Ann Inst Stat Math (2018) 70:807–853
https://doi.org/10.1007/s10463-017-0610-4

Inference for a change-point problem under a generalised Ornstein–Uhlenbeck setting

Fuqi Chen[1] · Rogemar Mamon[1,2] · Sévérien Nkurunziza[3]

Received: 14 June 2016 / Revised: 29 April 2017 / Published online: 19 June 2017

Abstract Determining accurately when regime and structural changes occur in various time-series data is critical in many social and natural sciences. We develop and show further the equivalence of two consistent estimation techniques in locating the change point under the framework of a generalised version of the one-dimensional Ornstein–Uhlenbeck process. Our methods are based on the least sum of squared error and the maximum log-likelihood approaches. The case where both the existence and the location of the change point are unknown is investigated and an informational methodology is employed to address these issues. Numerical illustrations are presented to assess the methods' performance.

Keywords Sequential analysis · Least sum of squared errors · Maximum likelihood · Consistent estimator · Existence of change point

✉ Rogemar Mamon
rmamon@stats.uwo.ca

Fuqi Chen
fchen66@uwo.ca

Sévérien Nkurunziza
severien@uwindsor.ca

1 Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond Street, London, ON N6A 5B7, Canada

2 Division of Physical Sciences and Mathematics, University of the Philippines Visayas, 5023 Miag-ao, Iloilo, Philippines

3 Department of Mathematics and Statistics, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B 3P4, Canada

1 Introduction

1.1 Background

In this paper, we consider the generalised Ornstein–Uhlenbeck (OU) process in one-dimensional set up. The OU process is commonly used to model the stochastic dynamics of various financial variables. Certain economic indicators also have stylised properties that are adequately captured by the OU process. Vasicek's (1977) pioneering work, employing an OU model to price a zero-coupon bond, inspired a multitude of research investigations from extensions addressing the model's weakness of constant mean-reverting level to various applications in economic and financial practice. The importance of this stochastic process is also demonstrated by its ubiquity in many fields. Amongst the finance and finance-related areas highlighting the usefulness of OU process are 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), centra-bank setting rate policy (e.g., Elliott and Wilson 2007), 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). Applications of the OU process could be as well found in biology (see Rohlfs et al. 2010), neurology (see Shinomoto et al. 1999), survival analysis (see Aalen and Gjessing 2004), physics (see Lánský and Sacerdote 2001), and chemistry (see Lu 2003, 2004).

To rectify the classical OU model's inability to capture the evolution of a process whose mean level varies with time, Dehling et al. (2010) introduced a generalised OU process in which a time-dependent function describes its mean-reverting level. Such a generalised version incorporates time-inhomogeneity and seasonality of mean reversion simultaneously. Moreover, the generalised OU process is capable of modelling drastic changes in certain time points (e.g., interest rates undergoing drastic moves due to financial crisis, war, etc). Dehling et al. (2014) developed the framework in the study of a change-point phenomenon in the generalised OU process to model the drastic change.

Our contributions in this paper hinged on the research results primarily from two research articles detailed as follows. The first is the paper of Dehling et al. (2010) in which a maximum likelihood estimator (MLE) for the drift parameters of the diffusion process is derived and the asymptotic properties, such as the asymptotic distribution of the proposed MLE, are studied. Dehling et al. (2014) considered an extended model, where there is one unknown change point and constructed a likelihood-ratio test statistic in determining a candidate change point. This line of enquiry was continued by Nkurunziza and Zhang (2016) who examined the asymptotic properties of both the unrestricted and restricted MLE for the drift parameters of the generalised OU process with a single change point. In particular, based on the established asymptotic distribution of the MLEs, a James-Stein-type shrinkage estimator for the drift parameters is proposed in Nkurunziza and Zhang (2016) as an improvement. In the estimation of the unknown change point, Nkurunziza and Zhang (2016) showed that the previously established asymptotic properties also hold for any consistent estimator for the rate of change point.

Nonetheless, both Dehling et al. (2014) and Nkurunziza and Zhang (2016) did not provide any explicit method to estimate the change point. This deficiency inspired the three main contributions of our paper. Firstly, we present two consistent methods to estimate the unknown change point. Secondly, we consider the case where the existence of the change point is uncertain and propose an informational approach to address this existence issue. Thirdly, the performance of the proposed methods is theoretically analysed and validated by a numerical implementation. In practice, many data series are characterised by some potential changes in 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 more efficient way.

Investigations concerning change-point problems are not new. Inaugural contributions to this field were spearheaded, for example, by Page (1954) and Shiryaev (1963). Recent developments have focused on (i) the estimation of change points and coefficients of linear regression models with multiple change points (cf. Bai and Perron 1998; Perron and Qu 2006; Lu and Lund 2007; Gombay 2010; Chen and Nkurunziza 2015); (ii) change-point testing for the drift parameters of a periodic mean-reverting process (cf. Dehling et al. 2014); (iii) change-point analysis involving stochastic differential equations (cf. De Gregorio and Iacus 2008; Iacus and Yoshida 2012; Lee 2011; Lee and Guo 2015); (iv) applications in finance (cf. Spokoiny 2009); (v) detection of malware within software (Yan et al. 2008); and (vi) climatology (Reeves et al. 2007; Robbins et al. 2011; Gallagher et al. 2012). In general, the analysis of change points could be described as a hypothesis-testing problem for the existence of change points in various locations. Alternatively, this could be viewed as a model selection problem that treats the change points as the additional unknown parameters to be estimated. However, unlike ordinary least-squares estimation, there is so far no closed-form estimation methods to calculate the change point directly or in a few steps. The existing change-point estimation approaches are predominantly designed to perform a search at every possible location of a change point with some efficient computational algorithms until some criteria are satisfied. The well-known algorithms for change-point detection are the (i) binary segmentation algorithm (Scott and Knott 1974; Sen and Srivastava 1975), (ii) segment-neighbourhood algorithm (Auger and Lawrence 1989; Bai and Perron 1998) with adaption to the restricted regression model (Perron and Qu 2006), and (iii) PELT algorithm (Killick et al. 2012).

There are two types of scenarios for which change-point problems are examined in the literature. In the first scenario, the number of change points is known, but their exact locations are unknown (see Perron and Qu 2006; Chen and Nkurunziza 2015). The second scenario covers the more general situation in which 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 performance assessment in the former scenario is relatively easier than that in the latter scenario.

1.2 Motivating examples

1.2.1 West Texas Intermediate (WTI) Cushing crude oil spot prices

The first motivating example of this paper is the West Texas Intermediate (WTI) Cushing crude oil spot prices, which is often being considered as a benchmark in oil pricing. The data set was compiled by Bloomberg with code "USCRWTIC" and covers approximately 4 years of daily prices from 09 November 2011 to 09 November 2015 (i.e., 1008 trading days). During this period (see Fig. 7), there is a price decline after September 2014 due to the conflict in the Middle East. It could be recalled that in September 2014, there was an increase in the OPEC oil production led by a rebound in the Libyan output. The dollar, on the other hand, continued to get stronger. These events caused the decline of the crude oil prices and suggested the potential existence of a change point in the data series.

Due to the potential existence of the change point, for this data set the classical OU process without change-point is inappropriate. Our findings show, as detailed in Sect. 6, that applying the classical method to the original data set produces very large Schwarz information criterion (SIC) as compared to the proposed method which takes into account the change point. Indeed, the proposed method increases the log-likelihood value from 0.72 to 11.89 and reduces the SIC by 69%.

1.2.2 XAU currency

The second motivating example of this paper is the XAU currency, which is the standard ticker symbol for one troy ounce of gold, considered as a currency to US dollar. This implementation is carried out to show the nuances in dealing with data sets or its transformed version whose change point is not clear-cut at the outset. The data set is also obtained from Bloomberg with code "XAU", and it is a 15-year data series ranging from 03 November 2000 to 04 November 2015 (i.e., 3913 trading days).

Descriptive analysis of the data set suggests a certain trend in the XAU currency series that changes over time. In particular, the currency was increasing since the beginning of the period until August 2011. Most notably after 2008 crisis, the increasing slope became sharper as the investors flocked to gold market. The price was close to \$1900 in August 2011, and it remained above \$1500 until April 2013. Then, the price plunged due to the banking crisis in Cyprus and increasing worries about an imminent change in the Federal Reserve's monetary policy. These features in the price movement suggest the potential existence of a change point.

By using the proposed method to the original data set, the log-likelihood is increased from 1.43 to 12.62 and reduces the SIC by 43%. However, by using the proposed method to the log-transformed data set, the proposed method does not allow to confirm the existence of a change point. Nonetheless, the proposed method preserves a good performance in terms of log-likelihood (7.06 vs. 3.31 for the classical method).

The remainder of this paper is organised as follows. In Sect. 2, we look at the formulation of the change-point problem. We recapitulate in Sect. 3 the results of both Dehling et al. (2014) and Nkurunziza and Zhang (2016) on MLE and the related asymptotic properties which are useful in delving into the asymptotic performance of our proposed methods. Section 4 considers the case where the existence of the change point is certain, but its location is unknown; two estimation methods are put forward to determine the unknown change point. The asymptotics of the estimators are also discussed and hence, the asymptotic properties established in Nkurunziza and Zhang (2016) also hold in our proposed techniques. The case where the existence and the location of the change point are both unknown is explored in Sect. 5. We develop an informational approach to detect the change point, and the consistency of our methods is likewise theoretically demonstrated. Section 6 provides the numerical implementation of our proposed methods on both simulated and observed financial market data. The final section gives some concluding remarks.

2 Description of the single-change-point problem

Our main consideration in this paper is the change-point estimation strategy under a generalised Ornstein–Uhlenbeck process with only one change point. We start with Nkurunziza and Zhang's (2016) framework , which assumes that a consistent estimator exists for the unknown change point $\tau \in [0, T]$. The model under examination is the generalised version of the Ornstein–Uhlenbeck (OU) process with SDE representation

$$\mathrm{d}X_t = \left(S(\theta^{(1)}, t, X_t)\mathbb{I}_{\{t \leqslant \tau\}} + S(\theta^{(2)}, t, X_t)\mathbb{I}_{\{t>\tau\}} \right) \mathrm{d}t + \sigma \mathrm{d}W_t, \quad 0 < t \leq T, \quad (1)$$

where $\mathbb{I}_A$ denotes the indicator function of the event A; τ is an unknown change point; $\theta^{(j)} = (\mu_1^{(j)}, \dots, \mu_p^{(j)}, -a^{(j)})'$ for $j = 1, 2$; the symbol $'$ refers to the transpose of a matrix; and $S(\theta^{(j)}, t, X_t) = L^{(j)}(t) - a^{(j)} X_t = \sum_{i=1}^{p} \mu_i^{(j)} \varphi_i(t) - a^{(j)} X_t$ for $i = 1, \dots, p$ and $j = 1, 2$. Also, W_t is a one-dimensional Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$.

In particular, we assume that there is one unknown change point $\tau = sT, 0 < s < 1$ such that

$$S(\theta^{(1)}, t, X_t) = \sum_{i=1}^{p} \mu_i^{(1)} \varphi_i(t) - a^{(1)} X_t, \quad 0 < t < \tau,$$
$$S(\theta^{(2)}, t, X_t) = \sum_{i=1}^{p} \mu_i^{(2)} \varphi_i(t) - a^{(2)} X_t, \quad \tau \leq t \leq T,$$

where $\theta^{(1)} = \left(\mu_1^{(1)}, \dots, \mu_p^{(1)}, -a^{(1)}\right)'$, for $0 < t < \tau$, and $\theta^{(2)} = \left(\mu_1^{(2)}, \dots, \mu_p^{(2)}, a^{(2)}\right)'$, for $\tau \leq t \leq T$.

For the case when there is no change point, maximum likelihood estimators for the drift parameters and their related asymptotic properties were derived in Dehling et al. (2010). These results are reviewed in the next section and serve as a springboard for our theoretical discussion. It has to be noted that our focus is the diffusion process described by (1) in continuous time. Thus, in deriving the theoretical results, we consider the diffusion parameter σ^2 to be known. In particular, σ^2 equals the quadratic variation of the process. In the data analysis presented in Sect. 6, the quadratic variation is taken as

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{N}\left(X_{t_i} - X_{t_{i-1}}\right)^2 \tag{2}$$

with $0 = t_0 < t_1 < \cdots < t_N = T$. In a continuous-time process, we have $\mathrm{P}\left\{\hat{\sigma}^2 = \sigma^2\right\} = 1$. The extended framework where the diffusion coefficient changes after the change point is a natural research direction that is being pursued in our on-going investigation.

3 Earlier MLE-based results and our new results

This section consists of two subsections: (i) review of the results for the MLE of the drift parameters (without change point) along with the related asymptotic properties demonstrated in Dehling et al. (2010); and (ii) review of the MLE for the drift parameters (with one change point) and the related asymptotic properties studied in Nkurunziza and Zhang (2016).

In Nkurunziza and Zhang (2016), however, asymptotic normality for the MLE estimator of the drift parameters is derived under the assumption that the estimator is already consistent. In our case, we shall show (instead of assume) that such an estimator of the change point is consistent, thereby proving the consistency properties of the proposed estimator.

Notation The expressions "$\xrightarrow[T\to\infty]{p}$", "$\xrightarrow[T\to\infty]{D}$", and "$\xrightarrow[T\to\infty]{a.s.}$" denote convergence in probability, convergence in distribution, and convergence almost surely, respectively. The "$O(\cdot)$" stands for "Big O" describing the asymptotic behaviour of functions; i.e., for a sequence 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 in the sense that $\forall\epsilon > 0,\ \exists M > 0,\ \ni\ P(|U_n/a_n| > M) < \epsilon,\ \ \forall n$. The symbol "small o" means $U_n = o_p(a_n)$, i.e., U_n/a_n converges in probability to zero as n approaches an appropriate limit. Considering $U_n = o_p(a_n)$ is equivalent to $U_n/a_n = O_p(1)$, convergence in probability is defined here as $\lim_{n\to\infty}(P(|U_n/a_n)| \geq \epsilon) = 0$.

3.1 Maximum likelihood estimator of the drift parameters

To gain some useful insights, we first consider the case where there is no change point ($\theta^{(1)} = \theta^{(2)}$). We review briefly the MLE of the drift parameters proposed in Dehling et al. (2010) under the following assumptions.

Assumption 1 $P\left(\int_0^T S^2(\theta, t, X_t) < \infty\right) = 1$, for all $0 < T < \infty$, for all $\theta \in \theta$.

With this assumption, Theorem 7.6 in Lipster and Shiryaev (2001) may be used to find an explicit expression for the corresponding likelihood function.

Suppose that there is no any change point in $[0, T]$. Then, let $\mathcal{C}[0, T]$ be the space of continuous, real-valued function on $[0, T]$ and let $\mathcal{B}[0, T]$ be the Borel σ-field associated with $\mathcal{C}[0, T]$. Let P_B be the probability measure generated by the Brownian motion on $(\mathcal{C}[0, T], \mathcal{B}[0, T])$, i.e., $P_B(A) = P\{\omega : B \in A\}$, $A \in \mathcal{B}[0, T]$. Suppose further that P_X is the probability measure generated by the observation X_T of the process with SDE specified in (1). Then, the likelihood function of X_T is

$$\mathcal{L}(\theta, X_T) = \frac{\mathrm{d}P_X}{\mathrm{d}P_B}(X_T) = \exp\left(\frac{1}{\sigma^2}\int_0^T S(\theta, t, X_t)\mathrm{d}X_t - \frac{1}{2\sigma^2}\int_0^T S^2(\theta, t, X_t)\mathrm{d}X_t\right). \quad (3)$$

Therefore, the MLE of the drift parameters is given by

$$\hat{\theta} = Q_{(0,T)}^{-1}\tilde{R}_{(0,T)} = \left(\frac{1}{T}Q_{(0,T)}\right)^{-1}\frac{1}{T}\tilde{R}_{(0,T)}, \quad (4)$$

where

$$Q_{(0,T)} = \begin{bmatrix} \int_0^T \varphi_1^2(t)\mathrm{d}t & \ldots \int_0^T \varphi_1(t)\varphi_p(t)\mathrm{d}t & -\int_0^T \varphi_1(t)X_t\mathrm{d}t \\ & \ldots & \\ -\int_0^T X_t\varphi_1(t)\mathrm{d}t & \ldots -\int_0^T X_t\varphi_p(t)\mathrm{d}t & \int_0^T X_t^2\mathrm{d}t \end{bmatrix},$$

and $\tilde{R}_{(0,T)} = (\int_0^T \varphi_1(t)\mathrm{d}X_t, \ldots, \int_0^T \varphi_p(t)\mathrm{d}X_t, -\int_0^T X_t\mathrm{d}X_t)'$.

Note that the MLE introduced above could be evaluated by applying the Euler's discretisation to (1), and then getting a linear model and applying the ordinary least-squares estimation method to provide an estimator containing the Riemann and Ito sums. Then, the OLS estimator will converge into the MLE estimator as $\Delta t \to 0$.

For the existence of $Q_{(0,T)}^{-1}$, it is shown in Remark 3 of Dehling et al. (2010) that $TQ_{(0,T)}^{-1}$ exists almost surely if T is large enough. Moreover, as stated in Nkurunziza and Zhang (2016), the positive definiteness of $\frac{1}{T}Q_{(0,T)}$ holds under the following assumption.

Assumption 2 For $T > 0$, the base function $\{\varphi_i(t), i = 1, \ldots, p\}$ is Riemann-integrable on $[0, T]$ and satisfies

1. Periodicity. That is, $\varphi_i(t + v) = \varphi_i(t)$, for all $i = 1, \ldots, p$ and v is the period observed in the data.
2. Orthogonality. That is, for all $j, k = 1, \ldots, p$, $\int_0^v \varphi_j(t)\varphi_k(t)\mathrm{d}t$ is equal to v if $j = k$ and 0 otherwise.
3. For large T, the family of base functions $\{\varphi_i(t), i = 1, \ldots, p\}$ is incomplete.

The first and second items in Assumption 2 correspond to similar assumptions in Dehling et al. (2014), and the third item is used to establish the positive definiteness of $\frac{1}{T}Q_{(0,T)}$. It should be noted that the link between the incomplete base functions and positive definiteness of $\frac{1}{T}Q_{(0,T)}$ is discussed in Zhang (2015), and Nkurunziza and Zhang (2016) applies this result directly. To provide a self-contained exposition, we recall below Proposition 2.1.1 of Zhang (2015), which characterises the positive definiteness of $\frac{1}{T}Q_{(0,T)}$.

Proposition 1 (Proposition 2.1.1 of Zhang (2015)) *The base functions* $\{\varphi_i(t), i = 1, \ldots, p\}$ *are incomplete if and only if* $\frac{1}{T}Q_{(0,T)}$ *is a positive definite matrix.*

Hence, for the rest of this paper, we assume 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 $v = 1$ and this implies that $\varphi_j(t+1) = \varphi_j(t)$.

