_j$, $j = 1, \ldots, m^0$ for case 2, (7.2) with $m^0 = 2$, true values of parameters as in Table 1

$\hat{s}_j$	$T=5$			$T=10$		
	Mean	95% C.I.	MSE	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.349	(0.333, 0.362)	6.97×10^{-5}	0.350	(0.341, 0.358)	1.51×10^{-5}
$\hat{s}_1$, (MLL)	0.349	(0.332, 0.362)	0.70×10^{-5}	0.350	(0.342, 0.358)	1.53×10^{-5}
$\hat{s}_2$, (LSSE)	0.701	(0.667, 0.734)	1.97×10^{-4}	0.700	(0.684, 0.718)	6.38×10^{-5}
$\hat{s}_2$, (MLL)	0.701	(0.667, 0.733)	2.01×10^{-4}	0.700	(0.684,0.718)	6.40×10^{-5}

$\hat{s}_j$	$T=20$		
	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.350	(0.344, 0.355)	7.63×10^{-6}
$\hat{s}_1$, (MLL)	0.350	(0.344, 0.355)	7.60×10^{-6}
$\hat{s}_2$, (LSSE)	0.70	(0.690, 0.707)	1.76×10^{-5}
$\hat{s}_2$, (MLL)	0.70	(0.690, 0.707)	1.80×10^{-5}

TABLE 4

Simulation results of $\hat{s}_j$, $j = 1, \ldots, m^0$ for case 1, (7.1) with $m^0 = 3$, true values of parameters as in Table 1

$\hat{s}_j$	$T=5$			$T=10$		
	Mean	95% C.I.	MSE	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.248	(0.217, 0.263)	1.23×10^{-4}	0.249	(0.231, 0.258)	3.95×10^{-5}
$\hat{s}_1$, (MLL)	0.248	(0.217, 0.264)	1.32×10^{-4}	0.249	(0.231, 0.258)	4.14×10^{-5}
$\hat{s}_2$, (LSSE)	0.502	(0.468, 0.549)	3.38×10^{-4}	0.502	(0.477, 0.532)	1.58×10^{-4}
$\hat{s}_2$, (MLL)	0.502	(0.468, 0.549)	3.39×10^{-4}	0.502	(0.477, 0.532)	1.58×10^{-4}
$\hat{s}_3$, (LSSE)	0.752	(0.725, 0.785)	1.74×10^{-4}	0.751	(0.740, 0.770)	4.78×10^{-5}
$\hat{s}_3$, (MLL)	0.752	(0.725, 0.785)	1.73×10^{-4}	0.751	(0.740, 0.770)	4.78×10^{-5}

$\hat{s}_j$	$T=20$		
	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.250	(0.242, 0.256)	2.66×10^{-5}
$\hat{s}_1$, (MLL)	0.250	(0.242, 0.256)	2.18×10^{-5}
$\hat{s}_2$, (LSSE)	0.501	(0.485, 0.522)	6.44×10^{-5}
$\hat{s}_2$, (MLL)	0.501	(0.486, 0.522)	6.23×10^{-5}
$\hat{s}_3$, (LSSE)	0.750	(0.746, 0.755)	5.65×10^{-6}
$\hat{s}_3$, (MLL)	0.750	(0.746, 0.755)	5.65×10^{-6}

TABLE 5

Simulation results of $\hat{s}_j$, $j = 1, \ldots, m^0$ for case 2, (7.2) with $m^0 = 3$, true values of parameters as in Table 1

$\hat{s}_j$	$T=5$			$T=10$		
	Mean	95% C.I.	MSE	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.248	(0.231, 0.262)	6.48×10^{-5}	0.250	(0.240, 0.259)	1.93×10^{-5}
$\hat{s}_1$, (MLL)	0.248	(0.233, 0.263)	6.37×10^{-5}	0.249	(0.242, 0.259)	1.83×10^{-5}
$\hat{s}_2$, (LSSE)	0.502	(0.475, 0.539)	1.94×10^{-4}	0.500	(0.483, 0.515)	5.09×10^{-5}
$\hat{s}_2$, (MLL)	0.502	(0.475, 0.539)	1.94×10^{-4}	0.500	(0.484, 0.517)	5.03×10^{-5}
$\hat{s}_3$, (LSSE)	0.750	(0.727, 0.774)	1.20×10^{-4}	0.750	(0.743, 0.760)	1.39×10^{-5}
$\hat{s}_3$, (MLL)	0.751	(0.727, 0.774)	1.17×10^{-4}	0.750	(0.743, 0.760)	1.42×10^{-5}

$\hat{s}_j$	$T=20$		
	Mean	95% C.I.	MSE
$\hat{s}_1$, (LSSE)	0.250	(0.244, 0.253)	5.29×10^{-6}
$\hat{s}_1$, (MLL)	0.250	(0.244, 0.254)	5.39×10^{-5}
$\hat{s}_2$, (LSSE)	0.499	(0.489, 0.506)	1.72×10^{-5}
$\hat{s}_2$, (MLL)	0.500	(0.489, 0.507)	1.73×10^{-5}
$\hat{s}_3$, (LSSE)	0.750	(0.746, 0.753)	3.20×10^{-6}
$\hat{s}_3$, (MLL)	0.750	(0.746, 0.754)	3.07×10^{-6}

histograms: Figure 1 presents the histograms of the estimated rates based on MLL method for Case 1 with $m^0 = 3$ when T increases from 5 to 20, which shows the behavior of the estimated rates as T increases. Figure 2 shows the histograms of the estimated rates for the two cases under the same conditions (i.e. $m^0 = 3$, $T = 10$ and the rates are estimated by LSSE method). Moreover, for the convenience of the reader, all histograms of the estimated rates for $m^0 = 3$ are provided in E (Figures 8–13) for reference.

From Tables 2–5 (along with Figures 1, 2 and 8–13 in E), we see that in Cases 1 and 2, both proposed methods (4.2) and (4.7) estimate very accurately the exact rates of change points. In particular, the sample means of the estimated change points' arrival rates are close to the exact values, and the results obtained by the 2 proposed methods are very close, which confirms Proposition 4.3. We also clearly observe that as T increases from 5 to 20, the lengths of the 95% empirical confidence intervals and MSEs of the two estimators all decrease. This is well substantiated for example by the pertinent histograms in Figure 1, which shows that when MLL method is employed to estimate the change points' arrival rates in Case 1 with $m^0 = 3$, the central tendencies of the estimated rates are all close to their exact values, and the sample variances decrease as T becomes larger. Similar evidences are shown for other choices of scenarios (different combinations of cases, methods and time period T) as illustrated in Figures 8–13. Although not shown in this paper, the histograms for the case $m^0 = 2$ exhibit similar features. These outcomes confirm the theoretical findings regarding the asymptotic consistency of our two proposed methods.

Also, from Tables 2–5 and Figure 1 (see E as well), the lengths of the 95% confidence intervals (C.I.'s), with corresponding MSEs, of the estimated rates for the first and the last unknown change points ($\hat{s}_1$ and $\hat{s}_3$) are more accurate than those of the middle change point ($\hat{s}_2$). Moreover, the improved accuracy of $\hat{s}_1$ and $\hat{s}_3$ is more sensitive to the increase of T as compared to that of $\hat{s}_2$. This is because the unknown change point's arrival rate s_j satisfies $s_{j-1} < s_j < s_{j+1}$ for $j = 1, \ldots, m$. For s_1 and s_m, one of their boundaries is known (which are $s_0 = 0$ and $s_{m+1} = 1$, respectively), whilst for the intermediate rates both the upper and lower bounds s_{j-1} and s_{j+1} are unknown. Therefore, under the same condition, the uncertainties of the first and the last change points' arrival rates would be lower than that in the middle.

For the comparisons between Case 1 and 2, we can see from the selected scenario ($m^0 = 3$, $T = 10$, LSSE method) shown in Figure 2, along with the results in Tables 2–5 that under the same conditions, the lengths of the 95% empirical C.I. and MSEs in Case 2 are all smaller than those in Case 1. However, since the simulated processes in the 2 cases are generated from different SDEs, it may be inappropriate to make conclusion based only on the results shown in the provided tables and figures. In fact, note that the main difference between (7.1) and (7.2) is the number of coefficients in the models. Therefore, the comparison between the 2 models when fitting them to the same process can be considered as a model selection problem that has been well studied in the literature.
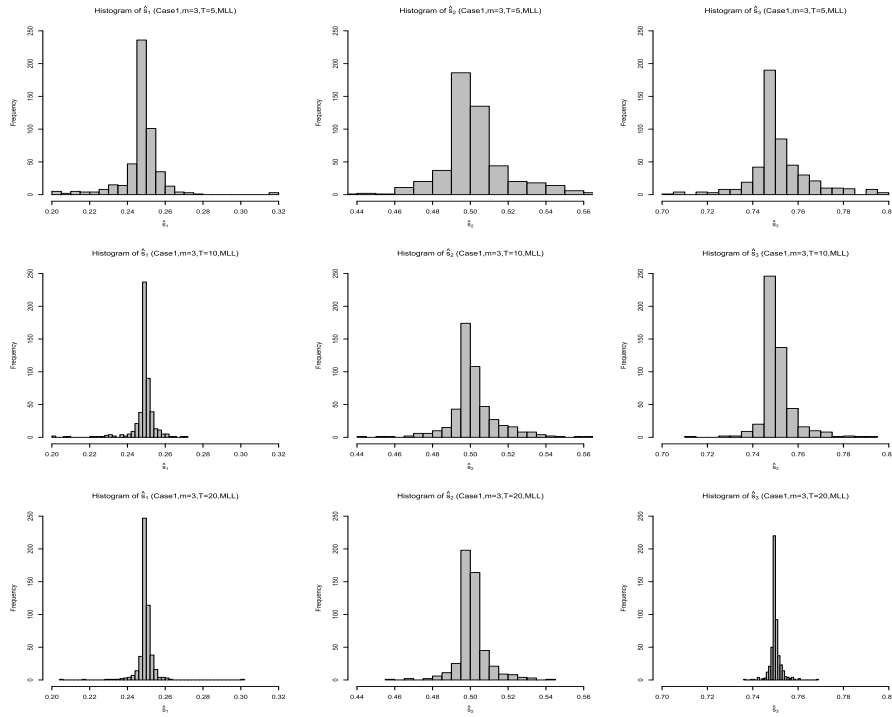
FIGURE 1. *Histogram of $\hat{s}$ based on MLL method for Case 1 with $m^0 = 3$, $T = (5, 10, 20)$ and exact value $s^0 = (0.25, 0.50, 0.75)$*
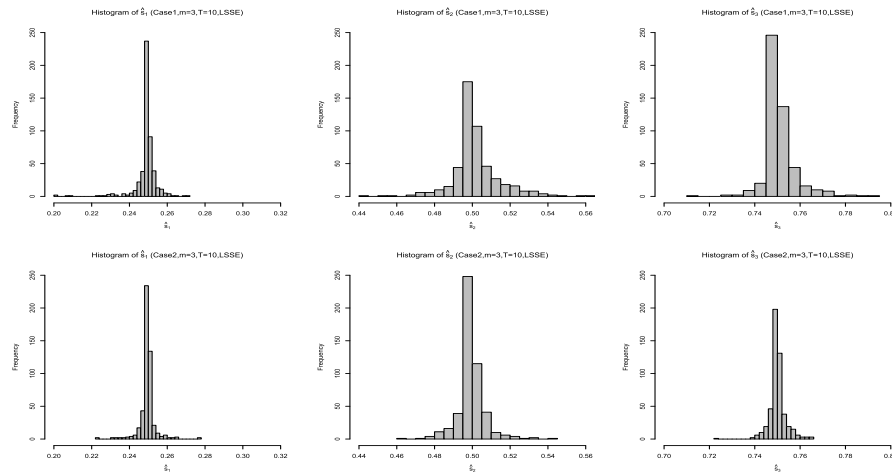

FIGURE 2. *Histogram of $\hat{s}$ based on LSSE method for Case 1 and 2, when $m^0 = 3$, $T = 10$ and exact value $s^0 = (0.25, 0.50, 0.75)$*

7.1.2. Estimating the number of change points

In this subsection, we study the performance of (5.2) in estimating the unknown number of the change points based on Algorithms 2 (SNS) and 3 (Modified PELT). For the simulation setup, we assume the exact value of $m^0 = 2$, with different time periods $T = 5, 10, 15, 20$. The pre-assigned coefficients are provided in Table 6.

Based on the simulated process, we apply Algorithms 2 and 3 with m ranging from 0 to 5 to estimate the unknown number of change points. In Tables 7, we count and report the cumulative frequency (CF) of 500 iterations that return the correct estimates $\left(\sum_{i=1}^{500} \mathbb{1}(\hat{m}_i = m^0)\right)$ and the relative frequency (RF) $\left(\frac{1}{500} \sum_{i=1}^{500} \mathbb{1}(\hat{m}_i = m^0) \times 100\%\right)$.

TABLE 6
Pre-assigned coefficients (with unknown number of change points)

Case	Coefficient	$j = 1$	$j = 2$	$j = 3$
1	$\mu^{(j)}$	0.08	2.50	0.08
	$\alpha^{(j)}$	0.10	1.00	0.50
2	$\mu_1^{(j)}$	0.08	2.50	0.08
	$\mu_2^{(j)}$	0.02	1.20	0.02
	$\alpha^{(j)}$	0.10	1.00	0.50

TABLE 7
Cumulative frequency and relative frequency in 500 iterations that return the correct estimates

Case	Algorithm	$T=5$		$T=10$		$T=15$		$T=20$	
		CF	RF	CF	RF	CF	RF	CF	RF
1	2 (SNS)	492	98.4%	498	99.7%	500	100.0%	500	100.0%
2	2 (SNS)	500	100.0%	500	100.0%	500	100.0%	500	100.0%
1	3 (PELT)	494	98.8%	499	99.8%	500	100.0%	500	100.0%
2	3 (PELT)	497	99.4%	500	100.0%	500	100.0%	500	100.0%

For the estimated number $\hat{m}$ of change points, one could see from Table 7 that, when $m^0 = 2$, the proposed methods perform very well in both cases with different time periods. Furthermore, the accuracy of the estimating results in different cases all increase as T increases. These results suggest that our proposed method is asymptotically consistent, which confirms the theoretical finding in Proposition 5.1.

7.2. Implementation on observed financial market data with discussion

We apply the estimation methods to the Brent oil one-month futures settlement daily price data for the period 18 March 1993 to 25 September 2015. The data set is available at www.quandl.com.

