Statistica Sinica **27** (2017), 575-605
doi: https://doi.org/10.5705/ss.202015.0359

SEQUENTIAL CHANGE-POINT DETECTION IN TIME SERIES MODELS BASED ON PAIRWISE LIKELIHOOD

Sze Him Leung, Wai Leong Ng and Chun Yip Yau

Chinese University of Hong Kong

Abstract: The paper proposes a sequential monitoring scheme for detecting changes in parameter values for general time series models using pairwise likelihood. Under this scheme, a change-point is declared when the cumulative sum of the first derivatives of pairwise likelihood exceeds a certain boundary function. The scheme is shown to have asymptotically zero Type II error with a prescribed level of Type I error. With the use of pairwise likelihood, the scheme is applicable to many complicated time series models in a computationally efficient manner. For example, the scheme covers time series models involving latent processes, such as stochastic volatility models and Poisson regression models with log link function.

Key words and phrases: Composite likelihood, on-line detection, Poisson regression model, quickest detection, sequential monitoring, stochastic volatility.

1. Introduction

Let $\{X_t\}$ be a stochastic process given by a certain statistical model with parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. We say that there is a change-point at t^* if the model for $X_1, X_2, \ldots, X_{t^*-1}$ has parameter $\boldsymbol{\theta_0}$, but the model for $X_{t^*}, X_{t^*+1}, \ldots$ has parameter $\boldsymbol{\theta_1}$. Suppose that $\{X_t\}$ is observed sequentially and the change-point t^* is unknown. The problem of declaring whether the change t^* has occurred is called the sequential monitoring or on-line monitoring problem.

Since Page (1954), sequential monitoring schemes for time series data have been restricted to a few specific models. Lai (1995) proposed the window limited GLR scheme for detecting changes in time series models when the conditional log-likelihoods given all the past observations are available. Gut and Steinebach (2002) developed a truncated sequential change-point detection scheme in renewal counting process. Berkes et al. (2004) used quasi-likelihood scores to monitor changes in GARCH(p,q) models. Fuh (2006) considered change point detection in state space models. Gombay and Serban (2009) monitored changes in parameters of AR models using CUSUM methods. Na, Lee and Lee (2011) proposed estimates-based and residual-based monitoring schemes for general time series model. Na, Lee and Lee (2012) considered monitoring changes for coupula ARMA-GARCH Models. Kirch and Tadjuidje (2015) developed a general sequential change-point test for time series based on estimating functions. For a recent overview of sequential change-point methods in time series, we refer to Aue and Horváth (2013). For more general time series models, the conditional likelihood may not be readily available. For example, for such nonlinear time series models as the Poisson regression models with log link function and stochastic volatility models, it is infeasible to compute the likelihood functions. Therefore, it will be useful to develop a sequential monitoring procedure that is applicable to more general time series models.

Composite likelihood, which is constructed by combining marginal likelihoods of small subsets of data, has been employed for statistical inference for complicated models in various fields, for example, correlated binary data (Kuk and Nott (2000)), image analysis (Nott and Ryden (1999)), binary spatial data (Heagerty and Lele (1998)), survival data (Paik and Ying (2012)), and model selection (Varin and Vidoni (2005)). One widely used special case of composite likelihood is the pairwise likelihood, which is a product of bivariate likelihoods from various pairs of observations. As an illustrative example, consider a sample $\mathbf{X} = (x_1, \ldots, x_n)$ from the stochastic volatility model

$$x_t = z_t e^{1/2(\alpha_t+\beta)}, \quad \text{where } \alpha_t = \eta\alpha_{t-1} + \sigma\epsilon_t, \quad z_t, \epsilon_t \overset{i.i.d.}{\sim} N(0,1) . \tag{1.1}$$

The series $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)$ is unobserved and $\{\alpha_t\}_{t=1,\ldots,}$ is known as the latent process.

Let $\theta = (\eta, \sigma^2, \beta)$ be the parameter vector, $f_\theta(\mathbf{X})$ and $g_\theta(\boldsymbol{\alpha})$ be the joint density of $\mathbf{X}$ and $\boldsymbol{\alpha}$, respectively, and $f_\theta(\mathbf{X}|\boldsymbol{\alpha})$ be the condition density of $\mathbf{X}$ given $\boldsymbol{\alpha}$. With the presence of the latent process, the log-likelihood of the sample involves an n-dimensional integral

$$\log f_\theta(\mathbf{X}) = \log \left(\int \cdots \int f_\theta(\mathbf{X}|\boldsymbol{\alpha}) g_\theta(\boldsymbol{\alpha}) d\alpha_1 \cdots d\alpha_n \right) ,$$

that is computationally infeasible. In this case, the so-called consecutive pairwise likelihood, which combines likelihoods from consecutive pairs of observations, is given by

$$\text{CPL}(\theta, \mathbf{X}) = \sum_{t=1}^{n-1} \log \left(\iint f_\theta(x_t, x_{t+1}|\alpha_t, \alpha_{t+1}) g_\theta(\alpha_t, \alpha_{t+1}) d\alpha_t d\alpha_{t+1} \right) .$$

The $\text{CPL}(\theta, \mathbf{X})$ involves $n-1$ two-dimensional integrals and is feasible for computation. It is found that pairwise likelihood yields consistent and asymptotic normal estimators with reasonable efficiency (see, e.g., Davis and Yau (2011) and Ng et al. (2011)).

In this paper, we propose a sequential monitoring scheme for general time series models based on pairwise likelihood. The scheme is based on the changes in the behavior of the score function of pairwise likelihood upon changes in parameter values. Specifically, it monitors the cumulative sum of the score function of pairwise likelihood when data-points are observed sequentially. A change-point is declared when the cumulative sum exceeds a pre-determined boundary function. With the specific structure of the pairwise likelihood, asymptotic results can be readily established. In particular, it is shown that the cumulative sum process converges to a functional of a multi-dimensional Wiener processes. Thus, a boundary function can be obtained such that the scheme has asymptotically correct size with power approaching 1. The use of pairwise likelihood allows the scheme to be implemented efficiently, and is applicable to many complicated time series models. Simulation studies demonstrate that the scheme has higher detection power than existing non-parametric monitoring schemes.

This paper is organized as follows. Section 2 describes the problem setting and assumptions. Section 3 presents the sequential monitoring scheme under general conditions. In Section 4, we show that the general conditions given in Section 2 cover the class of time series models with a latent autoregressive process, which includes the Poisson regression models with log link function and stochastic volatility models. Simulation experiments and data examples are given in Section 5 and 6, respectively. Proofs and technical details are given in the Appendix.

2. Problem Setting and Assumptions

2.1. Settings

Let $\{X_t\}$ be a stationary and ergodic sequence of random variables following a model with joint density $f_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta}$ is the parameter vector in a compact space $\boldsymbol{\Theta}$. Suppose that, before we begin to monitor changes in the sequence, a sample of m historical data points $\{x_1, x_2, \ldots, x_m\}$ is observed. Starting from time $t = m + 1$ onwards, we observe $\{X_t\}$ sequentially. We wish to test whether or not a change has occurred in $\boldsymbol{\theta}$ using the null hypothesis:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad \text{for } t = 1, 2, \ldots, m + mT$$

against the alternative hypothesis

$$H_A : \boldsymbol{\theta} = \begin{cases} \boldsymbol{\theta_0} & \text{for } t = 1, 2, \ldots, t^* - 1, \\ \boldsymbol{\theta_1} & \text{for } t = t^*, t^* + 1, \ldots, m + mT, \end{cases}$$

where $\boldsymbol{\theta_0} \neq \boldsymbol{\theta_1}$, $t^* > m$ is the unknown change-point, and mT the maximum number of observations that are to be inspected. A similar setting has been discussed in Gut and Steinebach (2002) and Gombay and Horváth (2009).

Let $f_{X_t, X_{t+j}}(x_t, x_{t+j}; \boldsymbol{\theta})$ be the bivariate density of the observations x_t and x_{t+j}. The l-th order consecutive pairwise likelihood for the data $(X_1, \ldots, X_m)$ is defined by

$$\mathrm{CPL}_m(l; \boldsymbol{\theta}) = \sum_{t=1}^{m-l} L_t(l; \boldsymbol{\theta}) \,,$$

where $L_t(l; \boldsymbol{\theta}) = \sum_{j=1}^{l} p_t(j, \boldsymbol{\theta})$ and $p_t(j; \boldsymbol{\theta}) = \log f_{X_t, X_{t+j}}(x_t, x_{t+j}; \boldsymbol{\theta})$. Thus, $\mathrm{CPL}_m(l; \boldsymbol{\theta})$ collects the likelihood of the pairs of observations that are within a time lag of l. Given the historical sample $(x_1, \ldots, x_m)$, the parameter $\boldsymbol{\theta}$ can be estimated by

$$\hat{\boldsymbol{\theta}}_m = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathrm{CPL}_m(l, \boldsymbol{\theta}) \,.$$

In practice, l has to be chosen such that the model is identifiable. When the number of model parameters involved in serial dependency is d, it is usually sufficient to choose $l \geq d$, see Davis and Yau (2011) and Ng et al. (2011). As an illustrative example, consider the stochastic volatility model (1.1). To ensure identifiability, it is sufficient to take the lag l greater than or equal to the order of the latent AR process, $l \geq d = 1$, see Assumption B1 of Ng et al. (2011). From the simulation experiments in Davis and Yau (2011) and Ng et al. (2011), it is found that $l = d$ usually gives satisfactory results for time series data. Variously, l can be chosen by information criteria, see Lindsay, Yi and Sun (2011).

Suppose that $\boldsymbol{\theta}_0$ is the true parameter value and $L'_t(l; \boldsymbol{\theta}) = \partial L_t(l; \boldsymbol{\theta}) / \partial \boldsymbol{\theta} = \sum_{j=1}^{l} \partial p_t(j; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ is the sum of score functions of pairwise likelihoods at time t up to lag l. Since each $p_t(j; \boldsymbol{\theta})$ is an exact likelihood function, it will be shown in Lemma 1 that $\hat{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_0$ and $\{L'_t(l; \boldsymbol{\theta}_0)\}_{t=m+1, m+2, \ldots}$ is a sequence of random variables with zero-mean under H_0. Also, with proper standardizations, the cumulative sum $S'_m(k, l, \hat{\boldsymbol{\theta}}_m) := \sum_{t=m+1}^{m+k} L'_t(l; \hat{\boldsymbol{\theta}}_m)$ converges to a Wiener process under some mixing conditions. Thus, sequential monitoring of changes can be developed based on these asymptotic results.

To standardize $S'_m(k, l, \hat{\boldsymbol{\theta}}_m)$ by its variance, define the long-run covariance matrix $M(l; \boldsymbol{\theta}_0)$ by

$$M(l;\boldsymbol{\theta}_0) = \lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-l} L_t'(l;\boldsymbol{\theta}_0)\right) = \sum_{j=-\infty}^{\infty} \mathbb{E}\left[L_0'(l;\boldsymbol{\theta}_0)(L_j'(l;\boldsymbol{\theta}_0))^T\right] .$$

Following Andrews (1991), $M(l;\boldsymbol{\theta}_0)$ can be estimated by

$$\widehat{M}_m(l) = \sum_{j=-\lceil m^{1/3}\rceil}^{\lceil m^{1/3}\rceil} \left(1 - \frac{|j|}{\lceil m^{1/3}\rceil}\right) \hat{\gamma}_l(j) ,$$

where $\hat{\gamma}_l(j) = 1/m \sum_{t=j+1}^{m} L_t'(l;\hat{\boldsymbol{\theta}}_m)(L_{t-j}'(l;\hat{\boldsymbol{\theta}}_m))^T$ and $\lceil x\rceil$ is the smallest integer greater than x. In Section 3, we develop a sequential monitoring scheme for changes in $\boldsymbol{\theta}$ based on the sequence $\left\{\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\}_{k=1,2,\ldots}$.