Inspired by the results of Dehling et al. (2010) and Dehling et al. (2014), Nkurunziza and Zhang (2016) first studied the case where the change point in (1) exists and known to be $\tau^0 = s^0 T$, $0 < s^0 < 1$ and derived the results in estimating $\theta^{(1)}$ and $\theta^{(2)}$. In particular,

$$\hat{\theta}^{(1)} = Q^{-1}_{(0,s^0T)}\tilde{R}_{(0,s^0T)} = \theta^{(1)} + \left(\frac{1}{T}Q_{(0,s^0T)}\right)^{-1}\frac{1}{T}R_{(0,s^0T)} \tag{5}$$

and

$$\hat{\theta}^{(2)} = Q^{-1}_{(s^0T,T)}\tilde{R}_{(s^0T,T)} = \theta^{(2)} + \left(\frac{1}{T}Q_{(s^0T,T)}\right)^{-1}\frac{1}{T}R_{(s^0T,T)}, \tag{6}$$

where $R_{(a,b)} = \left(\int_a^b \varphi_1(t)\mathrm{d}W_t, \ldots, \int_a^b \varphi_p(t)\mathrm{d}W_t, -\int_a^b X_t \mathrm{d}W_t\right)'$ for $0 \le a < b \le T$.

Also, the asymptotic properties of the above proposed MLEs are well studied in Dehling et al. (2010) for the case when there is no change point and Nkurunziza and Zhang (2016) for the case of a single change point. To summarise these results, we first go back to the case when there is no change point. By (1), we have

$$\int_0^T \varphi_i(t)\mathrm{d}X_t = \sum_{j=1}^p \mu_j \int_0^T \varphi_i(t)\varphi_j(t)\mathrm{d}t - a\int_0^T \varphi_i(t)X_t\mathrm{d}t + \sigma\int_0^T \varphi_i(t)\mathrm{d}W_t,$$

for $i = 1, \ldots, p$, and

$$\int_0^T X_t dX_t = \sum_{j=1}^p \mu_j \int_0^T X_t\varphi_j(t)\mathrm{d}t - a\int_0^T X_t^2\mathrm{d}t + \sigma\int_0^T X_t\mathrm{d}W_t.$$

It follows that

$$\hat{\theta} = Q^{-1}_{(0,T)}\tilde{R}_{(0,T)} = \theta + \sigma Q^{-1}_{(0,T)}R_{(0,T)} = \theta + \sigma T Q^{-1}_{(0,T)}\frac{1}{T}R_{(0,T)}.$$

By Ito's lemma, the SDE in (1) has the solution

$$X_t = \mathrm{e}^{-at}X_0 + h(t) + N_t, \tag{7}$$

where $h(t) = \mathrm{e}^{-at}\sum_{i=1}^p \mu_i \int_0^t \mathrm{e}^{as}\varphi_i(s)\mathrm{d}s$ and $N_t = \sigma\mathrm{e}^{-at}\int_0^t \mathrm{e}^{as}\mathrm{d}W_s$.

The uniform boundedness of solution (7) follows from the results in Nkurunziza and Zhang (2016). Using similar methods employed in the proof of Theorem 6.1 in

Nkurunziza and Zhang (2016) together with the mean-reversion property in the drift term of the OU process, one may verify that the SDE (1) admits a strong and unique solution that is uniformly bounded in L^2, and

$$\sup_{t\geq 0} \mathrm{E}(X_t^2) \leq K_1, \tag{8}$$

for $0 < K_1 < \infty$.

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

$$\tilde{X}_t = \tilde{h}(t) + \tilde{N}_t, \tag{9}$$

where $\tilde{h}(t) = \mathrm{e}^{-at} \sum_{i=1}^{p} \mu_i \int_{-\infty}^{t} \mathrm{e}^{as} \varphi_i(s)\mathrm{d}s$ and $\tilde{N}_t = \sigma \mathrm{e}^{-at} \int_{-\infty}^{t} \mathrm{e}^{as} \mathrm{d}\tilde{B}_s$, with $(\tilde{B}_s)_{s\in\mathbb{R}}$ denotes a bilateral Brownian motion, i.e.,

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

Here, $(B_s)_{s\geq 0}$ and $(\bar{B}_s)_{s\geq 0}$ are two independent standard Brownian motions and $\mathbf{1}_A$ stands for the indicator function over the set A. It follows from (8) and Lemma 4.3 in Dehling et al. (2010) that 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. Then, by Proposition 4.5 of Dehling et al. (2010),

$$\frac{1}{T} \int_0^T \tilde{X}_t \varphi_j(t)\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} \int_0^1 \tilde{h}(t)\varphi_j(t)\mathrm{d}t \tag{10}$$

and

$$\frac{1}{T} \int_0^T \tilde{X}_t^2 \mathrm{d}t \xrightarrow[T\to\infty]{a.s.} \int_0^1 \tilde{h}^2(t)\mathrm{d}t + \frac{\sigma^2}{2a}. \tag{11}$$

Moreover, it follows from Lemma 4.4 in Dehling et al. (2010) that under Assumption 2,

$$|\tilde{X}_t - X_t| \xrightarrow[t\to\infty]{a.s.} 0. \tag{12}$$

Using (12), the following properties hold:

$$\frac{1}{T} \int_0^T \tilde{X}_t \varphi_j(t)\mathrm{d}t - \frac{1}{T} \int_0^T X_t \varphi_j(t)\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} 0$$
$$\text{and } \frac{1}{T} \int_0^T \tilde{X}_t^2 \mathrm{d}t - \frac{1}{T} \int_0^T X_t^2 \mathrm{d}t \xrightarrow[T\to\infty]{a.s.} 0.$$

Then, it follows from (10) and (11) that

$$\frac{1}{T} \int_0^T X_t \varphi_j(t)\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} \int_0^1 \tilde{h}(t)\varphi_j(t)\mathrm{d}t$$

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

Hence,

$$TQ_{(0,T)}^{-1} \xrightarrow[T\to\infty]{a.s.} \Sigma_0^{-1}, \tag{13}$$

where

$$\Sigma_0 = \begin{bmatrix} I_p & \Lambda \\ \Lambda' & w \end{bmatrix},$$

with $\Lambda_{(0,T)} = (\int_0^1 \tilde{h}(t)\varphi_1(t)\mathrm{d}t, \ldots, \int_0^1 \tilde{h}(t)\varphi_p(t)\mathrm{d}t)'$ and $w = \int_0^1 \tilde{h}^2(t)\mathrm{d}t + \frac{\sigma^2}{2a}$.

Furthermore, under Assumptions 1–2, by following the techniques in Nkurunziza and Zhang (2016), one can verify that

$$\frac{1}{\sqrt{T}}(\hat{\theta}-\theta) \xrightarrow[T\to\infty]{D} \rho \sim \mathcal{N}_{p+1}(0, \Sigma_0^{-1}).$$

In Sect. 3 of Nkurunziza and Zhang (2016) (or see Chapter 2 and pertinent proofs in Appendix B of Zhang 2015), the above asymptotic properties are extended to the case of a single change point in the following way. To this end, we write

$$\tilde{h}^{(1)}(t) := \mathrm{e}^{-a^{(1)}t}\sum_{i=1}^{p}\mu_i^{(1)}\int_{-\infty}^{t}\mathrm{e}^{a^{(1)}s}\varphi_i(s)\mathrm{d}s$$

and

$$\tilde{h}^{(2)}(t) := \mathrm{e}^{-a^{(2)}t}\sum_{i=1}^{p}\mu_i^{(2)}\int_{-\infty}^{t}\mathrm{e}^{a^{(2)}s}\varphi_i(s)\mathrm{d}s.$$

Then

$$\frac{1}{T}\int_0^{s^0T}\tilde{X}_t\varphi_j(t)\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} s^0\int_0^1(\tilde{h}^{(1)})(t)\varphi_j(t)\mathrm{d}t \tag{14}$$

and

$$\frac{1}{T}\int_0^{s^0T}\tilde{X}_t^2\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} s^0\left(\int_0^1(\tilde{h}^{(1)})^2(t)\mathrm{d}t + \frac{\sigma^2}{2a^{(1)}}\right), \tag{15}$$

where $\tilde{X}_t$ is the process defined in (9). Similarly,

$$\frac{1}{T}\int_{s^0T}^{T}\tilde{X}_t\varphi_j(t)\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} (1-s^0)\int_0^1(\tilde{h}^{(2)})(t)\varphi_j(t)\mathrm{d}t \tag{16}$$

and

$$\frac{1}{T}\int_{s^0T}^{T}\tilde{X}_t^2\mathrm{d}t \xrightarrow[T\to\infty]{a.s.} (1-s^0)\left(\int_0^1(\tilde{h}^{(2)})^2(t)\mathrm{d}t + \frac{\sigma^2}{2a^{(2)}}\right). \tag{17}$$

Using (12), the following properties hold:

$$\frac{1}{T}\int_0^{s^0T}\tilde{X}_t\varphi_j(t)\mathrm{d}t-\frac{1}{T}\int_0^{s^0T}X_t\varphi_j(t)\mathrm{d}t\xrightarrow[T\to\infty]{a.s.}0,$$
$$\frac{1}{T}\int_0^{s^0T}\tilde{X}_t^2\mathrm{d}t-\frac{1}{T}\int_0^{s^0T}X_t^2\mathrm{d}t\xrightarrow[T\to\infty]{a.s.}0,$$
$$\frac{1}{T}\int_{s^0T}^{T}\tilde{X}_t\varphi_j(t)\mathrm{d}t-\frac{1}{T}\int_{s^0T}^{T}X_t\varphi_j(t)\mathrm{d}t\xrightarrow[T\to\infty]{a.s.}0,$$
$$\text{and }\frac{1}{T}\int_{s^0T}^{T}\tilde{X}_t^2\mathrm{d}t-\frac{1}{T}\int_{s^0T}^{T}X_t^2\mathrm{d}t\xrightarrow[T\to\infty]{a.s.}0.$$

Hence, it follows that

$$\frac{1}{T}Q_{(0,s^0T)}\xrightarrow[T\to\infty]{a.s.}s^0\Sigma_1, \tag{18}$$
$$TQ^{-1}_{(0,s^0T)}\xrightarrow[T\to\infty]{a.s.}\frac{1}{s^0}\Sigma_1^{-1}, \tag{19}$$
$$\frac{1}{T}Q_{(s^0T,T)}\xrightarrow[T\to\infty]{a.s.}(1-s^0)\Sigma_2, \tag{20}$$

and

$$TQ^{-1}_{(s^0T,T)}\xrightarrow[T\to\infty]{a.s.}\frac{1}{(1-s^0)}\Sigma_2^{-1}, \tag{21}$$

where

$$\Sigma_1=\begin{bmatrix}I_p & \Lambda_1\\ \Lambda_1' & w_1\end{bmatrix}\quad\text{and}\quad\Sigma_2=\begin{bmatrix}I_p & \Lambda_2\\ \Lambda_2' & w_2\end{bmatrix}$$

with $\Lambda_i=\left(\int_0^1\tilde{h}^{(i)}(t)\varphi_1(t)\mathrm{d}t,\ldots,\int_0^1\tilde{h}^{(i)}(t)\varphi_p(t)\mathrm{d}t\right)'$ and $w_i=\int_0^1(\tilde{h}^{(i)})^2(t)\mathrm{d}t+\frac{\sigma^2}{2a^{(i)}}$, $i=1,2$. Furthermore, it follows from Proposition 6.2 in Nkurunziza and Zhang (2016) (or from Proposition 3.1) that both Σ_1 and Σ_2 are positive definite provided that Assumptions 1–2 hold.

3.2 New results for the analysis of asymptotic properties

Based on the established results in Sect. 3.1, we provide a proposition which is useful in illustrating the asymptotic properties of the estimator for the change point.

Proposition 2 *For $\eta\in(0,s^0]$, we have*

$$\frac{1}{T}Q_{(0,\eta T)}\xrightarrow[T\to\infty]{a.s.}\eta\Sigma_1,\quad\frac{1}{T}Q_{(\eta T,s^0T)}\xrightarrow[T\to\infty]{a.s.}(s^0-\eta)\Sigma_1,$$
$$\text{and}\quad\frac{1}{T}Q_{(\eta T,T)}\xrightarrow[T\to\infty]{a.s.}(s^0-\eta)\Sigma_1+(1-s^0)\Sigma_2, \tag{22}$$

Furthermore, for $\eta \in (s^0, 1]$,

$$\frac{1}{T} Q_{(\eta T, T)} \xrightarrow[T\to\infty]{a.s.} (1-\eta)\Sigma_2, \quad \frac{1}{T} Q_{(s^0 T, \eta T)} \xrightarrow[T\to\infty]{a.s.} (\eta - s^0)\Sigma_2,$$
$$\text{and} \quad \frac{1}{T} Q_{(0, \eta T)} \xrightarrow[T\to\infty]{a.s.} s^0\Sigma_1 + (\eta - s^0)\Sigma_2. \tag{23}$$

The proof of Proposition 2 follows from Proposition 6.1 in Nkurunziza and Zhang (2016). Moreover, for the case where the change point is unknown, Nkurunziza and Zhang (2016) assumed that there exists a consistent estimator of the unknown change point and derived the same asymptotic properties for the drift parameters based on the consistent estimator assumption of a change point.

4 Two methods in the estimation of the change point

We develop the least sum of squared error (LSSE) and maximum log-likelihood (MLL) methods to yield an estimator for the unknown change point τ, and investigate its consistency under these two methods.

4.1 Least sum of squared error method

This subsection first introduces the LSSE method and then studies the consistency of the proposed estimator. To calculate the residuals, we apply the Euler–Maruyama discretisation method to (1). Consider a partition $0 = t_0 < \cdots < t_n = T$ on a given time period $[0, T]$ with a constant increment $\Delta_t = t_{i+1} - t_i$. Hence, $Y_i = X_{t_{i+1}} - X_{t_i}$ and $Z_i = (\varphi_1(t_i), \ldots, \varphi_p(t_i), -X_{t_i})(\Delta_t)$, and the discretized process is given by

$$Y_i = Z_i\theta + \epsilon_i, \quad t_i \in [0, T], \tag{24}$$

where ϵ is the error term given by $\sigma\sqrt{\Delta_t}N$, and N is the standard normal term. In this case, we could use the least-squared (LS) method to estimate the change point. The details of the LS method will be discussed in the next section. Based on (24), let τ^0 be the exact value of the unknown change-point τ. Then, τ^0 can be estimated using the least sum of squared errors (SSE) method described as

$$\hat{\tau} = \arg\min_{\tau} SSE(\tau), \tag{25}$$

where

$$SSE(\tau) = \sum_{t_i \in [0,T]} (Y_i - Z_i\hat{\theta}(\tau))'(Y_i - Z_i\hat{\theta}(\tau)) \tag{26}$$

and $\hat{\theta}(\tau)$ is the estimator of θ with the change point given by τ. More precisely, from Nkurunziza and Zhang (2016), $\hat{\theta} = (\hat{\theta}^{(1)}, \hat{\theta}^{(2)})$ where

$$\hat{\theta}^{(1)} = Q^{-1}_{(0,\hat{\tau})}\tilde{R}_{(0,\hat{\tau})} \quad \text{and} \quad \hat{\theta}^{(2)} = Q^{-1}_{(\hat{\tau},T)}\tilde{R}_{(\hat{\tau},T)}.$$

Consistency of the proposed estimator

Under Assumptions 1–2, $\sum_{t_i\in[0,\tau^0]} Z_i' Z_i$ and $\sum_{t_i\in(\tau^0,T]} Z_i' Z_i$, the respective discretised versions of $Q_{(0,\tau^0)}$ and $Q_{(\tau^0,T)}$ are both positive definite with probability 1 provided that the base functions $\{\varphi_i(t), i = 1,\ldots,p\}$ are incomplete. Moreover, it follows from Proposition 6.2 in Nkurunziza and Zhang (2016) (or from Proposition 3.1) that both $\frac{1}{s^0 T}Q_{(0,\tau^0)}$ and $\frac{1}{(1-s^0)T}Q_{(\tau^0,T)}$ converge in probability to some positive definite matrices for large T, and so are their respective discretised versions. Hence, for large T, it is reasonable to impose a useful assumption in proving the consistency of the estimator of the change point.

Assumption 3 Suppose that there exists an $L_0 > 0$ such that for all $L > L_0$ the minimum eigenvalues of $\frac{1}{L}\sum_{t_i\in(\tau^0,\tau^0+L]} Z_i' Z_i$ and of $\frac{1}{L}\sum_{t_i\in(\tau^0-L,\tau^0]} Z_i' Z_i$, as well as their respective continuous-time versions $\frac{1}{L}Q_{(\tau^0,\tau^0+L)}$ and $\frac{1}{L}Q_{(\tau^0-L,\tau^0]}$, are all bounded away from 0.

For more details about the above assumption, reader is referred to Perron and Qu (2006) (see also Chen and Nkurunziza 2015). Below are two propositions pertinent to the consistency of the rate of change point specified by $\hat{s} = \hat{\tau}/T$, where $\hat{\tau}$ is given by (25).

Proposition 3 *Suppose that $\theta^{(1)} - \theta^{(2)}$, the shift in the drift parameters, is of fixed nonzero magnitude independent of T. Then, under Assumptions 1–3, $\hat{s} - s^0 \xrightarrow[T\to\infty]{P} 0$.*

Proposition 4 *Suppose the conditions in Proposition 3 hold. Then, for every $\epsilon > 0$, there exists a $C > 0$ such that for large T, $P(T|\hat{s} - s| > C) < \epsilon$.*

The proofs of Propositions 3 and 4 are provided in Appendix A. Proposition 3 shows that the estimated rate $\hat{s}$ is consistent for s^0 and Proposition 4 shows that the rate of convergence is T. From these two propositions, we conclude that the proposed estimator satisfies the consistency assumption required in Nkurunziza and Zhang (2016). Hence, it follows from Proposition 2.1 in Nkurunziza and Zhang (2016) that

$$\sqrt{T}(\hat{\theta}(\hat{\tau}) - \theta^0) \xrightarrow[T\to\infty]{D} \mathcal{N}_{2p+2}(0, \sigma^2\tilde{\Sigma}^{-1}), \tag{27}$$

where $\tilde{\Sigma}^{-1} = \text{diag}\left(\frac{1}{s^0}\Sigma_1^{-1}, \frac{1}{1-s^0}\Sigma_2^{-1}\right)$ and $\hat{\tau}$ are obtained from (25).

Remark 1 Note that in this paper, we focus on the case where the shift in drift parameters $\theta^{(1)}$ and $\theta^{(2)}$ indicated in (1) is independent of time T. However, 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., $\theta^{(1)} - \theta^{(2)} = \mathbf{M}v_T$, where $\mathbf{M}$ is independent of T and $v_T \xrightarrow[T\to\infty]{} 0$. In this case, the validity of Proposition 3 and Proposition 4 depends on the speed v_T. In fact, using similar arguments as in the proofs of these two propositions (see Appendix A), one may show that if $v_T \xrightarrow[T\to\infty]{} 0$ and $T^{1/2-r^*}v_T \xrightarrow[T\to\infty]{} \infty$ for some $0 < r^* < 1/2$, then under Assumptions 1–3, 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, $P(T v_T^2 |\hat{s} - s| > C) < \epsilon$. (See Remark 3 in Appendix A).