The empirical studies of Schwartz (1997) and Chen (2010) showed that mean-reversion features hold for prices of several commodities including oil. Hence, we
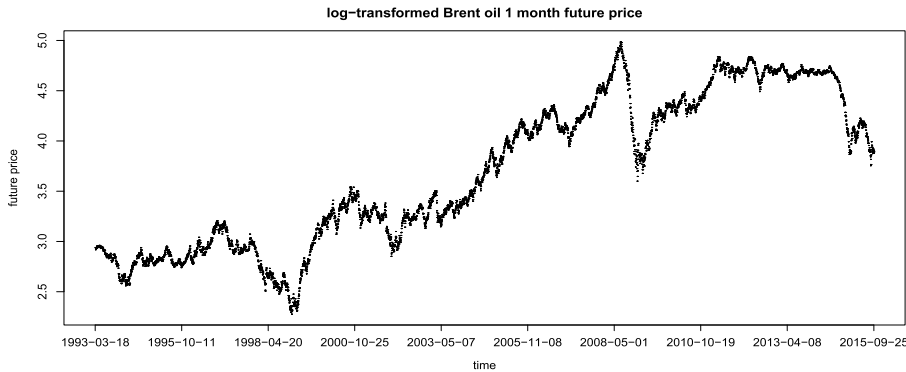
FIGURE 3. *Log-transformed Brent oil 1 month futures settlement prices: 18 March 1993 – 25 September 2015*

use OU-types processes to model such price behaviour. We first fit the classical OU process without any change point, i.e., using the dynamics $dX_t = (\mu - \alpha X_t)dt + \sigma dW_t)$, to the log-transformed data series with σ's estimate as the data's realised volatility $\left(\hat{\sigma} = \sqrt{\sum_{t_i \in [0,T]}(X_{t_{i+1}} - X_{t_i})^2/T}\right)$. The MLE of the drift parameters are given by $\hat{\mu} = 0.48$ and $\hat{\alpha} = 0.12$. Based on these MLE values, the log likelihood (via the Riemann-sum approximation) is about 1.02 and $\mathcal{IC}(m = 0) = 15.27$.

However, from the plot of the price series in Figure 3, there are several changes in the shapes of the price evolution. This observed feature suggests that it may be more appropriate to use (7.1) with unknown $(m > 0)$ change points.

The data set covers approximately 22.5 years giving a sample size of 5735 trading days. The recorded yearly number of trading days varies from year to year; so, for convenience, we let $\Delta_t = 22.5/5735$. Since the size of the data set is large, we apply Algorithm 3 using a minimum permissible regime-time length h of $0.25/\Delta_t \approx 63$ trading days (quarterly). The algorithm detects $\hat{m} = 2$ change points, which occurs on 24 September 2008 and 23 December 2008, with a corresponding log likelihood increase (via Riemann-sum approximation) of 26.81 and $\mathcal{IC}(m = 2) = -1.69$ lower than $\mathcal{IC}(m = 0)$. To confirm the results, we also apply Algorithm 2 with LSSE and MLL methods respectively and $m_{\max}$ set to be 10. The results are the same as that obtained from Algorithm 3.

The plot of the price series, with change points indicated, is depicted in Figure 4. It shows that, from September 2008 to March 2009, there is a decreasing trend in the log-transformed futures prices and then the trend is slightly increasing after this period. Further, based on the estimated change points, the MLE of the drift parameters, and the associated statistics such as (long-term) means (i.e., $\hat{\mu}^{(j)}/\hat{\alpha}^{(j)}$) and variances (i.e., $\hat{\sigma}^2/(2\hat{\alpha}^{(j)})$) are given in Table 8. From Table 8, there are huge changes in the MLEs of the drift parameters under different T's. Based on the MLEs, we also plot two simulated series based on the OU process with and without change points; see Figures 5 and 6, respectively. As
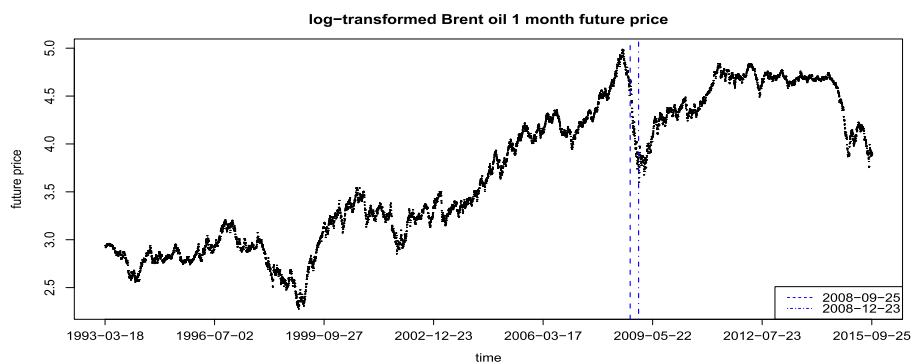
log−transformed Brent oil 1 month future price

FIGURE 4. *Highlighting the two detected change points in the log-transformed one-month futures settlement prices on Brent oil: 18 March 1993–25 September 2015*

most notably expected, the simulated series based on the OU process with two change points is closer to the original series, especially during the period spanning 25 September–23 December 2008, than the simulated series based on the OU process without change point. Judging from these observed characteristics and taking into account the above-mentioned substantial improvements in the log likelihood and SIC values, we conclude that the OU-process with two change points occurring on 25 September 2008 and 23 December 2008 is the appropriate model for the data set that we analysed. Moreover, we see from Figure 4 that from 18 March 1993 to 25 September 2008 (the first estimated change point), there are several noticeable changes in the series. For example, from 1996 to 1998 there was a fall in the futures price. However, based on the SIC, these changes are not significant enough to warrant the inference of a regime change and they are therefore ignored by the proposed methods. We take a closer look concentrating only on the 18-Mar-1993-to-25-Sep-2008 data set to see if there is any change point at all. This analysis is equivalent to reducing the sample size and the penalty term of the SIC accordingly. Algorithm 3 is re-applied to the reduced data set with a sample size 3928 and maintaining the same Δ_t and σ as those in our other experiments. The resulting SIC indicates still no change point during the shortened period. Furthermore, we employ the estimated parameters and sample size in our reduced data set to run a simulation similar to that in Subsection 7.1.2 in assessing the performance of the proposed methods, and we obtained an RF of 86.2% in producing the correct estimates.

TABLE 8

MLEs of the drift parameters, (long term) mean, variances and approximate log likelihoods for the OU process with two change points in modelling the Brent oil's one-month futures settlement prices from 18 March 1993 to 25 September 2015

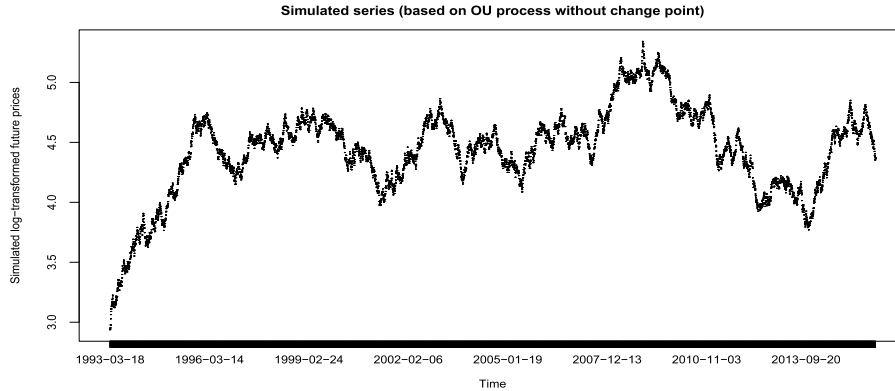| Time period	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\mu}/\hat{\alpha}$	$\hat{\sigma}$	$\hat{\sigma}^2/(2\hat{\alpha})$	$\log \ell$
18 Mar. 1993 to 25 Sept. 2008	0.128	0.005	25.794	0.328	10.889	0.884
26 Sept. 2008 to 23 Dec. 2008	5.501	2.418	2.275	0.328	0.022	23.367
24 Dec. 2008 to 25 Sept. 2015	3.977	0.879	4.524	0.328	0.061	2.557

FIGURE 5. *Simulated log-transformed one-month futures settlement prices on Brent oil based on OU process without change point: 18 March 1993–25 September 2015*
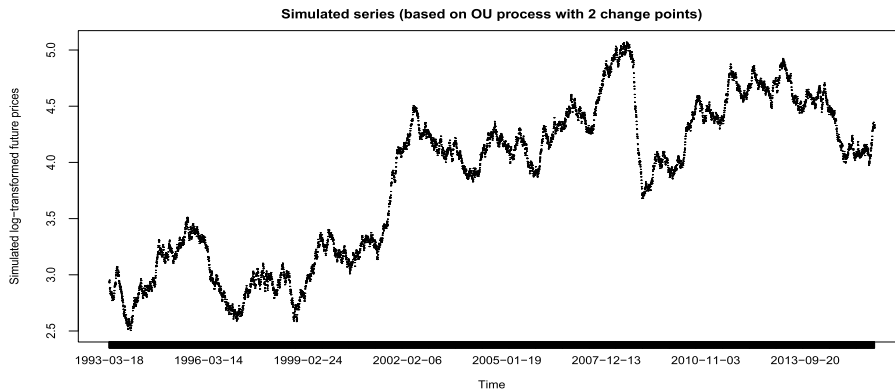

FIGURE 6. *Simulated log-transformed one-month futures settlement prices on Brent oil based on OU process with two change points: 18 March 1993–25 September 2015*

The above result tells us that an OU process without a change point would be appropriate to model the data series from 18 March 1993 to 25 September 2008. However, from Table 8 the respective long-term mean and variance (of values, not returns) $\dfrac{\hat{\mu}}{\hat{\alpha}}$ and $\dfrac{\sigma^2}{2\hat{\alpha}}$ are 25.794 and 10.889, which are both higher than those in the two other time periods. Such high statistics may be less preferable in practice, although they reasonably explain the increasing trend in the investigated time period. On the other hand, we know that imposing more change points, which is equivalent to increasing the number of coefficients in the model, can reduce the variance. Hence, we examine the potential reduction in the variance by imposing a change point into this period. To this end, we fit an OU process with one change point to the series and use Algorithm 1 to estimate the location in the OU process. The estimated change point is at 15 February 1999, which
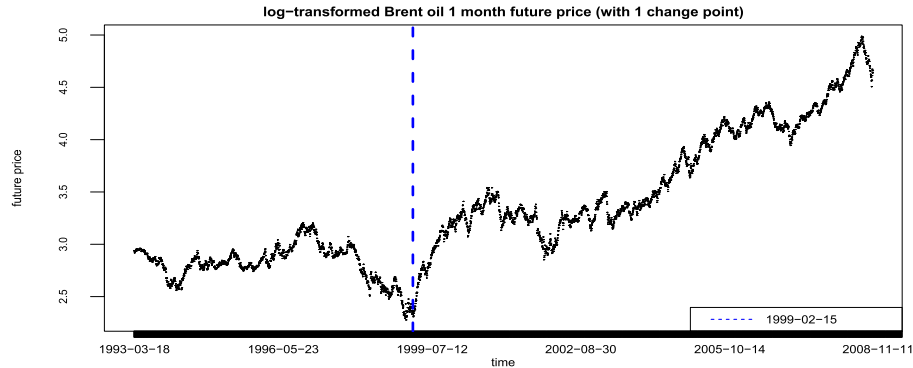
FIGURE 7. *log-transformed one-month futures settlement prices on Brent oil based on OU process with one change point: 18 March 1993–25 September 2008*

TABLE 9

MLEs of the drift parameters, (long-term) mean, variances and approximate log likelihoods for for the OU process with one change point in modeling Brent oil's one-month futures settlement prices from 18 March, 1993 to 25 September, 2008

Time period	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\mu}/\hat{\alpha}$	$\hat{\sigma}$	$\hat{\sigma}^2/(2\hat{\alpha})$	$\log \ell$
18 March 1993 to 15 February 1999	1.688	0.638	2.646	0.328	0.085	0.762
16 February 1999 to 25 September 2008	1.450	0.329	4.407	0.328	0.163	4.199

is near the bottom of the series; see Figure 7. Based on Table 9, by imposing a change point at 15 February 1999, the means $\dfrac{\hat{\mu}}{\hat{\alpha}}$ before and after the change point both strikingly decrease to 2.646 and 4.407, respectively, in comparison to the previous result of 25.794. Additionally, the variances $\dfrac{\sigma^2}{2\hat{\alpha}}$ for the time periods before and after the change point also markedly go down to respective values of 0.085 and 0.163. The approximate log likelihood increases correspondingly to 4.199 for the period 16 February 1999 – 25 September 2008. This implies that imposing a change point (15 February 1999) into the model does improve the accuracy. Nevertheless, recall that our SIC-based method shows no change point at this time period. Therefore, we demonstrated a strong potential for an over-fitting problem to arise when a change-point assumption is unnecessarily introduced into the model. This also reminds us of a trade-off between accuracy improvement and issue of over fitting that must be avoided whenever possible.

8. Conclusion

The main contribution of this paper is the development of MLL- and LSSE-based methods in detecting the unknown number of multiple change points along with the identification of their locations in a generalised univariate OU process. Additionally, we showed that our proposed estimators for the change points' locations satisfy the asymptotically consistent and normality properties under certain suitable conditions that we painstakingly imposed. These results

guided the design of three computing algorithms customised for the efficient implementation of our proposed methods. The numerical applications we showcased covering both simulated and observed data illustrated the excellent performance and accuracy of the estimation approaches that we created to handle change points detection. The usefulness of our results have relevance to regulatory authorities' policy-making, trading strategy's construction by investors and provider's of financial products and services, and other scientific endeavours in the natural and social sciences impacted by sudden and significant changes (e.g., break, jumps, shifts, etc) in the time-series data. This work provides impetus for the investigation and development of methodology suited in tackling further the multiple-change point problem for a multivariate OU process and other closely related modelling challenges in the research literature and practice that entail the statistical inference of stochastic processes.