2.2. Assumptions

The α-mixing coefficient between two σ-fields $\mathcal{F}_1, \mathcal{F}_2$ is defined by

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} .$$

We say that a process $\{Z_t\}$ is geometrically strongly mixing if $\alpha(k) := \alpha(\mathcal{A}_0, \mathcal{B}_k) = O(a^k)$ for some $a \in (0,1)$, where $\mathcal{A}_0$ and $\mathcal{B}_k$ are the σ-field generated by $\{Z_t; t \le 0\}$ and $\{Z_t; t \ge k\}$, respectively.

We need some assumptions for the properties of the sequential monitoring scheme. Let $\|c\|$ be the supremum norm of a vector c. When A is a matrix, let $\|A\| = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$.

Assumption A:

(A1) The true parameter $\boldsymbol{\theta}_0$ under H_0 is in the interior of $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^d$.

(A2) $\mathbb{E}\|L_t'(l;\boldsymbol{\theta}_0)\|^{4v} < \infty$ for some $v > 1$.

(A3) For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathbb{E}\|L_t''(l;\boldsymbol{\theta})\|^2 < \infty$.

(A4) $L_t(l;\boldsymbol{\theta})$, $L_t'(l;\boldsymbol{\theta})$ and $L_t''(l;\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$.

(A5) The model is identifiable: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ if and only if $L_t(l;\boldsymbol{\theta}_1) = L_t(l;\boldsymbol{\theta}_2)$ a.e.

(A6) $\{X_t\}$ is a sequence of geometrically strongly mixing random variables.

(A7) The change occurs after m, $t^* > m$.

(A8) When H_A is true, $\mathbb{E}\left[L_t'(l;\boldsymbol{\theta}_0)\right] \neq 0$, where the expectation is evaluated under $\boldsymbol{\theta}_1$.

Assumptions (A1) to (A6) are required to obtain the following results for the derivation of the monitoring scheme. The results are standard, and so the proofs are omitted. See Ng et al. (2011) and Andrews (1991).

Lemma 1.

a) *Under Assumptions (A1), (A4), (A5) and (A6), we have* $\hat{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_0$.

b) *Under Assumptions (A2), (A3), (A4) and (A6), we have* $\widehat{M}_m(l) \xrightarrow{p} M(l;\boldsymbol{\theta}_0)$.

c) *Under Assumptions (A2) and (A6), we have*

$$m^{-1/2}\sum_{t=1}^{\lfloor mt \rfloor} L_t'(l;\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}[0,T]} \mathbf{W}_{\mathbf{M}}(t) \qquad \text{as} \quad m \to \infty\,,$$

where $t \in [0,T]$, $\lfloor x \rfloor$ *is the largest integer smaller than* x *and* $\xrightarrow{\mathcal{D}[0,T]}$ *denotes weak convergence in* $\mathcal{D}[0,T]$, *the space of right-continuous function with left limit on* $[0,T]$, $\mathbf{W}_{\mathbf{M}}(s)$ *is a Gaussian process with* $\mathbb{E}\mathbf{W}_{\mathbf{M}}(s) = 0$ *and covariance function* $\mathbb{E}\mathbf{W}_{\mathbf{M}}(s)\mathbf{W}_{\mathbf{M}}^T(s') = \min(s,s')M(l;\boldsymbol{\theta}_0)$.

Assumption (A7) ensures that the change occurs after m, and hence the pre-change parameters can be consistently estimated. Assumption (A8) is needed to ensure that when there is a change-point, the cumulative sum process $\sum_{t=m+1}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m)$ has a drift. Hence, the change in $\boldsymbol{\theta}$ can be detected. This assumption is true when $\boldsymbol{\theta}_0$ is not a local maximizer of the expected $L_t(l;\boldsymbol{\theta})$ under H_A.

3. Monitoring Scheme

3.1. Pairwise likelihood sequential monitoring scheme (PLSMS)

As $\mathbb{E}[L_t'(l;\boldsymbol{\theta}_0)] = \mathbf{0}$ under H_0 and $\mathbb{E}[L_t'(l;\boldsymbol{\theta}_0)] \neq \mathbf{0}$ under H_A, by Lemma 1, the cumulative sum process $m^{-1/2}\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)$ is approximately a Wiener process under H_0 but, when H_A is true, the cumulative sum process diverges. Therefore, it is natural to declare that a change has occurred when $m^{-1/2}\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|$ exceeds some threshold or boundary. Define the stopping time

$$T_m(l) = \min\left\{\min\left\{k : \left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\left(1+\frac{k}{m}\right)c\right\}, mT+1\right\}, \tag{3.1}$$

where c is a constant, the decision boundary, and mT is the pre-specified maximum inspection time. We can also take T to infinite if we wish the inspection to continue until a change is declared. In this case, the stopping time is

$$T_m^*(l) = \min\left\{k : \left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\left(1+\frac{k}{m}\right)c\right\}. \quad (3.2)$$

For notational simplicity, suppose that the data $\{X_t\}$ arrive sequentially in a way that X_{t+m+l} is observed at time t, the pairwise likelihood sequential monitoring scheme (PLSMS) is as follows:

For the case in which $T > 0$ and T is finite, starting from time $k = 1$, check whether $\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\left(1+\frac{k}{m}\right)c$. If yes, then $T_m(l) = k$ and the scheme terminates; we reject H_0 and declare that a change in parameter had occurred at some time on or before k. Otherwise, proceed to time $k+1$ and repeat the same procedure. If the condition $\left\|\sum_{t=m+1}^{m+mT} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}(1+mT/m)\,c$ has not been met at time mT, then set $T_m(l) = mT+1$ and the scheme terminates; we declare that a change in parameter did not occur and conclude that H_0 is not rejected.

If $T = \infty$, then, starting from time $k = 1$, we have to check whether $\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\left(1+\frac{k}{m}\right)c$. If yes, then $T_m^*(l) = k$ and the scheme terminates; we reject H_0 and declare that a change in parameter had occurred at some time on or before k.

We assess the performance of the proposed sequential monitoring test based on asymptotic size and power. The size is the probability of declaring an occurrence of change when no change has occurred, and the power is the probability that a change-point is detected before mT given that the change had occurred at $t^* < mT$. We will derive a decision boundary $c = c(\alpha, l, T)$ such that the PLSMS has power approaching 1 with a pre-specified size α. Thus, the boundary $c = c(\alpha, l, T)$ is chosen such that $T_m(l)$ satisfies

$$\lim_{m\to\infty} P(T_m(l) \le mT | H_0) = \alpha\,, \quad \text{and} \quad (3.3)$$

$$\lim_{m\to\infty} P(T_m(l) \le mT | H_A) = 1\,, \quad (3.4)$$

for any fixed $0 < \alpha < 1$.

3.2. Asymptotic properties of $T_m(l)$ under H_0 and H_A

The proofs of results in this section are given in the Appendix.

Theorem 1. (Asymptotic size under H_0) *Assume (A1) to (A6) hold. Let $\{W(t), 0 \le t \le 1\}$ be a standard Wiener process and $\boldsymbol{\theta} \in \mathbb{R}^d$. The asymptotic size of the PLSMS with decision boundary c is given by*

$$\lim_{m\to\infty} P(T_m(l) \le mT|H_0) = 1 - \left[P\left(\sup_{0\le t<\frac{T}{1+T}} \frac{|W(t)|}{c} \le 1\right)\right]^d. \quad (3.5)$$

If $T = \infty$ and the ρ-mixing coefficient of $\{X_t\}$ satisfies

$$\rho(k) := \rho(\mathcal{A}_0, \mathcal{B}_k) = \sup_{f\in\mathcal{L}^2(\mathcal{A}_0), g\in\mathcal{L}^2(\mathcal{B}_k)} |\mathrm{Corr}(f,g)| = O(a^k), \quad (3.6)$$

where $0 < a < 1$, $\mathcal{A}_0$ and $\mathcal{B}_k$ are the σ-field generated by $\{X_t; t \le 0\}$ and $\{X_t; t \ge k\}$ respectively, $\mathcal{L}^2(\mathcal{A}_0)$ is the space of square-integrable $\mathcal{A}_0$-measurable random variables, and $\mathcal{L}^2(\mathcal{B}_k)$ is defined similarly (Doukhan (1994)), then

$$\lim_{m\to\infty} P(T_m^*(l) < \infty|H_0) = 1 - \left[P\left(\sup_{0\le t<1} \frac{|W(t)|}{c} \le 1\right)\right]^d. \quad (3.7)$$

The ρ-mixing assumption for the case $T = \infty$ is required to fulfill the conditions for using the ρ-mixing Hajek-Renyi inequality that controls the tail probability of the monitoring statistics when T goes to infinity. In practice, we always monitor the process of interest with finite T. Thus, if the ρ-mixing assumption is violated, we can approximate the decision boundaries using a large T and apply the monitor scheme in (3.5).

The probabilities in (3.5) and (3.7) can be evaluated explicitly using a result about the distribution function of $\sup_{0\le t<1}|W(t)|$.

Lemma 2. (Csorgo and Revesz (1981)) *For any $b > 0$, we have*

$$P\left(\sup_{0\le t<1} |W(t)| \le b\right) = \frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2}{8b^2}\right).$$

To choose c in (3.5) such that (3.3) holds, we combine Theorem 1 and Lemma 2 to give

$$\begin{aligned}
&\lim_{m\to\infty} P(T_m(l) \le mT|H_0) \\
&= 1 - \left(P\left\{\sup_{0\le s<1} \sqrt{\frac{1+T}{T}}\left|W\left(\frac{Ts}{1+T}\right)\right| \le c\sqrt{\frac{1+T}{T}}\right\}\right)^d \\
&= 1 - \left(P\left\{\sup_{0\le s<1} |W(s)| \le c\sqrt{\frac{1+T}{T}}\right\}\right)^d \\
&= 1 - \left(\frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2 T}{8c^2(1+T)}\right)\right)^d.
\end{aligned}$$

Therefore, for a given significant level α, we can choose c such that

$$1-\left(\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{\pi^2(2k+1)^2T}{8c^2(1+T)}\right)\right)^d=\alpha\,. \qquad (3.8)$$

A simple grid search or bisection algorithm can be performed to solve (3.8) for c.

Theorem 2. (Asymptotic power under H_A) *Assume that (A1) to (A8) hold. For the PLSMS using c that satisfies* (3.8) *at any given significance level $\alpha \in (0,1)$, we have*

$$\lim_{m\to\infty} P(T_m(l) \le mT|H_A) = 1\,, \qquad \text{and} \qquad (3.9)$$

$$\lim_{m\to\infty} P(T_m^*(l) < \infty|H_A) = 1\,, \qquad (3.10)$$

4. Time Series with a Latent Autoregressive Process

In Sections 2 and 3, general conditions were provided for the consistency of PLSMS. In this section, we verify these conditions for a class of time series models that feature a latent autoregressive process. Poisson regression models with log link function and stochastic volatility models fall into this class.

The process $\{X_t\}$ is said to be a time series with a latent Gaussian autoregressive process if

$$f_{X_t|\lambda_t}(x_t) \sim \exp(c_1(\lambda_t)T(x_t) + c_2(\lambda_t) + g(x_t))\,, \qquad (4.1)$$

where $\lambda_t = e^{\beta+\alpha_t}$ and $\{\alpha_t\}$ is the latent AR(p) process satisfying

$$\alpha_t = \eta_1\alpha_{t-1} + \ldots + \eta_p\alpha_{t-p} + \epsilon_t\,, \qquad \{\epsilon_t\} \overset{i.i.d.}{\sim} N(0,\sigma_\epsilon^2)\,.$$

We also assume that conditioned on $\{\alpha_t\}$, $\{X_t\}$ is an independent process. The parameter for $\{X_t\}$ is $\boldsymbol{\theta} = (\beta, \sigma_\epsilon^2, \eta_1, \ldots \eta_p)$, a $d = p+2$ dimensional vector. This definition covers some popular nonlinear time series models, as follows.

(i) The Stochastic volatility model: $c_1(\lambda_t) = -1/(2\lambda_t), T(x_t) = x_t^2, c_2(\lambda_t) = (-\log\lambda_t)/2,\ g(x_t) = 0$.