4.2 Maximum log-likelihood method

We introduce the maximum log-likelihood (MLL) method pertinent to the study of the consistency of the proposed estimator. By Theorem 7.6 in Lipster and Shiryaev (2001), the log-likelihood function of (1) (see also Dehling et al. 2010; Nkurunziza and Zhang 2016) is

$$\log \mathcal{L}^*((0,T),\theta) = \frac{1}{\sigma^2}\int_0^T S(\theta,t,X_t)\mathrm{d}X_t - \frac{1}{2\sigma^2}\int_0^T S^2(\theta,t,X_t)\mathrm{d}t. \quad (28)$$

Apparently, in practice the integrals $\int_0^T S(\theta,t,X_t)\mathrm{d}X_t$ and $\int_0^T S^2(\theta,t,X_t)\mathrm{d}t$ need to be approximated using appropriate finite sums that depend on some discrete sampling for which the discretisations are very small. It has to be noted as well that there is no mechanism that allows the collection of data in continuous time. To this end, consider the partition $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ with $t_k = \frac{Tk}{2^n}$, $k = 0, 1, \ldots, N$, and let $\delta_{k,N} = t_k - t_{k-1}$, $k = 1, 2, \ldots, 2^n$. As in Le Breton (1976), one can approximate $\log \mathcal{L}^*((0,T),\theta)$ by

$$\log \mathcal{L}_N(\tau,\theta) = \frac{1}{\sigma^2}\sum_{k=1}^{N} S(\theta,t_k,X_{t_k})\left(X_{t_k} - X_{t_{k-1}}\right) - \sum_{k=1}^{n_1} S^2(\theta,t_k,X_{t_k})\delta_{k,N}. \quad (29)$$

The following result is useful in proving that $\log \mathcal{L}_N^*((0,T),\theta)$ is a good approximation for $\log \mathcal{L}^*((0,T),\theta)$. Let $\varphi(t) = \left(\varphi_1(t), \varphi_2(t), \ldots, \varphi_p(t)\right)'$, $\Pi(t) = \left(\varphi'(t), X(t)\right)'$, and $\delta_N = \max_{1\leqslant k\leqslant N}\delta_{k,N}$. Denote the parameter space by Θ.

Proposition 5 *Let μ_θ^T be the probability measure induced by a generalised OU process. We have*

(i) $\left\| \sum_{k=1}^{N} \Pi(t_k)(X_{t_k} - X_{t_{k-1}}) - \int_0^T \Pi(t) dX_t \right\| \leqslant \delta_N^{1/2} V_N,$

(ii) $\left\| \sum_{k=1}^{N} \delta_{k,N}\Pi(t_k)\Pi'(t_k) - \int_0^T \Pi(t)\Pi'(t)\, dt \right\| \leqslant \delta_N^{1/2}\omega_N.$

In the above, $\{V_N\}_{N=1}^{\infty}$ and $\{\omega_N\}_{N=1}^{\infty}$ are sequences of random variables that are bounded in μ_θ^T-probability for all $\theta \in \Theta \subset \mathbb{R}^{p+1}$.

The proof of Proposition 5 follows directly from Lemma 6 in Le Breton (1976). The consequence of Proposition 5 is the following corollary that legitimises $\log \mathcal{L}_N(\tau,\theta)$ as a very good approximation for $\log \mathcal{L}^*((0,T),\theta)$ whenever the step of the discretisation is very small.

Corollary 1 *Suppose Θ_0 is a compact subset of the parameter space Θ. Then*

$$\| \log \mathcal{L}_N^*((0,T),\theta) - \log \mathcal{L}^*((0,T),\theta)\| \leqslant \delta_N^{1/2}\omega_N^*$$

and $\{\omega_N^*\}_{N=1}^{\infty}$ *is a sequence of random variables that is bounded in* μ_θ^T*-probability for all* $\theta \in \Theta_0$.

Proof By combining the triangle and Cauchy–Schwarz inequalities, we obtain

$$\begin{aligned}
&\|\log \mathcal{L}_N^*((0,T),\theta) - \log \mathcal{L}^*((0,T),\theta)\| \\
&\leqslant \frac{\|\theta\|}{\sigma^2}\left\|\sum_{k=1}^{N}\Pi(t_k)(X_{t_k}-X_{t_{k-1}}) - \int_0^T \Pi(t)\mathrm{d}X_t\right\| \\
&\quad + \frac{\|\theta\|^2}{2\sigma^2}\left\|\sum_{k=1}^{N}\delta_{k,N}\Pi(t_k)\Pi'(t_k) - \int_0^T \Pi(t)\Pi'(t)\,\mathrm{d}t\right\|,
\end{aligned}$$

for any $\theta \in \Theta_0$. In addition, since Θ_0 is a compact subset, there exists an $M_0 > 0$ such that $\|\theta\| + \|\theta\|^2 \leqslant M_0$ for all $\theta \in \Theta_0$. Then, employing Proposition 5,

$$\|\log \mathcal{L}_N^*((0,T),\theta) - \log \mathcal{L}^*((0,T),\theta)\| \leqslant \frac{M_0}{\sigma^2}\delta_N^{1/2}V_N + \frac{M_0}{2\sigma^2}\delta_N^{1/2}\omega_N = \delta_N^{1/2}\omega_N^*$$

with $\omega_N^* = M_0(2V_N + \omega_N)/\sigma^2$. This completes the proof. □

When a change point τ exists, an alternative method to estimate the unknown change point is via the maximum of the log-likelihood function

$$\log \mathcal{L}(\tau,\theta) = \log \mathcal{L}^*((0,\tau),\theta^{(1)}) + \log \mathcal{L}^*((\tau,T),\theta^{(2)}). \tag{30}$$

Relying on Corollary 1 along with the fact that the set of dyadic rationals on $[0,\ 1]$ is dense on $[0,\ 1]$, we propose to approximate $\log \mathcal{L}(\tau,\theta)$ by

$$\begin{aligned}
\log \hat{\mathcal{L}}_n(\tau,\theta) &= \frac{1}{\sigma^2}\sum_{k=1}^{n_1} S(\theta^{(1)},t_k,X_{t_k})\left(X_{t_k}-X_{t_{k-1}}\right) - \sum_{k=1}^{n_1} S^2(\theta^{(1)},t_k,X_{t_k})\frac{T}{2^n} \\
&\quad + \frac{1}{\sigma^2}\sum_{k=n_1+1}^{2^n} S(\theta^{(2)},t_k,X_{t_k})\left(X_{t_k}-X_{t_{k-1}}\right) \\
&\quad - \sum_{k=n_1+1}^{2^n} S^2(\theta^{(2)},t_k,X_{t_k})\frac{T}{2^n},
\end{aligned} \tag{31}$$

where $0 = t_0 < t_1 < t_2 < \cdots < t_{2^n} = T$ with $t_k = \frac{Tk}{2^n}$, $k = 0,1,\ldots,2^n$. That is, $t_k - t_{k-1} = \frac{T}{2^n}$, $k = 1,2,\ldots,2^n$, and n_1 is unknown integer such that $t_{n_1} \leqslant \tau = sT < t_{n_1+1}$.

Based on the approximated log-likelihood function in (31), we estimate τ by $\hat{\tau} = \frac{T\hat{n}_1}{2^n}$ where

$$\hat{n}_1 = \arg\max_{n_1} \log \hat{\mathcal{L}}_n(\tau,\hat{\theta}(\tau)). \tag{32}$$

Or equivalently,

$$\hat{n}_1 = \arg\min_{n_1} -2\log\hat{\mathcal{L}}_n(\tau, \hat{\theta}(\tau)), \tag{33}$$

where $\hat{\theta}$ is the MLE of θ based on $\hat{\tau}$. Thus, we estimate s by $\hat{s} = \frac{\hat{n}_1}{2^n}$.

Consistency of the proposed estimator

To investigate the consistency behaviour of the proposed estimator $\hat{\tau}$ obtained from (32), we update Assumption 3 to the following version:

Assumption 4 Suppose that there exists an $L_0^* > 0$ such that under Assumptions 1–3, for all $L > L_0^*$ the minimum eigenvalues of the following two symmetric matrices $\frac{1}{2L}[Q_{(0,\tau^0)}Q^{-1}_{(0,\tau^0+L)}Q_{(\tau^0,\tau^0+L)} + Q_{(\tau^0,\tau^0+L)}Q^{-1}_{(0,\tau^0+L)}Q_{(0,\tau^0)}]$ and $\frac{1}{2L}[Q_{(\tau^0,T)}Q^{-1}_{(\tau^0-L,T)}Q_{(\tau^0-L,\tau^0)} + Q_{(\tau^0-L,\tau^0)}Q^{-1}_{(\tau^0-L,T)}Q_{(\tau^0,T)}]$ are all bounded away from 0.

Then, similar to the LSSE method, for the estimator $\hat{\tau}$ given by (32), we also provide two propositions below regarding the consistency of $\hat{s}$.

Proposition 6 *Suppose that $\theta^{(1)} - \theta^{(2)}$, the shift in the drift parameters, is of fixed nonzero magnitude independent of T. Then, under Assumptions 1–2 and Assumption 4, $\hat{s} - s^0 \xrightarrow[T\to\infty]{P} 0$.*

The next proposition gives the rate of convergence, T, for $\hat{\tau}$.

Proposition 7 *Under the same conditions of Proposition 6, we have that for every $\epsilon > 0$, there exists a $C > 0$ such that for large T, $P(T|\hat{s} - s| > C) < \epsilon$.*

The proofs of Propositions 6 and 7 are provided in Appendix A. Proposition 6 establishes that the estimated rate $\hat{s}$ is consistent for s^0, and Proposition 7 shows that the rate of convergence is T. Moreover, in case the shift in the drift parameters is of shrinking magnitude, the discussions in Remark 1 also hold for this case. Propositions 6 and 7 imply that the proposed estimator satisfies the consistency assumption required in Nkurunziza and Zhang (2016). Thus, the asymptotic normality in (27) also holds when $\hat{\tau}$ are obtained from (32).

Remark 2 To see the connection between equations (32) and (25), one may apply the Riemann sum approximation with increment Δ_t to approximate the integrals inside the log-likelihood function $\log\mathcal{L}(\tau, \hat{\theta})$ specified in (30). The result of the approximation is of the form $\frac{1}{\sigma^2}\sum_{t_i\in[0,T]}\hat{\theta}(\tau)'V(t)'(X_{t_{i+1}} - X_{t_i}) - \frac{1}{2\sigma^2}\sum_{t_i\in[0,T]}(\hat{\theta}(\tau)'V(t)')^2\Delta_t$. Furthermore, if the increment Δ_t is same as that in the LSSE method, we have $X_{t_{i+1}} - X_{t_i} = Y_i$ and $V(t) = Z_i/\Delta_t$. Then, after some algebraic computations, such approximation can be transformed to $\frac{1}{2\Delta_t\sigma^2}(\sum_{t_i\in[0,T]} Y_i'Y_i - SSE(\tau))$. Hence, for an observed process X_t, $t \in [0, T]$ with same constant Δ_t and known σ, (25) and the Riemann sum approximation of (32) are equivalent. This finding will also be confirmed by the simulation results highlighted in Sect. 6.

5 Existence of a change point

In Sect. 4, we introduced two estimation methods for the case where the existence of the single change point is affirmative, that is, the number of change points is known to be 1. In this section, we shall deal with the extended change-point problem in which the number of change points may be either 0 or 1. In this case, it is of interest to test the existence of the change point and to determine its exact location if it exists. One popular methodology in the change-point literature in detecting the unknown number of change points is by treating it as a model selection problem. For instance, note that the existence of the change point in (1) also increases the number of drift parameters from $p+1$ to $2(p+1)$. Hence, detecting the existence of change points is equivalent to selecting a statistical model from two candidate models and this could be solved by using the informational approach. This approach deems that the most appropriate model is the one that minimises the log-likelihood-based information criterion

$$\mathcal{IC}(m) = -2\log\mathcal{L}(\hat{\tau}, \hat{\theta}) + (m+1)h(p)\phi(T). \tag{34}$$

In (34), $\log\mathcal{L}(\tau, \hat{\theta})$ is defined in (30); $\hat{\tau}$ is obtained via (32) corresponding to each m, where m is the potential number of change points to be determined ($m = 0$ or 1 in this case); $h(p) = p+1$ if there is no change in the diffusion coefficient σ before and after the change point or $p+2$ if there is a change in σ (i.e. $\sigma = \sigma^{(1)}$ for $t \in [0, \tau^0]$ and $\sigma = \sigma^{(2)}$ for $t \in (\tau^0, T]$); and $\phi(T)$ is a non-decreasing function of T.

Note that if the number of change points is known, then the term $(m+1)h(p)\phi(T)$ is fixed, and (34) is equivalent to the maximum log-likelihood method introduced in the previous section. The efficiency of the information criterion depends on the choice of the penalty criterion $\phi(T)$. For example, if $\phi(T) = 2$, then (34) 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 for example, Schwarz (1978). Many modified versions, thus, were proposed to overcome this problem. One of the modifications is the Schwarz information criterion (SIC) (Schwarz 1978) entails the setting of $\phi(T)$ as the log transform of the sample size. SIC has been successfully applied to the change-point analysis in the literature, and it gives an asymptotically consistent estimate of the order of the true model. Hence, we only focus on SIC on this particular theoretical development.

Further, it may be of interest to see which of the two penalty criteria we should use: $\phi(T) = \log(T)$ or $\phi(T) = \log(T/\Delta_t)$, where Δ_t is the increment defined in the previous section. Hence, in the ensuing discussion of our examples, we take into account these two criteria. (Note that $\log T$ is just a special case of $\log(T/\Delta_t) = \log(T) - \log(\Delta_t)$ with $\Delta_t = 1$).

Consider the hypothesis

$$\mathrm{H}_0: \quad m^0 = 0 \quad \text{versus} \quad \mathrm{H}_1: \quad m^0 = 1. \tag{35}$$

Based on (34), the rejection region for the null hypothesis in (35) is given by $\mathcal{IC}(m = 0) \geq \mathcal{IC}(m = 1)$. Moreover, the asymptotic significant level and power of the above test are investigated via the following results.

Proposition 8 *Suppose Assumptions 1–2 and 4 hold. Then, under H_0 in (35), $\lim_{T\to\infty} P\left(\mathcal{IC}(m=0) \geq \mathcal{IC}(m=1)\right) = 0$. Moreover, under H_1, $\lim_{T\to\infty} P\left(\mathcal{IC}(m=0) > \mathcal{IC}(m=1)\right) = 1$.*

The proof of Proposition 8 is presented in Appendix B. Let $\hat{m} = \arg\min_{m\in\{0,1\}} \mathcal{IC}(m)$ with $\phi(T) = \log T$ or $\log(T/\Delta_t)$. We then have the following.

Corollary 2 *Under Assumptions 1–2 and 4,* $\hat{m} - m^0 \xrightarrow[T\to\infty]{P} 0.$

The proof of Corollary 8 is immediate from Proposition 8. For a fixed Δ_t and large T, Corollary 2 shows that the two criteria, $\log(T)$ and $\log(T/\Delta_t)$, lead to asymptotically consistent estimate of the number of change points. For small T, we use Monte-Carlo simulation to compare the performance of these two criteria. Simulation results indicate that it would be more appropriate to use $\phi(T) = \log(T/\Delta_t)$ when T is small as $\phi(T) = \log T$ tends to over-estimate the number of change points, i.e., overfitting the model.

6 Numerical demonstrations

In this numerical work, we use in Sect. 6.1 the Monte-Carlo simulation technique to evaluate the performance of the (i) two estimation methods proposed in (25) and (32) in detecting the unknown location of a change point assumed to already exist, and (ii) method in (34) for testing the existence of a change point. In Sect. 6.2, we implement the above methods on some observed financial market data and illustrate the various implementation details.

6.1 Monte-Carlo simulation study

Our simulation considers two different scenarios. In the first scenario, we study the performance of the proposed methods under a classical OU process. In the second scenario, the performance evaluation of the proposed methods is applied to a periodic mean-reverting OU process. Each scenario consists of 1000 iterations. In each iteration, we first generate a desired simulated process based on the Euler-Maruyama discretisation scheme given a period T and pre-assigned "true" parameters such as the coefficients and rate of change point. Next, we estimate and record the rate of change points by applying (25) and (32) on the simulated process as well as the number of change points estimated by (34). To investigate the performance of (34), assuming there is no change point, we re-generate a simulated process with no change point and apply (34) to estimate and record again the number of change points. After 1000 iterations, we analyse the performance of the proposed methods based on the recorded results.

6.1.1 Simulation setup

Scenario 1: Classical OU process

Two classical OU processes are considered with stochastic dynamics

$$\mathrm{d}X_t = \begin{cases} (0.08 - 0.1X_t)\mathrm{d}t + 0.2\mathrm{d}W_t, & \text{if } 0 < t < 0.5T \\ (2.5 - 1X_t)\mathrm{d}t + 0.2\mathrm{d}W_t, & \text{if } 0.5T < t < T. \end{cases} \tag{36}$$

Equation (36) includes one change point occurring at $\tau^0 = 0.5T$ ($s^0 = 0.5$). This process is generated to determine the performance of the methods proposed in (25) and (32).

When there is one change point, we study the performance of (34) by considering the SDE

$$\mathrm{d}X_t = (2.5 - X_t)\mathrm{d}t + 0.2\mathrm{d}W_t, \quad 0 < t < T. \tag{37}$$

Scenario 2: Periodic mean-reverting OU process

For this scenario, we consider a mean-reverting OU process, with 2-dimensional periodic incomplete orthogonal set of functions $\left\{1, \sqrt{2}\cos\left(\frac{\pi t}{2\Delta_t}\right)\right\}$, given by

$$\mathrm{d}X_t = \begin{cases} \left[0.08 + 0.02\sqrt{2}\cos\left(\frac{\pi t}{2\Delta_t}\right) - 0.1X_t\right]\mathrm{d}t + 0.2\mathrm{d}W_t, & \text{if } 0 < t < 0.5T \\ \left[2.5 + 1.2\sqrt{2}\cos\left(\frac{\pi t}{2\Delta_t}\right) - 1X_t\right]\mathrm{d}t + 0.2\mathrm{d}W_t, & \text{if } 0.5T < t < T, \end{cases} \tag{38}$$

where $\Delta_t = t_{i+1} - t_i$ is the increment in the given time period $[0, T]$.