Appendix A: Proof of Proposition 3.1

Proof of Proposition 3.1. We first need to prove that the coefficients of (1.1) satisfy the space-variable Lipschitz condition. That is,

$$|\mu(t,x) - \mu(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \le K_a|x-y|^2, \quad \text{for some} \quad K_a > 0.$$

Given that $\sigma(t,x)$ is constant in our modelling framework, the second term above is 0. Hence,

$$|\mu(t,x) - \mu(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2$$

$$= \left(\sum_{j=1}^{m+1} \left(S\left(\theta^{(j)}, t, x\right) - S\left(\theta^{(j)}, t, y\right) \right) \mathbb{1}_{\{\tau_{j-1}^0 < t \le \tau_j^0\}} \right)^2$$

$$= \sum_{j=1}^{m+1} (L^{(j)}(t) - a^{(j)}x - L^{(j)}(t) + a^{(j)}y)^2 \mathbb{1}_{\{\tau_{j-1}^0 < t \le \tau_j^0\}}$$

$$= \sum_{j=1}^{m+1} (a^{(j)}(x-y))^2 \mathbb{1}_{\{\tau_{j-1}^0 < t \le \tau_j^0\}} \le \sum_{j=1}^{m+1} (a^{(j)}(x-y))^2 = (x-y)^2 \sum_{j=1}^{m+1} (a^{(j)})^2.$$

Since $a^{(j)} < \infty$ for $j = 1, \ldots, p$, there exists a $K_a > 0$ such that $\left(\sum_{j=1}^{m+1} (a^{(j)})^2 \right) \le K_a$. Then,

$$|\mu(t,x) - \mu(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \le K_a|x-y|^2.$$

Next, we prove the spatial growth condition. That is,

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 \le K_b(1+x^2), \quad \text{for some} \quad K_b > 0.$$

Note that

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 = \left[\sum_{j=1}^{m+1} S\left(\theta^{(j)}, t, x\right) \mathbb{1}_{\{\tau_{j-1}^0 < t \le \tau_j^0\}} \right]^2 + \sigma^2$$

$$= \sum_{j=1}^{m+1} \left(L^{(j)}(t) - a^{(j)}x \right)^2 \mathbb{1}_{\{\tau_{j-1}^0 < t \le \tau_j^0\}} + \sigma^2$$

$$\le \sum_{j=1}^{m+1} \left(L^{(j)}(t) - a^{(j)}x \right)^2 + \sigma^2.$$

Using the identity $(a + b)^2 \le 2a^2 + 2b^2$, we have

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 \le 2 \sum_{j=1}^{m+1} \left(L^{(j)}(t) \right)^2 + \sigma^2 + 2 \sum_{j=1}^{m+1} \left(a^{(j)} \right)^2 x^2.$$

Since $\{\varphi_k(t), \ i = 1, \ldots, p\}$ are bounded, we can find a constant $K_b > 0$ such that

$$\max \left(2 \sum_{j=1}^{m+1} \left(L^{(j)}(t) \right)^2 + \sigma^2, 2 \sum_{j=1}^{m+1} \left(a^{(j)} \right)^2 \right) \le K_b.$$

This gives

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 \le K_b(1 + x^2).$$

Moreover, let $K_c = \max(K_a, K_b)$. Then

$$|\mu(t,x) - \mu(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \le K_c|x - y|^2$$

and

$$|\mu(t,x)|^2 + |\sigma(t,x)|^2 \le K_c(1 + x^2). \qquad \square$$

Appendix B: Proof of the propositions in section 4

Proof of Proposition 4.1. Let $\hat{u}_i$ be the residual of the ith element based on the estimated change points $\{\hat{\tau}_j\}$, $j = 1, \ldots, m + 1$, i.e., $\hat{u}_i = Y_i - \mathbf{z}_i\hat{\boldsymbol{\theta}}_i = \mathbf{z}_i\boldsymbol{\theta}_i - \mathbf{z}_i\hat{\boldsymbol{\theta}}_i + u_i$, for $t_i \in [0, T]$, where $\boldsymbol{\theta}_i$ and $\hat{\boldsymbol{\theta}}_i$ are defined in Section 4 as the true value and MLE of the parameters based on the assigned estimated change points of the coefficients associated with the ith element. Also, let $\hat{u}_i^0$ be the residual of the ith element based on the exact change points $\{\tau_j^0, j = 1, \ldots, m + 1\}$ and $\hat{\boldsymbol{\theta}}_i^0 = \sum_{j=1}^{m+1} \hat{\boldsymbol{\theta}}^{(j,0)} \mathbb{1}(\tau_{j-1}^0 \le t_i \le \tau_j^0)$ with $\hat{\boldsymbol{\theta}}^{(j,0)} = \mathbf{Q}_{(\tau_j^0, \tau_{j-1}^0)}^{-1} \tilde{\mathbf{r}}_{(\tau_j^0, \tau_{j-1}^0)}$.

The proof relies on investigating the behaviour of

$$\frac{1}{T} \left(\sum_{t_i \in [0,T]} \hat{u}_i^2 - \sum_{t_i \in [0,T]} (\hat{u}_i^0)^2 \right). \tag{B.1}$$

By (4.2), (B.1) ≤ 0 with probability 1. Hence, it remains to show that if one of the change points is not consistently estimated, (B.1) > 0 with positive probability yielding a contradiction.

Using the quadratic expansion $(a + b)^2 = a^2 + 2ab + b^2$ and some algebraic computations, (B.1) can be expressed as

$$
(\text{B.1}) = \frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i \right) \right)^2 - \frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^0 \right) \right)^2
$$
$$
+ \frac{2}{T} \sum_{t_i \in [0,T]} \left(u_i \mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^0 \right) \right) - \frac{2}{T} \sum_{t_i \in [0,T]} \left(u_i \mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i \right) \right). \quad (\text{B.2})
$$

We aim to show that if there exists a change point τ_j^0, $j = 1 \ldots, m$, and it is not consistently estimated, the first term $\frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i \right) \right)^2$ is larger than a positive constant with positive probability, whilst the rest of terms is of $o_p(1)$ and hence, (B.1) > 0 with positive probability. To this end, we first provide a lemma, which will be useful in deriving the asymptotic consistency of the estimated change points.

Lemma B.1. *If at least one of the change points, say τ_j^0, can not be consistently estimated, then for large T,*

$$
\frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i \right) \right)^2 \geq C_0 \left\| \boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{(j+1)} \right\|^2 \quad \text{with positive probability.}
$$

Proof. If the change point τ_j^0 is not consistently estimated, then with some positive probability there exists an $\eta > 0$ such that there is no estimated change point in $[\tau_j^0 - \eta T, \tau_j^0 + \eta T]$ for some $\eta > 0$. Without loss of generality, let $\hat{\tau}_{k-1} \leq \tau_j^0 - \eta T \leq \tau_j^0 + \eta T \leq \hat{\tau}_k$. Then, since $(\mathbf{z}_i(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i))^2 \geq 0$ for each i,

$$
\frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i \right) \right)^2
$$
$$
\geq \eta \left(\left(\boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}}^{(k)} \right)^\top \frac{1}{\eta T} \sum_{t_i \in (\tau_j^0 - \eta T, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i \left(\boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}}^{(k)} \right) \right)
$$
$$
+ \eta \left(\left(\boldsymbol{\theta}^{(j+1)} - \hat{\boldsymbol{\theta}}^{(k)} \right)^\top \frac{1}{\eta T} \sum_{t_i \in (\tau_j^0, \tau_j^0 + \eta T]} \mathbf{z}_i^\top \mathbf{z}_i \left(\boldsymbol{\theta}^{(j+1)} - \hat{\boldsymbol{\theta}}^{(k)} \right) \right). \quad (\text{B.3})
$$

Let γ_1 and γ_2 be the smallest eigenvalues of $\frac{1}{\eta T} \sum_{t_i \in (\tau_j^0 - \eta T, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i$ and $\frac{1}{\eta T} \sum_{t_i \in (\tau_j^0, \tau_j^0 + \eta T]} \mathbf{z}_i^\top \mathbf{z}_i$, respectively. Then,

$$
(\text{B.3}) \geq \eta \gamma_1 \left\| \boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}}^{(k)} \right\|^2 + \eta \gamma_2 \left\| \boldsymbol{\theta}^{(j+1)} - \hat{\boldsymbol{\theta}}^{(k)} \right\|^2
$$
$$
\geq \eta \min(\gamma_1, \gamma_2) \left(\left\| \boldsymbol{\theta}^{(j)} - \hat{\boldsymbol{\theta}}^{(k)} \right\|^2 + \left\| \boldsymbol{\theta}^{(j+1)} - \hat{\boldsymbol{\theta}}^{(k)} \right\|^2 \right).
$$

Using the convexity of a quadratic function, we have

$$\left\| \theta^{(j)} - \hat{\theta}^{(k)} \right\|^2 + \left\| \theta^{(j+1)} - \hat{\theta}^{(k)} \right\|^2 \geq \frac{1}{2} \left\| \theta^{(j)} - \theta^{(j+1)} \right\|^2.$$

Hence,

$$\frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\theta_i - \hat{\theta}_i \right) \right)^2 \geq \eta \frac{\min(\gamma_1, \gamma_2)}{2} \left\| \theta^{(j)} - \theta^{(j+1)} \right\|^2.$$

Under Assumption 4, γ_1 and γ_2 are both bounded away from 0 and $\min(\gamma_1, \gamma_2)$ is also bounded away from 0. Therefore, the right-hand side of the above inequality is positive. Then, the proof is complete by letting $C_0 = \eta \frac{\min(\gamma_1, \gamma_2)}{2}$. □

Lemma B.2. *Under Assumptions 1–3,* $\frac{1}{T} \mathbf{Q}_{(\hat{s}_{j-1}T, \hat{s}_j T)} \xrightarrow[T \to \infty]{a.s.} (\hat{s}_j - \hat{s}_{j-1}) \boldsymbol{\Sigma}_j$ *for* $s^0_{j-1} \leq \hat{s}_{j-1} < \hat{s}_j \leq s^0_j$.

For the proof of Lemma B.2, one may refer to Proposition 6.2 in Nkurunziza and Zhang (2016). To emphasise again, both lemmas B.1 and B.2 are key in proving the asymptotic properties of the proposed estimators.

Next, note that for $t_i \in (\tau^0_{j-1}, \tau^0_j]$, $j = 1, \ldots, m$, we have $\theta_i = \theta^{(j)}$ and $\hat{\theta}^0_i = \theta^{(j)} + \mathbf{Q}^{-1}_{(\tau^0_{j-1}, \tau^0_j)} \sigma \mathbf{r}_{(\tau^0_{j-1}, \tau^0_j)}$. Substituting them into $\frac{1}{T} \sum_{t_i \in [0,T]} \left(\mathbf{z}_i \left(\theta_i - \hat{\theta}^0_i \right) \right)^2$, we get

$$\sum_{j=1}^{m} \frac{1}{T} \mathbf{r}^\top_{(\tau^0_{j-1}, \tau^0_j)} \mathbf{Q}^{-1}_{(\tau^0_{j-1}, \tau^0_j)} \sum_{t_i \in (\tau^0_{j-1}, \tau^0_j]} \mathbf{z}^\top_i \mathbf{z}_i \mathbf{Q}^{-1}_{(\tau^0_{j-1}, \tau^0_j)} \mathbf{r}_{(\tau^0_{j-1}, \tau^0_j)}. \tag{B.4}$$

To proceed further, we first prove the following inequality. Suppose $0 < \tau^*_1 < \tau^*_2 \leq T$. By the Markov inequality, Itô's isometry and (3.8), we have

$$\mathrm{P} \left(\frac{| \int_{\tau^*_1}^{\tau^*_2} X_t dW_t |}{\sqrt{\tau^*_2 - \tau^*_1}} > K^* \right) \leq \frac{\mathrm{E} \left(| \int_{\tau^*_1}^{\tau^*_2} X_t dW_t |^2 \right)}{(\tau^*_2 - \tau^*_1)(K^*)^2} = \frac{\int_{\tau^*_1}^{\tau^*_2} E(X^2_t) dt}{(\tau^*_2 - \tau^*_1)(K^*)^2}$$

$$\leq \frac{K_1(\tau^*_2 - \tau^*_1)}{(\tau^*_2 - \tau^*_1)(K^*)^2} = \frac{K_1}{(K^*)^2}. \tag{B.5}$$

Therefore, by letting $K^* = (\log T)^{a^*}$, for some $0 < a^* < 1/2$, the above probability tends to 0 as T tends to infinity. This implies that for some $0 < a^* < 1/2$,

$$\frac{1}{\sqrt{\tau^*_2 - \tau^*_1}} \left\| \mathbf{r}_{(\tau^*_1, \tau^*_2)} \right\| = O_p((\log T)^{a^*}) \quad \text{for any} \quad 0 < \tau^*_1 < \tau^*_2 \leq T. \tag{B.6}$$

Now, continuing the proof of Proposition 4.1, we note that under Assumption 3, $\sup_{0 \leq t \leq T} |\varphi_k(t)| \leq K_{\varphi_k} < \infty$ for $i = 1, \ldots, p$. By similar argument used to

obtain (B.6), we have $\frac{1}{\sqrt{\tau_2^* - \tau_1^*}} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \varphi_k(t) dW_t = O_p\left((\log T)^{a^*}\right)$ for $i = 1, \ldots, p$

so that $\left\| \frac{1}{\sqrt{\tau_2^* - \tau_1^*}} \mathbf{r}_{(\hat{\tau}_{j-1}, \hat{\tau}_j)} \right\| = O_p\left((\log T)^{a^*}\right)$. Moreover, since $\sum_{t_i \in (\tau_1^*, \tau_2^*]} u_i \mathbf{z}_i$ is

the discretised versions of $\mathbf{r}_{(\tau_1^*, \tau_2^*)}$, it follows that $\left\| \frac{1}{\sqrt{\tau_2^* - \tau_1^*}} \sum_{t_i \in (\tau_1^*, \tau_2^*]} u_i \mathbf{z}_i \right\| =$

$O_p\left((\log T)^{a^*}\right)$. Also, $\sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i$ is the discretised version of $\mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}$,

thus the asymptotic results in (3.13) and (3.15) also hold for $\sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i$.
Therefore, by the Cauchy-Schwarz inequality, (3.13) and (3.15) we have, after
some algebraic manipulations,

$$(\text{B.4}) \leq \sum_{j=1}^m \frac{1}{T} \left\| \frac{1}{\sqrt{(s_j^0 - s_{j-1}^0)T}} \mathbf{r}_{(\tau_{j-1}^0, \tau_j^0)} \right\|^2 \left\| (s_j^0 - s_{j-1}^0) T \mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}^{-1} \right\|^2$$

$$\times \left\| \frac{1}{(s_j^0 - s_{j-1}^0)T} \sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \mathbf{z}_i^\top \mathbf{z}_i \right\| = o_p(1). \tag{B.7}$$

It remains to investigate the quantity

$$\frac{2}{T} \sum_{t_i \in [0,T]} (u_i \mathbf{z}_i (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^0)) - \frac{2}{T} \sum_{t_i \in [0,T]} (u_i \mathbf{z}_i (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)) = \frac{2}{T} \sum_{t_i \in [0,T]} (u_i \mathbf{z}_i (\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_i^0)).$$
$$\tag{B.8}$$

Note that the structure of $\hat{\boldsymbol{\theta}}_i$ is affected by the location of the estimated
change points. It is, therefore, difficult to substitute the expressions for $\hat{\boldsymbol{\theta}}_i$ into
(B.8) directly. Without loss of generality, we consider $m = 2$ (but the procedure
can be extended to the general case $m > 0$), and assume that $0 = \tau_0 < \hat{\tau}_1 <$
$\tau_1^0 < \tau_2^0 < \hat{\tau}_2 < \tau_3^0 = T$. Other cases can be analysed in a similar manner. With
$m = 2$, (B.8) reduces to

$$(\text{B.8}) = \frac{2}{T} \sum_{t_i \in (0, \hat{\tau}_1]} \left(u_i \mathbf{z}_i (\hat{\boldsymbol{\theta}}^{(1,0)} - \hat{\boldsymbol{\theta}}^{(1)}) \right) + \frac{2}{T} \sum_{t_i \in (\hat{\tau}_1, \tau_1^0]} \left(u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(1,0)} - \hat{\boldsymbol{\theta}}^{(2)} \right) \right)$$

$$+ \frac{2}{T} \sum_{t_i \in (\tau_1^0, \tau_2^0]} \left(u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(2,0)} - \hat{\boldsymbol{\theta}}^{(2)} \right) \right) + \frac{2}{T} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \left(u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \hat{\boldsymbol{\theta}}^{(2)} \right) \right)$$

$$+ \frac{2}{T} \sum_{t_i \in (\hat{\tau}_2, T]} \left(u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \hat{\boldsymbol{\theta}}^{(3)} \right) \right), \tag{B.9}$$

where

$$\hat{\boldsymbol{\theta}}^{(1)} = \boldsymbol{\theta}^{(1)} + \sigma \mathbf{Q}_{(0,\hat{\tau}_1)}^{-1} \mathbf{r}_{(0,\hat{\tau}_1)},$$

$$\hat{\boldsymbol{\theta}}^{(2)} = \mathbf{Q}_{(\hat{\tau}_1,\hat{\tau}_2)}^{-1} (\mathbf{Q}_{(\hat{\tau}_1,\tau_1^0)} \boldsymbol{\theta}^{(1)} + \mathbf{Q}_{(\tau_1^0,\tau_2^0)} \boldsymbol{\theta}^{(2)} + \mathbf{Q}_{(\tau_2^0,\hat{\tau}_2)} \boldsymbol{\theta}^{(3)} + \sigma \mathbf{r}_{(\hat{\tau}_1,\hat{\tau}_2)}),$$

$$\hat{\boldsymbol{\theta}}^{(3)} = \mathbf{Q}_{(\hat{\tau}_2,T)}^{-1} (\mathbf{Q}_{(\hat{\tau}_2,T)} \boldsymbol{\theta}^{(3)} + \sigma \mathbf{r}_{(\hat{\tau}_2,T)}),$$

$$\hat{\boldsymbol{\theta}}^{(1,0)} = \mathbf{Q}_{(0,\tau_1^0)}^{-1}(\mathbf{Q}_{(0,\tau_1^0)}\boldsymbol{\theta}^{(1)} + \sigma\mathbf{r}_{(0,\tau_1^0)}),$$

$$\hat{\boldsymbol{\theta}}^{(2,0)} = \mathbf{Q}_{(\tau_1^0,\tau_2^0)}^{-1}(\mathbf{Q}_{(\tau_1^0,\tau_2^0)}\boldsymbol{\theta}^{(2)} + \sigma\mathbf{r}_{(\tau_1^0,\tau_2^0)}), \text{ and}$$

$$\hat{\boldsymbol{\theta}}^{(3,0)} = \mathbf{Q}_{(\tau_2^0,T)}^{-1}(\mathbf{Q}_{(\tau_2^0,T)}\boldsymbol{\theta}^{(2)} + \sigma\mathbf{r}_{(\tau_2^0,T)}).$$