(ii) The Poisson regression model with log link function: $c_1(\lambda_t) = \log\lambda_t, T(x_t) = x_t, c_2(\lambda_t) = \lambda_t, g(x_t) = -\log\Gamma(x_t+1)$. The model was proposed by Zeger (1988) for time series of counts.

(iii) The Bernoulli logit model: $c_1(\lambda_t) = \log\lambda_t, T(x_t) = x_t, c_2(\lambda_t) = \log(1+\lambda_t), g(x_t) = 0$.

For model (4.1), the bivariate density of the observations (X_t, X_{t+j}) is given by

$$f_{X_t,X_{t+j}}(x_t,x_{t+j};\boldsymbol{\theta}) = \frac{1}{2\pi\left|\boldsymbol{\Sigma}_j\right|^{1/2}}\int f_{X_t|\lambda_t}(x_t)f_{X_{t+j}|\lambda_{t+j}}(x_{t+j}) \exp\left\{-0.5\boldsymbol{\alpha}_j^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\alpha}_j\right\}d\boldsymbol{\alpha}_j\,,$$

where $\boldsymbol{\alpha}_j = (\alpha_t, \alpha_{t+j})$ and

$$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_j(\boldsymbol{\theta}) = \begin{pmatrix}\gamma_0 & \gamma_j\\ \gamma_j & \gamma_0\end{pmatrix}$$

is the covariance matrix of $\boldsymbol{\alpha}_j$, where $\gamma_k = \gamma_k(\boldsymbol{\theta})$ is the autocovariance function of $\{\alpha_t\}$ at lag k which depends on the parameter $\boldsymbol{\theta}$.

Some assumptions on the model (4.1) are required for applying PLSMS.

Assumption B:

(B1) The true parameter $\boldsymbol{\theta}_0$ under H_0 is in the interior $\boldsymbol{\Theta}$, which is a compact set in $\mathbb{R}^d$.

(B2) For $1 \le j \le l$, $\|\boldsymbol{\Sigma}_j\| > a$ for some $a > 0$.

(B3) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and some $v > 1$, both $\boldsymbol{\Sigma}_{0,j}^{-1} \pm 8v\left(\boldsymbol{\Sigma}_{0,j}^{-1} - \boldsymbol{\Sigma}_j^{-1}\right)$ are positive definite, where $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_j(\boldsymbol{\theta})$ is the covariance matrix of the latent process $\boldsymbol{\alpha}_j$ under parameter value $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}_{0,j} = \boldsymbol{\Sigma}_j(\boldsymbol{\theta}_0)$.

(B4) The model is identifiable: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ if and only if $L_t(l;\boldsymbol{\theta}_1) = L_t(l;\boldsymbol{\theta}_2)$ a.e..

(B5) When H_A is true, $t^* > m$ and $\mathbb{E}\left[L_t'(l;\boldsymbol{\theta}_0)\right] \neq 0$, where the expectation is evaluated under $\boldsymbol{\theta}_1$.

We show that Assumption A is satisfied for model (4.1) under Assumption B, which establishes the applicability of PLSMS. First, note that (A1), (A5), (A7), and (A8) are guaranteed by (B4) and (B5). Also, that (A4) follows from Lemma A.1 of Ng et al. (2011).

To show (A6), we can write

$$X_t = F_{X_t|\alpha_t}^{-1}(U_t;\alpha_t,\boldsymbol{\theta}_0)\,, \tag{4.2}$$

where $\{U_t\}$ is an independent sequence of uniform random variables, and $F_{X_t|\alpha_t}^{-1}$ is the inverse of the cumulative distribution function corresponding to $f_{X_t|\lambda_t}(x)$ in (4.1). From Theorem 1 of Mokkadem (1988), the stationary ARMA process $\{\alpha_t\}$ is geometrically completely regular. That is, the β-mixing coefficient defined by

$$\beta(\mathcal{A}_0,\mathcal{B}_k) = \sup\frac{1}{2}\sum_{i=1}^{I}\sum_{j'=1}^{J}\left|\mathbb{P}(A_iB_{j'}) - \mathbb{P}(A_i)\mathbb{P}(B_{j'})\right|,$$

satisfies $\beta(k) = \beta(\mathcal{A}_0, \mathcal{B}_k) = O(a^k)$ for some $0 < a < 1$, where $\mathcal{A}_0$ and $\mathcal{B}_k$ are the σ-field generated by $\{\alpha_t; t \leq 0\}$ and $\{\alpha_t; t \geq k\}$ respectively, and the supremum is taken all over finite partitions $(A_1, A_2, \ldots, A_I)$ and $(B_1, B_2, \ldots, B_J)$ of the probability space Ω, with $\{A_i\} \in \mathcal{A}_0$ and $\{B_j\} \in \mathcal{B}_k$. Since $\{U_t\}$ is an independent process, (α_t, U_t) is also geometrically completely regular. From (4.2), X_t is a measurable transform of (α_t, U_t). Therefore, the σ fields $\sigma\{X_t; t \leq 0\}$ and $\sigma\{X_t; t \geq k\}$ are proper subsets of $\sigma\{(\alpha_t, U_t); t \leq 0\}$ and $\sigma\{(\alpha_t, U_t); t \geq k\}$, respectively, and the β-mixing coefficient of $\{X_t\}$ is not greater than that of $\{(\alpha_t, U_t)\}$. Thus, the process $\{X_t\}$ is also geometrically completely regular. Since the mixing coefficients satisfy $\alpha(k) < \beta(k)$ by Doukhan (1994), $\{X_t\}$ is strongly mixing with geometrical rate and (A6) is verified. On the other hand, model (4.1) is a Gaussian process, and we have $\rho(k) \leq 2\pi\alpha(k)$ for any Gaussian process by Kolmogorov and Rozanov (1960). Hence, $\{X_t\}$ is ρ-mixing with geometrical rate, and (3.6) is verified. Therefore, the PLSMS monitoring time T can be chosen to be infinite, and the inspection can continue until a change is declared.

The verification of (A2) and (A3) are more technical. We put this as a lemma and provide the proof in the Appendix.

Lemma 3. *For all $t > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have*

a) $\mathbb{E}\,\|L_t'(l;\boldsymbol{\theta})\|^{4v} < \infty$ *for some $v > 1$.*

b) $\mathbb{E}\,\|L_t''(l;\boldsymbol{\theta})\|^{2} < \infty$.

5. Simulation Experiments

We studied the finite sample performance of the proposed monitoring scheme through simulations. We focused on the stochastic volatility models with autoregressive order p, SV(p), given by

$$\begin{aligned} x_t &= z_t e^{1/2(\alpha_t+\beta)}\,, \qquad (5.1)\\ \alpha_t &= \eta_1\alpha_{t-1} + \eta_2\alpha_{t-2} + \ldots + \eta_p\alpha_{t-p} + \epsilon_t\,, \end{aligned}$$

where $\{z_t\} \overset{i.i.d.}{\sim} N(0,1)$, $\{\epsilon_t\} \overset{i.i.d.}{\sim} N(0,\sigma^2)$ and $|\eta_i| < 1$ for $i = 1, \ldots, p$. For simplicity, we considered the SV(1) with parameter vector $\boldsymbol{\theta} = (\eta, \sigma, \beta)$ and $|\eta| < 1$. For the lag of the consecutive pairwise likelihood l, it was sufficient to choose $l = 1$ for model identifiability (see Ng et al. (2011) or Davis and Yau (2011)). The pre-change historical data were taken to be $m = 500$, 1,000 and 5,000 for a reasonably accurate estimate of $\boldsymbol{\theta}$ and the long run covariance matrix $M(l;\boldsymbol{\theta}_0)$. In practice, we recommend that $m \geq 500$ should be used since time series with latent process are difficult to estimate. Given α and T, the boundary function c for the stopping time $T_m(l)$ in (3.1) was obtained by solving (3.8) with $d = 3$.

Due to the difficulty in evaluating the full likelihood function, one can resort to non-parametric methods that focus on the changes in mean or covariance structure. Standard procedures that detect changes in mean or covariance structure are not applicable to SV models and, since a change in parameter values in (5.1) induces a change in the variance of $\{x_t\}$, sequential monitoring can be based on cumulative sum of the squared data. We call this monitoring scheme the cusum sequential monitoring scheme (CSSMS).

The CSSMS is based on the stopping time

$$N_m = \min\left\{\min\left\{k : \sum_{t=m+1}^{m+k} (X_t^2 - \hat{\mu}_m)\hat{\sigma}_m^{-1/2} > m^{1/2}\left(1 + \frac{k}{m}\right)c\right\}, mT + 1\right\},$$

where

$$\hat{\mu}_m = \frac{1}{m}\sum_{t=1}^{m} X_t^2, \quad \text{and} \quad \hat{\sigma}_m^2 = \hat{\gamma}_m(0) + 2\sum_{j=1}^{\lceil m^{1/3}\rceil} \hat{\gamma}_m(j),$$

and $\hat{\gamma}_m(j)$ is the sample auto-covariance function of $\{X_t^2\}$ with lag j. We reject the null hypothesis of no change point when $N_m < mT+1$. Similar to PLSMS, for given T and α, the decision boundary c can be obtained by solving (3.8) with $d = 1$. See Gombay and Horváth (2009) for details.

In financial time series, the GARCH model is commonly used as an alternative to stochastic volatility models, see Tsay (2010, 2012). In particular, Berkes et al. (2004) proposed a monitoring scheme for GARCH(p,q) models. For simplicity, we call it the GARCH(p,q) sequential monitoring scheme (GASMS). The GASMS is based on the stopping time

$$G_m = \min\left\{\min\left\{k : \left\|\sum_{t=m+1}^{m+k} \widehat{D}_m(\hat{\theta}_m)^{-1/2} p_t'(\hat{\theta}_m)\right\| > m^{1/2}\left(1+\frac{k}{m}\right)c\right\}, mT+1\right\},$$

where

$$p_t(\theta) = -\frac{1}{2}\left\{\log \widehat{w}_t(\theta) + \frac{X_t^2}{\widehat{w}_t(\theta)}\right\} \quad \text{and} \quad \widehat{D}_m(\hat{\theta}_m) = \frac{1}{m}\sum_{t=2}^{m} p_t'(\hat{\theta}_m) p_t'(\hat{\theta}_m)^T.$$

Here, $p_t(\theta)$ is the conditional likelihood at θ and $\widehat{D}_m$ is the covariance matrix estimate for $p_t(\hat{\theta}_m)$. For the definition of $\widehat{w}_t(\theta)$ and the asymptotic results, see Berkes et al. (2004). We reject the null hypothesis of no change point when $G_m < mT+1$. For given T and α, the decision boundary c is the same as PLSMS since

Table 1. Decision boundaries c under different α and T for $T_m(1)$ in PLSMS and G_m in GASMS (The numbers with brackets are for N_m in CSSMS and R_m in RVSMS).

α	$T = 1$	$T = 2$	$T = \infty$
0.05	1.861(1.585)	2.149(1.83)	2.632(2.241)
0.1	1.684(1.386)	1.944(1.6)	2.382(1.96)

they have the same asymptotic distribution.

Besides GASMS, Na, Lee and Lee (2011) proposed another monitoring procedure for GARCH(p,q) models by using a residual variance approach. For simplicity, we call it the Residual Variance sequential monitoring scheme (RVSMS). The RVSMS for the GARCH(1,1) model is based on the stopping time

$$R_m = \min\left\{\min\left\{k : \frac{1}{s_m^2}\left|\frac{1}{m+k}\sum_{t=1}^{m+k}\hat{\epsilon}_t^2 - \frac{1}{m}\sum_{t=1}^{m}\hat{\epsilon}_t^2\right| > \frac{c}{\sqrt{m}}\right\}, mT+1\right\},$$

where

$$\hat{\epsilon}_t = \frac{X_t}{\tilde{\sigma}_t} \quad \text{and} \quad s_m^2 = \frac{1}{m}\sum_{t=1}^{m}\hat{\epsilon}_t^4 - \left(\frac{1}{m}\sum_{t=1}^{m}\hat{\epsilon}_t^2\right)^2.$$

Here, the $\tilde{\sigma}_t^2$ are defined recursively by $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\hat{\omega}, \hat{\alpha}, \hat{\beta}) = \hat{\omega} + \hat{\alpha}X_{t-1}^2 + \hat{\beta}\tilde{\sigma}_{t-1}^2$, where $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ is the QMLE of the parameters of GARCH(1,1) models. See Na, Lee and Lee (2011) for details. We reject the null hypothesis of no change point when $R_m < mT+1$. For given T and α, the decision boundary c is the same as CSSMS since they have the same asymptotic distribution.

