Statistica Sinica Preprint No: SS-2018-0269	
Title	A Self-Normalized Approach to Sequential Change-point Detection for Time Series
Manuscript ID	SS-2018-0269
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0269
Complete List of Authors	Ngai Hang Chan Wai Leong Ng and Chun Yip Yau
Corresponding Author	Ngai Hang Chan
E-mail	nhchan@sta.cuhk.edu.hk
Notice: Accepted version subject to English editing.	

Statistica Sinica

A Self-Normalized Approach to Sequential Change-point Detection for Time Series

Ngai Hang Chan[1,2], Wai Leong Ng[3] and Chun Yip Yau[2]

Southwestern University of Finance and Economics[1],

The Chinese University of Hong Kong[2],

Hang Seng University of Hong Kong[3]

Abstract:

This paper proposes a self-normalization sequential change-point detection method for time series. In testing for parameter changes, most of the traditional sequential monitoring tests utilize a CUSUM-based test statistic, which involves a long-run variance estimator. However, the commonly used long-run variance estimators require the choice of bandwidth parameter which could be sensitive to the performance. Moreover, the traditional tests usually suffer from severe size distortion due to the slow convergence rate to the limit distribution in the early monitoring stage. In this article, a self-normalization method is proposed to tackle these issues. We establish null asymptotic and the consistency of the proposed sequential change-point test under general regularity conditions. Simulation experiments and applications to railway bearing temperature data are conducted for illustrations.

Key words and phrases: ARMA-GARCH model, on-line detection, pairwise likelihood, quickest detection, sequential monitoring, stochastic volatility model.

1. Introduction

Let $\{X_t\}$ be a sequence of observations governed by a statistical model with a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. We say that there is a change-point at time t^* if the model for $\{X_1, X_2, \ldots, X_{t^*-1}\}$ has a parameter $\boldsymbol{\theta}_0$, but the model for $\{X_{t^*}, X_{t^*+1}, \ldots\}$ has a parameter $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$. 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 known as the sequential change-point detection, sequential monitoring or on-line monitoring problem.

Sequential change-point detection has been receiving considerable attention with the advance in serial data collection in engineering, econometrics, finance and statistics. Quick detection of changes in the underlying process structure with controlled false alarm rate is crucial as it allows practitioners to make immediate decisions and necessary adjustments in a timely manner. Sequential change-point detection schemes can mainly be classified into two categories. The first category involves an average run length (ARL)-type constraint on false alarms. For example, Lai (1995) proposed the window limited generalized likelihood ratio scheme for detecting changes in time series models with a minimal detection delay for a given ARL to false alarm; see also Lai (1998), Yakir (1997), Polunchenko and Tartakovsky (2010), and Han, Tsung and Xian (2017). However, as Lai (1995) pointed out, although the ARL constraint stipulates a long expected duration to false alarm, it does not necessarily imply that the probability of having a false alarm before some specified time is small, and hence false alarms at the initial stage cannot be controlled. In practice, frequent false alarms will result in a waste of resource in repairing and replacement.

The second category utilizes a hypothesis testing framework so that the probability of false alarms can be controlled with Type I error. To control Type I error in a sequential setting, one straightforward approach is to employ retrospective change-point tests repeatedly via a multiple testing procedure such as simulation-based adjustment, Bonferroni adjustment or false discovery rate (FDR) method; see Hawkins, Qiu and Kang (2003), Hawkins and Zamba (2005) and Choi, Ombao and Ray (2008). However, the simulation-based adjustment only works in simple models with independence assumption, and Bonferroni or FDR adjustment will lead to conservative results. A more sophisticated approach to control the Type I error is by establishing the asymptotic distribution of the running maximum of a cumulative sum process which is standardized by a long-run variance estimator; see Chu, Stinchcombe and White (1996) and Zeileis *et al.* (2005) in a linear regression setting; Gut and Steinebach (2002) for renewal counting process; Berkes *et al.* (2004) for GARCH models; Fuh (2006) for state space models; Gombay and Serban (2009) for AR models; Gombay and Horváth (2009) for covariance structure of weakly stationary time series; Na, Lee and Lee (2011), Kirch and Tadjuidje Kamgaing (2015), and Leung, Ng and Yau (2017) for general time series model. However, one challenge for this class for sequential change-point detection scheme is that a consistent long-run variance estimator is involved in the test statistic. The commonly used lag-window type and kernel type long-run variance estimators require an appropriate choice of bandwidth parameter, which is a difficult issue in practice. Moreover, severe size distortions are observed in the empirical studies of Berkes *et al.* (2004), Zeileis *et al.* (2005), Na, Lee and Lee (2011), among others.