Using the Cauchy-Schwarz inequality, (B.6), Lemma B.2 along with the Continuous Mapping Theorem, the first term in (B.8) is bounded above, i.e.,

$$\frac{2}{T} \sum_{t_i \in [0,\hat{\tau}_1]} (u_i \mathbf{z}_i(\hat{\boldsymbol{\theta}}^{(1,0)} - \hat{\boldsymbol{\theta}}^{(1)}))$$

$$\leq \frac{2\sqrt{\hat{s}_1}}{\sqrt{s_1^0 T}} \left\| \frac{1}{\sqrt{\hat{s}_1 T}} \sum_{t_i \in [0,\hat{\tau}_1]} u_i \mathbf{z}_i \right\| \left\| s_1^0 T \mathbf{Q}_{(0,\tau_1^0)}^{-1} \right\| \left\| \frac{1}{\sqrt{s_1^0 T}} \mathbf{r}_{(0,\tau_1^0)} \right\|$$

$$- \frac{2}{T} \left\| \frac{1}{\sqrt{\hat{s}_1 T}} \sum_{t_i \in [0,\hat{\tau}_1]} u_i \mathbf{z}_i \right\| \left\| \hat{s}_1 T \mathbf{Q}_{(0,\hat{\tau}_1)}^{-1} \right\| \left\| \frac{1}{\sqrt{\hat{s}_1 T}} \mathbf{r}_{(0,\hat{\tau}_1)} \right\| = o_p(1).$$

Similarly, one can show that the rest of the terms in (B.8) are all of $o_p(1)$. Following these arguments, it can be shown that in the general case when $m > 0$, the terms in (B.8) are all of order $o_p(1)$. So, if one of the change points' arrival rates, say s_j^0, is not consistently estimated, (B.1) is dominated by the first term, which is larger than 0 with positive probability. This gives a contradiction. Therefore, $\hat{s}_j - s_j^0 \xrightarrow[T\to\infty]{p} 0$ for every $j = 1, \ldots, m$.   □

Proof of Proposition 4.2. Without loss of generality, assume that $m = 3$. We provide an explicit proof dealing with the rate T-consistency for $\hat{\tau}_2$ only. The consistency analysis for $\hat{\tau}_1$ and $\hat{\tau}_3$ can be similarly completed. By Proposition 4.1, $\hat{\tau}_j \in \{\tau : |\tau_j - \tau_j^0| \leq \eta T, 1 \leq j \leq m\}$ for each $\eta > 0$. For $C > 0$, define the set $V_\eta(C) := \{\tau : |\tau_j - \tau_j^0| \leq \eta T, 1 \leq j \leq m; \ |\tau_2 - \tau_2^0| > C\}$.

Let $SSE_1 = SSE(\tau_1, \tau_2, \tau_3) = \sum_{t_i \in [0,T]} (Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i)^2$, where $\hat{\boldsymbol{\theta}}_i$ is the MLE of $\boldsymbol{\theta}$ associated with the ith element under the change point (τ_1, τ_2, τ_3), and let $SSE_2 = SSE(\tau_1, \tau_2^0, \tau_3)$. So, we have $\min_{(\tau_1,\tau_2,\tau_3)}(SSE_1 - SSE_2) \leq 0$ with probability 1. If we can show that for each $\epsilon > 0$, there exists a $C > 0, 0 < \eta < 1$ such that for large T and any $\tau \in V_\eta(C)$, $P(\min_{\tau \in V_\eta(C)}(SSE_1 - SSE_2) > 0) < \epsilon$, this would imply that for some $C > 0$, the global optimisation can not be achieved on the set $V_\eta(C)$. Thus with large probability, $|\hat{\boldsymbol{\tau}} - \tau^0| \leq C$.

Let $\hat{\boldsymbol{\tau}} = \arg\min_{\tau \in V_\eta(C)}(SSE_1 - SSE_2)$. Assume, without loss of generality, that $\hat{\tau}_1 \leq \tau_1^0 < \tau_2^0 < \hat{\tau}_2 < \hat{\tau}_3 \leq \tau_3^0$. We now focus on the behaviour of

$$(SSE_1 - SSE_2)/(\hat{\tau}_2 - \tau_2^0). \tag{B.10}$$

Applying the identity $(a + b)^2 = a^2 + 2ab + b^2$ into (B.10) and then breaking the time period $[0, T]$ into different intervals, we have

$$SSE_1 - SSE_2$$

$$= -\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2,0)}\right)\right)^2 - \sum_{t_i \in (\tau_1^0, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \hat{\boldsymbol{\theta}}^{(2,0)}\right)\right)^2$$

$$- \sum_{t_i \in (\tau_2^0, \hat{\tau}_3]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(3)} - \hat{\boldsymbol{\theta}}^{(3,0)}\right)\right)^2 \tag{B.11}$$

$$+ \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(2,0)} - \hat{\boldsymbol{\theta}}^{(2)}\right) + \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} 2u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \hat{\boldsymbol{\theta}}^{(2)}\right)$$

$$+ \sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} 2u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \hat{\boldsymbol{\theta}}^{(3)}\right) \tag{B.12}$$

$$+ \sum_{t_i \in (\hat{\tau}_1, \tau_1^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}\right)\right)^2 + \sum_{t_i \in (\tau_1^0, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \hat{\boldsymbol{\theta}}^{(2)}\right)\right)^2$$

$$+ \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(3)} - \hat{\boldsymbol{\theta}}^{(2)}\right)\right)^2 + \sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(3)} - \hat{\boldsymbol{\theta}}^{(3)}\right)\right)^2 \tag{B.13}$$

where

$$\hat{\boldsymbol{\theta}}^{(2)} = \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \left(\mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \boldsymbol{\theta}^{(1)} + \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \boldsymbol{\theta}^{(2)} + \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \boldsymbol{\theta}^{(3)} + \sigma \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}\right),$$

$$\hat{\boldsymbol{\theta}}^{(3)} = \mathbf{Q}_{(\hat{\tau}_2, \hat{\tau}_3)}^{-1} \left(\mathbf{Q}_{(\hat{\tau}_2, \hat{\tau}_3)} \boldsymbol{\theta}^{(3)} + \sigma \mathbf{r}_{(\hat{\tau}_2, \hat{\tau}_3)}\right),$$

$$\hat{\boldsymbol{\theta}}^{(2,0)} = \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \left(\mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \boldsymbol{\theta}^{(1)} + \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \boldsymbol{\theta}^{(2)} + \sigma \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}\right), \text{ and}$$

$$\hat{\boldsymbol{\theta}}^{(3,0)} = \mathbf{Q}_{(\tau_2^0, \hat{\tau}_3)}^{-1} \left(\mathbf{Q}_{(\tau_2^0, \hat{\tau}_3)} \boldsymbol{\theta}^{(3)} + \sigma \mathbf{r}_{(\tau_2^0, \hat{\tau}_3)}\right).$$

It remains to show that for large T, $\dfrac{\text{(B.13)} + \text{(B.11)} + \text{(B.12)}}{\hat{\tau}_2 - \tau_2^0}$ is positive with probability 1. Note that this depends on the choice of $\hat{\tau}_2 \in V_\eta(C)$; (B.11)/$(\hat{\tau}_2 - \tau_2^0)$ can be $O_p(1)$ instead of $o_p(1)$, and thus the arguments we used in the proof of Proposition 4.1 cannot be applied here directly. To overcome this problem, we need to expand each term in (B.13). We note that $\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}$ is equal to

$$\mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \left(\mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\right) + \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)\right) + \sigma \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}.$$

So,

$$\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \left(\mathbf{z}_i (\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)})\right)^2$$

$$= \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\right)$$

$$+ 2 \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ 2 \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right)$$

$$+ 2\sigma \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$+ 2\sigma \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right) \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$+ \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}^{\top} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)},$$

and

$$\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(1)} - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(2,0)} \right) \right)^2$$

$$= \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

$$+ 2\sigma \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}$$

$$+ \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}^{\top} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}.$$

By Tobing and McGilchrist (1992), we have

$$\mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} = \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} + O_p \left(\frac{\hat{\tau}_2 - \tau_2^0}{T^2} \right). \tag{B.14}$$

Hence,

$$\left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

$$- \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

$$= 2 O_p \left(\frac{\hat{\tau}_2 - \tau_2^0}{T^2} \right) \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

$$+ O_p \left(\frac{(\hat{\tau}_2 - \tau_2^0)^2}{T^4} \right) \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right). \tag{B.15}$$

Using the asymptotic results in Proposition 4.1, we have that $\frac{\tau_1^0 - \hat{\tau}_1}{T} \xrightarrow[T \to \infty]{p} 0$. Therefore,

$$\frac{2}{\hat{\tau}_2 - \tau_2^0} O_p \left(\frac{\hat{\tau}_2 - \tau_2^0}{T^2} \right) \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^{\top} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i' \mathbf{z}_i \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

$$\leq 2 O_p \left(\frac{\hat{\tau}_2 - \tau_2^0}{T^2} \right) \frac{(\tau_2^0 - \tau_1^0)^2 (\tau_1^0 - \hat{\tau}_1)}{(\tau_2^0 - \hat{\tau}_1)} \left\| \boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right\|^2 \left\| \frac{1}{(\tau_2^0 - \tau_1^0)} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \right\|^2$$

$$\times \left\| (\tau_2^0 - \hat{\tau}_1) \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \right\| \frac{1}{(\tau_1^0 - \hat{\tau}_1)} \left\| \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i' \mathbf{z}_i \right\| = o_p(1).$$

Similarly,

$$\frac{1}{\hat{\tau}_2 - \tau_2^0} O_p \left(\frac{(\hat{\tau}_2 - \tau_2^0)^2}{T^4} \right) \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)$$

is of order $o_p(1)$. Thus, (B.15)/$(\hat{\tau}_2 - \tau_2^0) = o_p(1)$. Furthermore, using the same argument above together with (B.6), we have

$$\frac{2\sigma}{\hat{\tau}_2 - \tau_2^0} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$- \frac{2\sigma}{\hat{\tau}_2 - \tau_2^0} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}$$

$$= O_p \left(\frac{1}{T^2} \right) 4\sigma \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}$$

$$+ O_p \left(\frac{(\hat{\tau}_2 - \tau_2^0)}{T^4} \right) 2\sigma \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)} \right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i' \mathbf{z}_i \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)} = o_p(1).$$

With the aid of the identity $\mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)} = \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)} + \mathbf{r}_{(\tau_2^0, \hat{\tau}_2)}$, (B.14) and the Cauchy-Schwarz inequality, we get

$$\frac{\sigma^2}{\hat{\tau}_2 - \tau_2^0} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}^\top \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$- \frac{\sigma^2}{\hat{\tau}_2 - \tau_2^0} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)}^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)} = o_p(1)$$

and the following two terms are of order $o_p(1)$, respectively:

$$\frac{1}{\hat{\tau}_2 - \tau_2^0} 2 \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right)^\top \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right),$$

$$\frac{2\sigma}{\hat{\tau}_2 - \tau_2^0} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right) \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)},$$

Combining the above results, we have

$$\frac{\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)} \right) \right)^2 - \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(1)} - \hat{\boldsymbol{\theta}}^{(2,0)} \right) \right)^2}{\hat{\tau}_2 - \tau_2^0} = o_p(1).$$

Similarly,

$$\sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} (\mathbf{z}_i(\theta^{(3)} - \hat{\theta}^{(3)}))^2 - \sum_{t_i \in (\tau_2^0, \hat{\tau}_3]} (\mathbf{z}_i(\theta^{(3)} - \hat{\theta}^{(3,0)}))^2$$

$$= \sigma^2 \mathbf{r}_{(\hat{\tau}_2, \hat{\tau}_3)}^\top \mathbf{Q}_{(\hat{\tau}_2, \hat{\tau}_3)}^{-1} \sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_2, \hat{\tau}_3)}^{-1} \mathbf{r}_{(\hat{\tau}_2, \hat{\tau}_3)}$$

$$- \sigma^2 \mathbf{r}_{(\tau_2^0, \hat{\tau}_3)}^\top \mathbf{Q}_{(\tau_2^0, \hat{\tau}_3)}^{-1} \sum_{t_i \in (\tau_2^0, \hat{\tau}_3]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\tau_2^0, \hat{\tau}_3)}^{-1} \mathbf{r}_{(\tau_2^0, \hat{\tau}_3)}. \qquad \text{(B.16)}$$

Note that $\mathbf{r}_{(\tau_2^0, \hat{\tau}_3)} = \mathbf{r}_{(\tau_2^0, \hat{\tau}_2)} + \mathbf{r}_{(\hat{\tau}_2, \hat{\tau}_3)}$. So, from (B.5), $\mathbf{r}_{(\tau_2^0, \hat{\tau}_2)}/\sqrt{\hat{\tau}_2 - \tau_2^0} = O_p\left((\hat{\tau}_2 - \tau_2^0)^{a^*}\right)$ for some $0 < a^* < 1/2$. Moreover,

$$\sum_{t_i \in (\tau_2^0, \hat{\tau}_3]} \mathbf{z}_i^\top \mathbf{z}_i = \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i + \sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} \mathbf{z}_i^\top \mathbf{z}_i,$$

and by Tobing and McGilchrist (1992), $\mathbf{Q}_{(\tau_2^0, \hat{\tau}_3)}^{-1} = \mathbf{Q}_{(\hat{\tau}_2, \hat{\tau}_3)}^{-1} + O_p\left(\frac{\hat{\tau}_2 - \tau_2^0}{T^2}\right)$. Again, when these results are combined, one may verify that, with a suitable choice of C, for large T the order of $\frac{\text{(B.16)}}{(\hat{\tau}_2 - \tau_2^0)}$ is $o_p(1)$.