Since GASMS and RVSMS are derived for sequential change-point detection in GARCH(p,q) models, one is not theoretically justified in applying them to data following stochastic volatility models. Nevertheless, in view of the similarity between GARCH and stochastic volatility models, it was of interest to explore the performance of GASMS and RVSMS in our simulations.

We compared the asymptotic sizes and power of PLSMS, CSSMS, GASMS, and RVSMS on SV models. Table 1 summarizes the decision boundaries c for PLSMS, CSSMS, GASMS, and RVSMS.

5.1. Simulation results under H_0

To investigate the empirical sizes of PLSMS, CSSMS, GASMS, and RVSMS under H_0, we carried out simulations based on the SV(1) models

Model 1: $\eta = 0.6$, $\beta = 3.5$, $\sigma_\epsilon = 1$, 2 and 3;

Model 2: $\eta = 0.3$, $\beta = 3$, $\sigma_\epsilon = 1$, 2 and 3;

Model 3: $\eta = 0.15$, $\beta = 2.5$, $\sigma_\epsilon = 1$, 2 and 3.

Table 2. Empirical sizes for Model 1 of PLSMS, CSSMS, GASMS, and RVSMS in different SV(1) models with m = 500, 1,000 and 5,000 when $l = 1.300$ replications were carried out for each pair (α, T).

			Model 1								
			$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
α	T	Method	m=500	1,000	5,000	m=500	1,000	5,000	m=500	1,000	5,000
0.05	1	PLSMS	0.13	0.1	0.03	0.113	0.11	0.06	0.123	0.09	0.07
	1	CSSMS	0.131	0.12	0.065	0.253	0.239	0.21	0.306	0.03	0.266
	1	GASMS	0.31	0.3	0.173	0.557	0.51	0.473	0.673	0.677	0.58
	1	RVSMS	0.213	0.157	0.08	0.29	0.263	0.18	0.383	0.253	0.267
	2	PLSMS	0.147	0.137	0.06	0.14	0.1	0.073	0.117	0.137	0.07
	2	CSSMS	0.155	0.131	0.09	0.322	0.294	0.24	0.43	0.406	0.353
	2	GASMS	0.357	0.333	0.177	0.617	0.63	0.603	0.767	0.763	0.713
	2	RVSMS	0.253	0.133	0.087	0.373	0.31	0.243	0.47	0.363	0.37
	∞	PLSMS	0.107	0.083	0.063	0.12	0.08	0.043	0.107	0.1	0.06
	∞	CSSMS	0.173	0.132	0.064	0.419	0.382	0.3	0.596	0.555	0.513
	∞	GASMS	0.477	0.363	0.17	0.83	0.757	0.733	0.897	0.88	0.843
	∞	RVSMS	0.257	0.14	0.067	0.523	0.393	0.27	0.643	0.58	0.493
0.1	1	PLSMS	0.243	0.16	0.147	0.183	0.123	0.123	0.167	0.187	0.103
	1	CSSMS	0.198	0.156	0.106	0.293	0.261	0.243	0.34	0.344	0.306
	1	GASMS	0.413	0.303	0.24	0.56	0.613	0.57	0.69	0.667	0.673
	1	RVSMS	0.29	0.183	0.127	0.367	0.317	0.213	0.38	0.39	0.323
	2	PLSMS	0.253	0.14	0.113	0.23	0.177	0.1	0.17	0.153	0.137
	2	CSSMS	0.176	0.17	0.097	0.353	0.331	0.289	0.447	0.453	0.401
	2	GASMS	0.443	0.397	0.24	0.687	0.697	0.657	0.78	0.73	0.81
	2	RVSMS	0.293	0.24	0.143	0.47	0.367	0.247	0.483	0.447	0.37
	∞	PLSMS	0.19	0.11	0.1	0.203	0.16	0.103	0.177	0.127	0.073
	∞	CSSMS	0.223	0.173	0.107	0.437	0.444	0.322	0.625	0.6	0.52
	∞	GASMS	0.473	0.407	0.237	0.843	0.833	0.817	0.897	0.87	0.877
	∞	RVSMS	0.34	0.227	0.113	0.54	0.467	0.34	0.65	0.577	0.543

The models with different values of η represent different degrees of correlation in the latent autoregressive process. Within each model, different values of σ_ϵ represent different volatilities of the latent autoregressive process. We also considered the combinations of $\alpha = 0.05, 0.1$ and $T = 1, 2, \infty$. For $T = \infty$, since it is impossible to monitor the scheme for an unlimited time horizon, the inspection time was chosen to be $10m$. Figure 1(a) provides time series plots of some realizations of these three models with $\sigma_\epsilon = 1$. Figure 1(b) provides time series plots of some realizations of Model 2 with different σ_ϵ values. Notice that spikes occur more frequently under larger variance in the latent process. Tables 2 to 4 report the proportion of rejection of H_0 for the models when m = 500, 1,000, and 5,000.

Table 3. Empirical sizes for Model 2 of PLSMS, CSSMS, GASMS, and RVSMS in different SV(1) models with $m = 500$, 1,000 and 5,000 when $l = 1.300$ replications were carried out for each pair (α, T).

			Model 2								
			$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
α	T	Method	m=500	1,000	5,000	m=500	1,000	5,000	m=500	1,000	5,000
0.05	1	PLSMS	0.107	0.1	0.073	0.1	0.07	0.043	0.113	0.093	0.057
	1	CSSMS	0.095	0.077	0.045	0.211	0.201	0.148	0.29	0.259	0.252
	1	GASMS	0.253	0.25	0.157	0.557	0.463	0.317	0.697	0.587	0.533
	1	RVSMS	0.117	0.083	0.053	0.277	0.213	0.203	0.36	0.327	0.223
	2	PLSMS	0.097	0.83	0.05	0.08	0.08	0.077	0.093	0.077	0.06
	2	CSSMS	0.13	0.075	0.05	0.258	0.266	0.177	0.396	0.343	0.282
	2	GASMS	0.31	0.22	0.18	0.58	0.447	0.373	0.757	0.727	0.643
	2	RVSMS	0.163	0.147	0.07	0.35	0.287	0.21	0.41	0.34	0.323
	∞	PLSMS	0.087	0.063	0.05	0.077	0.063	0.033	0.08	0.06	0.057
	∞	CSSMS	0.113	0.091	0.052	0.363	0.299	0.2222	0.553	0.499	0.422
	∞	GASMS	0.3	0.203	0.107	0.727	0.59	0.48	0.883	0.857	0.767
	∞	RVSMS	0.22	0.163	0.053	0.453	0.36	0.243	0.57	0.543	0.44
0.1	1	PLSMS	0.153	0.143	0.1	0.15	0.163	0.137	0.143	0.14	0.09
	1	CSSMS	0.121	0.134	0.093	0.231	0.232	0.188	0.318	0.31	0.258
	1	GASMS	0.297	0.227	0.177	0.5	0.433	0.36	0.667	0.677	0.533
	1	RVSMS	0.143	0.16	0.127	0.31	0.287	0.23	0.373	0.303	0.267
	2	PLSMS	0.163	0.16	0.13	0.137	0.127	0.113	0.153	0.14	0.09
	2	CSSMS	0.167	0.109	0.097	0.303	0.31	0.197	0.318	0.31	0.258
	2	GASMS	0.347	0.287	0.18	0.59	0.52	0.44	0.827	0.72	0.673
	2	RVSMS	0.223	0.173	0.15	0.433	0.317	0.263	0.45	0.437	0.403
	∞	PLSMS	0.153	0.13	0.127	0.137	0.097	0.08	0.113	0.097	0.107
	∞	CSSMS	0.164	0.136	0.075	0.393	0.329	0.273	0.558	0.52	0.477
	∞	GASMS	0.363	0.207	0.167	0.76	0.65	0.533	0.92	0.833	0.837
	∞	RVSMS	0.257	0.19	0.083	0.497	0.4	0.263	0.57	0.583	0.433

From Tables 2 to 4, the size distortion of PLSMS is reduced if a larger m or larger T is chosen. It can also be seen that the size distortion of PLSMS increases with η, which represents the correlation of the latent autoregressive process. This is consistent with the observation in Ng et al. (2011) that pairwise likelihood estimation is less efficient when there is a high latent correlation.

Comparing PLSMS, CSSMS, GASMS, and RVSMS, Tables 2 to 4 suggest that the proposed PLSMS generally has a empirical size closest to the significance level α, and GASMS has the heaviest size distortion. In particular, the size distortions of CSSMS, GASMS, and RVSMS increase with the variance of latent AR(p) processes, σ_ϵ^2. This may be related to the rapid change in variability of the sequence $\{X_t^2\}$ due to the occurrence of spikes. On the other hand, the size

Table 4. Empirical sizes for Model 3 of PLSMS, CSSMS, GASMS, and RVSMS in different SV(1) models with $m = 500$, 1,000 and 5,000 when $l = 1.300$ replications were carried out for each pair (α, T).

			Model 3								
			$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
α	T	Method	m=500	1,000	5,000	m=500	1,000	5,000	m=500	1,000	5,000
0.05	1	PLSMS	0.087	0.083	0.07	0.077	0.077	0.087	0.093	0.083	0.57
	1	CSSMS	0.112	0.087	0.051	0.178	0.203	0.149	0.282	0.271	0.232
	1	GASMS	0.2	0.15	0.123	0.413	0.33	0.25	0.64	0.58	0.47
	1	RVSMS	0.12	0.097	0.073	0.253	0.23	0.14	0.31	0.313	0.223
	2	PLSMS	0.083	0.093	0.07	0.093	0.08	0.053	0.083	0.09	0.063
	2	CSSMS	0.117	0.079	0.051	0.268	0.2	0.173	0.35	0.337	0.298
	2	GASMS	0.243	0.18	0.093	0.467	0.48	0.337	0.717	0.687	0.553
	2	RVSMS	0.167	0.137	0.053	0.303	0.273	0.183	0.393	0.33	0.277
	∞	PLSMS	0.087	0.07	0.037	0.077	0.05	0.053	0.073	0.047	0.057
	∞	CSSMS	0.114	0.092	0.05	0.358	0.287	0.186	0.497	0.471	0.432
	∞	GASMS	0.223	0.167	0.08	0.62	0.497	0.43	0.863	0.847	0.727
	∞	RVSMS	0.193	0.087	0.04	0.393	0.377	0.163	0.52	0.543	0.43
0.1	1	PLSMS	0.143	0.157	0.093	0.173	0.14	0.107	0.117	0.11	0.12
	1	CSSMS	0.141	0.134	0.1	0.224	0.218	0.197	0.311	0.314	0.255
	1	GASMS	0.237	0.2	0.133	0.45	0.38	0.343	0.68	0.623	0.477
	1	RVSMS	0.143	0.123	0.12	0.327	0.21	0.217	0.39	0.32	0.233
	2	PLSMS	0.143	0.15	0.093	0.147	0.12	0.127	0.133	0.107	0.12
	2	CSSMS	0.142	0.14	0.087	0.315	0.259	0.218	0.398	0.376	0.339
	2	GASMS	0.247	0.197	0.16	0.61	0.487	0.35	0.73	0.713	0.563
	2	RVSMS	0.193	0.197	0.113	0.347	0.297	0.253	0.43	0.423	0.34
	∞	PLSMS	0.143	0.137	0.093	0.12	0.123	0.087	0.13	0.123	0.107
	∞	CSSMS	0.143	0.141	0.082	0.382	0.338	0.235	0.542	0.51	0.425
	∞	GASMS	0.283	0.193	0.103	0.653	0.567	0.38	0.867	0.847	0.793
	∞	RVSMS	0.277	0.127	0.063	0.483	0.383	0.3	0.553	0.5	0.537