Similarly, we study the performance of (34), under the assumption that there is no change point, using the SDE

$$\mathrm{d}X_t = \left(2.5 + 1.2\sqrt{2}\cos\left(\frac{\pi t}{2\Delta_t}\right) - X_t\right)\mathrm{d}t + 0.2\mathrm{d}W_t, \quad \text{if } 0 < t < T. \tag{39}$$

We choose $T = 5,\ 10,\ 20$ and 50, and $\Delta_t = 1/252$ and the starting point $X_0 = 0.05$. Simulated sample paths for the processes (36)–(39) with different time periods ($T = 5$ and 50) are shown in Figs. 1, 2, 3 and 4.

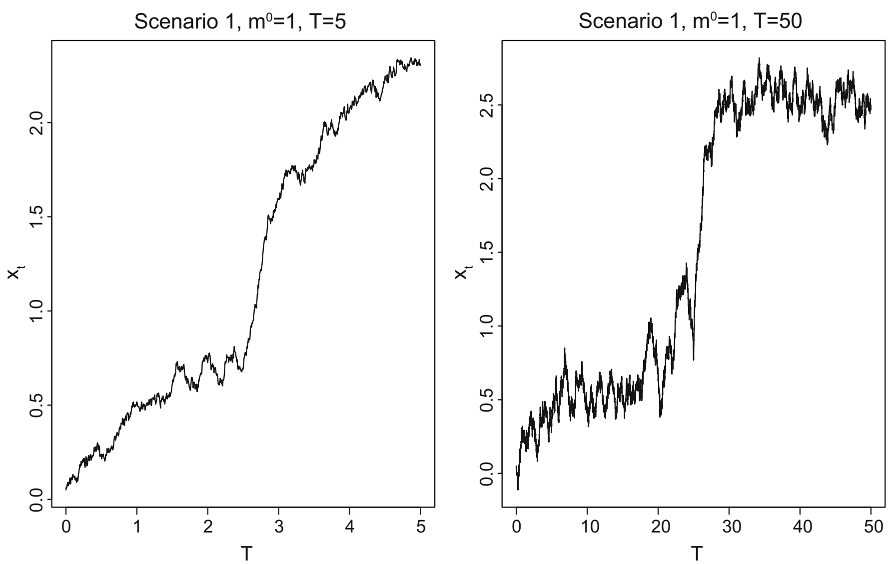

Fig. 1 Sample series for (36) under scenario 1 with one change point

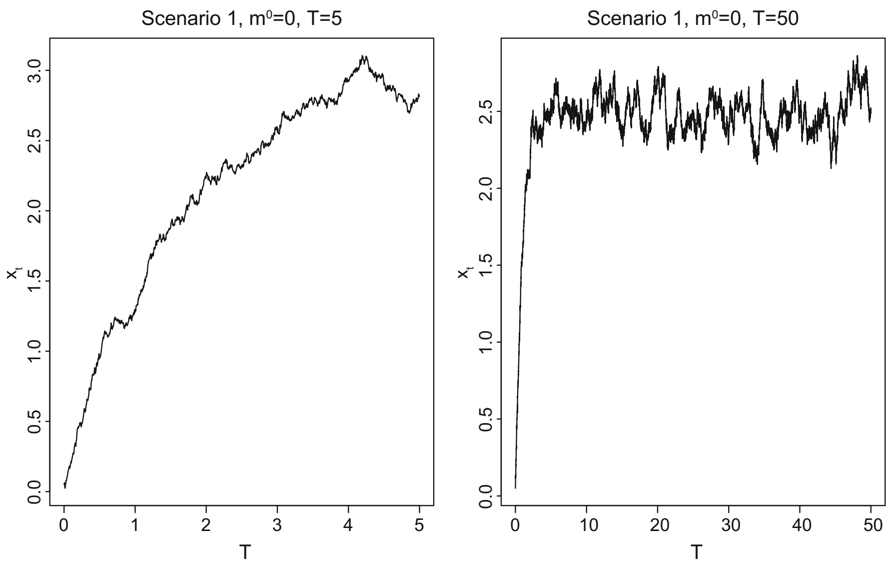

Fig. 2 Sample series for (37) under scenario 1 without a change point

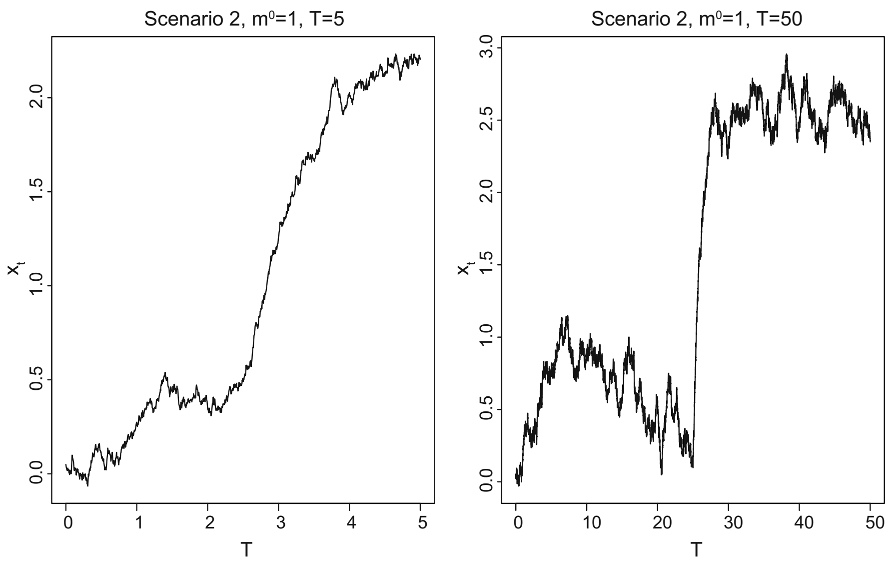

Fig. 3 Sample series for (38) under scenario 2 with one change point

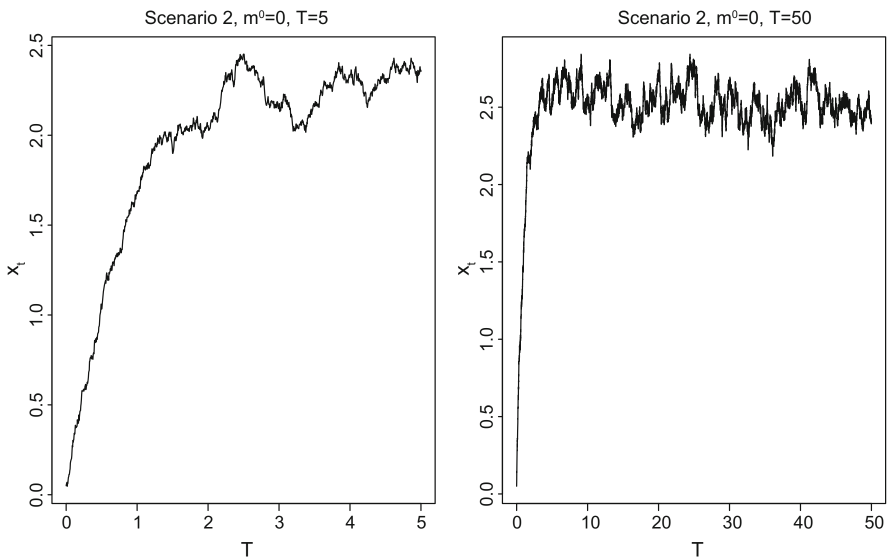

Fig. 4 Sample series for (39) under scenario 2 without a change point

Table 1 Mean and MSE of $\hat{s}$ under scenario 1 (36)

T	LSSE method		MLL method	
	Mean	MSE	Mean	MSE
5	0.4986794	0.0003834606	0.4987587	0.0003823898
10	0.499711	0.0001417382	0.4997503	0.0001417427
20	0.5004065	8.622138×10^{-5}	0.5004069	8.623254×10^{-5}
50	0.4999693	7.983829×10^{-6}	0.4999696	7.982264×10^{-6}

6.1.2 Discussion of simulation results

Estimating the rate of change point

We first look at the performance of (25) and (32) in estimating the corresponding rate of change point in (36) and (38), respectively. Note that in each iteration, we apply (25) and (32) to the same simulated process; hence, results from these two methods are expected to be close to each other. The mean and mean-squared error (MSE) of the estimated rate of change point $\hat{s}$ for (36) based on LSSE and MLL methods are summarised in Table 1, and the results for (38) are displayed in Table 2.

To illustrate further the simulated results, the corresponding histograms for $\hat{s}$ in (36) are depicted in Fig. 5 under scenario 1, and in Fig. 6 under scenario 2.

Estimating and testing the number of change points

Here, we evaluate the performance of the proposed method by utilising the percent accuracy (PA) metric defined by

Table 2 Mean and MSE of $\hat{s}$ under scenario 2 (38)

T	LSSE method		MLL method	
	Mean	MSE	Mean	MSE
5	0.4992968	0.0001146825	0.4996579	0.0001042964
10	0.5003373	1.884511×10^{-5}	0.500502	2.130748×10^{-5}
20	0.5002268	6.765125×10^{-6}	0.5002948	6.991292×10^{-6}
50	0.5001443	1.750852×10^{-6}	0.5001291	1.742431×10^{-6}

$$\text{PA}(m^0) = \frac{1}{1000} \sum_{i=1}^{1000} I_{(\hat{m}_i = m^0)} \times 100\%,$$

where $\hat{m}_i$ is the estimated number of change points in the ith iteration. Note that

$$1 - \text{PA}(0) = \frac{1}{1000} \sum_{i=1}^{1000} I_{(\mathcal{IC}(\hat{m}_i=0) \geq \mathcal{IC}(\hat{m}_i=1))},$$

which is the empirical significance level. Further,

$$\text{PA}(1) = \frac{1}{1000} \sum_{i=1}^{1000} I_{(\mathcal{IC}(\hat{m}_i=0) \geq \mathcal{IC}(\hat{m}_i=1))},$$

which is the empirical power of the proposed test.

As stated in the previous subsection, we aim to assess the performance of (34); and for this purpose we use the criteria $\phi(T) = \log T$ and $\phi(T) = \log(T/\Delta_t)$. Moreover, for $h(p)$, we also consider two cases: $h(p) = p + 1$ and $h(p) = p + 2$ for whether or not there is a potential change in the diffusion coefficient σ specified in (1). Consequently, we make a comparison on the basis of four penalty criteria: $(p + 1) \log T$, $(p + 2) \log T$, $(p + 1) \log(T/\Delta_t)$ and $(p + 2) \log(T/\Delta_t)$. The results are reported in Tables 3, 4, 5 and 6 for each of the scenarios.

From Tables 1 and 2 as well as the plotted histograms, we see that both proposed methods (25) and (32) estimate very accurately the exact rate of change point ($s^0 = 0.5$). In addition, as the time period T increases from 10 to 50, the MSEs of the two estimators decrease. These outcomes confirm the theoretical findings for the asymptotic consistency of our two proposed methods.

For the estimated number $\hat{m}$ of change points, one could see that, when there exists one change point in the model, (34) gives a high empirical power in both scenarios with different penalty criteria and time periods; see Tables 3 and 5. Within the penalty criteria employed, $\phi(T) = \log T$ provides slightly better empirical power than that of $\phi(T) = \log(T/\Delta_t)$. When there is no change point, Tables 4 and 6 reveal that the empirical significance levels, under different penalty criteria, decrease as T increases. These results also imply that our proposed method is asymptotically consistent.

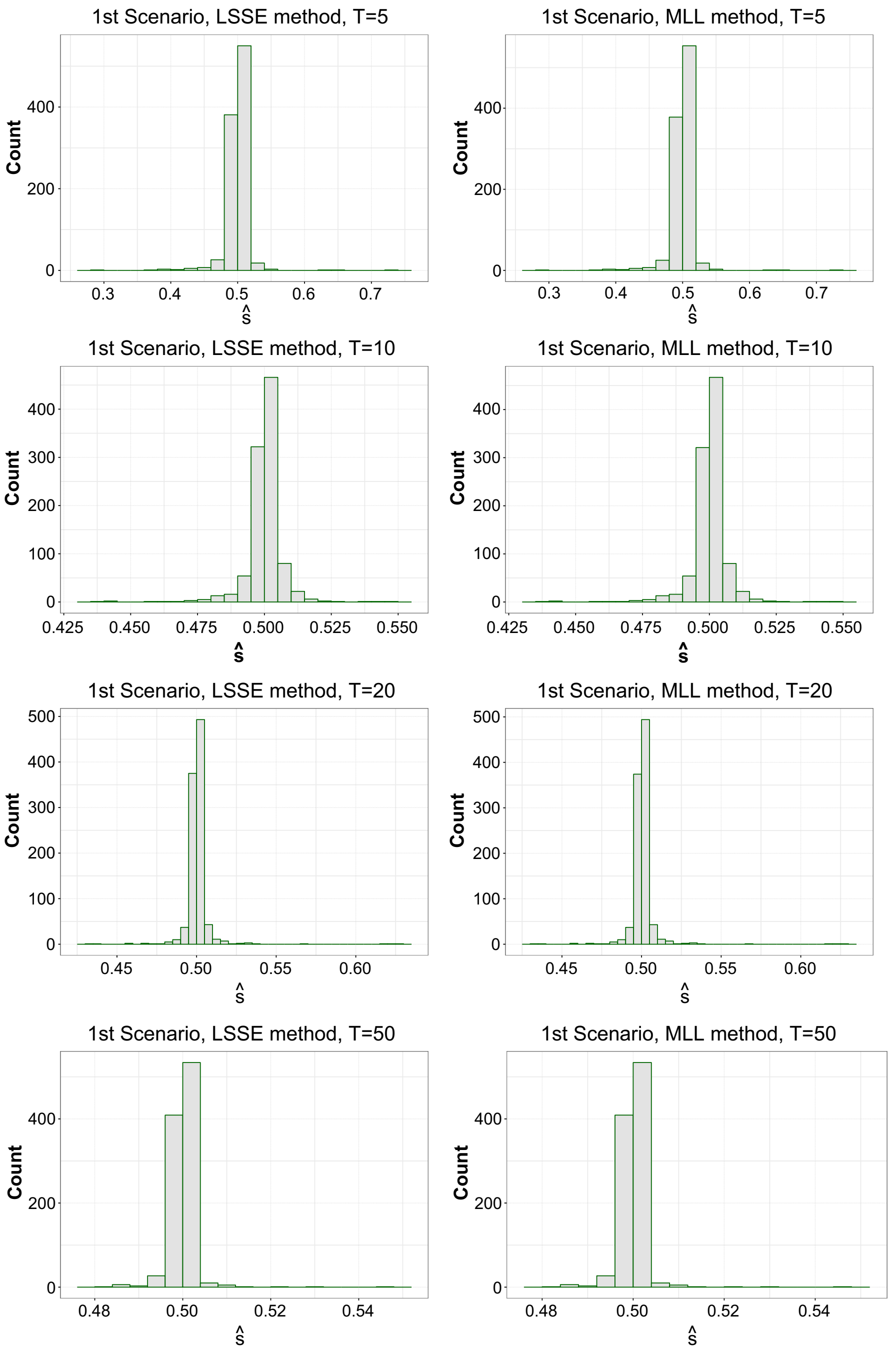

Fig. 5 Histogram of $\hat{s}$ for scenario 1 (36)

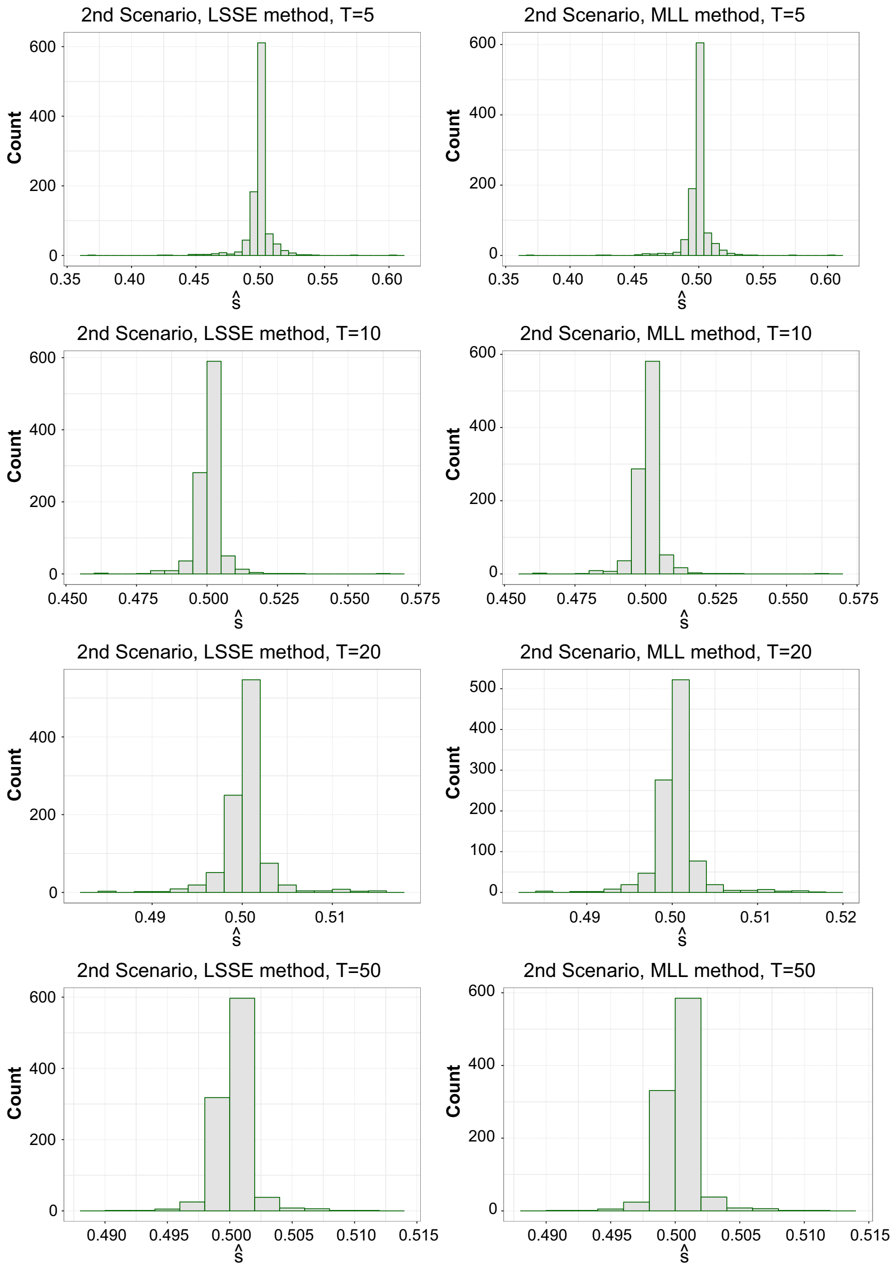

Fig. 6 Histogram of $\hat{s}$ for scenario 2 (38)

Table 3 Empirical power of the test (in %), under scenario 1 (36)

T	$(p+1)\log T$	$(p+2)\log T$	$(p+1)\log(T/\Delta_t)$	$(p+2)\log(T/\Delta_t)$
$T=5$	100	100	98.9	96.5
$T=10$	100	100	99.4	97.3
$T=20$	100	100	99.7	97.5
$T=50$	100	100	99.6	97.2

Table 4 Empirical significance level (in %), under scenario 1 (37)

T	$(p+1)\log T$	$(p+2)\log T$	$(p+1)\log(T/\Delta_t)$	$(p+2)\log(T/\Delta_t)$
$T=5$	83.1	67.6	1.5	0.1
$T=10$	67.5	46.8	0.6	0
$T=20$	54.4	17.9	0	0
$T=50$	19.8	5.2	0	0

Table 5 Empirical power of the test (in %), under scenario 2 (38)

T	$(p+1)\log T$	$(p+2)\log T$	$(p+1)\log(T/\Delta_t)$	$(p+2)\log(T/\Delta_t)$
$T=5$	100	100	100	100
$T=10$	100	100	100	100
$T=20$	100	100	100	100
$T=50$	100	100	100	100

Table 6 Empirical significance level (in %), under scenario 2 (39)

T	$(p+1)\log T$	$(p+2)\log T$	$(p+1)\log(T/\Delta_t)$	$(p+2)\log(T/\Delta_t)$
$T=5$	78.7	55.4	0.4	0
$T=10$	57.2	29.5	0	0
$T=20$	38.5	13.8	0	0
$T=50$	10.8	2.6	0	0

Amongst the 4 penalty criteria, we observe that when $\phi(T) = \log T$, the empirical significance level is relatively high when T is small, whilst the empirical significance level decreases when we change $h(p)$ from $p+1$ to $p+2$. This outcome tells us that it would be more appropriate in this case to use a penalty criterion that is larger than $h(p)\phi(T) = (p+1)\log T$ for better estimation. On the other hand, when using $\phi(T) = \log(T/\Delta_t)$, the performance is significantly improved compared to that of $\phi(T) = \log T$. In both scenarios, one could see that when the time period is small ($T = 5$ and $T = 10$), the empirical significance level of the proposed method is relatively high when using $\phi(T) = \log T$, but decreases to almost 0% when $\phi(T) = \log(T/\Delta_t)$.