The above procedure can also be applied to investigate the behaviour of (B.12). For example,

$$\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2 u_i \mathbf{z}_i \left(\theta^{(2,0)} - \mathbf{z}_i \hat{\theta}^{(2)}\right)$$

$$= \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2 u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\theta^{(1)} - \theta^{(2)}\right)$$

$$- \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2 u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\theta^{(1)} - \theta^{(2)}\right)$$

$$+ \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2 u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\theta^{(1)} - \theta^{(3)}\right)$$

$$+ 2\sigma \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$- 2\sigma \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)},$$

with

$$\frac{2}{\hat{\tau}_2 - \tau_2^0} \left(\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\theta^{(1)} - \theta^{(2)}\right) \right.$$

$$\left. - \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\theta^{(1)} - \theta^{(2)}\right) \right)$$

$$= O_p\left(\frac{1}{T^2}\right) \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}\right) = o_p(1),$$

$$\frac{2}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)$$

$$\leq \left\| \frac{1}{T} \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \right\| \left\| T \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \right\| \left\| \frac{2}{\hat{\tau}_2 - \tau_2^0} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \right\| \left\| \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right) \right\| = o_p(1)$$

and

$$\frac{2\sigma}{\hat{\tau}_2 - \tau_2^0} \left(\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)} - \sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} u_i \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \tau_2^0)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \tau_2^0)} \right) = o_p(1).$$

Hence, for large T,

$$\sum_{t_i \in (\hat{\tau}_1, \tau_2^0]} 2 u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(2,0)} - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(2)} \right) / (\hat{\tau}_2 - \tau_2^0) = o_p(1).$$

In an analogous manner, one can show that, with a suitable choice of C, for large T,

$$\frac{\sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} 2 u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(2)} \right) \hat{\tau}_2 - \tau_2^0 + \sum_{t_i \in (\hat{\tau}_2, \hat{\tau}_3]} 2 u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}^{(3,0)} - \hat{\boldsymbol{\theta}}^{(3)} \right)}{\hat{\tau}_2 - \tau_2^0}$$

is of order $o_p(1)$. Finally, we investigate the behaviour of the remaining term.

$$\frac{\sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \left(\mathbf{z}_i \left(\boldsymbol{\theta}^{(3)} - \hat{\boldsymbol{\theta}}^{(2)} \right) \right)^2}{\hat{\tau}_2 - \tau_2^0}$$

$$= \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ 2 \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ 2 \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$+ 2 \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

$$+ \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}^\top \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i' \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}.$$

By Proposition 4.1, $s_1^0 - \hat{s}_1 \xrightarrow[T \to \infty]{p} 0$. Hence, applying again the Cauchy-Schwarz inequality and similar reasoning as before, we have

$$\left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)$$

$$\leq (s_1^0 - \hat{s}_1)^2 \left\| \boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right\| \left\| T\mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \right\|^2 \left\| \frac{1}{(s_1^0 - \hat{s}_1)T} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \right\|^2 \left\| \boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)} \right\|$$

$$\times \left\| \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \right\| = o_p(1).$$

Also, the following terms

$$2\left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right),$$

$$2\left(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)},$$

$$2\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)},$$

and

$$\mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}^\top \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{r}_{(\hat{\tau}_1, \hat{\tau}_2)}$$

are all of order $o_p(1)$, respectively.

Moreover, it follows from the identity $\mathbf{Q}_{(\tau_1^0, \tau_2^0)} = \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)} - \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} - \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)}$ that

$$\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_1^0, \tau_2^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$= \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$+ 2\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$- 2\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)^\top \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^\top \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}\right)$$

$$- 2 \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right),$$

with the following terms

$$\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right),$$

$$\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right),$$

$$2 \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right),$$

$$2 \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\tau_2^0, \hat{\tau}_2)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right),$$

$$2 \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)^{\top} \mathbf{Q}_{(\hat{\tau}_1, \tau_1^0)} \mathbf{Q}_{(\hat{\tau}_1, \hat{\tau}_2)}^{-1} \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)$$

are all of order $o_p(1)$. Now, suppose γ_4 be the smallest eigenvalue of $\frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i$. With a suitable C, γ_4 is bounded away from 0. Hence,

$$\left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right)' \frac{1}{\hat{\tau}_2 - \tau_2^0} \sum_{t_i \in (\tau_2^0, \hat{\tau}_2]} \mathbf{z}_i^{\top} \mathbf{z}_i \left(\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)} \right) \geq \quad \gamma 4 ||\boldsymbol{\theta}^{(2)} - \boldsymbol{\theta}^{(3)}||^2 > 0$$

with probability 1 and this dominates the rest of the terms in (B.10) when T is large. This implies that (B.10) is positive with probability 1, which gives a contradiction and it indicates that with large probability $\hat{\tau}_2$ cannot be in the set $V_\eta(C)$. □

Proof of Corollary 4.1. Part (i) of Corollary 4.1 follows from the same arguments utilised in the proof of Proposition 4.1 with the term $\frac{1}{T}$ in (B.1) replaced by $\frac{1}{T^{2r^*}}$, together with the fact that $\frac{T}{T^{2r^*}}||\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}||^2 = (T^{1-2r^*}v_T^2||\mathbf{M}||^2)$ $\xrightarrow[T\to\infty]{} \infty$ and $\log T/T^{2r^*} \xrightarrow[T\to\infty]{} 0$. On the other hand, part (ii) may be verified by employing similar arguments as in the proof of Proposition 4.2 in investigating the set $V_\eta(C, v_T) = \{\tau : C/v_T^2 < |\tau - \tau^0| < \eta T\}$ instead of $V_\eta(C)$. □

Proof of Proposition 4.3. To examine the behaviour of (4.7), we first define $Y_i^* = X_{t_{i+1}^*} - X_{t_i^*}$ and $\mathbf{z}_i^* = V(t = t_i^*)\Delta_t^*$. Then,

$$\log \ell^*(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau})) = \sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \frac{\hat{\boldsymbol{\theta}}^{(j)\top} \mathbf{z}_i^{*\top} Y_i^*}{\Delta_t^* \sigma^2} - \frac{1}{2\Delta_t^* \sigma^2} \sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \left(\hat{\boldsymbol{\theta}}^{(j)'} \mathbf{z}_i^{*\top} \right)^2$$

$$- \frac{1}{2\Delta_t^* \sigma^2} \sum_{t_i^* \in [0, T]} Y_i^{*\top} Y_i^* + \frac{1}{2\Delta_t^* \sigma^2} \sum_{t_i^* \in [0, T]} Y_i^{*\top} Y_i^*.$$

After some computations, we get

$$\log \ell^*(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau})) = \frac{1}{2\Delta_t^* \sigma^2} \left(\sum_{t_i^* \in [0,T]} Y_i^{*\top} Y_i^* - \sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \left(Y_i^* - \mathbf{z}_i^* \hat{\boldsymbol{\theta}}^{(j)} \right)^2 \right).$$

The term $\sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \left(Y_i^* - \mathbf{z}_i^* \hat{\boldsymbol{\theta}}^{(j)} \right)^2$ is non-negative. For an observed process $X_t = x_t$ with constant Δ_t^* and known σ, $\sum_{t_i^* \in [0,T]} Y_i^{*\top} Y_i^*$ is fixed and does not depend on the change points τ. Hence, finding the change points $\tau = (\tau_1, \ldots, \tau_m)$ that maximise (4.7) is equivalent to the minimisation of the term

$$\sum_{j=1}^{m+1} \sum_{t_i^* \in (\tau_{j-1}, \tau_j]} \left(Y_i^* - \mathbf{z}_i^* \hat{\boldsymbol{\theta}}_i \right)^2. \tag{B.17}$$

If $\Delta_t^* = \Delta_t$, the structure of (B.17) is the same as $SSE(T, \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}(\boldsymbol{\tau}))$ in (4.3). The rest of the proof for Proposition 4.3 follows directly via the same arguments used in establishing Proposition 4.1 and 4.2. $\qquad \square$

Appendix C: Proof of the properties in section 4.3

Proof of Proposition 4.4. Note that by (3.11) and (3.12),

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt \xrightarrow[T \to \infty]{a.s.} (s_j - s_{j-1}) \int_0^1 \tilde{h}(t) \varphi_k(t) dt$$

and

$$\frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t^2 dt \xrightarrow[T \to \infty]{a.s.} (s_j^0 - s_{j-1}^0) \left(\int_0^1 (\tilde{h}^{(j)}(t))^2 dt + \frac{\sigma^2}{2a^{(j)}} \right). \tag{C.1}$$

In addition, invoking similar statements as in Theorem 6.2 in Nkurunziza and Zhang (2016), we have

$$\frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t \varphi_k(t) dt - \frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t \varphi_k(t) dt \xrightarrow[T \to \infty]{P} 0 \tag{C.2}$$

and

$$\frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t^2 dt - \frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t^2 dt \xrightarrow[T \to \infty]{P} 0. \tag{C.3}$$

Therefore, it suffices to prove that

$$\frac{1}{T} \int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t \varphi_k(t) dt - \frac{1}{T} \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt \xrightarrow[T \to \infty]{p} 0. \tag{C.4}$$

Similarly,

$$\frac{1}{T}\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t^2 dt - \frac{1}{T}\int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t^2 dt \xrightarrow[T\to\infty]{p} 0. \tag{C.5}$$

Let $0 < \delta_j < \min((s_{j+1}^0 - s_j^0), (s_j^0 - s_{j-1}^0))/2$. Then, it follows form Proposition 4.2 that

$$\lim_{T\to\infty} P(|\hat{s}_j - s_j^0| > \delta_j) = 0, \quad j = 1, \ldots, m.$$

Therefore, we have

$$P\left(\frac{1}{T}|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t \varphi_k(t) dt - \frac{1}{T}\int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt| > \epsilon\right)$$

$$= P\left(\frac{1}{T}|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t \varphi_k(t) dt - \frac{1}{T}\int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt| > \epsilon, |\hat{s}_j - s_j^0| \le \delta_j,\right.$$

$$\left. |\hat{s}_{j-1} - s_{j-1}^0| \le \delta_{j-1}\right). \tag{C.6}$$

Since $|\hat{s}_j - s_j^0| \le \delta_j$ is equivalent to $s_j^0 - \delta_j \le \hat{s}_j \le s_j^0 + \delta_j$, and it follows that, for every j,

$$\left\|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} \tilde{X}_t \varphi_k(t) dt - \int_{\tau_{j-1}^0}^{\tau_j^0} \tilde{X}_t \varphi_k(t) dt\right\| \le \int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t \varphi_k(t)| dt$$

$$+ \int_{\tau_j^0 - \delta_j T}^{\tau_j^0 + \delta_j T} |\tilde{X}_t \varphi_k(t)| dt.$$

Hence,

$$(\text{C.6}) \le P\left(\frac{1}{T}\left(\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t \varphi_k(t)| dt + \int_{\tau_j^0 - \delta_j T}^{\tau_j^0 + \delta_j T} |\tilde{X}_t \varphi_k(t)| dt\right) > \epsilon\right)$$

$$\le P\left(\frac{1}{T}\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t \varphi_k(t)| dt > \epsilon/2\right)$$

$$+ P\left(\frac{1}{T}\int_{\tau_j^0 - \delta_j T}^{\tau_j^0 + \delta_j T} |\tilde{X}_t \varphi_k(t)| dt > \epsilon/2\right). \tag{C.7}$$

It then follows from the Markov's inequality and Jensen's inequality that

$$P\left(\frac{1}{T}\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t \varphi_k(t)| dt > \epsilon/2\right) \le \frac{4\mathrm{E}\left[\left(\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t||\varphi_k(t)| dt\right)^2\right]}{\epsilon^2 T^2}$$

$$\le \frac{4\mathrm{E}\left(\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0 + \delta_{j-1}T} |\tilde{X}_t|^2 |\varphi_k(t)|^2 dt\right)}{\epsilon^2 T^2}.$$

In the same vein,

$$P\left(\frac{1}{T}\int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}|\tilde{X}_t\varphi_k(t)|dt > \epsilon/2\right) \leq \frac{4\mathrm{E}\left[\left(\int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}|\tilde{X}_t||\varphi_k(t)|dt\right)^2\right]}{\epsilon^2 T^2}$$

$$\leq \frac{4\mathrm{E}\left(\int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}|\tilde{X}_t|^2|\varphi_k(t)|^2 dt\right)}{\epsilon^2 T^2}.$$

Consequently,

$$(\text{C.7}) \leq \frac{4\mathrm{E}\left(\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0+\delta_{j-1}T}|\tilde{X}_t|^2|\varphi_k(t)|^2 dt + \int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}|\tilde{X}_t|^2|\varphi_k(t)|^2 dt\right)}{\epsilon^2 T^2}. \quad (\text{C.8})$$

Then, it follows from (3.8) that $E(|\tilde{X}_t|^2) < K_1 < \infty$, for all $t \geq 0$. Also, $|\varphi_k(t)| \leq K_\varphi$. Therefore,

$$(\text{C.8}) \leq \frac{8\left(\delta_{j-1}+\delta_j\right)TK_1^2 K_\varphi^2}{\epsilon^2 T^2}.$$

Hence,

$$P\left(\frac{1}{T}\left|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j}\tilde{X}_t\varphi_k(t)dt - \frac{1}{T}\int_{\tau_{j-1}^0}^{\tau_j^0}\tilde{X}_t\varphi_k(t)dt\right| > \epsilon\right)$$

$$\leq \lim_{T\to\infty}P\left(\frac{1}{T}\left(\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0+\delta_{j-1}T}|\tilde{X}_t\varphi_k(t)|dt + \int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}|\tilde{X}_t\varphi_k(t)|dt\right) > \epsilon/2\right)$$

$$\leq \lim_{T\to\infty}\frac{8(\delta_{j-1}+\delta_j)TK_1^2 K_\varphi^2}{\epsilon^2 T^2} = 0.$$

Similarly, we have

$$P\left(\frac{1}{T}\left|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j}\tilde{X}_t^2 dt - \int_{\tau_{j-1}^0}^{\tau_j^0}\tilde{X}_t^2 dt\right| > \epsilon, |\hat{s}_j - s_j^0| \leq \delta_j, |\hat{s}_{j-1} - s_{j-1}^0| \leq \delta_{j-1}\right)$$

$$\leq P\left(\frac{1}{T}\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0+\delta_{j-1}T}\tilde{X}_t^2 dt > \epsilon/2\right) + P\left(\frac{1}{T}\int_{\tau_j^0-\delta_j T}^{\tau_j^0+\delta_j T}\tilde{X}_t^2 dt > \epsilon/2\right). \quad (\text{C.9})$$

Again, using the Markov's inequality,

$$(\text{C.9}) \leq \frac{4K_1(\delta_{j-1}+\delta_j)T}{\epsilon T} = \frac{4K_1(\delta_{j-1}+\delta_j)}{\epsilon}. \quad (\text{C.10})$$

Using the consistency properties of the estimators $\hat{s}_j$ provided in Section 4, i.e., $\hat{s}_j - s_j^0 \xrightarrow[T\to\infty]{P} 0$, we can choose arbitrarily small δ_{j-1} and δ_j such that

$$\lim_{T\to\infty}(\text{C.9}) \leq \lim_{T\to\infty}\frac{4K_1(\delta_{j-1}+\delta_j)}{\epsilon} = 0. \qquad \square$$

Proof of Proposition 4.6. By Propositions 4.4–4.5 (see also Proposition 6.2 in Nkurunziza and Zhang 2016), it may be shown that

$$\frac{1}{T}\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)} \xrightarrow[T\to\infty]{P} (s_j^0 - s_{j-1}^0)\boldsymbol{\Sigma}_j, j = 1, \ldots, m. \tag{C.11}$$

Under Assumptions 2–4, $\frac{1}{T}\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}$ is positive definite. Then, by the Continuous Mapping Theorem,

$$T\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}^{-1} = g\left(\frac{1}{T}\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}\right) \xrightarrow[T\to\infty]{P} \frac{1}{s_j^0 - s_{j-1}^0}\boldsymbol{\Sigma}_j^{-1} = g\left((s_j^0 - s_{j-1}^0)\boldsymbol{\Sigma}_j\right),$$
$$\tag{C.12}$$

where $j = 1, \ldots, m$ and $g(X) = X^{-1}$ for any positive definite matrix X. $\quad\square$

Proof of Proposition 4.7. Here we only prove that

$$\frac{1}{\sqrt{T}}\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t dW_t - \frac{1}{\sqrt{T}}\int_{\tau_{j-1}^0}^{\tau_j^0} X_t dW_t \xrightarrow[T\to\infty]{P} 0. \tag{C.13}$$

The convergence of the remaining components may be proved analogously by the same approach.