In this paper, we propose a new sequential monitoring scheme for detecting change-points in general time series models that achieves asymptotically exact Type I error without estimating the long-run variance. The key idea of the proposed method is to introduce the self-normalization (SN) concept to the sequential monitoring scheme. The self-normalization concept has found success in various statistical problems, for example, see Lobato (2001), Shao (2010), Shao and Zhang (2010), Zhang, Li and Shao (2014), Huang, Volgushev and Shao (2015) and Zhang and Lavitas (2018). In particular, the self-normalization method utilizes a self-normalizer to replace the long-run variance estimator for standardizing the test statistic. Thus, the estimation of the long-run variance is bypassed and hence no user-chosen bandwidth parameter is involved. As a result, the asymptotic null distribution of SN-based test is pivotal, and the critical values can be tabulated through simulations. Moreover, the SN-based test has the "better size but less power" phenomenon, see Lobato (2001) and Shao and Zhang (2010), which is a desirable feature to resolve the aforementioned size distortion problem. For a detailed review on self-normalization for time series and its recent developments, see Shao (2015). In this paper, we derive a SN-based statistic for sequentially monitoring change-points for general time series models. We study the null asymptotic distribution and the consistency of the proposed SN-based sequential change-point test under general regularity conditions. The asymptotic null distribution of the SN-based test statistic is quite different from that of test statistic using long-run variance estimator since it accounts for the extra randomness due to the self-normalizer. The test is shown to have asymptotically zero Type II error with a prescribed level of Type I error. Simulation studies demonstrate that the scheme has a significant improvement in size distortion and still maintains a good power.

This paper is organized as follows. Section 2 describes the problem setting and the proposed SN-based sequential monitoring scheme. Section 3 presents the general regularity assumptions and establishes the asymptotic behavior of the scheme under the null and alternative hypotheses. Simulation experiments and empirical studies on application to monitoring of railway bearing temperature are provided in Section 4. Section 5 concludes. Technical proofs are provided in the online supplement.

2. Problem Setting and Sequential monitoring scheme

In this section, we first introduce the setting of the sequential change-point problem. Then, we propose the SN-based sequential monitoring scheme.

2.1 Problem Settings

Assume that $\{X_t\}_{t=1,2,\ldots}$ is a stationary and ergodic random process with joint density $f_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta}$ is the parameter vector in a compact space $\boldsymbol{\Theta}$ and $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ can be regarded as a class of parametric model with parameter $\boldsymbol{\theta}$. Some examples include ARMA(p,q)-GARCH(r,s) models and stochastic volatility models SV(p) with autoregressive order p, which will be discussed in details in the simulation studies in Section 4.

It is essential to assume that before we start monitoring changes in the sequentially collected data sequence, a training sample of m historical stationary data points $\{x_1, x_2, \ldots, x_m\}$ has been observed for initial estimation of pre-change parameter value, say $\boldsymbol{\theta}_0$, since it is necessary to know the initial values of the parameters that are subjected to change; see Chu, Stinchcombe and White (1996), Gut and Steinebach (2002), Berkes *et al.* (2004), Gombay and Horváth (2009), Na, Lee and Lee (2011), Kirch and Tadjuidje Kamgaing (2015), among others.