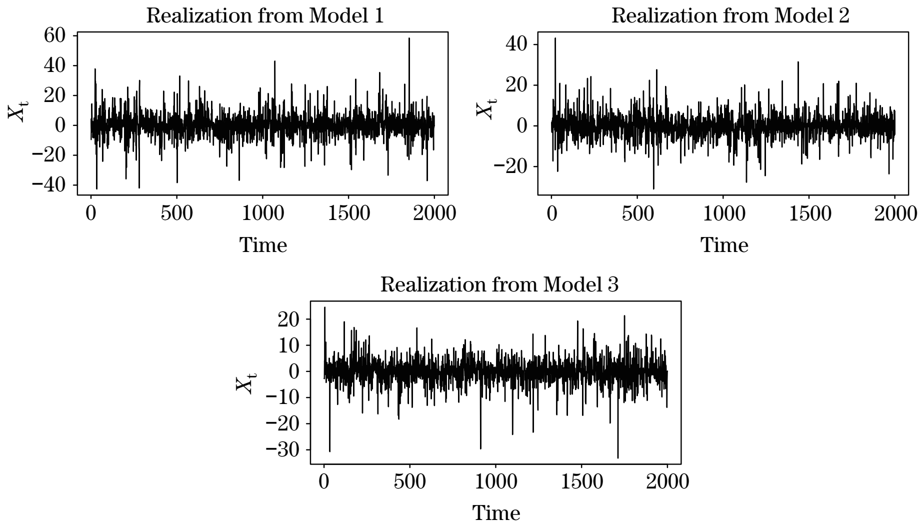
of PLSMS is not sensitive to changes in σ_ϵ. One possible reason is that the use of likelihood accounts for the effect of σ_ϵ and gives accurate critical values for the test.

5.2. Simulation results under H_A

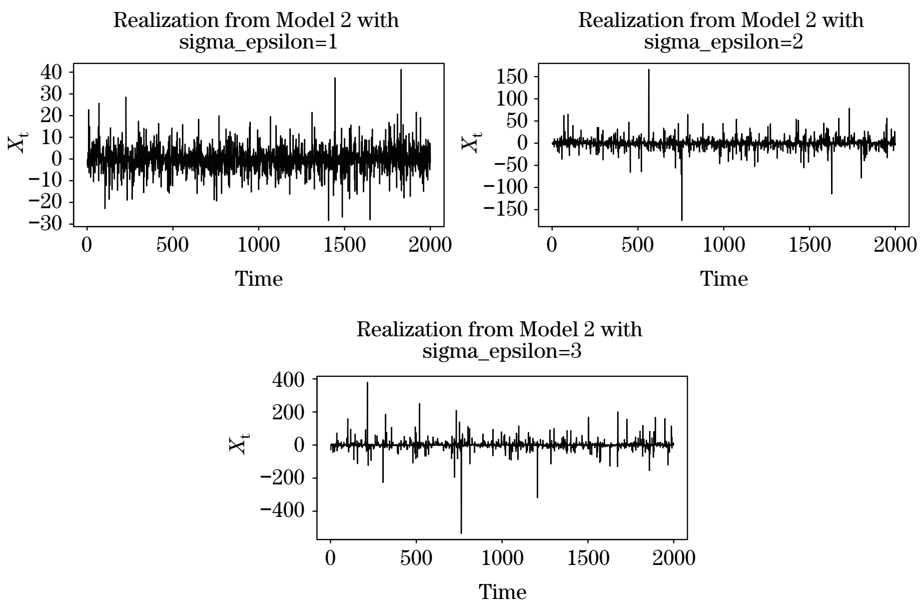
To investigate the empirical power under H_A, we carried out simulations based on three change-point models:

Model 1: $\eta_0 = 0.2$, $\sigma_{\epsilon,0} = 1.2$, $\beta_0 = -0.45$, changed to $\eta_A = 0.6$, $\sigma_{\epsilon,A} = 1.2$, $\beta_A = -0.45$ at t^*;

Model 2: $\eta_0 = 0.7$, $\sigma_{\epsilon,0} = 0.2$, $\beta_0 = -0.1$, changed to $\eta_A = 0.5$, $\sigma_{\epsilon,A} = 0.7$, $\beta_A = -0.3$ at t^*;



(a) Realizations of Model 1, 2 and 3 with $\sigma_\epsilon = 1$ in Section 5.1.



(b) Realizations of Model 2 with $\sigma_\epsilon = 1$, 2 and 3 in Section 5.1.

Figure 1. Realizations for Models in Section 5.1 with $m = 1,000$ and $T = 1$ without any change-points.

Model 3: $\eta_0 = 0.2$, $\sigma_{\epsilon,0} = 1$, $\beta_0 = 0.1$, changed to $\eta_A = 0.65$, $\sigma_{\epsilon,A} = 0.7756$, $\beta_A = 0.1$ at t^*.

Model 2 has all parameters changed, while only η is changed in Model 1. In

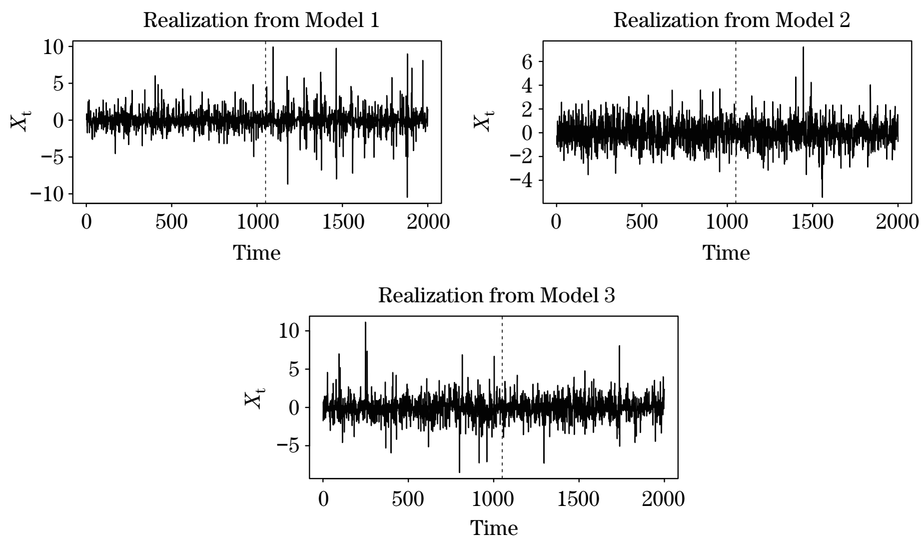


Figure 2. Realizations for Model 1, 2 and 3 in Section 5.2 with $m = 1,000$ and $t^* = 50$. The change-points are represented by the vertical dash lines.

Model 3, the parameters are restricted to change in a way such that the variance of the observed process $\{X_t\}$ remains unchanged. Tables 5 and 6 report the empirical power for Model 1 to 3 with $t^* = 50$ and 250 when $m = 500$ and 1,000, respectively. The numbers in the table are the proportion of simulation trials that PLSMS/CSSMS/GASMS/RVSMS rejects H_0. Figure 2 provides time series plots of some realizations of the three change point models with $m = 1,000$ and $t^* = 50$.

From Tables 5 and 6, PLSMS generally outperforms CSSMS, GASMS, and RVSMS for all three models. The change-points in Model 1 are well-detected by PLSMS with power greater than 0.7 in general for $m = 500$, and greater than 0.9 in general for $m = 1,000$. For Model 2, the change-points are well-detected by PLSMS with power greater than 0.9 in general, while the power of CSSMS and RVSMS are below 0.6 and the power of GASMS is below 0.7 in general. By construction, the change-point in Model 3 is difficult to be detected since the parameters change in a way such that the variance stays constant. Nevertheless, the performance of PLSMS and GASMS are comparable with power exceeding 0.5, much higher than the power of CSSMS and RVSMS.

Overall, the detection rule based on PLSMS has a higher empirical power in all the three models compared with that of CSSMS, GASMS, and RVSMS.

Table 5. Empirical power for Model 1 to 3 with m=500 and t^*=50 and 250.

			Model 1		Model 2		Model 3	
α	T	Method	$t^* = 50$	$t^* = 250$	$t^* = 50$	$t^* = 250$	$t^* = 50$	$t^* = 250$
0.05	1	PLSMS	0.777	0.437	0.98	0.773	0.41	0.233
	1	CSSMS	0.557	0.365	0.308	0.142	0.139	0.127
	1	GASMS	0.787	0.6	0.56	0.357	0.48	0.35
	1	RVSMS	0.56	0.35	0.323	0.193	0.163	0.16
	2	PLSMS	0.883	0.787	0.993	0.967	0.523	0.373
	2	CSSMS	0.64	0.512	0.331	0.254	0.144	0.149
	2	GASMS	0.87	0.827	0.667	0.513	0.533	0.46
	2	RVSMS	0.673	0.553	0.36	0.31	0.153	0.193
	∞	PLSMS	0.97	0.97	1	1	0.58	0.54
	∞	CSSMS	0.807	0.778	0.379	0.317	0.156	0.142
	∞	GASMS	0.943	0.937	0.713	0.62	0.69	0.617
	∞	RVSMS	0.77	0.753	0.467	0.37	0.183	0.187
0.1	1	PLSMS	0.833	0.503	0.963	0.837	0.51	0.313
	1	CSSMS	0.583	0.384	0.333	0.206	0.172	0.151
	1	GASMS	0.807	0.647	0.617	0.33	0.533	0.397
	1	RVSMS	0.62	0.43	0.413	0.26	0.227	0.213
	2	PLSMS	0.913	0.833	0.99	0.973	0.627	0.46
	2	CSSMS	0.693	0.573	0.413	0.319	0.184	0.202
	2	GASMS	0.903	0.83	0.697	0.583	0.607	0.527
	2	RVSMS	0.733	0.58	0.483	0.363	0.243	0.223
	∞	PLSMS	0.98	0.973	1	1	0.693	0.637
	∞	CSSMS	0.821	0.811	0.47	0.389	0.178	0.206
	∞	GASMS	0.97	0.937	0.717	0.687	0.737	0.697
	∞	RVSMS	0.77	0.77	0.513	0.48	0.383	0.313

6. Data Examples

6.1. Sequential monitoring of the S&P 500 log-return series

We applied PLSMS, CSSMS, GASMS, and RVSMS to the S&P 500 index data and compared their performances. Figure 3 plots the daily closing values and daily log returns of the S&P 500 index from January 2004 to December 2009.

In applying the PLSMS, CSSMS, GASMS, and RVSMS, the daily log return series from January 2004 to December 2005, which has 504 observations ($m = 504$), is used as the training data. From the time series plot of the daily log return data in Figure 3, the 504 training data points appear to be stationary and free from structural break. We then fit an ARSV(1) model to the training data and used maximum lag $l = 1$ to apply PLSMS. From January 2006 to December 2009, the total number of observations in the daily log return series was 1,006, about twice the number of observations in the training data set. Therefore, we

Table 6. Empirical power for Model 1 to 3 with $m = 1{,}000$ and $t^* = 50$ and 250.