Since by Proposition 4.2, $\lim_{T\to\infty} P\left(|\hat{s}_j - s_j^0| > \delta_j\right) = 0, \quad j = 1, \ldots, m$,

$$P\left(\frac{1}{T}\left|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t dW_t - \frac{1}{T}\int_{\tau_{j-1}^0}^{\tau_j^0} X_t dW_t\right| > \epsilon\right).$$

is equal to

$$P\left(\frac{1}{T}\left|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t dW_t - \int_{\tau_{j-1}^0}^{\tau_j^0} X_t dW_t\right| > \epsilon, |\hat{s}_j - s_j^0| \le \delta_j, |\hat{s}_{j-1} - s_{j-1}^0| \le \delta_{j-1}\right).$$
$$\tag{C.14}$$

Without loss of generality, we assume that $\hat{\tau}_{j-1} < \tau_{j-1}^0 < \tau_j^0 < \hat{\tau}_j$. Then,

$$\left|\int_{\hat{\tau}_{j-1}}^{\hat{\tau}_j} X_t dW_t - \int_{\tau_{j-1}^0}^{\tau_j^0} X_t dW_t\right| = \left|\int_{\hat{\tau}_{j-1}}^{\tau_{j-1}^0} X_t dW_t + \int_{\tau_j^0}^{\hat{\tau}_j} X_t dW_t\right|$$

$$\le \left|\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\hat{\tau}_{j-1}} X_t dW_t\right| + \left|\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0} X_t dW_t\right| + \left|\int_{\tau_j^0}^{\tau_j^0 + \delta_j T} X_t dW_t\right|$$

$$+ \left|\int_{\hat{\tau}_j}^{\tau_j + \delta_j T} X_t dW_t\right|. \tag{C.15}$$

With the same arguments as in the proof of (C.7), together with above inequality, (C.14) is not greater than

$$P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\hat{\tau}_{j-1}} X_t dW_t\right| > \epsilon/4\right) + P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_{j-1}^0 - \delta_{j-1}T}^{\tau_{j-1}^0} X_t dW_t\right| > \epsilon/4\right)$$

$$+ P\left(\frac{1}{\sqrt{T}}\left|\int_{\hat{\tau}_j}^{\tau_j^0+\delta_j T} X_t dW_t\right| > \epsilon/4\right) + P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_j^0}^{\tau_j^0+\delta_j T} X_t dW_t\right| > \epsilon/4\right).$$

(C.16)

Then, by Markov inequality and Itô's isometry,

$$P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0} X_t dW_t\right| > \epsilon/4\right) \leq \frac{16\mathrm{E}\left[\left(\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0} X_t dW_t\right)^2\right]}{\epsilon^2 T}$$

$$= \frac{16\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0} \mathrm{E}\left[X_t^2\right] dt}{T\epsilon^2}$$

$$\leq \frac{16 K_1 \delta_{j-1} T}{\epsilon^2 T}$$

$$= \frac{16 K_1 \delta_{j-1}}{\epsilon^2},$$

which tends to 0 for an infinitesimal δ_{j-1}. Similarly, the probability

$P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_j^0}^{\tau_j^0+\delta_j T} X_t dW_t\right| > \epsilon/4\right)$ also tends to 0 by choosing an infinitesimal

δ_j. Further, note that

$$\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\hat{\tau}_{j-1}} X_t dW_t = \int_0^T X_t \mathbb{1}(\tau_{j-1}^0 - \delta_{j-1}T \leq t \leq \hat{\tau}_{j-1}) dW_t.$$

So, again, by the Markov inequality and Itô's isometry,

$$P\left(\frac{1}{\sqrt{T}}\left|\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\hat{\tau}_{j-1}} X_t dW_t\right| > \epsilon/4\right)$$

$$= P\left(\frac{1}{\sqrt{T}}\left|\int_0^T X_t \mathbb{1}(\tau_{j-1}^0 - \delta_{j-1}T \leq t \leq \hat{\tau}_{j-1}) dW_t\right| > \epsilon/4\right)$$

$$\leq \frac{16\mathrm{E}\left[\left(\int_0^T X_t \mathbb{1}(\tau_{j-1}^0 - \delta_{j-1}T \leq t \leq \hat{\tau}_{j-1}) dW_t\right)^2\right]}{\epsilon^2 T}$$

$$= \frac{16\int_{\tau_{j-1}^0-\delta_{j-1}T}^{\tau_{j-1}^0+\delta_{j-1}T} \mathrm{E}\left[X_t^2\right] dt}{T\epsilon^2}$$

$$\leq \frac{32 K_1 \delta_{j-1} T}{\epsilon^2 T}$$

$$= \frac{32 K_1 \delta_{j-1}}{\epsilon^2},$$

which goes to 0 for an infinitesimal δ_{j-1}. Similarly, $P\left(\frac{1}{\sqrt{T}}\left|\int_{\hat{\tau}_j}^{\tau_j^0+\delta_j T} X_t dW_t\right| > \epsilon/4\right)$

also approaches 0 by choosing an infinitesimal δ_j. This implies that (C.16) tends to 0 for an infinitesimal δ_{j-1} and δ_j.   □

Proof of Corollary 4.2. Let $\tilde{\mathbf{r}}_T(\hat{\boldsymbol{\tau}}) = \left(\tilde{\mathbf{r}}_{(0,\hat{\tau}_1)}^{\top}, \ldots, \tilde{\mathbf{r}}_{(\hat{\tau}_m,T)}^{\top} \right)^{\top}$,

$\mathbf{r}_T(\hat{\boldsymbol{\tau}}) = \left(\mathbf{r}_{(0,\hat{\tau}_1)}^{\top}, \ldots, \mathbf{r}_{(\hat{\tau}_m,T)}^{\top} \right)^{\top}$ and let $\tilde{\mathbf{Q}}(\hat{\boldsymbol{\tau}})^{-1} = \text{diag}\left(\mathbf{Q}_{(0,\hat{\tau}_1)}^{-1}, \ldots, \mathbf{Q}_{(\hat{\tau}_m,T)}^{-1} \right)$.
Then, we have $\hat{\boldsymbol{\theta}} = \tilde{\mathbf{Q}}(\hat{\boldsymbol{\tau}})^{-1}\tilde{\mathbf{r}}_T(\hat{\boldsymbol{\tau}})$. By Propositions 4.6–4.8 and application of
Slutsky's Theorem,

$$\sqrt{T}\tilde{\mathbf{Q}}(\hat{\boldsymbol{\tau}})^{-1}\tilde{\mathbf{r}}_T(\hat{\boldsymbol{\tau}}) = T\tilde{\mathbf{Q}}(\hat{\boldsymbol{\tau}})^{-1}\frac{1}{\sqrt{T}}\mathbf{r}_T(\hat{\boldsymbol{\tau}}) \xrightarrow[T\to\infty]{D} \rho \sim \mathcal{N}_{((m+1)(p+1))}\left(0, \tilde{\boldsymbol{\Sigma}}^{-1}\right).$$

Next, we investigate the asymptotic behaviour of $\hat{\boldsymbol{\theta}}$ based on $\hat{\boldsymbol{\tau}}$. Without
loss of generality, we assume that for the jth block, $1 \leq j \leq m$, we have
$\hat{\tau}_{j-1} < \tau_{j-1}^0 < \hat{\tau}_j < \tau_j^0$. In this case, we have

$$\hat{\boldsymbol{\theta}}^{(j)} = \mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}^{-1}\left(\mathbf{Q}_{(\hat{\tau}_{j-1},\tau_{j-1}^0)}\boldsymbol{\theta}^{(j-1)} + \mathbf{Q}_{(\tau_{j-1}^0,\hat{\tau}_j)}\boldsymbol{\theta}^{(j)} + \sigma\mathbf{r}_{(\hat{\tau}_{j-1},\hat{\tau}_j)} \right).$$

Hence,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}^{(j)} - \boldsymbol{\theta}^{(j)} \right)$$

$$= T\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}^{-1}\left(\frac{1}{\sqrt{T}}\mathbf{Q}_{(\hat{\tau}_{j-1},\tau_{j-1}^0)}\left(\boldsymbol{\theta}^{(j-1)} - \boldsymbol{\theta}^{(j)} \right) + \sigma\frac{1}{\sqrt{T}}\mathbf{r}_{(\hat{\tau}_{j-1},\hat{\tau}_j)} \right).$$

By Proposition 4.2, $|\hat{\tau}_{j-1} - \tau_{j-1}^0| \leq C$ for some $C > 0$ with probability 1.
Invoking the Markov inequality,

$$P\left(\frac{1}{\sqrt{T}}\int_{\tau_{j-1}^0}^{\hat{\tau}_{j-1}} X_t^2 dt > \epsilon \right) \leq \frac{2\mathrm{E}\left[\int_{\tau_{j-1}^0}^{\hat{\tau}_{j-1}} X_t^2 dt \right]}{\epsilon\sqrt{T}} \leq \frac{2K_1 C}{\epsilon\sqrt{T}}$$

and

$$P\left(\frac{1}{\sqrt{T}}\int_{\tau_{j-1}^0}^{\hat{\tau}_{j-1}} X_t\varphi_k(t) dt > \epsilon \right) \leq \frac{4\mathrm{E}\left[\int_{\tau_{j-1}^0}^{\hat{\tau}_{j-1}} X_t\varphi_k(t) dt \right]^2}{\epsilon^2 T} \leq \frac{4K_1 K_\varphi C}{\epsilon^2 T}.$$

Therefore, $\left\| \frac{1}{\sqrt{T}}\mathbf{Q}_{(\hat{\tau}_{j-1},\tau_{j-1}^0)} \right\| \xrightarrow[T\to\infty]{p} 0$, which means

$$\sqrt{T}(\hat{\boldsymbol{\theta}}^{(j)} - \boldsymbol{\theta}^{(j)}) - \sigma T\mathbf{Q}_{(\hat{\tau}_{j-1},\hat{\tau}_j)}^{-1}\frac{1}{\sqrt{T}}\mathbf{r}_{(\hat{\tau}_{j-1},\hat{\tau}_j)} \xrightarrow[T\to\infty]{p} 0.$$

So,

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \sigma T\tilde{\mathbf{Q}}(\hat{\boldsymbol{\tau}})^{-1}\frac{1}{\sqrt{T}}\mathbf{r}_T(\hat{\boldsymbol{\tau}}) \xrightarrow[T\to\infty]{p} 0$$

and

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \sigma T\tilde{Q}(\hat{\boldsymbol{\tau}})^{-1}\frac{1}{\sqrt{T}}\mathbf{r}_T(\hat{\boldsymbol{\tau}}) + \sigma T\tilde{Q}(\hat{\boldsymbol{\tau}})^{-1}\frac{1}{\sqrt{T}}\mathbf{r}_T(\hat{\boldsymbol{\tau}}) \xrightarrow[T\to\infty]{D} \rho$$

$$\sim \mathcal{N}_{((m+1)(p+1))}\left(0, \sigma^2\tilde{\boldsymbol{\Sigma}}^{-1}\right).$$

$\square$

Appendix D: Proof of Proposition 5.1

The proof of Proposition consists of two parts. In Part (i), we prove that $\mathcal{IC}(m = m^0) < \mathcal{IC}(m < m^0)$, whilst in Part (ii) we prove that $\mathcal{IC}(m = m^0) > \mathcal{IC}(m < m^0)$.

Part (i): $\mathcal{IC}(m = m^0) < \mathcal{IC}(m < m^0)$. From Proposition 4.3,

$$
\mathcal{IC}(m = m^0) = -2 \frac{1}{2\Delta_t^* \sigma^2} \left(\sum_{t_i \in [0,T]} Y_i' Y_i - \sum_{j=1}^{m^0+1} \sum_{t_i \in (\hat{\tau}_{j-1}, \hat{\tau}_j]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(j)} \right)^2 \right)
$$
$$
+ (m^0 + 1)(p+1) \log(T/\Delta_t), \tag{D.1}
$$

where $\hat{\tau}_j$, $j = 1 \ldots, m^0$ are obtained via (4.5). Next, we define

$$
\mathcal{IC}^0(m = m^0) = -2 \frac{1}{2\Delta_t \sigma^2} \left(\sum_{t_i \in [0,T]} Y_i' Y_i - \sum_{j=1}^{m^0+1} \sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(j,0)} \right)^2 \right)
$$
$$
+ (m^0 + 1)(p+1) \log(T/\Delta_t), \tag{D.2}
$$

where $\hat{\boldsymbol{\theta}}^{(j,0)}$ was given in Appendix B. Since $\hat{\tau}_j$, $j = 1, \ldots, m^0$, are obtained by maximising $\log \ell^*(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}})$, or equivalently minimising $-2 \log \ell^*(\boldsymbol{\tau}, \hat{\boldsymbol{\theta}})$, we have that $\mathcal{IC}(m = m^0) \leq \mathcal{IC}^0(m = m^0)$ with probability 1. Hence, we must show that

$$
\mathcal{IC}(m < m^0) > \mathcal{IC}^0(m = m^0) \tag{D.3}
$$

with probability 1.