Overall, based on the results in Tables 3, 4, 5 and 6 for different cases, we find that for a bigger T, $\hat{m}$ obtained via (34) under four different penalty criteria all perform consistently in estimating the number of change points. However, when T is small, the performance based on the criterion $h(p)\phi(T) = (p+1)\log(T/\Delta_t)$ is efficient and stable vis-á-vis the other criteria in each case. This suggests that $h(p)\phi(T) = (p+1)\log(T/\Delta_t)$ is appropriate for this simulation study.

6.2 Implementation on observed financial market data with discussion

We apply the estimation methods (25), (32) and (34) to two different financial market data series. For each series, we fit the process with the following two different mean-reverting OU processes with one change point.

$$\mathrm{d}X_t = \begin{cases} (\mu_1^{(1)} - a^{(1)} X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, & \text{if } 0 < t < sT, \\ (\mu_1^{(2)} - a^{(2)} X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, & \text{if } sT < t < T, \end{cases} \tag{40}$$

$$\mathrm{d}X_t = \begin{cases} (\mu_1^{(1)} + \sqrt{2}\cos(\frac{\pi t}{2\Delta_t})\mu_2^{(1)} - a^{(1)} X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, & \text{if } 0 < t < sT \\ (\mu_1^{(2)} + \sqrt{2}\cos(\frac{\pi t}{2\Delta_t})\mu_2^{(2)} - a^{(1)} X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, & \text{if } sT < t < T, \end{cases} \tag{41}$$

where X_t is the target of interest (i.e., spot price, log-transformed spot price, daily return, etc.) at time t. The no-change-point versions of (40) and (41) are

$$\mathrm{d}X_t = (\mu_1 - aX_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad \text{if } 0 < t < T, \tag{42}$$

$$\mathrm{d}X_t = \left(\mu_1 + \sqrt{2}\cos\left(\frac{\pi t}{2\Delta_t}\right)\mu_2 - aX_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad \text{if } 0 < t < T. \tag{43}$$

For (40) and (41), we apply (25) and (32) to estimate the unknown change point, whilst for (42) and (43), we train the MLE of drift parameters based on the entire time period. Then, we use (34) to test the existence of a change point. In our formulation, σ is assumed to remain unchanged for the entire time period, and thus it may be estimated using the data's realised volatility, i.e., $\hat{\sigma} = (\sum_{t_i \in [0,T]} (X_{t_{i+1}} - X_{t_i})^2 / T)^{1/2}$. Alternatively, one may fit the data series to the model, take the standard error of the residuals and then divide it by $\sqrt{\Delta_t}$ (see Smith 2010).

When σ is time dependent, EWMA- and GARCH-type volatility estimation methods may be used. However, under this situation the MLE and related asymptotic properties established in Dehling et al. (2010) as well as the asymptotic properties derived in this paper may need re-evaluation as they are all based on the time-independent assumption of the diffusion coefficient σ. In this paper, we consider σ to be time independent and after the estimated results are obtained, the most suitable model is chosen as the one yielding the least SIC value. In addition, we also report the log-likelihood for comparison. To this end, we let $\mathcal{LL}_0$ and $\mathcal{LL}_1$ denote the log-likelihood under H_0 and H_1, respectively.

Table 7 Change-point detection results for the WTI crude oil prices

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2014-09-26	2014-09-26	1	11.89	0.72	3.87	12.39
(41)	2014-09-26	2014-09-26	1	9.96	3.99	9.96	12.76

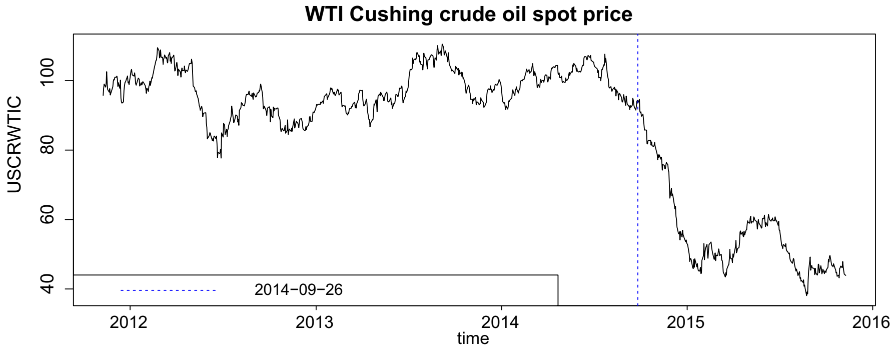

Fig. 7 WTI Cushing crude oil spot prices (09 November 2011–09 November 2015)

6.2.1 Application to West Texas Intermediate Cushing crude oil spot price

The first data set is the time series of West Texas Intermediate Cushing crude oil spot prices, which was first described in Sect. 1.2.1. Our interest in this data set is justified by the fact that, as mentioned in Sect. 1.2.1, it is often being considered as a benchmark in oil pricing. Furthermore, the modelling of and point-change detection in commodity prices are important in the valuation of commodity derivatives and risk management of portfolios with large commodity holdings.

For our preliminary attempt, we set the WTI crude oil spot price to be our target of interest. To fit the model, we choose $T = 4$ and so $\Delta_t = 4/1008$. With the data fitted to the two candidate models using the proposed methods, we display the results shown in Table 7 and Fig. 7.

Looking at Table 7 and Fig. 7, one can see that the value of the log-likelihood increases as the number of the coefficients in the model increases. Further, both candidate models confirm the existence of a change point ($\hat{m} = 1$) during this time period. Under both models, the detected change point is the same (i.e., 26 September 26 2014) for the proposed methods. On the other hand, despite the log-likelihood comparison showing that the periodic mean-reverting model (41) produces higher log-likelihood than the classical OU process with change point (40), the SIC, nonetheless, suggests that (40) is more appropriate than (41) for this data series.

As suggested in Chen (2010), we also analyse the log-transformed WTI crude oil spot prices. We examine the log-transformed WTI crude oil spot prices as our target of interest and re-apply the proposed techniques. The results are shown in Table 8 and Fig. 8. It is worth noting that the detected change points for this log-transformed spot prices are still the same (i.e., 26 September 2014), as well as the behaviour of log-

Table 8 Change-point detection for the log-transformed WTI crude oil prices

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2014-09-26	2014-09-26	1	8.15	0.82	11.37	12.18
(41)	2014-09-26	2014-09-26	0	13.68	4.26	14.12	12.20

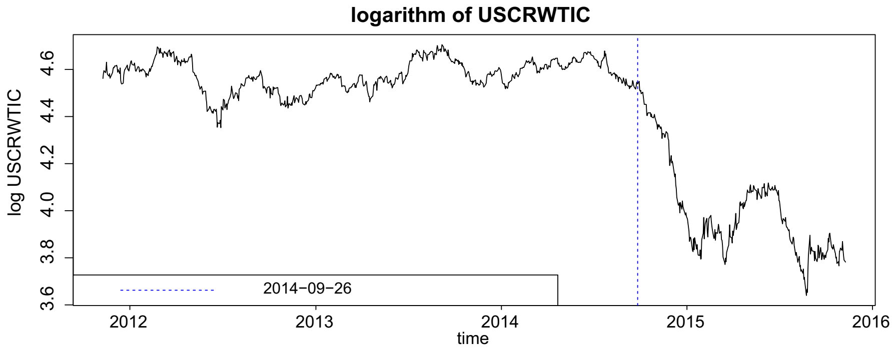

Fig. 8 Log-transformed WTI Cushing crude oil spot prices (09 November 2011–09 November 2015)

likelihood (increase as the number of coefficients in the model increases); although, this time around, the results based on (41) fail to pass the test on the existence of a change point ($\hat{m} = 0$). Judging from the SIC numbers, model (40) is again better than model (41). This suggests that (40), with the change point occurring on 26 September 2014, is more suitable for both the WTI Cushing crude oil price data and its log-transformed series.

Based on our *primary statistics of interest*, defined here as the estimated change point $\hat{\tau}$ and the associated MLE $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$) obtained from the proposed methods, we use model (40) to generate a simulated crude oil price data set and log-transformed crude oil price data series. Additionally, we also generate a simulated data series based on (42) for comparison. The two simulated series are presented side-by-side in Fig. 9. By comparing the simulated series in Figs. 9 and 10 to the original series (shown in Figs. 7, 8), we see that the simulated series generated by (40), with the change point occurring on 26 September 2014, is closer to the original series than the one generated by (42). This confirms the efficiency of the proposed methods for this series. In forecasting the future value (after 26 September 2014) of WTI crude oil spot prices, the practitioner should use the second relation in Eq. (40).

6.2.2 Application to XAU currency

We apply as well the proposed methods to the XAU currency data, which are described in Sect. 1.2.2. This data set refers to the time series of prices, in US dollars, for a troy ounce of gold. Our interest in this kind of data set is motivated by the pricing of currency swaps, futures and options.

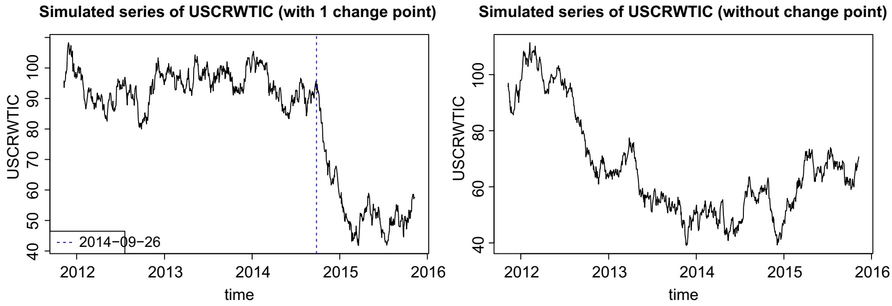

Fig. 9 Simulated series of WTI Cushing crude oil spot price (09 November 2011–09 November 2015)

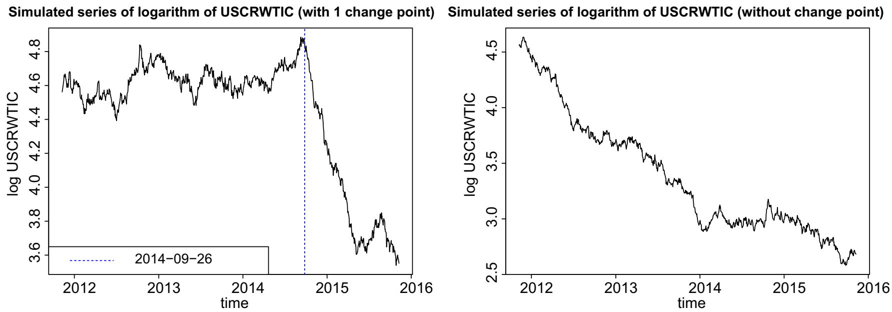

Fig. 10 Simulated series of log-transformed WTI Cushing crude oil spot price (09 November 2011–09 November 2015)

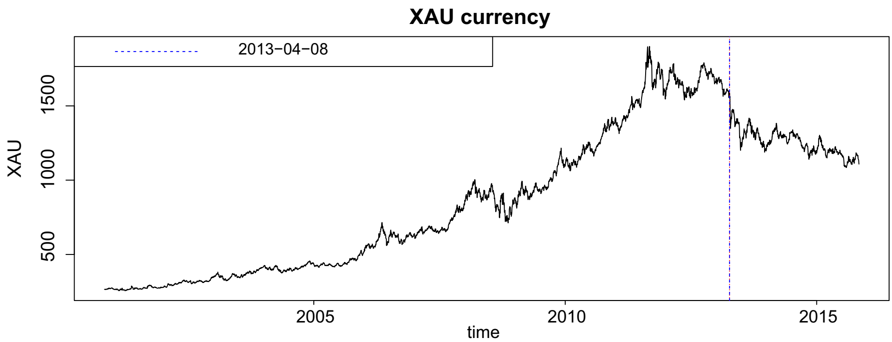

Fig. 11 Evolution of the XAU currency (03 November 2000–04 November 2015)

Figure 11 shows that the trend of XAU currency series changed over time. As in the previous study, we take both the XAU currency and its logarithm as our targets of interest. To fit the data with the candidate models, we choose $T = 15$ and $\Delta_t = 15/3913$. The results for the XAU currency are depicted in Table 9 and Fig. 11, whilst the results for the log-transformed XAU currency are shown in Table 10 and Fig. 12.

For the original XAU currency, one can see that log-likelihood increases as the number of coefficients in the model increases. Further, (40) successfully detects one

Table 9 Change-point detection for XAU currency (15 years)

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2013-04-08	2013-04-08	1	12.62	1.43	7.83	13.67
(41)	2013-04-08	2013-04-08	0	13.31	1.81	23.01	21.18

Table 10 Change-point detection for the log-transformed XAU currency (15 years)

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2011-08-19	2011-08-19	0	7.08	3.25	18.92	10.04
(41)	2011-08-19	2011-08-19	0	7.41	3.46	34.80	17.89

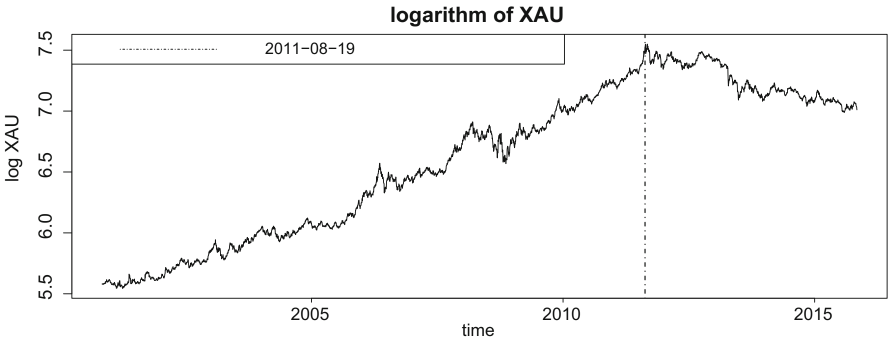

Fig. 12 Log-transformed XAU currency (03 November 2000–04 November 2015)

change point on 08 April 2013, whilst (41) fails to detect the change point ($\hat{m} = 0$). However, the SIC comparison shows that the classical OU process (SIC= 7.83) is more appropriate than the periodic mean-reverting model (41) (SIC= 21.18); we, therefore, select (40) with the change point on 08 April 2013 as the suitable model for this series.

For the log-transformed XAU currency, Fig. 12 illustrates visually that the series is smoother than the original series (see Fig. 11), and the potential change in the series becomes less clear. Applying the proposed methods, both (40) and (41) detect the same change point on 19 August 2011, which is the time when the log-transformed XAU currency almost reaches the highest value. Although the comparison of log-likelihood functions indicates that imposing a change point in the model can produce higher log-likelihood and (40) is still more suitable than (41) in terms of the SIC, both models fail to pass the test for the existence of change point ($\hat{m} = 0$). These results suggest that for this log-transformed data series, imposing a change point into the model is not as efficient as compared to the original XAU currency series.

Similar to the previous study, we also generate, employing the estimated values of the primary statistics of interest, some simulated series based on (40) and (41) for both XAU currency and its log transform. The results are given in Fig. 13 for the

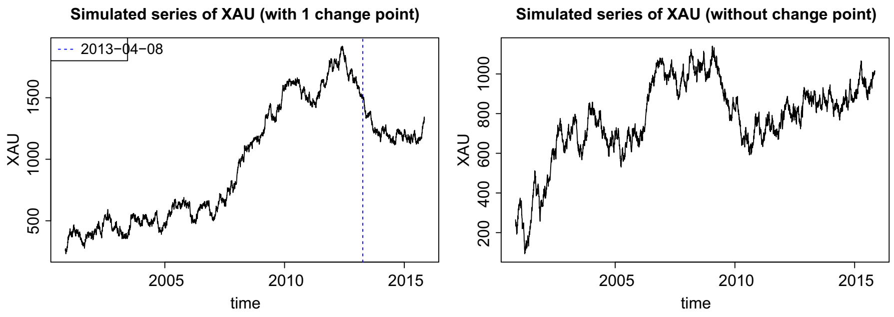

Fig. 13 Simulated series of XAU currency (03 November 2000–04 November 2015)

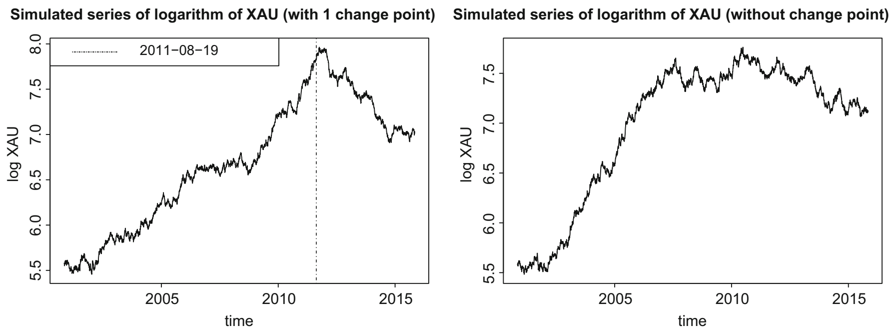

Fig. 14 Simulated series of log-transformed XAU currency (03 November 2000–04 November 2015)

original XAU currency and Fig. 14 for the log-transformed XAU currency, respectively.