			Model 1		Model 2		Model 3	
α	T	Method	$t^*=50$	$t^*=250$	$t^*=50$	$t^*=250$	$t^*=50$	$t^*=250$
0.05	1	PLSMS	0.96	0.867	1	0.997	0.573	0.423
	1	CSSMS	0.686	0.567	0.384	0.264	0.121	0.112
	1	GASMS	0.897	0.807	0.667	0.533	0.56	0.46
	1	RVSMS	0.72	0.563	0.353	0.307	0.16	0.13
	2	PLSMS	0.977	0.917	1	1	0.713	0.63
	2	CSSMS	0.787	0.732	0.44	0.362	0.13	0.117
	2	GASMS	0.927	0.92	0.777	0.723	0.653	0.583
	2	RVSMS	0.793	0.737	0.48	0.39	0.13	0.163
	∞	PLSMS	1	1	1	1	0.89	0.863
	∞	CSSMS	0.896	0.881	0.504	0.47	0.127	0.103
	∞	GASMS	0.983	0.983	0.803	0.797	0.81	0.777
	∞	RVSMS	0.847	0.873	0.447	0.467	0.15	0.12
0.1	1	PLSMS	0.973	0.953	1	0.993	0.667	0.527
	1	CSSMS	0.743	0.617	0.459	0.324	0.144	0.143
	1	GASMS	0.9	0.817	0.737	0.597	0.613	0.52
	1	RVSMS	0.68	0.62	0.433	0.373	0.217	0.207
	2	PLSMS	1	0.977	1	1	0.81	0.707
	2	CSSMS	0.853	0.793	0.508	0.436	0.158	0.159
	2	GASMS	0.957	0.957	0.833	0.767	0.73	0.707
	2	RVSMS	0.837	0.827	0.543	0.477	0.207	0.18
	∞	PLSMS	1	1	1	1	0.927	0.903
	∞	CSSMS	0.925	0.916	0.616	0.567	0.166	0.155
	∞	GASMS	0.983	0.99	0.873	0.88	0.847	0.877
	∞	RVSMS	0.907	0.88	0.517	0.583	0.193	0.177

Table 7. Performance of the different monitoring schemes for the S&P 500 index.

		$T = 2$			
		PLSMS ($T_m(1)$)	CSSMS (N_m)	GASMS (G_m)	RVSMS (R_m)
α	0.05	229	460	59	406
	0.1	204	428	54	402

chose $T = 2$. We performed the detection schemes under $\alpha = 0.05$ and 0.1.

In Table 7, T_m represents the time point where a change in parameter was declared by PLSMS. Similarly, N_m, G_m, and R_m represent the time point where a change in parameter was declared by CSSMS, GASMS, and RVSMS, respectively. By visual inspection of the daily log return data in Figure 3, a change is likely to occur at some moment prior to the dotted line when a steady increase of the daily closing index is observed. The steady, non-volatile growth in mid 2006, in contrast to the volatile daily return during 2004 to 2005, matched with the

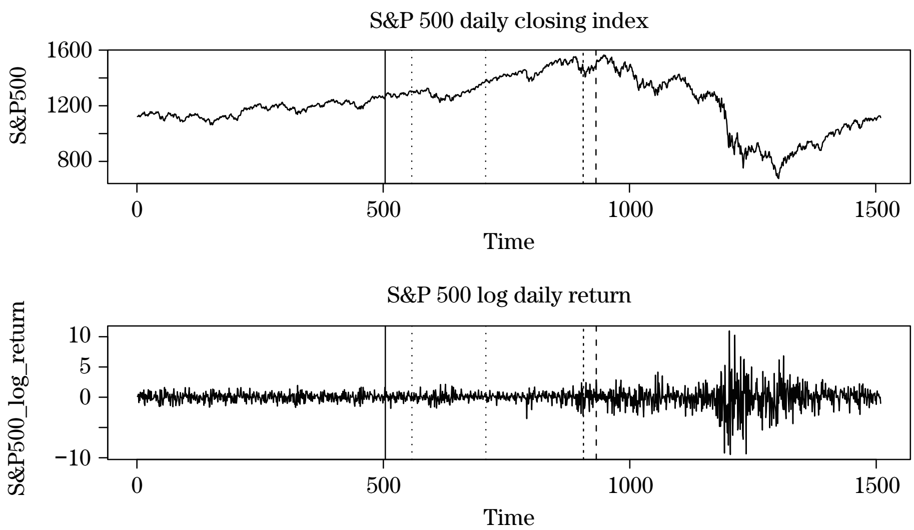


Figure 3. Plots of daily closing index and daily log return of the S&P 500 from 2004 to 2009, the observations on the left-hand side of the solid line are the training data. The dotted, dashed, dot-dashed and two-dashed line represent the time points at which T_m, N_m, G_m and R_m declare a change at $\alpha = 0.1$, respectively.

stable market condition in US and the low volatility phenomenon reflected by the CBOE Volatility Index, which was as low as 8.6 in 2006, preceded the 2007 crash. Thus, the change point was quickly detected by PLSMS. On the other hand, both CSSMS and RVSMS had detected another change point at time around 900, August 2007. This change can be attributed to the beginning of the subprime mortgage hedge fund crisis, with the major investment bank Bear Stearns revealing, in July 2007, that their two subprime hedge funds had lost nearly all of their value. Also, it can be seen that GASMS declared a change much earlier than PLSMS, CSSMS, and RVSMS. Based on visual inspection and the large size distortions in simulation in Section 5.1, we suspected that the change point detected by GASMS was probably a false alarm. Overall, the PLSMS gave the quickest detection of the change in the log-return series.

6.2. Sequential monitoring time series of counts

In recent years, the modeling of integer-valued time series has received considerable attention. For example, Kang and Lee (2009, 2014) studied retrospective change point analysis for time series of counts. However, the literature on sequential change-point detection in time series of counts remains unexplored. To demonstrate the usefulness and generality of the use of pairwise likelihood

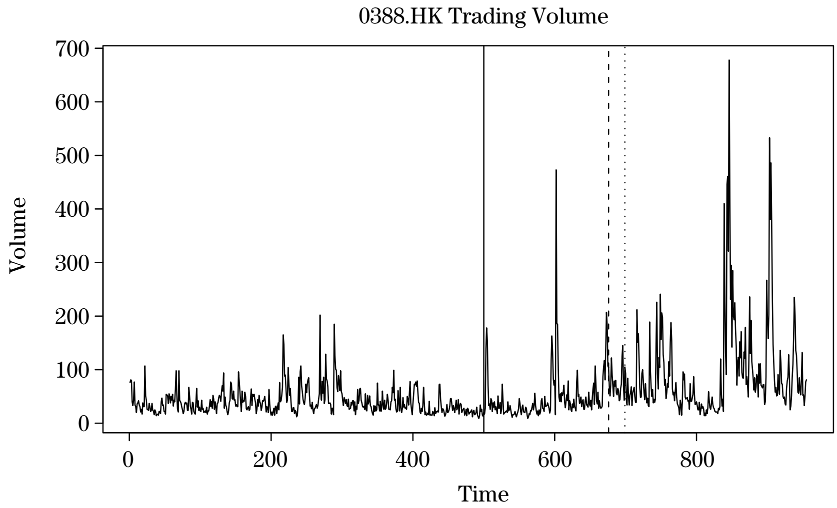


Figure 4. Plots of trading volume (per 100,000) of 0388.HK from November 2011 to September 2015, the observations on the left-hand side of the solid line are the training data. The dotted and the dashed line represent the time points at which T_m declare a change at $\alpha = 0.05$ and $\alpha = 0.1$, respectively.

in change-point analysis, we applied PLSMS to the daily trading volume (per 100,000) of Hong Kong Exchanges and Clearing Limited (0388.HK) from November 2011 to September 2015, which is depicted in Figure 4. The series is a sequence of integers with a majority of observations less than 30 and a few values as large as hundreds, which lead to some spikes in the time series plot. This spiky structure motivates the use of a latent process model, such as the Poisson regression model with AR(1) log link function.

From November 2011 to September 2015, the total number of observations in the dataset was 956 and we used the first 500 data points, which seemed stationary, as training dataset. Therefore, we chose $T = 0.912$. We then fit a Poisson regression model with AR(1) log link function to the training data and used maximum lag $l = 1$ in our monitoring scheme.

PLSMS based on T_m declares a change at time 699 under significance level $\alpha = 0.05$ and at time 676 under significance level $\alpha = 0.1$. Both are detected after a sudden large trading volume. On the other hand, by visual inspection of the daily trading volume in Figure 4, a change is likely to occur at some moment after time 600 prior to the dotted line in which a more volatile trading volume is observed. This change could be attributed to the announcement of the establishment of the Shanghai-Hong Kong Stock Connect at time 602, i.e., April 10th, 2014. The change point is quickly detected by PLSMS.

7. Conclusion

The paper proposes a sequential monitoring scheme for detecting changes in parameter values using pairwise likelihood. The scheme is shown to have asymptotically zero Type II error for any prescribed level of Type I error. With the use of pairwise likelihood, the scheme is applicable to many complicated time series models in a computationally efficient manner. For example, the scheme covers time series models involving latent processes, such as stochastic volatility models and Poisson regression models with log link function, in which the evaluation of full likelihood is computationally inefficient or intractable. Simulation and empirical studies show that the proposed monitoring scheme works well.

Acknowledgment

We would like to thank an associate editor and the two anonymous referees for their helpful suggestions and comments. Research has been supported in part by HKSAR-RGC-ECS 405012 and HKSAR-RGC-GRF 405113, 14601015.

Appendix

A. Proofs of Lemmas and Theorems

Proof of Theorem 1. Using a Taylor series expansion, we have

$$\frac{\left\|\sum_{t=m+1}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m) - \left(\sum_{t=m+1}^{m+k} L_t'(l;\boldsymbol{\theta}_0) + \sum_{t=m+1}^{m+k} L_t''(l;\boldsymbol{\theta}_c)(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)\right)\right\|}{m^{1/2}\left(1+k/m\right)c} = 0\,, \tag{A.1}$$

where $\boldsymbol{\theta}_c$ is between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_m$. By Assumptions (A3), (A4), (A7), and Lemma 1, we have

$$\begin{aligned}&\sup_{1\le k\le mT} \frac{\left\|\left[\sum_{t=m+1}^{m+k} L_t''(l;\boldsymbol{\theta}_c) - k\mathbb{E}L_t''(l;\boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)\right\|}{m^{1/2}\left(1+k/m\right)c}\\ &\le \frac{1}{c}\sup_{1\le k\le mT}\left\|\frac{\sum_{t=m+1}^{m+k} L_t''(l;\boldsymbol{\theta}_c) - k\mathbb{E}L_t''(l;\boldsymbol{\theta}_0)}{m+k}\sqrt{m}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)\right\| \xrightarrow{p} 0\,, \end{aligned}\tag{A.2}$$

as $m\to\infty$. Combining (A.1) and (A.2), we have, as $m\to\infty$,

$$\sup_{1\le k<mT} \frac{\left\|\sum_{t=m+1}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m) - \left(\sum_{t=m+1}^{m+k} L_t'(l;\boldsymbol{\theta}_0) + k\mathbb{E}L_t''(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)\right)\right\|}{m^{1/2}\left(1+k/m\right)c} \xrightarrow{p} 0\,. \tag{A.3}$$

By a Taylor series expansion, we have that for some $\boldsymbol{\theta}_{\tilde{c}}$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_m$,

$$\left\| \frac{\sum_{t=1}^{m} L_t'(l;\hat{\boldsymbol{\theta}}_m)}{m} - \left(\frac{\sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0)}{m} + \frac{\sum_{t=1}^{m} L_t''(l;\boldsymbol{\theta}_{\tilde{c}})}{m} (\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) \right) \right\| = 0 \,. \quad \text{(A.4)}$$

As $\sum_{t=1}^{m} L_t'(l;\hat{\boldsymbol{\theta}}_m) = 0$ by definition, combining with Assumptions (A1) and (A4), (A.4) is

$$m^{1/2} \left\| \hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0 + (\mathbb{E}(L_t''(l;\boldsymbol{\theta}_0))^{-1} \frac{\sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0)}{m} \right\| \xrightarrow{a.s.} 0, \quad \text{(A.5)}$$

as $m \to \infty$. Multiplying $(\mathbb{E}L_t''(\boldsymbol{\theta}_0)k/m)/(\big(1+k/m\big)c)$ on both sides of (A.5) and taking supremum over k, we obtain

$$\sup_{1 \le k < mT} \frac{\left\| \Big(k\mathbb{E}L_t''(\boldsymbol{\theta}_0)\Big)(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0) + (k/m)\sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0) \right\|}{m^{1/2}\Big(1+k/m\Big)c} \xrightarrow{p} 0. \quad \text{(A.6)}$$