Starting from time $t = m + 1$ onwards, we then observe fresh data $\{X_t\}_{t=m+1,m+2,\dots}$ sequentially and monitor whether or not a change has occurred in $\boldsymbol{\theta}$ using the null hypothesis:

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

against the alternative hypothesis:

$$H_1 : \boldsymbol{\theta} = \begin{cases} \boldsymbol{\theta}_0 \,, & \text{for } t = 1, 2, \dots, t^* - 1 \,, \\ \boldsymbol{\theta}_1 \,, & \text{for } t = t^*, t^* + 1, \dots, m + mT \,, \end{cases}$$

where $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$, $t^* > m$ is the unknown change-point and $t^* = m + k^*$ for some $k^* > 0$, and mT is the monitoring horizon which is the maximum number of observations to be inspected. The quantity $T \in (0, \infty]$, which is a user-specified positive number (possibly infinite), can be regarded as the ratio of the monitoring horizon to the size of the training sample. In the asymptotic analysis of the proposed monitoring scheme, we let the size of the training sample m grow to infinity with T remains fixed such that the monitoring horizon mT grows to infinity. The asymptotic distribution of the monitoring test statistic depends on T, see Section 3 for details.

By using the stationary pre-change training sample $\{x_1, x_2, \ldots, x_m\}$, we can consistently estimate the parameter of interest $\boldsymbol{\theta}$ using some objective function $L(\mathbf{X}_t, \boldsymbol{\theta})$. This general framework includes all classical estimation methods such as maximum likelihood estimators, M-estimators, least-squares estimators and generalized moment estimators. The parameter estimates $\hat{\boldsymbol{\theta}}_m$ will then satisfy the following system of equations:

$$\sum_{t=1}^{m} L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m) = 0\,, \tag{2.1}$$

where $\{\mathbf{X}_t\}_{t=1,\ldots,m}$ are the training sample and L is the objective function taking values in $\mathbb{R}^d$ with $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}_0)) = 0$, where the expectation is taken under H_0, and $\boldsymbol{\theta}_0$ is the unique solution. Note that $\{\mathbf{X}_t\}_{t=1,\ldots,m}$ can be univariate (i.e. $\mathbf{X}_t = x_t$) or multivariate (i.e. $\mathbf{X}_t = (x_{t-r}, x_{t-r+1}, \ldots, x_{t-1}, x_t)$) depending on the objective function L and the underlying time series models, for example, $\mathbf{X}_t = x_t$ for i.i.d. data, and $\mathbf{X}_t = (x_{t-p}, x_{t-p+1}, \ldots, x_{t-1}, x_t)$ if L is the score function of an AR(p) model. In Section 3, regularity conditions on the objective function L are provided to ensure the consistency of the parameter estimators $\hat{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_0$ as $m \to \infty$.

2.2 SN-based sequential monitoring scheme (SNSMS)

In this subsection, we derive the sequential change-point monitoring scheme based on the self-normalization principle. As $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}_0)) = 0$ under H_0, if $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}_0)) \neq 0$ under H_1, then it is reasonable to monitor the new incoming data points $\{\mathbf{X}_t\}_{t=m+1,m+2,\ldots}$ by the following cumulative sum (CUSUM) statistics:

$$S_m(k, \hat{\boldsymbol{\theta}}_m) = \sum_{t=m+1}^{m+k} L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\,,$$

where $\hat{\boldsymbol{\theta}}_m$ is the parameter estimate in (2.1). The rationale behind is as follows. If the process remains unchanged, then $\{L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\}_{t=m+1,m+2,\ldots}$ should have expectation close to 0. Hence, under weak dependence condition on $\{L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\}$, the cumulative sum behaves like a Wiener process by the weak invariance principle, see Lin and Lu (1996) for various mixing and moment conditions. On the other hand, if the process has a structural change at t^*, then all of the additional terms $\{L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\}_{t>t^*}$ have a non-zero expectation. Hence, the cumulative sum $S_m(k, \hat{\boldsymbol{\theta}}_m)$ diverges as $k \to \infty$. This difference can be used to distinguish between the null and alternative hypothesis under an appropriate standardization of $S_m(k, \hat{\boldsymbol{\theta}}_m)$. Traditional sequential tests or monitoring schemes require the use of a long-run variance estimator to standardize the cumulative sum process. It is difficult to choose an appropriate choice of bandwidth parameter for the lag-window type and kernel type long-run variance estimators in practice. Moreover, large size distortion is commonly found in empirical studies among the existing literature; see, for example, Berkes *et al.* (2004), Zeileis *et al.* (2005), and Na, Lee and Lee (2011). To tackle these problems, we employ the self-normalization method by replacing the long-run variance estimator by the self-normalization factor, which does not require the choice of bandwidth parameters. Specifically, we consider the self-normalized process $S_m(k, \hat{\boldsymbol{\theta}}_m)' D_m(\hat{\boldsymbol{\theta}}_m)^{-1} S_m(k, \hat{\boldsymbol{\theta}}_m)$, where