For the original XAU currency, the comparison shows that the simulated series generated by (40) with the change point on 08 April 2013 is closer to the original series, especially around the location where the change point happens; this is in contrast to the model without a change point. This result confirms the efficiency of employing the change-point process (40) in improving the estimation of this series. However, the change in the log-transformed series is not clear as the difference between the two simulated series becomes small; they are both close to the original log-transformed series. This means that there may be no need to impose a change point for this log-transformed series; this is because the improvement is not as significant to the one obtained in the original series.

To probe matters on this series further, we change the starting date of the data series to 04 November 2008. This reduces the time period from 15 years to 7 years, and one can see from Fig. 15 that the XAU currency levels in this time period are all higher than \$700. We re-apply the methods to the XAU currency data and its log-transformed series. Table 11 and Fig. 15 show the results for the original XAU currency, and Table 12 and Fig. 16 display the results for the log-transformed XAU currency.

As one can see from the results, for the original XAU currency, the detected change points based on (25) and (32) in both models are still the same as in the 15 years' time

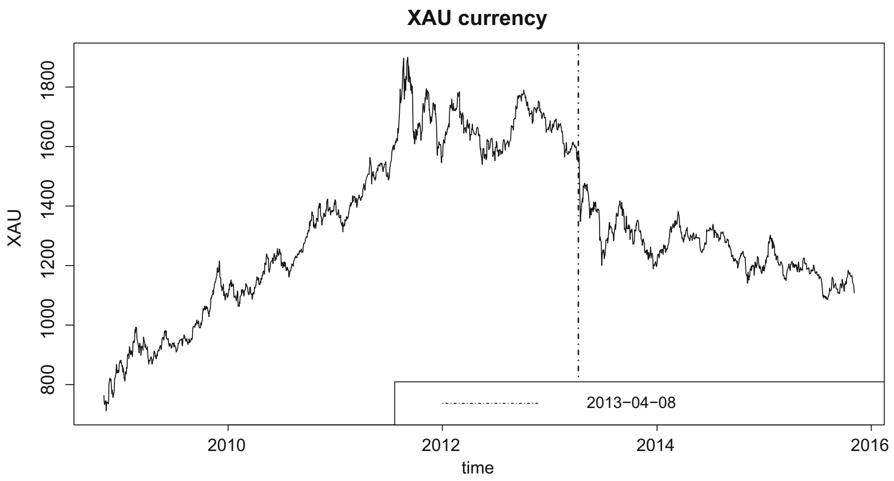

Fig. 15 XAU currency (04 November 2008–04 November 2015)

Table 11 Change-point detection for XAU currency (7 years)

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2013-04-08	2013-04-08	0	9.26	2.28	11.50	10.47
(41)	2013-04-08	2013-04-08	0	9.62	2.56	25.80	17.40

Table 12 Change-point detection for the log-transformed XAU currency (7 years)

Model	LSSE method	MLL method	$\hat{m}$	$\mathcal{LL}_1$	$\mathcal{LL}_0$	$\mathcal{IC}(m=1)$	$\mathcal{IC}(m=0)$
(40)	2011-08-19	2011-08-19	0	6.06	2.95	17.90	9.11
(41)	2009-02-23	2009-02-23	0	7.06	3.31	30.92	15.89

period, suggesting the consistency of the proposed methods. Further, the comparisons of log-likelihood suggest that increasing the number of coefficients in the model is useful in producing higher log-likelihood. This finding is consistent with the previous study. However, after applying (34) to test the existence of change point, we find that both models fail to pass the test. This tells us that using (42) would be enough for modelling the series for this time period. Accordingly, in forecasting the future values of XAU currency, the practitioner does not need to take into account the existence of the change point; hence, he should use model (42).

On the other hand, after applying the log transformation to the XAU currency, the SIC reveals that the model (40) is still more suitable than (41) for this case. Moreover, as one can see from the results shown in Table 12, together with the simulated series provided in Figs. 17 and 18, similar outcome as in the case of log-transformed XAU currency with 15 years' time period again suggests that for this log-transformed

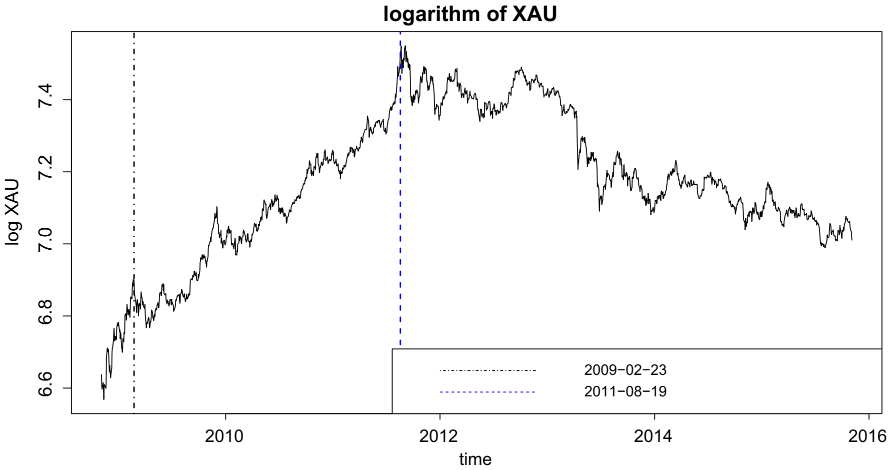

Fig. 16 XAU currency (04 November 2008–04 November 2015)

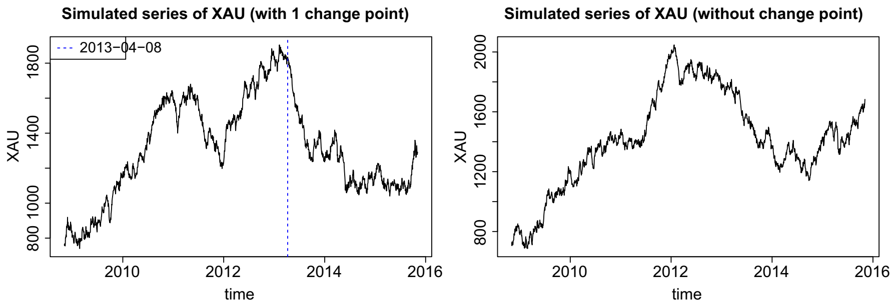

Fig. 17 Simulated series of XAU currency (04 November 2008–04 November 2015)

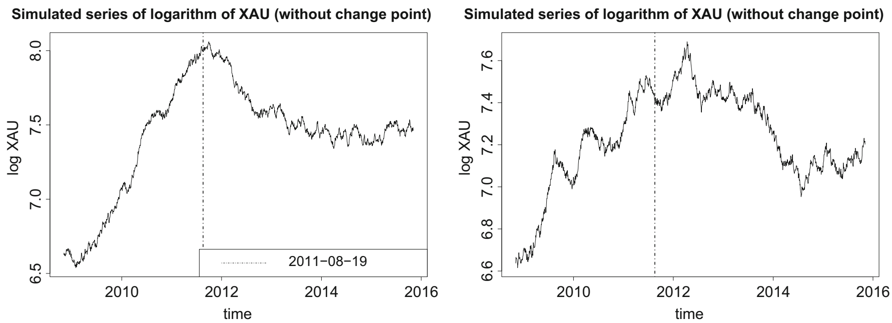

Fig. 18 Simulated series of log-transformed XAU currency (04 November 2008–04 November 2015)

data series, the improvement from employing a model with a change point is not as significant as the one in the original XAU currency series with a 15-years time period.

7 Conclusion

The theoretical results and illustrative examples, involving both simulated and observed data, of this paper were motivated by the practical considerations of point-change detection. Such motivations are driven by applications centred on confirming significant changes in the time-series data so that proper responses and policies could be put in place as in the case of interest rate-setting behaviour of monetary regulatory authority, design of trading strategies in hedging and speculation, appropriate calibration of models in financial product valuation, amongst others. Our main contribution highlighted the development of MLL- and LSSE-based methods in showing the existence or non-existence of a change point and in determining the unknown location whenever such a change point exists. We established the equivalence of the estimators for the change point under the two methods. In addition, we provided conditions so that our estimators for the change-point location are asymptotically consistent, which in turn aided the design of an efficient implementation algorithm. Our work certainly gives impetus for the investigation and development of methodology suited in tackling the multiple-change-point problem, which is what we commonly encounter in practice. Our research results aim to lay down the groundwork so that further modifications could be made and new ideas could be adopted in making further progress in point-change detection involving time series models with more complex and elaborate dynamics and stylised features.

Acknowledgements F. Chen and R. Mamon would like to acknowledge the financial support, via a Fields Postdoctoral Fellowship for F. Chen, provided by the Fields Institute for Research in Mathematical Sciences, Toronto, Ontario, Canada, where this research was initially conceived and partially conducted. R. Mamon wishes to express his appreciation for the hospitality of the Division of Physical Sciences and Mathematics, University of the Philippines Visayas, where certain revisions of this paper were made during his academic visit both as an Adjunct Professor and a DOST-PCIEERD Balik Scientist for the Philippine government. Both S. Nkurunziza and R. Mamon thank the Natural Sciences and Engineering Research Council of Canada for the support of their research. Likewise, all authors are grateful to Charmaine Dean (Western University's Dean of Science) and Matt Davison (Canada Research Chair in Quantitative Finance) for their generous financial support on this research collaboration. Some valuable feedback from Charmaine Dean and Reg Kuperger are very much appreciated. Finally, the authors sincerely thank the Associate Editor and an anonymous referee for their helpful comments.

Appendix A Proofs of Propositions 3, 4, 6 and 7

Appendix A.1 Preliminaries for the proofs of Propositions 3 and 4

In this section, we let $||A||$ denote the Euclidean norm for a vector A and $||B|| = \sqrt{\text{trace}(B'B)}$ for a matrix B. Further, let $\hat{u}_i$ be the residual of the ith element in (24) based on the estimated change point $\hat{\tau}$. That is, $\hat{u}_i = Y_i - Z_i\hat{\theta}^{(j)} = Z_i\theta^{(j)} - Z_i\hat{\theta}^{(j)} + u_i$, $j = 1$ for $0 < i \leq \hat{\tau}$ and $\tau = \hat{\tau}$ estimated by (25). Similarly, let $\hat{u}_i^0$ be the residual of the ith element in (24) based on the exact change point τ^0 and the associated MLE of $\theta^{(j)}$ denoted by $\theta^{(j,0)}$, $j = 1, 2$. Without loss of generality, we assume that $\hat{\tau} > \tau^0$.

The proofs of Propositions 3 and 4 both rely on investigating the behaviour of

$$\phi\left(\sum_{t_i\in[0,T]}\hat{u}_i'\hat{u}_i-\sum_{t_i\in[0,T]}\hat{u}_i^{0\prime}\hat{u}_i^0\right). \tag{44}$$

We divide the time period $[0, T]$ involved in (44) into 3 sub-intervals: $[0, \tau^0]$, $(\tau^0, \hat{\tau}]$ and $(\hat{\tau}, T]$. Then, by substituting the expressions for $\hat{u}_i$ and $\hat{u}_i^0$ into (44) and applying the identity $(a+b)^2 = a^2 + 2ab + b^2$ to expand the quadric error terms, we have

$$\begin{aligned}\text{Expression}(44) = \phi\Bigg(&\sum_{t_i\in[0,\tau^0]}(Z_i(\theta^{(1)}-\hat{\theta}^{(1)}))^2+\sum_{t_i\in(\tau^0,\hat{\tau}]}(Z_i(\theta^{(2)}-\hat{\theta}^{(1)}))^2\\&+\sum_{t_i\in(\hat{\tau},T]}(Z_i(\theta^{(2)}-\hat{\theta}^{(2)}))^2\Bigg) &(45)\\&-\phi\sum_{t_i\in[0,\tau^0]}(Z_i\theta^{(1)}-Z_i\hat{\theta}^{(1,0)})^2-\phi\sum_{t_i\in(\tau^0,T]}(Z_i\theta^{(2)}-Z_i\hat{\theta}^{(2,0)})^2\\& &(46)\\&+2\phi\Bigg(\sum_{t_i\in[0,\tau^0]}u_iZ_i'(\hat{\theta}^{(1,0)}-\hat{\theta}^{(1)})+\sum_{t_i\in(\tau^0,\hat{\tau}]}u_iZ_i'(\hat{\theta}^{(2,0)}-\hat{\theta}^{(1)})\\&+\sum_{t_i\in(\hat{\tau},T]}u_iZ_i'(\theta^{(2,0)}-\hat{\theta}^{(2)})\Bigg), &(47)\end{aligned}$$

where ϕ is a positive scalar to be defined later, $\hat{\theta}^{(1)} = Q^{-1}_{(0,\hat{\tau})}\tilde{R}_{(0,\hat{\tau})}$, $\hat{\theta}^{(2)} = Q^{-1}_{(\hat{\tau},T)}\tilde{R}_{(\hat{\tau},T)}$, $\hat{\theta}^{(1,0)} = Q^{-1}_{(0,\tau^0)}\tilde{R}_{(0,\tau^0)}$, and $\hat{\theta}^{(2,0)} = Q^{-1}_{(\tau^0,T)}\tilde{R}_{(\tau^0,T)}$, respectively.

It follows as well from the proof of Corollary 3.1 in Nkurunziza and Zhang (2016) (see also Proposition 4.1 in Dehling et al. 2010) that $\tilde{R}_{(0,\tau^0)} = Q_{(0,\tau^0)}\theta^{(1)} + \sigma R_{(0,\tau^0)}$, $\tilde{R}_{(\tau^0,\hat{\tau})} = Q_{(\tau^0,\hat{\tau})}\theta^{(2)} + \sigma R_{\tau^0,\hat{\tau}}$, $\tilde{R}_{(\tau^0,T)} = Q_{(\tau^0,T)}\theta^{(2)} + \sigma R_{(\tau^0,T)}$ and $\tilde{R}_{(\hat{\tau},T)} = Q_{(\hat{\tau},T)}\theta^{(2)} + \sigma R_{(\hat{\tau},T)}$. Therefore, $\hat{\theta}^{(1,0)} - \theta^{(1)} = \sigma Q^{-1}_{(0,\tau^0)}R_{(0,\tau^0)}$, and $\hat{\theta}^{(2,0)} - \theta^{(2)} = \sigma Q^{-1}_{(\tau^0,T)}R_{(\tau^0,T)}$. In this case, we have expression

$$\begin{aligned}(46) = -\phi\sigma^2\Bigg(&R'_{(0,\tau^0)}Q^{-1}_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]}Z_i'Z_iQ^{-1}_{(0,\tau^0)}R_{(0,\tau^0)}\\&+R'_{(\tau^0,T)}Q^{-1}_{(\tau^0,T)}\sum_{t_i\in(\tau^0,T]}Z_i'Z_iQ^{-1}_{(\tau^0,T)}R_{(\tau^0,T)}\Bigg).\end{aligned} \tag{48}$$

Similarly,

$$\hat{\theta}^{(1,0)} - \hat{\theta}^{(1)} = \sigma Q^{-1}_{(0,\tau^0)} R_{(0,\tau^0)} - Q^{-1}_{(0,\hat{\tau})} Q_{(\tau^0,\hat{\tau})}(\theta^{(2)} - \theta^{(1)}) - \sigma Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})}, \tag{49}$$

$$\hat{\theta}^{(2,0)} - \hat{\theta}^{(1)} = \sigma Q^{-1}_{(\tau^0,T)} R_{(\tau^0,T)} - Q^{-1}_{(0,\hat{\tau})} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) - \sigma Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})}, \tag{50}$$

$$\hat{\theta}^{(2,0)} - \hat{\theta}^{(2)} = \sigma Q^{-1}_{(\tau^0,T)} R_{(\tau^0,T)} - \sigma Q^{-1}_{(\hat{\tau},T)} R_{(\hat{\tau},T)}, \tag{51}$$

and expression

$$(47) = 2\phi \left(\sum_{t_i \in [0,\tau^0]} u_i Z_i'(49) + \sum_{t_i \in (\tau^0,\hat{\tau}]} u_i Z_i'(50) + \sum_{t_i \in (\hat{\tau},T]} u_i Z_i'(51) \right). \tag{52}$$

Proof of Proposition 3 Take $\phi = \frac{1}{T}$. In general, (44) is non-positive with probability 1 since by (25), $\hat{\tau}$ is chosen from all possible values in $[0, T]$ to minimise the SSR, whilst τ^0 is just a particular value in $[0, T]$. Hence, it suffices to show that if the rate of τ^0, given by $s^0 = \tau^0/T$, can not be consistently estimated by $\hat{s} = \hat{\tau}/T$, then (44) > 0 with positive probability and thus we have a contradiction.

First, note that in case that the rate of the change point τ^0 cannot be consistently estimated, then with positive probability there exists an $\eta > 0$ such that $\hat{s}T - s^0T > \eta T > L^0$ for large T. In this case, we have (45) $\geq C_1||\theta^{(1)} - \theta^{(2)}||^2$ for some $C_1 > 0$ with positive probability (see Bai and Perron 1998, Lemma 2).