Hence, combining (A.3) and (A.6), we have

$$\sup_{1 \le k < mT} \frac{\left\| \sum_{t=m+1}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m) - \Big(\sum_{t=m+1}^{m+k} L_t'(l;\boldsymbol{\theta}_0) - (k/m)\sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0) \Big) \right\|}{m^{1/2}\Big(1+k/m\Big)c} \xrightarrow{p} 0. \quad \text{(A.7)}$$

Using Lemma 1c), we have

$$\begin{pmatrix} m^{-1/2}\sum_{t=m}^{m(1+t)} L_t'(l;\boldsymbol{\theta}_0) \\ m^{-1/2} t \sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow{\mathcal{D}[0,T]} \begin{pmatrix} \mathbf{W}_{\mathbf{M}}(1+t) - \mathbf{W}_{\mathbf{M}}(1) \\ t\mathbf{W}_{\mathbf{M}}(1) \end{pmatrix}.$$

By the Continuous Mapping Theorem, we have

$$m^{-1/2} \left\{ \sum_{t=m}^{m(1+t)} L_t'(l;\boldsymbol{\theta}_0) - t\sum_{t=1}^{m} L_t'(l;\boldsymbol{\theta}_0) \right\} \xrightarrow{\mathcal{D}[0,T]} \mathbf{W}_{\mathbf{M}}(1+t) - (1+t)W_{\mathbf{M}}(1) \,. \quad \text{(A.8)}$$

Combining (A.7) and (A.8), we have

$$\begin{aligned} &\sup_{1 \le k \le mT} \frac{\left\| \sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m) \right\|}{m^{1/2}\Big(1+k/m\Big)c} \\ &\xrightarrow{\mathcal{D}} \sup_{0<t<T} \frac{\left\| \mathbf{M}^{-1/2}\left(\mathbf{W}_{\mathbf{M}}(1+t) - (1+t)\mathbf{W}_{\mathbf{M}}(1)\right) \right\|}{(1+t)c} \,, \end{aligned} \quad \text{(A.9)}$$

as $m \to \infty$. Next, the covariance of $\mathbf{W}_{\mathbf{M}}(1+t) - (1+t)\mathbf{W}_{\mathbf{M}}(1)$ is $t(1+t)\mathbf{M}$, which implies that for $t \ge 0$,

$$\Big\{ \mathbf{M}^{-1/2}\left(\mathbf{W}_{\mathbf{M}}(1+t) - (1+t)\mathbf{W}_{\mathbf{M}}(1)\right) \Big\}$$

$$\stackrel{\mathcal{D}}{=} \left\{(1+t)\mathcal{W}_1\left(\frac{t}{1+t}\right), (1+t)\mathcal{W}_2\left(\frac{t}{1+t}\right), \ldots, (1+t)\mathcal{W}_d\left(\frac{t}{1+t}\right)\right\}^T,$$

where $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_d$ are independent standard Wiener processes. Thus, we have

$$\sup_{0<t<T} \frac{\left\|\mathbf{M}^{-1/2}\left(\mathbf{W}_{\mathbf{M}}(1+t) - (1+t)\mathbf{W}_{\mathbf{M}}(1)\right)\right\|}{(1+t)c} \stackrel{\mathcal{D}}{=} \max_{1\le i\le d} \sup_{0<s<\frac{T}{1+T}} \frac{|\mathcal{W}_i(s)|}{c}. \tag{A.10}$$

From (A.9) and (A.10), we obtain

$$\sup_{1\le k\le mT} \frac{\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+k/m\Big)c} \stackrel{\mathcal{D}}{\to} \max_{1\le i\le d} \sup_{0<s<\frac{T}{1+T}} \frac{|\mathcal{W}_i(s)|}{c}, \tag{A.11}$$

as $m \to \infty$. Finally,

$$\begin{aligned}
&\lim_{m\to\infty} P(T_m(l) \le mT | H_0) \\
&= \lim_{m\to\infty} P\left(\sup_{1\le k\le mT} \left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\left(1+k/m\right)c \,\middle|\, H_0\right) \\
&= 1 - P\left(\max_{1\le i\le d} \sup_{0\le s\le T/(1+T)} \frac{|\mathcal{W}_i(s)|}{c} \le 1 \,\middle|\, H_0\right) \\
&= 1 - \left(P\left\{\sup_{0\le t\le T/(1+T)} \frac{|W(t)|}{c} \le 1\right\}\right)^d,
\end{aligned}$$

yielding (3.5).

To prove (3.7), similar to (A.9), we have to show that

$$\begin{aligned}
&\sup_{1\le k<\infty} \frac{\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+k/m\Big)c} \\
&\stackrel{\mathcal{D}}{\to} \sup_{0<t<\infty} \frac{\left\|\mathbf{M}^{-1/2}\left(\mathbf{W}_{\mathbf{M}}(1+t) - (1+t)\mathbf{W}_{\mathbf{M}}(1)\right)\right\|}{(1+t)c}.
\end{aligned} \tag{A.12}$$

By the ρ-mixing Hajek-Renyi inequality, Theorem 1 of Wan (2013) and (3.6), we get for $\epsilon > 0$ and $\{c_j\}$ a sequence of non-decreasing real numbers,

$$\begin{aligned}
&P\left(\max_{m\le k\le n} \left|\frac{1}{c_k}\sum_{t=1}^{k} L_t'(l;\theta_0)\right| \ge \epsilon\right) \\
&\le \frac{1}{\epsilon^2}\left(\sum_{j=1}^{m} \frac{Var(L_0'(l;\theta_0))}{c_m^2} + \sum_{j=m+1}^{n} \frac{Var(L_0'(l;\theta_0))}{c_j^2}\right).
\end{aligned} \tag{A.13}$$

Replacing m by $m+mT$, n by $n+m$, and then taking $\lim_{T\to\infty}\lim_{m\to\infty}\lim_{n\to\infty}$ on both sides of (A.13), and putting $c_k = m^{1/2}(1+\frac{k-m}{m})c$, we get for $\epsilon > 0$,

$$\lim_{T\to\infty}\lim_{m\to\infty}\lim_{n\to\infty} P\left(\sup_{mT\le k\le n}\left|\frac{1}{m^{1/2}(1+k/m)c}\sum_{t=1}^{m+k} L_t'(l;\theta)\right| \ge \epsilon\right) = 0. \tag{A.14}$$

By the Law of Iterated Logarithm, we have

$$\sup_{T\le t<\infty}\frac{\|\mathbf{W}_{\mathbf{M}}(1+t)\|}{(1+t)c} \overset{a.s.}{\to} 0, \qquad \text{as } T\to\infty. \tag{A.15}$$

From Berkes et al. (2004), we have that (A.12) follows from (A.8), (A.14), and (A.15).

Our proof of Theorem 2 is based on Berkes et al. (2004).

Proof of Theorem 2. Since

$$\begin{aligned}&\lim_{m\to\infty} P(T_m(l)\le mT|H_A)\\ &= \lim_{m\to\infty} P\left(\sup_{1\le k\le mT}\left\|\sum_{t=m+1}^{m+k}\widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\| > m^{1/2}\Big(1+k/m\Big)c\,\middle|\, H_A\right),\end{aligned}$$

it suffices to show that

$$\sup_{1\le k\le mT}\frac{\left\|\widehat{M}_m(l)^{-1/2}\sum_{t=m+1}^{m+k}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+k/m\Big)c} \overset{p}{\to} \infty\,. \tag{A.16}$$

When $k < t^* - m$, we have $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Thus, (A.11) implies that

$$\sup_{1\le k< t^*-m}\frac{\left\|\sum_{t=m+1}^{m+k}\widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+k/m\Big)c} = O_p(1). \tag{A.17}$$

On the other hand, as $\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0 = o_p(1)$ and, for sufficiently large $k \ge t^* - m$, we have $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. Thus, the Ergodic Theorem implies that

$$\sum_{t=t^*}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m) = (m+k-t^*+1)\left[\mathbb{E}_{\boldsymbol{\theta}_1}[L_t'(l;\boldsymbol{\theta}_0)] + o_p(1)\right]\,.$$

Hence, taking $k = mT$, we have

$$\begin{aligned}&\frac{\left\|\sum_{t=t^*}^{m+mT}\widehat{M}_m(l)^{-1/2}L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+mT/m\Big)c}\\ &= \frac{\left\|\widehat{M}_m(l)^{-1/2}\left[(m+mT-t^*+1)\mathbb{E}_{\boldsymbol{\theta}_1}[L_t'(l;\boldsymbol{\theta}_0)] + o_p(m)\right]\right\|}{m^{1/2}\Big(1+mT/m\Big)c}\end{aligned}$$

$$\geq c_1 m^{1/2} \xrightarrow{p} \infty, \tag{A.18}$$

for some $c_1 > 0$, since $\mathbb{E}_{\boldsymbol{\theta}_1}(L_t'(l;\boldsymbol{\theta}_0)) \neq \mathbf{0}$ by Assumption (A8). Combining (A.17) and (A.18), we have

$$\begin{aligned}\sup_{1\leq k\leq mT} \frac{\left\|\sum_{t=m+1}^{m+k} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+k/m\Big)} &\geq \frac{\left\|\sum_{t=m+1}^{m+mT} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+mT/m\Big)c} \\ &= O_p(1) + \frac{\left\|\sum_{t=t^*}^{m+mT} \widehat{M}_m(l)^{-1/2} L_t'(l;\hat{\boldsymbol{\theta}}_m)\right\|}{m^{1/2}\Big(1+mT/m\Big)c} \\ &\geq c_2 m^{1/2} \xrightarrow{p} \infty\,,\end{aligned}$$

for some $c_2 > 0$, yielding (A.16). A similar argument can be applied for the case of $T_m^*(l)$.

Proof of Lemma 3. For a), let $u(a_t, a_{t+1})$ be a polynomial in (a_t, a_{t+1}), and

$$Q(u, x_t, x_{t+1}) = \frac{\int\int u(a_t,a_{t+1}) k_\beta(x_t;a_t) k_\beta(x_{t+1};a_{t+1}) f_{\alpha_t,\alpha_{t+1}}(a_t,a_{t+1}) da_t da_{t+1}}{\int\int k_\beta(x_t;a_t) k_\beta(x_{t+1};a_{t+1}) f_{\alpha_t,\alpha_{t+1}}(a_t,a_{t+1}) da_t da_{t+1}}\,,$$

where $k_\beta(x_t;\alpha_t) = f_{X_t|\lambda_t}(x_t)$ and $\lambda_t = e^{\beta+\alpha_t}$. Take $\boldsymbol{\theta} = (\beta, \sigma_\epsilon, \eta_1, \ldots \eta_p)$. Let $\partial(\cdot)/\partial\theta_i$, $i = 1, 2, \ldots, p+2$, be the partial derivative operator with respect to the i-th element of $\boldsymbol{\theta}$. For $l = 1$, we have

$$\begin{aligned}&\mathbb{E}\left(\left|\frac{\partial}{\partial\theta_i} L_t(l;\boldsymbol{\theta})\right|^{4v}\right) \\ &= \mathbb{E}\left(|\partial/(\partial\theta_i) p_t(1;\boldsymbol{\theta})|^{4v}\right) = \mathbb{E}\left(\left|\frac{\partial/(\partial\theta_i) f_{X_t,X_{t+1}}(x_t,x_{t+1};\boldsymbol{\theta})}{f_{X_t,X_{t+1}}(x_t,x_{t+1};\boldsymbol{\theta})}\right|^{4v}\right) \\ &\leq \mathbb{E}\left(\left|\frac{\int\int |p_i(a_t,a_{t+1})| k_\beta(x_t;a_t) k_\beta(x_{t+1};a_{t+1}) f_{\alpha_t,\alpha_{t+1}}(a_t,a_{t+1}) da_t da_{t+1}}{\int\int k_\beta(x_t;a_t) k_\beta(x_{t+1};a_{t+1}) f_{\alpha_t,\alpha_{t+1}}(a_t,a_{t+1}) da_t da_{t+1}}\right|^{4v}\right) \\ &= \mathbb{E}\left(Q(|p_i|, x_t, x_{t+1})^{4v}\right)\,,\end{aligned} \tag{A.19}$$

where $f_{\alpha_t,\alpha_{t+1}}(a_t,a_{t+1})$ is the joint density of (α_t, α_{t+1}) and $p_i(a_t, a_{t+1})$ is a polynomial in (a_t, a_{t+1}).