$$D_m(\hat{\boldsymbol{\theta}}_m) = \frac{1}{m^2} \sum_{t=1}^{m} \left\{ \left(\sum_{j=1}^{t} L(\mathbf{X}_j, \hat{\boldsymbol{\theta}}_m) \right) \left(\sum_{j=1}^{t} L(\mathbf{X}_j, \hat{\boldsymbol{\theta}}_m) \right)' \right\}, \quad (2.2)$$

is the self-normalization factor, and define the monitoring test statistic $\mathbb{M}_m(k)$ at time $m + k$ as

$$\mathbb{M}_m(k) = \frac{S_m(k, \hat{\boldsymbol{\theta}}_m)' D_m(\hat{\boldsymbol{\theta}}_m)^{-1} S_m(k, \hat{\boldsymbol{\theta}}_m)}{m(1 + \frac{k}{m})^2}.$$

The proposed monitoring scheme is defined based on the stopping time

$$T_m = \begin{cases} \min\{k : \mathbb{M}_m(k) > c, 1 \leq k \leq mT\}, \\ mT + 1, \text{if } \mathbb{M}_m(k) \leq c, \text{for all } 1 \leq k \leq mT, \end{cases} \quad (2.3)$$

where c is the decision boundary determined by the asymptotic distribution derived in Section 3. This asymptotic distribution is shown to be pivotal so that estimation of long-run variance is avoided.

The case $T \in (0, \infty)$ is called a closed-end monitoring scheme where we stop monitoring after a fixed number of observations mT, and the case $T = \infty$ is called an open-end monitoring scheme where we always continue monitoring before a change is found. Starting from time $k = 1$, we check whether $\mathbb{M}_m(k) > c$. If yes, then we set $T_m = k$ and the monitoring scheme terminates by rejecting H_0 and declaring that a change in parameter had occurred at some time on or before k. Otherwise, we proceed to time $k+1$ and continue the monitoring. For closed-end monitoring, if the condition $\mathbb{M}_m(k) > c$ has not been met at time mT, then set $T_m = mT + 1$ and the monitoring scheme terminates by declaring that a change in parameter did not occur and concluding that H_0 is not rejected. Note that the update at each time point involves neither numerical optimization nor resampling, and hence can be computed efficiently.

The performance of the proposed sequential monitoring scheme can be assessed by studying its empirical size, power and run length. The size is the probability of declaring an occurrence of a change when no change has occurred, the power is the probability that a change-point is declared on or before mT given that the change had occurred at $k^* \leq mT$, and the run length is the time until an alarm is given, i.e., T_m.

3. Regularity assumptions and asymptotic theory

The asymptotic analysis of the proposed monitoring scheme is based on letting the size of the training sample m grow to infinity. Intuitively, with a larger training sample size m, we can estimate the unknown parameters more accurately. Hence, with a more accurate estimated pre-change model, the monitoring scheme will be more sensitive to the deviations from the incoming data following a post-change model with different parameter values. Similar asymptotic settings have been studied in Chu, Stinchcombe and White (1996), Gut and Steinebach (2002), Berkes *et al.* (2004), Gombay and Horváth (2009), Na, Lee and Lee (2011), Kirch and Tadjuidje Kamgaing (2015), among others.