To proceed further, we prove the following inequality. Note that by (8), $\sup_{t\geq 0} E((X_t)^2) \leq K_1$ for some K_1 with $0 < K_1 < \infty$. This implies that for $0 < \tau_1^* < \tau_2^* \leq T$,

$$\int_{\tau_1^*}^{\tau_2^*} E(X_t^2)\mathrm{d}t \leq K_1(\tau_2^* - \tau_1^*). \tag{53}$$

Further, by the Markov inequality and Ito isometry, together with inequality (53) and Assumption 2, we have

$$\begin{aligned}
&\mathrm{P}\left(\frac{1}{\sqrt{\tau_2^* - \tau_1^*}} \left| \int_{\tau_1^*}^{\tau_2{}^*} X_t \mathrm{d}W_t \right| > K^* \right) \leq \frac{E\left(\left| \int_{\tau_1^*}^{\tau_2^*} X_t \mathrm{d}W_t \right|^2 \right)}{(\tau_2^* - \tau_1^*)(K^*)^2} = \frac{\int_{\tau_1^*}^{\tau_2^*} E(X_t^2)\mathrm{d}t}{(\tau_2^* - \tau_1^*)(K^*)^2} \\
&\leq \frac{K_1(\tau_2^* - \tau_1^*)}{(\tau_2^* - \tau_1^*)(K^*)^2} = \frac{K_1}{(K^*)^2}, \mathrm{P}\left(\frac{1}{\sqrt{\tau_2^* - \tau_1^*}} \left| \int_{\tau_1^*}^{\tau_2{}^*} \varphi_i(t)\mathrm{d}W_t \right| > K^* \right) \\
&\leq \frac{E\left(\left| \int_{\tau_1^*}^{\tau_2^*} \varphi_i(t)\mathrm{d}W_t \right|^2 \right)}{(\tau_2^* - \tau_1^*)(K^*)^2} \leq \frac{1}{(K^*)^2}.
\end{aligned}$$

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

$$\frac{1}{\sqrt{\tau_2^* - \tau_1^*}} ||R_{(\tau_1^*, \tau_2^*)}|| = O_p(\log^{a^*} T) \quad \text{or} \quad O_p((\hat{\tau} - \tau^0)^{a^*}) \quad \text{for any } 0 < \tau_1^* < \tau_2^* \le T. \tag{54}$$

Similarly, for the discretised case, $\frac{1}{\sqrt{\tau_2^* - \tau_1^*}} || \sum_{t_i \in (\tau_1^*, \tau_2^*]} Z_i u_i || = O_p(\log^{a^*} T)$ or $O_p((\hat{\tau} - \tau^0)^{a^*})$. Furthermore, it follows from Proposition 2 and the Continuous Mapping Theorem that

$$\left|\left| \frac{1}{T} Q_{(\tau_1^*, \tau_2^*)} \right|\right| = O_p(1) \quad \text{and} \quad ||T Q^{-1}_{(\tau_1^*, \tau_2^*)}|| = O_p(1) \tag{55}$$

for any $\tau_1^*, \tau_2^* \in \{0, \tau^0, \hat{\tau}, T\}$ such that $\tau_1^* < \tau_2^*$. Then, applying the Cauchy–Schwarz inequality to (48) and (52), together with above asymptotic results, we have that (46) and (47) are both $o_p(1)$. Hence, (44) is dominated by (45), which is positive, and thus gives a contradiction. Therefore, $\hat{s} - s^0 \xrightarrow[T \to \infty]{P} 0$. □

Proof of Proposition 4 Write $\phi := \frac{1}{\hat{\tau} - \tau^0}$ and $V_\eta := \{\tau : |\tau - \tau^0| \le \eta T\}$. It follows from Proposition 3 that for each $\eta > 0$, $P(\hat{\tau} \in V_\eta) \xrightarrow[T \to \infty]{} 1$. Therefore, we only need to investigate the sum of squared error $SSE(\hat{\tau})$ for those $\hat{\tau} \in V_\eta$. For $C > 0$, define the set $V_\eta(C) = \{\tau : C < |\tau - \tau^0| < \eta T\}$ and let $\hat{\tau}$ be the estimated change point with the minimum taken over the set $V_\eta(C)$. Then, it suffices to show that the order of these three terms, or one of the three terms is larger than all of the remaining terms in (44), and that leads to a contradiction since the term (44) ≤ 0 with probability 1.

First notice that (45) is $O_p(1)$ instead of $o_p(1)$. Hence, it is difficult to compare it with (44) directly. In this case, we need to factorise the term (44). We observe that

$$\sum_{t_i \in [0, \tau^0]} (Z_i(\theta^{(1)} - \hat{\theta}^{(1)}))^2 = (a1) + (a2) + (a3),$$

where

$$(a1) = (\theta^{(1)} - \theta^{(2)})' Q_{(\tau^0, \hat{\tau})} Q^{-1}_{(0, \hat{\tau})} \sum_{t_i \in [0, \tau^0]} Z_i' Z_i Q^{-1}_{(0, \hat{\tau})} Q_{(\tau^0, \hat{\tau})} (\theta^{(1)} - \theta^{(2)}),$$

$$(a2) = R_{(0, \hat{\tau})} \sigma^2 Q^{-1}_{(0, \hat{\tau})} \sum_{t_i \in [0, \tau^0]} Z_i' Z_i Q^{-1}_{(0, \hat{\tau})} R_{(0, \hat{\tau})}$$

$$(a3) = 2\sigma(\theta^{(1)} - \theta^{(2)})' Q_{(\tau^0, \hat{\tau})} Q^{-1}_{(0, \hat{\tau})} \sum_{t_i \in [0, \tau^0]} Z_i' Z_i Q^{-1}_{(0, \hat{\tau})} R_{(0, \hat{\tau})}.$$

Similarly,

$$\sum_{t_i \in (\tau^0, \hat{\tau}]} (Z_i(\theta^{(2)} - \hat{\theta}^{(1)}))^2 = (a4) + (a5) + (a6),$$

where

$$(a4) = (\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} \sum_{t_i \in (\tau^0, \hat{\tau}]} Z_i' Z_i Q^{-1}_{(0,\hat{\tau})} Q_{(0,\tau^0)} (\theta^{(1)} - \theta^{(2)}),$$
$$(a5) = R_{(0,\hat{\tau})} \sigma^2 Q^{-1}_{(0,\hat{\tau})} \sum_{t_i \in (\tau^0, \hat{\tau}]} Z_i' Z_i Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})},$$
$$(a6) = 2\sigma (\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} \sum_{t_i \in (\tau^0, \hat{\tau}]} Z_i' Z_i Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})}.$$

and

$$\sum_{t_i \in (\hat{\tau}, T]} (Z_i \theta^{(2)} - Z_i \hat{\theta}^{(2)})' (Z_i \theta^{(2)} - Z_i \hat{\theta}^{(2)})$$
$$= \sigma^2 R_{(\hat{\tau}, T)} Q^{-1}_{(\hat{\tau}, T)} \sum_{t_i \in (\hat{\tau}, T]} Z_i' Z_i Q^{-1}_{(\hat{\tau}, T)} R_{(\hat{\tau}, T)} = (a7).$$

One could see that $\phi(a1)$, $\phi(a4)$ and $\phi(a6)$ are all of order $o_p(1)$. Additionally, by Theorem A.1 in Tobing and McGilchrist (1992), we have that for large T,

$$Q^{-1}_{(0,\hat{\tau})} = Q^{-1}_{(0,\tau^0)} + O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right). \tag{56}$$

Therefore,

$$R'_{(0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} \sum_{t_i \in [0,\tau^0]} Z_i' Z_i Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} - R'_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} \sum_{t_i \in [0,\tau^0]} Z_i' Z_i Q^{-1}_{(0,\tau^0))} R_{(0,\tau^0))}$$
$$= (R_{(0,\tau^0)} + R_{(\tau^0,\hat{\tau})})' \left(Q^{-1}_{(0,\tau^0)} + O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right)\right) \sum_{t_i \in [0,\tau^0]} Z_i' Z_i \left(Q^{-1}_{(0,\tau^0)}\right.$$
$$\left. + O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right)\right) (R_{(0,\tau^0)} + R_{(\tau^0,\hat{\tau})})$$
$$- R_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} \sum_{t_i \in [0,\tau^0]} Z_i' Z_i Q^{-1}_{(0,\tau^0))} R_{(0,\tau^0)}$$
$$= O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right) R_{(0,\tau^0)} \sum_{t_i \in [0,\tau^0]} Z_i' Z_i Q^{-1}_{(0,\tau^0))} R_{(0,\tau^0)}$$
$$+ O_p\left(\frac{(\hat{\tau} - \tau^0)^2}{T^4}\right) R_{(0,\tau^0)} \sum_{t_i \in [0,\tau^0]} Z_i' Z_i R_{(0,\tau^0)}$$

$$
\begin{aligned}
&+2R_{(0,\tau^0)}Q^{-1}_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i Q^{-1}_{(0,\tau^0)}R_{(\tau^0,\hat{\tau})}\\
&+O_p\left(\frac{\hat{\tau}-\tau^0}{T^2}\right)R_{(0,\tau^0)}Q^{-1}_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i R_{(\tau^0,\hat{\tau})}\\
&+O_p\left(\frac{\hat{\tau}-\tau^0}{T^2}\right)R_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i Q^{-1}_{(0,\tau^0)}R_{(\tau^0,\hat{\tau})}\\
&+O_p\left(\frac{(\hat{\tau}-\tau^0)^2}{T^4}\right)R_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i R_{(\tau^0,\hat{\tau})}\\
&+R_{(\tau^0,\hat{\tau})}Q^{-1}_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i Q^{-1}_{(0,\tau^0)}R_{(\tau^0,\hat{\tau})}\\
&+O_p\left(\frac{\hat{\tau}-\tau^0}{T^2}\right)R_{(\tau^0,\hat{\tau})}Q^{-1}_{(0,\tau^0)}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i R_{(\tau^0,\hat{\tau})}\\
&+O_p\left(\frac{(\hat{\tau}-\tau^0)^2}{T^4}\right)R_{(\tau^0,\hat{\tau})}\sum_{t_i\in[0,\tau^0]} Z_i'Z_i R_{(\tau^0,\hat{\tau})}. \qquad (57)
\end{aligned}
$$

Since $\hat{\tau}-\tau^0 \leq \eta T$ for each $\eta > 0$, and using the asymptotic results in the proof of Proposition 3 with small enough η, we have $\phi\sigma^2$ (57) $= o_p(1)$. Similarly,

$$
\begin{aligned}
\phi\sigma^2 \Bigg(& R_{(\hat{\tau},T)}Q^{-1}_{(\hat{\tau},T)}\sum_{t_i\in(\hat{\tau},T]} Z_i'Z_i Q^{-1}_{(\hat{\tau},T)}R_{(\hat{\tau},T)}\\
&- R_{(\tau^0,T)}Q^{-1}_{(\tau^0,T)}\sum_{t_i\in(\tau^0,T]} Z_i'Z_i Q^{-1}_{(\tau^0,T)}R_{(\tau^0,T)}\Bigg) = o_p(1).
\end{aligned}
$$

Consequently, $(a2)+(a4)+(45)=o_p(1)$. For $(a3)$, we have

$$
\begin{aligned}
&(\theta^{(1)}-\theta^{(2)})'Q_{(0,\tau^0)}Q^{-1}_{(0,\hat{\tau})}\frac{\sum_{t_i\in(\tau^0,\hat{\tau}]} Z_i'Z_i}{\hat{\tau}-\tau^0}Q^{-1}_{(0,\hat{\tau})}Q_{(0,\tau^0)}(\theta^{(1)}-\theta^{(2)})\\
&=(\theta^{(1)}-\theta^{(2)})'Q_{(0,\tau^0)}\left(Q^{-1}_{(0,\tau^0)}+O_p\left(\frac{\hat{\tau}-\tau^0}{T^2}\right)\right)\frac{\sum_{t_i\in(\tau^0,\hat{\tau}]} Z_i'Z_i}{\hat{\tau}-\tau^0}\Bigg(Q^{-1}_{(0,\tau^0)}\\
&\quad+O_p\left(\frac{\hat{\tau}-\tau^0}{T^2}\right)\Bigg)Q_{(0,\tau^0)}(\theta^{(1)}-\theta^{(2)})\\
&=(\theta^{(1)}-\theta^{(2)})'\frac{\sum_{t_i\in(\tau^0,\hat{\tau}]} Z_i'Z_i}{\hat{\tau}-\tau^0}(\theta^{(1)}-\theta^{(2)})+o_p(1),
\end{aligned}
$$

with $(\theta^{(1)}-\theta^{(2)})'\frac{\sum_{t_i\in(\tau^0,\hat{\tau}]} Z_i'Z_i}{\hat{\tau}-\tau^0}(\theta^{(1)}-\theta^{(2)}) \geq \gamma_1||\theta^{(1)}-\theta^{(2)}||^2$, where γ_1 is the minimum eigenvalue of $\frac{\sum_{t_i\in(\tau^0,\hat{\tau}]} Z_i'Z_i}{\hat{\tau}-\tau^0}$. Under Assumption 4 with a suitable choice of C, we have that for $\hat{\tau} \in V_\eta(C)$, γ_1 is bounded away from 0. Hence, $\phi(a3) \geq C_2||\theta^{(1)} - \theta^{(2)}||^2$ for some $C_2 > 0$. Moreover, applying (56) to (47), together with some factorisations, we have that (47) $= o_p(1)$. Therefore, the term $\phi(a3)$ dominates all others and it is positive with probability 1 for large T. This implies that with large probability, (44) > 0, which gives a contradiction. This indicates that with large probability $\hat{\tau}$ cannot be in the set $V_\eta(C)$ and hence $P(T|\hat{s} - s^0| \geq C) \leq \epsilon$ when T is large. □

Remark 3 As discussed in Remark 1 in Sect. 4, in case the shift is of shrinking magnitude with shrinking speed v_T, the asymptotic behaviour (i) discussed in Remark 1 may be verified by following the same arguments in the proof of Proposition 3 with $\phi = \frac{1}{T^{2r*}}$, together with the fact that $T\phi||\theta^{(1)} - \theta^{(2)}||^2 = (T^{1-2r^*}v_T^2||\mathbf{M}||^2) \xrightarrow[T\to\infty]{} \infty$ and $\log T/T^{2r^*} \xrightarrow[T\to\infty]{} 0$. On the other hand, (ii) may be verified by following similar arguments as in the proof of Proposition 4 to investigate the set $V_\eta(C, v_T) = \{\tau : C/v_T^2 < |\tau - \tau^0| < \eta T\}$ instead of $V_\eta(C)$.

Appendix A.2 Proofs of Propositions 6 and 7

Write $V(t) := (\varphi_1(t), \ldots, \varphi_p(t), -X_t)$ and let $\log \mathcal{L}_1$ be the log-likelihood function based on the estimated change point $\hat{\tau}$ and the associated MLE $(\hat{\theta}^{(1)}, \hat{\theta}^{(2)})$. That is, $\log \mathcal{L}_1 = \frac{1}{\sigma^2}(\int_0^{\hat{\tau}} S(\hat{\theta}^{(1)}, t, X_t)\mathrm{d}X_t + \int_{\hat{\tau}}^T S(\hat{\theta}^{(2)}, t, X_t)\mathrm{d}X_t) - \frac{1}{2\sigma^2}(\int_0^{\hat{\tau}} S^2(\hat{\theta}^{(1)}, t, X_t)$ $\mathrm{d}t + \int_{\hat{\tau}}^T S^2(\hat{\theta}^{(2)}, t, X_t)\mathrm{d}t$. Similarly, we let $\log \mathcal{L}_0$ be the log-likelihood function based on the exact value of change point τ^0 and the associated MLE $(\hat{\theta}^{(1,0)}, \hat{\theta}^{(2,0)})$. Then, the proofs of Propositions 6 and 7 rely on the behaviour of

$$\phi(\log \mathcal{L}_1 - \log \mathcal{L}_0). \tag{58}$$

Now, without loss of generality, suppose first that $\hat{\tau} > \tau^0$ and divide the time period $[0, T]$ into three sub-intervals: $[0, \tau^0]$, $(\tau^0, \hat{\tau}]$ and $(\hat{\tau}, T]$. Using relation (1), along with some algebraic manipulation, we have expression

$$\begin{aligned}
(58) = &\frac{1}{2\sigma^2}\left(\int_0^{\tau^0}(V(t)\hat{\theta}^{(1)})^2\mathrm{d}t + \int_{\tau^0}^{\hat{\tau}}(V(t)\hat{\theta}^{(1)})^2\mathrm{d}t + \int_{\hat{\tau}}^{T}(V(t)\hat{\theta}^{(2)})^2 dt\right)\\
&- \frac{1}{2\sigma^2}\left(\int_0^{\tau^0}(V(t)\hat{\theta}^{(1,0)})^2\mathrm{d}t + \int_{\tau^0}^{\hat{\tau}}(V(t)\hat{\theta}^{(2,0)})^2\mathrm{d}t + \int_{\hat{\tau}}^{T}(V(t)\hat{\theta}^{(2,0)})^2 dt\right)\\
&+ \frac{1}{\sigma}\left(\int_0^{\tau^0}(V(t)\hat{\theta}^{(1)})\mathrm{d}W_t + \int_{\tau^0}^{\hat{\tau}}(V(t)\hat{\theta}^{(1)})\mathrm{d}W_t + \int_{\hat{\tau}}^{T}(V(t)\hat{\theta}^{(2)})\mathrm{d}W_t\right)\\
&- \frac{1}{\sigma}\left(\int_0^{\tau^0}(V(t)\hat{\theta}^{(1,0)})\mathrm{d}W_t + \int_{\tau^0}^{\hat{\tau}}(V(t)\hat{\theta}^{(2,0)})\mathrm{d}W_t + \int_{\hat{\tau}}^{T}(V(t)\hat{\theta}^{(2,0)})\mathrm{d}W_t\right).
\end{aligned}$$

We plug in the expressions of the MLEs, and after some algebraic computations, we get

$$(58) = \frac{\phi}{\sigma}\Big(2(\theta^{(1)}-\theta^{(2)})' R_{(\tau^0,\hat{\tau})} - 2(\theta^{(1)}-\theta^{(2)})' Q_{(\tau^0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})}\Big) \tag{59}$$

$$+ \frac{3\phi}{2}\Big(R'_{(0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} - R'_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} R_{(0,\tau^0)} + R'_{(\hat{\tau},T)} Q^{-1}_{(\hat{\tau},T)} R_{(\hat{\tau},T)}\Big) \tag{60}$$

$$- \frac{3\phi}{2} R'_{(\tau^0,T)} Q^{-1}_{(\tau^0,T)} R_{(\tau^0,T)} \tag{61}$$

$$- \frac{\phi}{2\sigma^2}(\theta^{(1)}-\theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} Q_{(\tau^0,\hat{\tau})}(\theta^{(1)}-\theta^{(2)}). \tag{62}$$

The remaining parts of the proofs for Propositions 6 and 7 depend on investigating the asymptotic behaviours of (59)–(62).

Proof of Proposition 6. Let $\phi = \frac{1}{T}$. Note that $\log \mathcal{L}_1$ is taken to be the maximum of the log-likelihood function from all possible choices of $\tau \in [0, T]$, whilst $\log \mathcal{L}_0$ is based on one particular change point $\tau^0 \in [0, T]$. It follows from the definition of MLE that (58) ≥ 0 with probability 1. However, if the rate of change point s^0 is not consistently estimated by $\hat{s}$, then with positive probability, we have $\hat{\tau} - \tau^0 = (\hat{s} - s^0)T > L_0^*$, where L_0^* is defined in Assumption 4. In this case, under Assumptions 3 and 4, the minimum eigenvalue of $D^* = \frac{1}{2(\hat{s}-s^0)T}(Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} Q_{(\tau^0,\hat{\tau})} + Q_{(\tau^0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} Q_{(0,\tau^0)})$ is bounded away from 0. Denoting this mimimum eigenvalue by γ_2, we have

$$\begin{aligned}&\frac{\phi}{2\sigma^2}(\theta^{(1)}-\theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} Q_{(\tau^0,\hat{\tau})}(\theta^{(1)}-\theta^{(2)})\\ &= \frac{(\hat{s}-s^0)}{4\sigma^2}(\theta^{(1)}-\theta^{(2)})' D^*(\theta^{(1)}-\theta^{(2)}) \geq \frac{(\hat{s}-s^0)\gamma_2}{4\sigma^2}||\theta^{(1)}-\theta^{(2)}||^2.\end{aligned} \tag{63}$$

So, (62) $\leq -C_3||\theta^{(1)}-\theta^{(2)}||^2$ with $C_3 = \frac{(\hat{s}-s^0)\gamma_2}{4\sigma^2} > 0$.