For any positive integer m^* such that $0 < m^* < m^0$, suppose that the estimated locations of these m^* change points are $\hat{\tau}_1^*, \ldots, \hat{\tau}_{m^*}^*$, and the MLE of drift parameters associated with the ith observation is $\hat{\boldsymbol{\theta}}_i^* = \sum_{j=1}^{m^*+1} \hat{\boldsymbol{\theta}}^{(j,*)} \mathbb{1}(\hat{\tau}_{j-1}^* \leq t_i \leq \hat{\tau}_j^*)$, where $\hat{\boldsymbol{\theta}}^{(j,*)} = \mathbf{Q}_{(\hat{\tau}_{j-1}^*, \hat{\tau}_j^*)}^{-1} \tilde{\mathbf{r}}_{(\hat{\tau}_{j-1}^*, \hat{\tau}_j^*)}$. Furthermore,

$$
\frac{1}{T}(\mathcal{IC}(m = m^*) - \mathcal{IC}^0(m = m^0))
$$
$$
= \frac{1}{T\Delta_t \sigma^2} \left(\sum_{j=1}^{m^*+1} \sum_{t_i^* \in (\hat{\tau}_{j-1}, \hat{\tau}_j]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^* \right)^2 - \sum_{j=1}^{m^0+1} \sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}^{(j,0)} \right)^2 \right)
$$
$$
- \frac{(m^0 - m^*)(p+1) \log(T/\Delta_t^*)}{T}, \tag{D.4}
$$

Since $m^* < m^0$, there exists at least one change point that cannot be consistently estimated. Without loss of generality, let τ_j^0 be the change point. With similar arguments utilised in the proof of Lemma B.1, we get

$$
\frac{1}{T}(\mathcal{IC}(m < m^0) - \mathcal{IC}^0(m = m^0)) \geq C^* \|\boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{(j+1)}\|^2 + o_p(1) \tag{D.5}
$$

for some $C^* > 0$ with probability 1. Therefore, $\mathcal{IC}(m < m^0) > \mathcal{IC}^0(m = m^0)$ with probability 1. This completes the proof of part (i). □

Part (ii): $\mathcal{IC}(m = m^0) < \mathcal{IC}(m > m^0)$. Since $\mathcal{IC}(m = m^0) \leq \mathcal{IC}^0(m = m^0)$, where $\mathcal{IC}^0(m = m^0)$ is defined in Part (i), it remains to show that the difference $\mathcal{IC}(m = m^* > m^0) - \mathcal{IC}^0(m = m^0)$ is positive with probability 1.

Note that for the case where $m = m^* > m^0$ and the estimated locations of the m^* change points are given by $\hat{\tau}_1 < \hat{\tau}_2 <, \ldots, < \hat{\tau}_{m^*}$, we have

$$\mathcal{IC}(m = m^*) - \mathcal{IC}^0(m = m^0)$$

$$= \frac{1}{\Delta_t \sigma^2} \left(\sum_{t_i \in [0,T]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^* \right)^2 - \sum_{t_i \in [0,T]} \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^0 \right)^2 \right)$$

$$+ (m^* - m^0)(p+1) \log(T/\Delta_t^*), \tag{D.6}$$

where $\hat{\boldsymbol{\theta}}_i^* = \sum_{j=1}^{m^*+1} \boldsymbol{\theta}^{(j,*)} \mathbb{1}(\tau_{j-1}^* \leq t_i \leq \tau_j^*)$ with $\hat{\boldsymbol{\theta}}^{(j,*)} = \mathbf{Q}_{(\hat{\tau}_j^*, \hat{\tau}_{j-1}^*)}^{-1} \tilde{\mathbf{r}}_{(\hat{\tau}_j^*, \hat{\tau}_{j-1}^*)}$. We note that $m^* > m^0$, and from m^* of these estimated change points, there are $m^* - m^0$ estimated change points that divide the time interval $[0, T]$ into $m^* - m^0 + 1$ regimes such that within each regime, the number of estimated change points is equal to the number of exact change points. For example, suppose that $m^0 = 2$ and $m^* = 3$ with $0 < \hat{\tau}_1^* < \tau_1^0 < \hat{\tau}_2^* < \tau_2^0 < \hat{\tau}_3^* < T$. Then, if we divide the given time interval into $[0, \hat{\tau}_2^*]$ and $(\hat{\tau}_2^*, T]$, we can see that within these two intervals, the number of estimated change points is equal to the number of exact change points.

Denote the particular $m^* - m^0$ estimated change points by $\{\tilde{\tau}_j^*, j = 1, \ldots, m^* - m^0\}$. Also, let $\tilde{\tau}_0^* = 0$ and $\tilde{\tau}_{m^*-m^0+1}^* = T$. Then,

$$(\text{D.6}) = \frac{1}{\Delta_t \sigma^2} \sum_{j=1}^{m^*-m^0+1} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} \left[\left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^* \right)^2 \right.$$

$$\left. - \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^0 \right)^2 + \frac{(m^* - m^0)(p+1) \log(T/\Delta_t^*)}{m^* - m^0 + 1} \right].$$

Thus, it remains to show that in each regime $(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]$, $j = 1, \ldots, m^* - m^0 + 1$,

$$\frac{1}{\Delta_t \sigma^2} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} \left[\left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^* \right)^2 - \left(Y_i - \mathbf{z}_i \hat{\boldsymbol{\theta}}_i^0 \right)^2 + \frac{(m^* - m^0)(p+1) \log(T/\Delta_t^*)}{m^* - m^0 + 1} \right]$$

$$= \frac{1}{\Delta_t \sigma^2} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} \left[\left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^* \right) \right)^2 - \left(\mathbf{z}_i \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^0 \right) \right)^2 + 2u_i \mathbf{z}_i \left(\hat{\boldsymbol{\theta}}_i^* - \hat{\boldsymbol{\theta}}_i^0 \right) \right.$$

$$\left. + \frac{(m^* - m^0)(p+1) \log(\frac{T}{\Delta_t^*})}{m^* - m^0 + 1} \right] \tag{D.7}$$

is positive with probability 1.

Since within $(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]$, the number of estimated change points and the number of exact change points are the same, we first consider the case where there is no any change points within $(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]$. In this case, we have $\tau_{k^*-1}^0 < \tilde{\tau}_{j-1}^* < \tilde{\tau}_j^* < \tau_{k^*}^0$ for some j and k^*. Then, $\hat{\theta}_i^* = \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^{-1} \left(\mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)} \theta_i + \sigma \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)} \right) = \theta_i + \sigma \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^{-1} \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}$ and $\hat{\theta}_i^0 = \theta_i + \sigma \mathbf{Q}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}^{-1} \mathbf{r}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}$. Substituting the above expressions into (D.7), we have

$$
\begin{aligned}
(\text{D.7}) = \frac{1}{\Delta_t \sigma^2} &\left[\sigma^2 \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^\top \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^{-1} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^{-1} \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)} \right. \\
&- \sigma^2 \mathbf{r}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}^\top \mathbf{Q}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}^{-1} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}^{-1} \mathbf{r}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)} \\
&+ 2 \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} u_i \mathbf{z}_i \sigma (\mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)}^{-1} \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*)} - \mathbf{Q}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}^{-1} \mathbf{r}_{(\tau_{k^*-1}^0, \tau_{k^*}^0)}) \\
&+ \left. \frac{(m^* - m^0)(p+1) \log(T/\Delta_t^*)}{m^* - m^0 + 1} \right].
\end{aligned}
\tag{D.8}
$$

From the approach used in the proof of Propositions 4.1 and 4.2, we have

$$
\begin{aligned}
&\left\| \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)}^\top \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)}^{-1} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*]} \mathbf{z}_i^\top \mathbf{z}_i \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)}^{-1} \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)} \right\| \\
&\leq \left\| \frac{1}{\sqrt{T}} \mathbf{r}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)} \right\|^2 \left\| T \mathbf{Q}_{(\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*)}^{-1} \right\|^2 \left\| \frac{1}{T} \sum_{t_i \in (\tilde{\tau}_{j-1}^*, \hat{\tau}_2^*]} \mathbf{z}_i^\top \mathbf{z}_i \right\| \\
&= O_p \left(\log^{2a^*}(T) \right)
\end{aligned}
$$

for some $0 < a^* < 1/2$. Similar results also hold for the second and the third terms of (D.8).

Therefore, for large T, (D.8) is dominated by $\dfrac{(m^* - m^0)(p+1) \log(T/\Delta_t^*)}{m^* - m^0 + 1}$, which is positive. This implies that for large T, (D.6) is positive with probability 1.

Now, consider the case where there exist m_j $(0 < m_j \leq m^0)$ exact change points (so are the estimated change points) in $(\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]$. We label these m_j exact change points by $\tilde{\tau}_{j-1}^* < \tau_{(m_j,1)}^0 <, \ldots, < \tau_{(m_j, m_j)}^0 < \tilde{\tau}_j^*$ and similarly for the estimated change points, $\tilde{\tau}_{j-1}^* < \hat{\tau}_{(m_j,1)}^* <, \cdots, < \hat{\tau}_{(m_j, m_j)}^* < \tilde{\tau}_j^*$.

By the quadratic structure, $\sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} (Y_i - \mathbf{z}_i \hat{\theta}_i^0)^2 \leq \sum_{t_i \in [0,T]} (Y_i - \mathbf{z}_i \hat{\theta}_i^0)^2$, and

$$
\sum_{t_i \in [0,T]} (Y_i - \mathbf{z}_i \hat{\theta}_i^0)^2 = \sum_{j=1}^m \mathbf{r}_{(\tau_{j-1}^0, \tau_j^0)}^\top \mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}^{-1} \sum_{t_i \in (\tau_{j-1}^0, \tau_j^0]} \mathbf{z}_i' \mathbf{z}_i \mathbf{Q}_{(\tau_{j-1}^0, \tau_j^0)}^{-1} \mathbf{r}_{(\tau_{j-1}^0, \tau_j^0)}.
$$

It follows from (B.7) that

$$\sum_{j=1}^{m} \mathbf{r}_{(\tau_{j-1}^0,\tau_j^0)}^{\top} \mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)}^{-1} \sum_{t_i \in (\tau_{j-1}^0,\tau_j^0]} \mathbf{z}_i^{\top} \mathbf{z}_i \mathbf{Q}_{(\tau_{j-1}^0,\tau_j^0)}^{-1} \mathbf{r}_{(\tau_{j-1}^0,\tau_j^0)} = O_p\left((\log T)^{2a^*}\right)$$

for some $0 < a^* < 1/2$. By similar methods employed in the proof of Proposition 4.1, we have $\sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} 2u_i \mathbf{z}_i (\hat{\boldsymbol{\theta}}_i^* - \hat{\boldsymbol{\theta}}_i^0) = O_p\left((\log T)^{2a^*}\right)$. Since for large T, $(\log T)^{2a^*} < \dfrac{(m^* - m^0)(p+1)\log(T/\Delta_t^*)}{m^* - m^0 + 1}$, we have that for large T, (D.7) is dominated by either $\dfrac{(m^* - m^0)(p+1)\log(T/\Delta_t^*)}{(m^* - m^0 + 1)\Delta_t \sigma^2}$ or $\dfrac{1}{\Delta_t \sigma^2} \displaystyle\sum_{t_i \in (\tilde{\tau}_{j-1}^*, \tilde{\tau}_j^*]} (\mathbf{z}_i(\boldsymbol{\theta}_i -$

$\hat{\boldsymbol{\theta}}_i^*))^2$ and they are both positive. This implies that for large T, (D.6) is positive with probability 1. $\quad\square$