To simplify the notation, let $(x, \tilde{x}, a, \tilde{a}) = (x_t, x_{t+1}, a_t, a_{t+1})$, $\int_{[k]}$ be the k-th folded integral on the real line, $K_\beta = K_\beta(x, \tilde{x}, a, \tilde{a}) = k_\beta(x, a) k_\beta(\tilde{x}, \tilde{a})$, $f_{\mu,\Sigma}(\cdot,\cdot)$ be the bivariate normal density function with mean μ and variance Σ, and $g_{\beta,\mu,\Sigma}(x,\tilde{x}) = \int_{[2]} K_\beta f_{\mu,\Sigma}(a,\tilde{a}) da d\tilde{a}$. Then (A.19) reduces to

$$\mathbb{E}\left(Q(|p_i|, x_t, x_{t+1})^{4v}\right)$$

$$
\begin{aligned}
&= \int_{[2]} \left[\frac{\int_{[2]} |p_i(a,\tilde{a})| K_\beta f_{0,\Sigma}(a,\tilde{a}) da d\tilde{a}}{g_{\beta,0,\Sigma}(x,\tilde{x})} \right]^{4v} g_{\beta_0,0,\Sigma_0}(x,\tilde{x}) dx d\tilde{x} \\
&= \int_{[2]} \left[\frac{\int_{[2]} |p_i(a,\tilde{a})| K_0[(f_{\beta,\Sigma}(a,\tilde{a}))/(f_{\beta_0,\Sigma_0}(a,\tilde{a}))] f_{\beta_0,\Sigma_0}(a,\tilde{a}) da d\tilde{a}}{g_{0,\beta,\Sigma}(x,\tilde{x})} \right]^{4v} \\
&\qquad g_{0,\beta_0,\Sigma_0}(x,\tilde{x}) dx d\tilde{x} \\
&= \int_{[2]} \frac{\left\{ \int_{[2]} [|p_i(a,\tilde{a})| [(f_{\beta,\Sigma}(a,\tilde{a}))/(f_{\beta_0,\Sigma_0}(a,\tilde{a}))]]^{8v} K_0 f_{\beta_0,\Sigma_0}(a,\tilde{a}) da d\tilde{a} \right\}^{1/2}}{g^{4v}_{0,\beta,\Sigma}(x,\tilde{x})} \\
&\qquad g^{4v+1/2}_{0,\beta_0,\Sigma_0}(x,\tilde{x}) dx d\tilde{x} \\
&\le \int_{[4]} \left[|p_i(a,\tilde{a})| \frac{f_{\beta,\Sigma}(a,\tilde{a})}{f_{\beta_0,\Sigma_0}(a,\tilde{a})} \right]^{8v} K_0 f_{\beta_0,\Sigma_0}(a,\tilde{a}) da d\tilde{a} dx d\tilde{x} + \int_{[2]} \frac{g^{8v+1}_{0,\beta_0,\Sigma_0}(x,\tilde{x})}{g^{8v}_{0,\beta,\Sigma}(x,\tilde{x})} dx d\tilde{x} \\
&= \mathbb{E} \left\{ \left[|p_i(a,\tilde{a})| \frac{f_{\beta,\Sigma}(a,\tilde{a})}{f_{\beta_0,\Sigma_0}(a,\tilde{a})} \right]^{8v} \right\} + \int_{[2]} \frac{g^{8v+1}_{0,\beta_0,\Sigma_0}(x,\tilde{x})}{g^{8v}_{0,\beta,\Sigma}(x,\tilde{x})} dx d\tilde{x} \\
&\le \mathbb{E} \left\{ \left[|p_i(a,\tilde{a})| \frac{f_{\beta,\Sigma}(a,\tilde{a})}{f_{\beta_0,\Sigma_0}(a,\tilde{a})} \right]^{8v} \right\} + \int_{[4]} K_0 f_{\beta,\Sigma}(a,\tilde{a}) \left[\frac{f_{\beta_0,\Sigma_0}(a,\tilde{a})}{f_{\beta,\Sigma}(a,\tilde{a})} \right]^{8v+1} dx d\tilde{x} da d\tilde{a} \\
&=: A + B,
\end{aligned}
$$

where

$$
\begin{aligned}
A &= \int_{[2]} \left[|p_i(a,\tilde{a})| \frac{f_{\beta,\Sigma}(a,\tilde{a})}{f_{\beta_0,\Sigma_0}(a,\tilde{a})} \right]^{8v} f_{\beta_0,\Sigma_0}(a,\tilde{a}) da d\tilde{a}, \\
B &= \int_{[2]} \left[\frac{f_{\beta_0,\Sigma_0}(a,\tilde{a})}{f_{\beta,\Sigma}(a,\tilde{a})} \right]^{8v} f_{\beta_0,\Sigma_0}(a,\tilde{a}) da d\tilde{a}.
\end{aligned}
$$

The two inequalities follow from the Holder's inequality. The integrand of A can be expressed as

$$
\begin{aligned}
&\left[|p_i(a,\tilde{a})| \frac{f_{\beta,\Sigma}(a,\tilde{a})}{f_{\beta_0,\Sigma_0}(a,\tilde{a})} \right]^{8v} f_{\beta_0,\Sigma_0}(a,\tilde{a}) \\
&= C \frac{|p_i(a,\tilde{a})|^{8v} \left(\exp\left(-1/2 \mathbf{a}^T \Sigma^{-1} \mathbf{a} \right) \right)^{8v}}{\left(\exp\left(-1/2 \mathbf{a}^T \Sigma_0^{-1} \mathbf{a} \right) \right)^{8v}} \exp\left(-\frac{1}{2} \mathbf{a}^T \Sigma_0^{-1} \mathbf{a} \right) \\
&= C \, |p_i(a,\tilde{a})|^{8v} \exp\left(-\frac{1}{2} \mathbf{a}^T \left(\Sigma_0^{-1} - 8v \left(\Sigma_0^{-1} - \Sigma^{-1} \right) \right) \mathbf{a} \right), \qquad \text{(A.20)}
\end{aligned}
$$

for some constant C, where $\boldsymbol{a} = (a, \tilde{a})^T$. Similarly, the integrand of B can be expressed as

$$C_2 \exp\left(-\frac{1}{2}\mathbf{a}^T\left(\Sigma_0^{-1} + 8v\left(\Sigma_0^{-1} - \Sigma^{-1}\right)\right)\mathbf{a}\right) \tag{A.21}$$

for some constant C_2. Combining (A.20), (A.21), and Assumption (B3), both A and B are integrable, implying that $\mathbb{E}\|L_t'(1;\boldsymbol{\theta})\|^{4v}$ exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and is finite.

Similar arguments can be employed to show $\mathbb{E}\left(|\partial/(\partial\theta_i)p_t(j;\boldsymbol{\theta})|^{4v}\right) < \infty$ for any fixed $j \geq 1$. Thus, for $l \geq 1$, we have $\mathbb{E}\left[\|L_t'(l;\boldsymbol{\theta})\|^{4v}\right] \leq \sum_{j=1}^{l}\mathbb{E}\left[\|l_t'(j;\boldsymbol{\theta})\|^{4v}\right] < \infty$, completing the proof of part a).

For b), from Ng et al. (2011), for $l = 1$, we have

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\partial}{\partial\theta_i\partial\theta_j}L_t(l;\boldsymbol{\theta})\right|^2\right] &= \mathbb{E}\left([Q(p_{i,j,1},x_t,x_{t+1})Q(p_{i,j,2},x_t,x_{t+1})]^2\right)\\ &\leq \mathbb{E}\left([Q(|p_{i,j,1}|,x_t,x_{t+1})]^4\right) + \mathbb{E}\left([Q(|p_{i,j,2}|,x_t,x_{t+1})]^4\right)\\ &< \infty\,,\end{aligned}$$

where $p_{i,j,1}(a_t,a_{t+1})$ and $p_{i,j,2}(a_t,a_{t+1})$ are two polynomials in (a_t,a_{t+1}) and the last inequality is established in the proof of a). Similar arguments yield $\mathbb{E}\left[|\partial L_t(l;\boldsymbol{\theta})/(\partial\theta_i\partial\theta_j)|^2\right] < \infty$ for $l > 1$. Thus, the proof of Lemma 3 is complete.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis* **34**, 1-16.

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH(p, q) models. *Econometric Theory* **20**, 1140-67.

Csorgo, M. and Revesz, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press.

Davis, R. A. and Yau, C. Y. (2011). Comment on pairwise likelihood in time series models. *Statistica Sinica* **21**, 255-77.

Doukhan, P. (1994). *Mixing Properties and Examples*. Lecture Notes in Statistics.

Fuh, C. D. (2006). Optimal Change Point Detection in State Space Models. Academia Sinica.

Gombay, E. and Horváth, L. (2009). Sequential tests and change detection in the covariance structure of weakly stationary time series. *Communications in Statistics - Theory and Methods* **38**, 2872-83.

Gombay, E. and Serban, D. (2009). Monitoring parameter change in AR(p) time series models. *Journal of Multivariate Analysis* **100**, 715-25.

Gut, A. and Steinebach, J. (2002). Truncated sequential change-point detection based on renewal counting processes. *Scandinavian Journal of Statistics* **29**, 693-719.

Heagerty, J. and Lele R. (1998). Composite likelihood approach to binary spatial data. *Journal of the American Statistical Association* **93**, 1099-1111.

Kang, J. and Lee, S. (2009). Parameter change test for random coefficient integer-valued autoregressive processes with application to polio data analysis. *Journal of Time Series Analysis* **30**, 239-258.

Kang, J. and Lee, S. (2014). Parameter change test for Poisson autoregressive models. *Scandinavian Journal of Statistics* **41**, 1136-1152.

Kirch, C. and Tadjuidje Kamgaing, J. (2015). On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference* **161**, 25-49.

Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability and Its Applications* **5**, 204-208.

Kuk, A. and Nott, D. J. (2000). A pairwise likelihood approach to analyzing correlated binary data. *Statistics and Probability Letters* **47**, 329-35.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of Royal Statistical Society B* **57**, 613-58.

Lindsay, B., Yi, G. Y. and Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* **21**, 71-105.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and Their Applications* **29**, 309-15.

Na, O., Lee, J. and Lee, S. (2012). Change point detection in copula ARMA-GARCH models. *Journal of Time Series Analysis* **33**, 554-69.

Na, O., Lee, Y. and Lee, S. (2011). Monitoring parameter change in time series models. *Statistical Methods & Applications* **20**, 171-199.

Ng, C. T., Joe, H., Karlis, D. and Liu, J. (2011). Composite likelihood for time series models with a latent autoregressive process. *Statistica Sinica* **21**, 279-305.

Nott, D. J. and Ryden, T. (1999). Pairwise likelihood methods for inference in image models. *Biometrika* **86**, 661-76.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.

Paik, J. and Ying, Z. (2012). A composite likelihood approach for spatially correlated survival data. *Computational Statistics and Data Analysis* **56**, 209-16.

Tsay, R. S. (2010). *Analysis of Financial Time Series.* Third edition. John Wiley & Sons.

Tsay, R. S. (2012). *An Introduction to Analysis of Financial Data with R.* John Wiley & Sons.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519-28.

Wan, Y. (2013). On the rho-mixing Hajek-Renyi inequality. *Journal of Jianghan University Natural Science Edition* **41**, No. 1.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621-629.

Department of Statistics, Lady Shaw Building, Chinese University of Kong Kong, Shatin, Hong Kong

E-mail: isaacleung201506@gmail.com

Department of Statistics, Lady Shaw Building, Chinese University of Kong Kong, Shatin, Hong Kong

E-mail: s1009613922@sta.cuhk.edu.hk

Department of Statistics, Lady Shaw Building, Chinese University of Kong Kong, Shatin, Hong Kong

E-mail: cyyau@sta.cuhk.edu.hk

(Received January 2015; accepted February 2016)