Using (54) and (55), together with the Cauchy–Schwarz inequality, along with some algebraic computations, we find that (59)–(61) are all of order $o_p(1)$. Hence, (58) is dominated by (62), which is negative. This means that (58) < 0 with positive probability, which is a contradiction. Therefore, for large T and $\forall \epsilon > 0$, $\hat{s} - s^0 < \epsilon$. This implies that $\hat{s} - s^0 \xrightarrow[T\to\infty]{P} 0$, which completes the proof. □

Proof of Proposition 7 Write $\phi := \frac{1}{\hat{\tau} - \tau^0}$ and $V_\eta := \{\tau : |\tau - \tau^0| \leq \eta T\}$. Then, it follows from Proposition 6 that for each $\eta > 0$, $P(\hat{\tau} \in V_\eta) \xrightarrow[T\to\infty]{} 1$. Therefore, we only need to investigate the asymptotic behaviour of (59)–(62) for those $\hat{\tau} \in V_\eta$. For $C > 0$, define the set $V_\eta(C) = \{\tau : C < |\tau - \tau^0| < \eta T\}$ and let $\hat{\tau}$ be the estimated change point with the minimum taken over the set $V_\eta(C)$. Then, it suffices to show that for some $C > 0$, such that for any $\hat{\tau} \in V_\eta(C)$, (44) < 0 with positive probability, and this leads to a contradiction since the term (44) ≤ 0 with probability 1. This would imply that for some $C > 0$ and any $0 < \eta < 1$, the global optimisation cannot be achieved on the set $V_\eta(C)$. Thus, with large probability, $|\hat{\tau} - \tau^0| \leq C$.

First note that under Assumption 4, with a suitable C such that $C > L_0$, it follows from (63) that (62) $\leq -C_3||\theta^{(1)} - \theta^{(2)}||^2$ for some $C_3 > 0$.

Next, for (59), by (54) and the Cauchy–Schwarz inequality, we have

$$\frac{2(\theta^{(1)} - \theta^{(2)})' R_{(\tau^0,\hat{\tau})}}{\sigma(\hat{\tau} - \tau^0)} \leq \frac{2}{\sigma\sqrt{\hat{\tau} - \tau^0}} ||\theta^{(1)} - \theta^{(2)}|| \left\| \frac{1}{\sqrt{\hat{\tau} - \tau^0}} R_{(\tau^0,\hat{\tau})} \right\|$$
$$= (\hat{\tau} - \tau^0)^{a^*-1/2} O_p(1),$$

where $0 < a^* < 1/2$, and

$$\frac{2(\theta^{(1)} - \theta^{(2)})' Q_{(\tau^0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})}}{\sigma(\hat{\tau} - \tau^0)} \leq \frac{2}{\sigma\sqrt{T}} ||\theta^{(1)} - \theta^{(2)}||$$
$$\left\| \frac{Q_{(\tau^0,\hat{\tau})}}{(\hat{\tau} - \tau^0)} \right\| ||T Q^{-1}_{(0,\hat{\tau})}|| \left\| \frac{1}{\sqrt{T}} R_{(0,\hat{\tau})} \right\| = o_p(1).$$

For (61), applying again Theorem A.1 in Tobing and McGilchrist (1992), we have that for large T,

$$Q^{-1}_{(\tau^0,T)} = Q^{-1}_{(\hat{\tau},T)} + O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right). \tag{64}$$

Together with (56), we obtain

$$R'_{(0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} - R'_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} R_{(0,\tau^0)}$$
$$= R'_{(\tau^0,\hat{\tau})} Q^{-1}_{(0,\tau^0)} R_{(\tau^0,\hat{\tau})} - 2R'_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} R_{(\tau^0,\hat{\tau})}$$
$$+ O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right) R'_{(0,\hat{\tau})} R_{(0,\hat{\tau})}$$

and

$$R'_{(\tau^0,T)} Q^{-1}_{(\tau^0,T)} R_{(\tau^0,T)} - R'_{(\hat{\tau},T)} Q^{-1}_{(\hat{\tau},T)} R_{(\hat{\tau},T)}$$
$$= R'_{(\tau^0,\hat{\tau})} Q^{-1}_{(\hat{\tau},T)} R_{(\tau^0,\hat{\tau})} - 2R'_{(\hat{\tau},T)} Q^{-1}_{(\hat{\tau},T)} R_{(\tau^0,\hat{\tau})}$$
$$+ O_p\left(\frac{\hat{\tau} - \tau^0}{T^2}\right) R'_{(\tau^0,T)} R_{(\tau^0,T)}.$$

From (54), (55), together with the Cauchy–Schwarz inequality, we have (61) $= o_p(1)$ for large T. Therefore, by choosing a suitable large C, we obtain (62) + (59) < 0, which implies that (58) < 0 and this gives a contradiction. Therefore, with large probability, $\hat{\tau}$ cannot be in the set $V_\eta(C)$ and hence $P(T|\hat{s} - s^0| \geq C) \leq \epsilon$ when T is large. □

Appendix B Proof details of Proposition 8

This proof can be completed by comparing $\mathcal{IC}(m=0)$ and $\mathcal{IC}(m=1)$ under H_0 and H_1, respectively. Moreover, note that $\log(T/\Delta_t) = \log T - \log(\Delta_t)$ and $\log T$ is just a special case of $\log(T/\Delta_t)$ with $\Delta_t = 1$. Hence, in the succeeding proof, we only prove the case $\phi(T) = \log(T/\Delta_t)$ with Δ_t a fixed constant.

Proof of Proposition 8 Under H_0, $m^0 = 0$ and $\theta^{(1)} = \theta^{(2)}$. In this case, for $\mathcal{IC}(m = 0)$, it follows from (1) that

$$\mathcal{IC}(m=0) = -2\left(\frac{1}{\sigma^2}\int_0^T S(\hat{\theta}^{(1)}, t, X_t)\mathrm{d}X_t - \frac{1}{2\sigma^2}\int_0^T S(\hat{\theta}^{(1)}, t, X_t)^2\mathrm{d}t\right) + h(p)\log(T/\Delta_t) = -2\left(\frac{1}{2\sigma^2}\theta^{(1)\prime}Q_{(0,T)}\theta^{(1)} + \frac{2}{\sigma}\theta^{(1)\prime}R_{(0,T)} + \frac{3\sigma^2}{2}R'_{(0,T)}Q^{-1}_{(0,T)}R_{(0,T)}\right) + h(p)\log(T/\Delta_t).$$

Assume further that we get $\hat{\tau}$ from (32) when m is set to 1 (note that in this case $\theta^{(2)}$ is still equal to $\theta^{(1)}$). Then,

$$\mathcal{IC}(m=1) = -\frac{1}{\sigma^2}\theta^{(1)\prime}Q_{(0,T)}\theta^{(1)} - \frac{4}{\sigma}\theta^{(1)\prime}R_{(0,T)} - 3R'_{(0,\hat{\tau})}Q^{-1}_{(0,\hat{\tau})}R_{(0,\hat{\tau})} - 3R'_{(\hat{\tau},T)}Q^{-1}_{(\hat{\tau},T)}R_{(\hat{\tau},T)} + 2h(p)(\log T - \log(\Delta_t))$$

so that

$$\mathcal{IC}(m=1) - \mathcal{IC}(m=0) = 3R'_{(0,T)}Q^{-1}_{(0,T)}R_{(0,T)} - 3R'_{(0,\hat{\tau})}Q^{-1}_{(0,\hat{\tau})}R_{(0,\hat{\tau})} - 3R'_{(\hat{\tau},T)}Q^{-1}_{(\hat{\tau},T)}R_{(\hat{\tau},T)} + h(p)(\log T - \log(\Delta_t)). \tag{65}$$

Moreover, by (54) and (55), together with the Cauchy–Schwarz inequality,

$$R'_{(0,T)}Q^{-1}_{(0,T)}R_{(0,T)} \le ||\frac{1}{\sqrt{T}}R_{(0,T)}||^2||TQ^{-1}_{(0,T)}|| = O_p(\log^{2a^*}T),$$

where $0 < a^* < 1/2$. Similarly, $R'_{(0,\hat{\tau})}Q^{-1}_{(0,\hat{\tau})}R_{(0,\hat{\tau})}$ and $R'_{(\hat{\tau},T)}Q^{-1}_{(\hat{\tau},T)}R_{(\hat{\tau},T)}$ are also of order $O_p(\log^{2a^*}T)$. Therefore, for large T and fixed Δ_t, (65) is dominated by $h(p)\log T$, which is positive. This implies that under H_0, the probability of $\mathcal{IC}(m = 0) > \mathcal{IC}(m = 1)$ tends to 0 as T tends to ∞.

Under H_1, let τ^0 be the exact value of the change point and $\hat{\tau}$ be the estimator of τ^0 obtained from (32). Without loss of generality, we assume that $\hat{\tau} > \tau^0$. So,

$$\mathcal{IC}(m=0) = -\frac{1}{\sigma^2}\theta^{(2)\prime}Q_{(0,T)}\theta^{(2)} - \frac{1}{\sigma^2}(\theta^{(1)} - \theta^{(2)})'Q_{(0,\tau^0)}Q^{-1}_{(0,T)}Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)})$$

$$
\begin{aligned}
&- 3R'_{(0,T)} Q^{-1}_{(0,T)} R_{(0,T)} - \frac{4}{\sigma}\theta^{(2)\prime} R_{(0,T)} \\
&- \frac{4}{\sigma}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,T)} R_{(0,T)} \\
&- \frac{4}{\sigma^2}\theta^{(2)\prime} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) + h(p)(\log T - \log(\Delta_t)),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{IC}(m=1) = &-\frac{1}{\sigma^2}\theta^{(2)\prime} Q_{(0,T)}\theta^{(2)} \\
&- \frac{1}{\sigma^2}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) \\
&- 3R'_{(0,\hat{\tau})} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} - 3R'_{(\hat{\tau},T)} Q^{-1}_{(\hat{\tau},T)} R_{(\hat{\tau},T)} \\
&- \frac{4}{\sigma}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} \\
&- \frac{4}{\sigma}\theta^{(2)\prime} R_{(0,T)} - \frac{4}{\sigma^2}\theta^{(2)\prime} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) \\
&+ 2h(p)(\log T - \log(\Delta_t)).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathcal{IC}(m=1) - \mathcal{IC}(m=0) = &-\frac{1}{\sigma^2}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)}(Q^{-1}_{(0,\hat{\tau})} \\
&- Q^{-1}_{(0,T)}) Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) \\
&- 3R'_{(0,\tau^0)} Q^{-1}_{(0,\tau^0)} R_{(0,\tau^0)} - 3R'_{(\tau^0,T)} Q^{-1}_{(\tau^0,T)} R_{(\tau^0,T)} \\
&+ 3R'_{(0,T)} Q^{-1}_{(0,T)} R_{(0,T)} - \frac{4}{\sigma}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} \\
&\times Q^{-1}_{(0,\hat{\tau})} R_{(0,\hat{\tau})} - \frac{4}{\sigma}(\theta^{(1)} - \theta^{(2)})' Q_{(0,\tau^0)} Q^{-1}_{(0,T)} R_{(0,T)} \\
&+ h(p)(\log T - \log(\Delta_t)).
\end{aligned}
$$

Multiplying both sides of the above identity by $\frac{1}{T}$, and using (54), (55) and the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&\frac{1}{T}(\mathcal{IC}(m=1) - \mathcal{IC}(m=0)) \\
&= \frac{1}{\sigma^2}(\theta^{(1)} - \theta^{(2)})' \frac{1}{T} Q_{(\tau^0,\hat{\tau})} T Q^{-1}_{(0,\hat{\tau})} \frac{1}{T} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) \\
&\quad - \frac{(1-s^0)}{\sigma^2}(\theta^{(1)} - \theta^{(2)})' \frac{1}{(1-s^0)T} Q_{(\tau^0,T)} Q^{-1}_{(0,T)} Q_{(0,\tau^0)}(\theta^{(1)} - \theta^{(2)}) \\
&\quad + o_p(1). \qquad (66)
\end{aligned}
$$

Note that the second term in (66) is equal to

$$-\frac{(1-s^0)}{\sigma^2}(\theta^{(1)}-\theta^{(2)})'\frac{1}{2(1-s^0)T}(Q_{(\tau^0,T)}Q^{-1}_{(0,T)}Q_{(0,\tau^0)} \\ + Q_{(0,\tau^0)}Q^{-1}_{(0,T)}Q_{(\tau^0,T)})(\theta^{(1)}-\theta^{(2)}).$$

Using the similar argument as in the proof of Proposition 6, we have that under Assumption 4, the second term is less than $-C_4||\theta^{(1)}-\theta^{(2)}||^2$ for some $C_4 > 0$. It also follows from Proposition 7 that $||\frac{1}{T}Q_{(\tau^0,\hat{\tau})}|| = \frac{\hat{\tau}-\tau^0}{T}||\frac{1}{\hat{\tau}-\tau^0}Q_{(\tau^0,\hat{\tau})}|| = o_p(1)$ for large T. Thus, by (55), together with the Cauchy–Schwarz inequality we have

$$\frac{1}{\sigma^2}(\theta^{(1)}-\theta^{(2)})'\frac{1}{T}Q_{(\tau^0,\hat{\tau})}TQ^{-1}_{(0,\hat{\tau})}\frac{1}{T}Q_{(0,\tau^0)}(\theta^{(1)}-\theta^{(2)}) = o_p(1).$$

This tells us that (66) is dominated by the first term for large T and it is negative. Therefore, $\mathcal{IC}(m=0) > \mathcal{IC}(m=1)$ with probability 1. □

References

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

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov, F. Csáki (Eds.), *2nd international symposium on information theory*, Tsahkadsor, Armenia, USSR, September 2–8, 1971, Budapest: Akadémiai Kiadó, pp. 267–281.

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

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

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

Chen, S. (2010). *Modelling the dynamics of commodity prices for investment decisions under uncertainty*. Ph.D. dissertation, University of Waterloo, Canada.

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

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

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

De Gregorio, A., Iacus, S. (2008). Least squares volatility change point estimation for partially observed diffusion processes. *Communications in Statistics*: *Theory and Methods*, *37*, 2342–2357.

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

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

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

Elliott, R., Wilson, C. (2007). The term structure of interest rates in a hidden Markov setting. In R. Mamon, R. Elliott (Eds.), *Hidden Markov models in finance* (pp. 14–31). New York, NY: Springer.

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

Gallagher, C., Lund, R., Robbins, M. (2012). Changepoint detection in daily precipitation data. *Environmetrics*, *23*, 407–419.
Gombay, E. (2010). Change detection in linear regression with time series errors. *Canadian Journal of Statistics*, *38*, 65–79.
Howell, S., Duck, P., Hazel, A., Johnson, P., Pinto, H., Strbac, G., et al. (2011). A partial differential equation system for modelling stochastic storage in physical systems with applications to wind power generation. *IMA Journal of Management Mathematics*, *22*, 231–252.
Iacus, S., Yoshida, N. (2012). Estimation for the change point of volatility in a stochastic differential equation. *Stochastic Processes and Applications*, *122*, 1068–1092.
Killick, R., Fearnhead, P., Eckley, I. (2012). Optimal detection of change points with a linear computational cost. *Journal of the American Statistical Association*, *107*, 1590–1598.
Lánský, P., Sacerdote, L. (2001). The Ornstein-Uhlenbeck neuronal model with signal-dependent noise. *Physics Letters A*, *285*, 132–140.
Le Breton, A. (1976). On continuous and discrete sampling for parameter estimation in diffusion type processes. *Mathematical Programming Studies*, *5*, 124–144.
Lee, S. (2011). Change point test for dispersion parameter based on discretely observed sample from SDE models. *Bulletin of the Korean Mathematical Society*, *48*, 839–845.
Lee, A., Guo, M. (2015). Monitoring change point for diffusion parameter based on discretely observed sample from stochastic differential equation models. *Applied Stochastic Models in Business and Industry*, *31*, 609–625.
Liang, Z., Yuen, K., Guo, J. (2011). Optimal proportional reinsurance and investment in a stock market with Ornstein-Uhlenbeck process. *Insurance*: *Mathematics and Economics*, *49*, 207–215.
Lipster, R., Shiryaev, A. (2001). *Statistics of random processes I*. Berlin: Springer.
Lu, Q., Lund, R. (2007). Simple linear regression with multiple level shifts. *Canadian Journal of Statistics*, *35*, 447–458.
Lu, S. (2003). Ornstein-Uhlenbeck diffusion quantum Monte Carlo calculations for small first-row polyatomic molecules. *Journal of Chemical Physics*, *118*, 9528–9532.
Lu, S. (2004). 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*. doi:10.1063/1.1639370.
Nkurunziza, S., Zhang, P. (2016). Estimation and testing in generalized mean-reverting processes with change-point. *Statistical Inference for Stochastic Processes. to appear*. doi:10.1007/s11203-016-9151-3.
Page, E. (1954). Continuous inspection schemes. *Biometrika*, *41*, 100–115.
Perron, P., Qu, Z. (2006). Estimating restricted structural change models. *Journal of Econometrics*, *134*, 373–399.
Reeves, J., Chen, J., Wang, X., Lund, R., Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, *46*, 900–915.
Robbins, M., Lund, R., Gallagher, C., Lu, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *Journal of the American Statistical Association*, *106*, 89–99.
Rohlfs, R., Harrigan, P., Nielsen, R. (2010). Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Scandinavian Journal of Statistics*, *37*, 200–220.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Scott, A., Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, *30*, 507–512.
Sen, A., Srivastava, M. (1975). On tests for detecting change in mean. *Annals of Statistics*, *3*, 98–108.
Shinomoto, S., Sakai, Y., Funahashi, S. (1999). The Ornstein-Uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. *Neural Computation*, *11*, 935–951.
Shiryaev, A. (1963). On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, *8*, 26–51.
Smith, W. (2010). On the simulation and estimation of the mean-reverting Ornstein-Uhlenbeck process. *CommodityModels.com*. Version 1.01.
Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Annals of Statistics*, 1405–1436.
Tobing, H., McGilchrist, C. (1992). Recursive residuals for multivariate regression models. *Australian Journal of Statistics*, *34*, 217–232.

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

Yan, G., Xiao, Z., Eidenbenz, S. (2008). Catching instant messaging worms with change-point detection techniques. In *Proceedings of the 1st usenix workshop on large-scale exploits and emergent threats*, vol. 6, pp. 1–10.

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