Appendix E: Histograms of the estimated change points' arrival rates (Section 7.1.1)
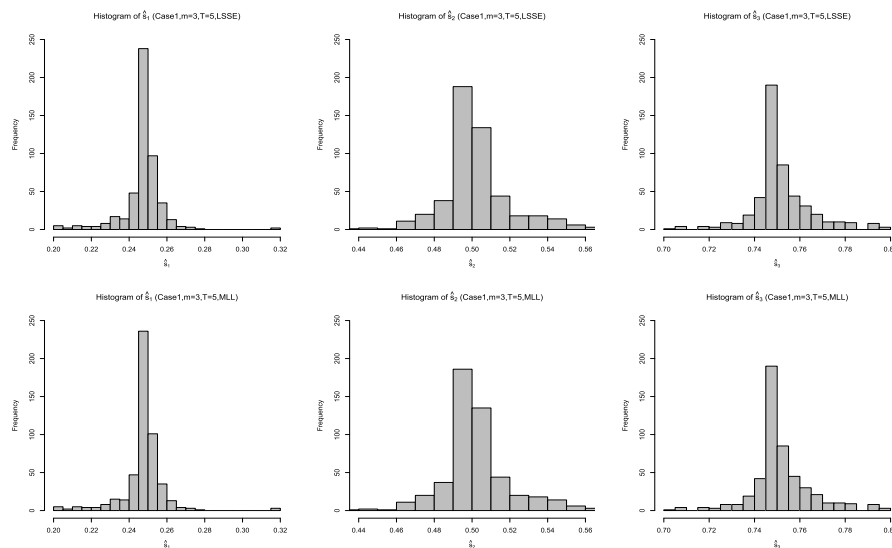
FIGURE 8. *Histogram of $\hat{s}$ for Case 1, $m^0 = 3$ and $T = 5$ with exact value $s^0 = (0.25, 0.50, 0.75)$*
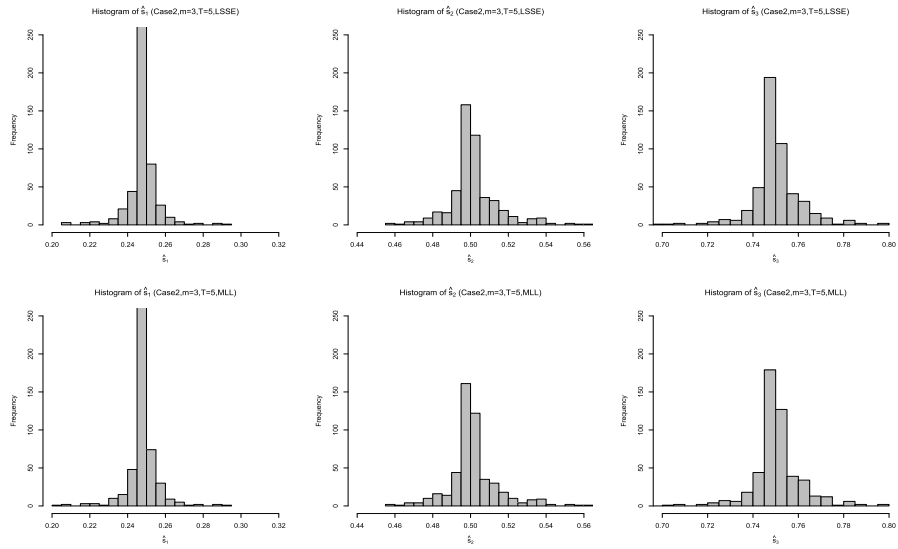
FIGURE 9. *Histogram of $\hat{s}$ for Case 2, $m^0 = 3$ and $T = 5$ with exact value $s^0 = (0.25, 0.50, 0.75)$*
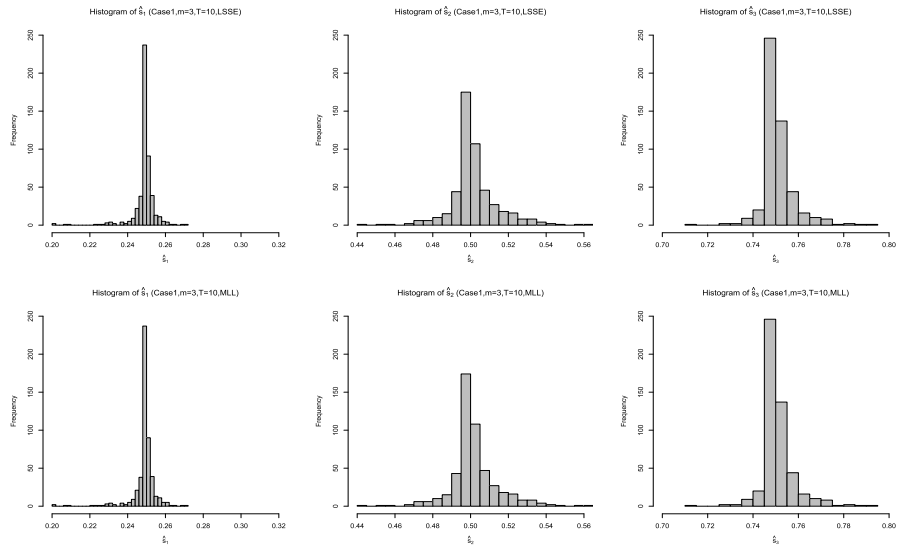
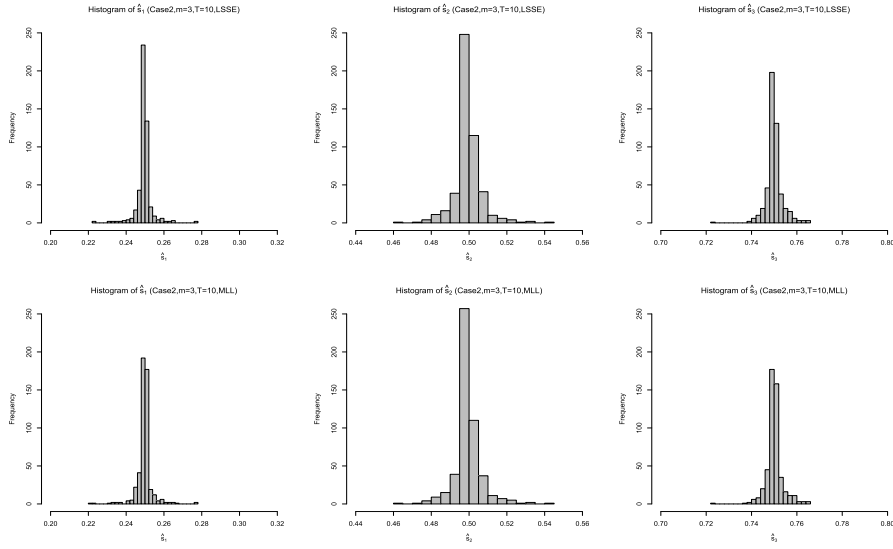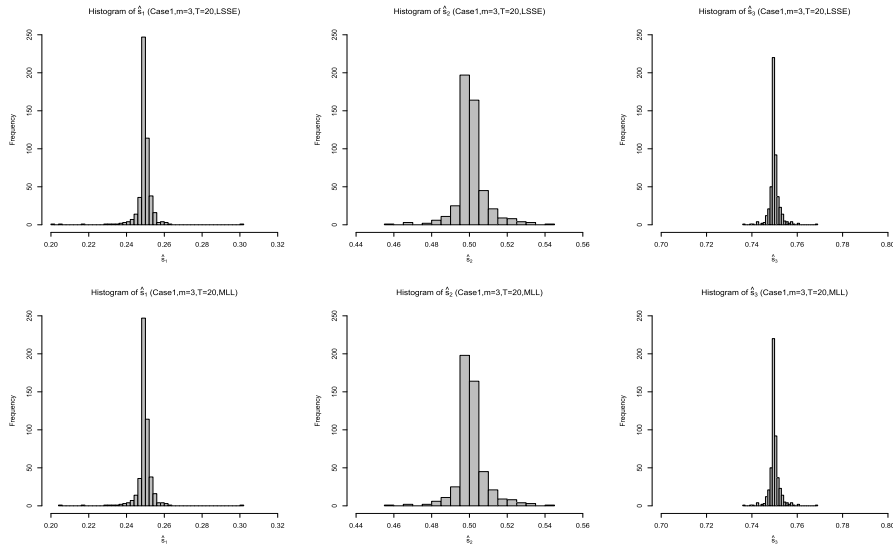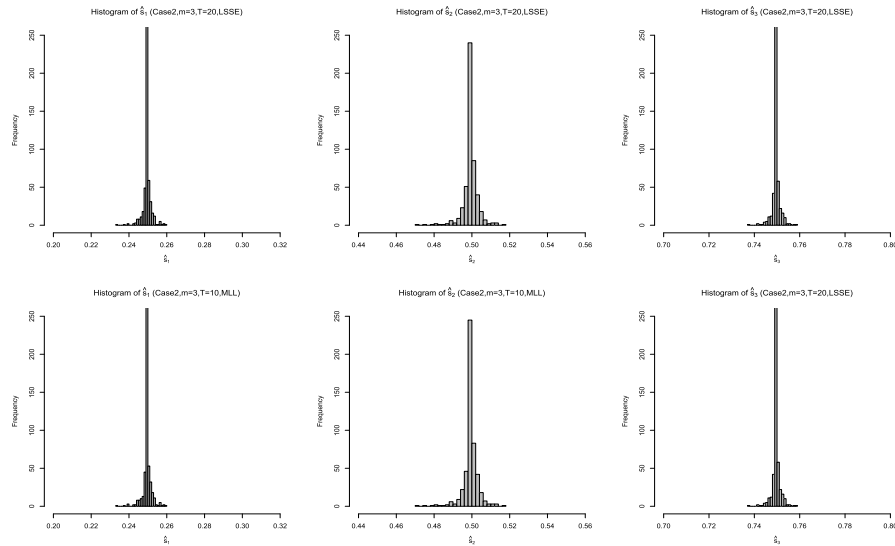FIGURE 10. *Histogram of $\hat{s}$ for Case 1, $m^0 = 3$ and $T = 10$ with exact value $s^0 = (0.25, 0.50, 0.75)$*

FIGURE 11. *Histogram of $\hat{s}$ for Case 2, $m^0 = 3$ and $T = 10$ with exact value $s^0 = (0.25, 0.50, 0.75)$*

FIGURE 12. *Histogram of $\hat{s}$ for Case 1, $m^0 = 3$ and $T = 20$ with exact value $s^0 = (0.25, 0.50, 0.75)$*

FIGURE 13. *Histogram of $\hat{s}$ for Case 2, $m^0 = 3$ and $T = 20$ with exact value $s^0 = (0.25, 0.50, 0.75)$*

Acknowledgements

References

Aalen, O., Gjessing, H., Survival models based on the Ornstein-Uhlenbeck process, Lifetime Data Analysis. 10(4), (2004), 407–423. MR2125423

Akaike, H., Information theory and an extension of the maximum likelihood principle, in Petrov, B., Csáki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971, Budapest: Akadémiai Kiadó, (1973), 267–281. MR0483125

Auger, I., Lawrence, C., Algorithms for the optimal identification of segment neighborhoods, Bulletin of Mathematical Biology. 51(1), (1989), 39–54. MR0978902

Bai, J., Perron, P., Estimating and testing linear models with multiple structural changes, Econometrica. 66(1), (1998) 47–78. MR1616121

Benth, F., Koekebakker, S., Taib, C., Stochastic dynamical modelling of spot freight rates, IMA Journal of Management Mathematics. 26(3), (2015), 273–297. MR3365478

Chen, S., Modelling the dynamics of commodity prices for investment decisions under uncertainty, PhD Dissertation. University of Waterloo, (2010). https://uwspace.uwaterloo.ca/bitstream/handle/10012/5504/Chen_Shan.pdf

Chen, F., Nkurunziza, S., Optimal method in multiple regression with structural changes, Bernoulli. 21(4), (2015), 2217–2241. MR3378465

Date, P., Bustreo, R., Value-at-risk for fixed-income portfolios: a Kalman filtering approach, IMA Journal of Management Mathematics. 27(4), (2016), 557–573. MR3565094

Date, P., Mamon, R., Tenyakov, A., Filtering and forecasting commodity futures prices under an HMM framework, Energy Economics. 40, (2013), 1001–1013.

Dehling, H., Franke, B., Kott, T., Drift estimation for a periodic mean reversion process, Statistical Inference for Stochastic Processes. 13, (2010), 175–192. MR2729648

Dehling, H., Franke, B., Kott, T., Kulperger, R., Change point testing for the drift parameters of a periodic mean reversion process, Statistical Inference for Stochastic Process. 17(1), (2014), 1–18. MR3178011

Ditlevsen, S., Lansky, P., Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model, Physical Review E. 71, (2005), 011907. MR2139313

Elias, R., Wahab, M., Fung, F., A comparison of regime-switching temperature modeling approaches for applications in weather derivatives, European Journal of Operational Research. 232(3), (2014), 549–560.

Elliott, R., van der Hoek, J., Malcolm, P., Pairs trading, Quantitative Finance, 5(3) (2005), 271–276. MR2238537

Elliott, P., Wilson, C., The term structure of interest rates in a hidden Markov setting, in Mamon, R., Elliott, R. (eds), Hidden Markov Models in Finance. Springer, New York, (2007), 14–31. MR2405580

Erlwein, C., Benth, F., Mamon, R.S., HMM filtering and parameter estimation of an electricity spot price model, Energy Economics, 32(5), (2010), 1034–1043.

Gallagher, C., Lund, R., Robbins, M., Changepoint detection in daily precipitation data, Environmetrics, 23(5), (2012), 407–419. MR2958921

Gombay, E., Change detection in linear regression with time series errors, Canadian Journal of Statistics, 38(1), (2010), 65–79. MR2676930

Hull, J., White, A., Pricing interest-rate-derivative securities, Review of Financial Studies. 3(4), (1990), 573–592.

HOWELL, S., DUCK, P., HAZEL, A., JOHNSON, P., PINTO, H., STRBAC, G., PROUDLOVE, N., BLACK, M., A partial differential equation system for modelling stochastic storage in physical systems with applications to wind power generation, IMA Journal of Management Mathematics, 22(3), (2011), 231–252. MR2806201

JACKSON, B., SCARGLE, J., BARNES, D., ARABHI, S., ALT, A., GIOUMOUSIS, P., GWIN, E., SANGTRAKULCHAROEN, P., TAN, L., TSAI, T., An algorithm for optimal partitioning of data on an interval, IEEE Signal Processing Letters, 12, (2002), 105–108.

KILLICK, R., FEARNHEAD, P., ECKLEY, I., Optimal detection of change points with a linear computational cost, Journal of the American Statistical Association. 107(500), (2012), 1590–1598. MR3036418

KUTOYANTS, Y.A., Statistical Inference for Ergodic Diffusion Processes, Springer Series in Statistics, London, (2004). MR2144185

LANSKY, P., SACERDOTE, L., The Ornstein-Uhlenbeck neuronal model with signal-dependent noise, Physics Letters A. 285(3–4), (2001), 132–140. MR1844801

LIANG, Z., YUEN, K., GUO, J., Optimal proportional reinsurance and investment in a stock market with Ornstein-Uhlenbeck process, Insurance: Mathematics and Economics. 49(2), (2011), 207–215. MR2811989

LIPSTER, R., SHIRYAEV, A., Statistics of Random Processes I. Springer-Verlag, Berlin-Heidelber-New York (2001). MR1800858

LU, Q., LUND, R., Simple linear regression with multiple level shifts, Canadian Journal of Statistics. 35 (3), (2007), 447–458. MR2396030

LU, S., Ornstein-Uhlenbeck diffusion quantum Monte Carlo calculations for small first-row polyatomic molecules, Journal of Chemical Physics. 118(21), (2003), 9528–9532.

LU, S., Ornstein-Uhlenbeck diffusion quantum Monte Carlo study on the bond lengths and harmonic frequencies of some first-row diatomic molecules, Journal of Chemical Physics. 120, (2004), http://dx.doi.org/ 10.1063/1.1639370.

MAIDSTONE, R., HOCKING, T., RIGAILL, G., FEARNHEAD, P., On optimal multiple changepoint algorithms for large data, ArXiv e-prints. (2014).

NKURUNZIZA, S., ZHANG, P., Estimation and testing in generalized mean-reverting processes with change-point. Statistical Inference for Stochastic Processes. (2016), in press. DOI 10.1007/s11203-016-9151-3.

PAGE, E., Continuous inspection schemes, Biometrika. 41, (1954), 100–115. MR0088850

PERRON, P., QU, Z., Estimating restricted structural change models, Journal of Econometrics. 134(2), (2006), 373–399. MR2328414

REEVES, J., CHEN, J., WANG, X., LUND, R., LU, Q., A review and comparison of changepoint detection techniques for climate data, Journal of Applied Meteorology and Climatology. 46(6), (2007), 900–915.

ROBBINS, M., LUND, R., GALLAGHER, C., LU, Q., Changepoints in the North Atlantic tropical cyclone record, Journal of the American Statistical Association, 106 (493), (2011), 89–99. MR2816704

ROHLFS, R., HARRIGAN, P., NIELSEN, R., Modeling gene expression evolution

with an extended Ornstein-Uhlenbeck process accounting for within-species variation, Scandinavian Journal of Statistics. 37(2), (2010), 200–220.

Schwartz, E., The stochastic behavior of commodity prices: implications for valuation and hedging, Journal of Finance 52. (1997), 923–973.

Sen, A., Srivastava, M., On tests for detecting change in mean, Annals of Statistics. 3(1), (1975), 98–108. MR0362649

Scott, A., Knott, M., A cluster analysis method for grouping means in the analysis of variance, Biometrics. 30(3), (1974), 507–512.

Schwarz, G., Estimating the dimension of a model, Annals of Statistics. 6(2), (1978), 461–464. MR0468014

Shinomoto, S., Sakai, Y., Funahashi, S., The Ornstein-Uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex, Neural Computation. 11(4), (1999), 935–951.

Shiryaev, A., On optimum methods in quickest detection problems, Theory of Probability and Its Applications. 8, (1963), 26–51. MR0155708

Smith, W., On the simulation and estimation of the mean-reverting Ornstein-Uhlenbeck process. Version 1.01, CommodityModels.com, (2010).

Spokoiny, V., Multiscale local change point detection with applications to value-at-risk, Annals of Statistics. 37(3), (2009), 1405–1436. MR2509078

Tenyakov, A., Mamon, R., Davison, M., Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach, Knowledge-Based Systems. 101, (2016), 142–155.

Tenyakov, A., Mamon, R., Davison, M., Filtering of a discrete-time HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity regimes, IEEE Journal of Selected Topics in Signal Processing. 10(6), (2016), 994–1005.

Tobing, H., McGilchrist, C., Recursive residuals for multivariate regression models, Australian Journal of Statistics. 34(2), (1992), 217–232. MR1193774

Vasiček, O., An equilibrium characterisation of the term structure, Journal of Financial Economics. 5, (1977), 177–188.

Yan, G., Xiao, Z., Eidenbenz, S., Catching instant messaging worms with change-point detection techniques, Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats. 6(8), (2008), 1–10.

Yu, X., Baron, M., Choudhary, P., Change-point detection in binomial thinning processes with applications in epidemiology, Sequential Analysis 32 (3), (2013), 350–367. MR3175827

Zhang, P., On Stein-rules in generalized mean-reverting processes with change point, Master's Thesis, University of Windsor, (2015).