In this section, we derive the null and alternative asymptotics of the sequential monitoring test under some regularity assumptions on the underlying process and the objective function. Let $L'(\mathbf{X}_t, \boldsymbol{\theta})$ be the gradient matrix for the objective function $L(\mathbf{X}_t, \boldsymbol{\theta})$ with respect to the parameter $\boldsymbol{\theta}$. Define the vector norm $\|c\|$ as the supremum norm of a vector c. When A is a matrix, define the matrix norm $\|A\| = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$. Denote $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$ and $\lceil x \rceil = \min\{z \in \mathbb{Z} : z \geq x\}$.

3.1 Regularity conditions under null hypothesis H_0

In this subsection, we give regularity conditions under which we can derive asymptotic results of the initial estimation of the parameters, the monitoring test statistic and the stopping time of the proposed procedure.

Assumption $\mathcal{A}$.1. The true parameter value $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^d$.

Assumption $\mathcal{A}$.2. The process $\{\mathbf{X}_t\}$ is stationary and ergodic.

Assumption $\mathcal{A}$.3. $\mathbb{E}(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|L(\mathbf{X}_t, \boldsymbol{\theta})\|) < \infty$ and $\boldsymbol{\theta}_0$ is the unique zero of $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}))$, i.e., for all $\epsilon > 0$, there exists a $\kappa > 0$ such that $\|\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}))\| > \kappa$ for all $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \epsilon$.

Assumption $\mathcal{A}$.4. $\mathbb{E}(\|L(\mathbf{X}_t, \boldsymbol{\theta}_0)\|^{2+\delta}) < \infty$ for some $\delta > 0$ and $\{\mathbf{X}_t\}$ is a strong mixing sequence with mixing coefficients $\alpha(n)$ satisfying

$$\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty\,.$$

For the definition of strong mixing coefficient $\alpha(n)$, see Lin and Lu (1996).

Assumption $\mathcal{A}$.5. $L(\mathbf{X}_t, \boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ in a neighborhood $V_{\boldsymbol{\theta}_0}$ of $\boldsymbol{\theta}_0$. Also, $\mathbb{E}(L'(\mathbf{X}_t, \boldsymbol{\theta}_0))$ is positive definite, and $\mathbb{E}(\sup_{\boldsymbol{\theta}\in V_{\boldsymbol{\theta}_0}} \|L'(\mathbf{X}_t, \boldsymbol{\theta})\|) < \infty$.

3.2 Regularity conditions under alternative hypothesis H_1

In this subsection, we give some mild conditions on the post-change process such that the proposed monitoring scheme will stop in finite time with probability approaching one as $m \to \infty$ if a change occurs.

Assumption $\mathcal{B}$.1. The change-point occurs after time m, i.e., $t^* = \lfloor m\phi \rfloor$ for $1 < \phi < 1 + T$ such that the observations after the change-point are $\{\mathbf{X}_t^*\}_{t\geq t^*}$ where $\{\mathbf{X}_t^*\}$ is a stationary and ergodic process with parameter value $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$.

Assumption $\mathcal{B}$.2. (a) $\mathbb{E}(L(\mathbf{X}_t^*, \boldsymbol{\theta}_0)) = c \neq 0$ for some constant $c \in \mathbb{R}^d$.

(b) $\mathbb{E}(\sup_{\boldsymbol{\theta} \in U_{\boldsymbol{\theta}_0}} \|L(\mathbf{X}_t^*, \boldsymbol{\theta})\|) < \infty$ for some neighborhood $U_{\boldsymbol{\theta}_0}$ of $\boldsymbol{\theta}_0$.

Assumption $\mathcal{B}$.1 is a condition on the location of the change-point to ensure that enough post-change observations will be accumulated for the monitoring. Assumption $\mathcal{B}$.2 on moment conditions for the post-change process $\{\mathbf{X}_t^*\}$ is crucial for the consistency of the sequential monitoring scheme since it indicates that $L(\mathbf{X}_t, \boldsymbol{\theta}_0)$ behaves differently before and after the change-point.

3.3 Asymptotic properties of SNSMS under H_0

The following lemma shows the consistency of the parameter estimation and the invariance principle of the partial sum process. These results are standard and thus we refer the proofs to Theorem 3 of Kirch and Tadjuidje Kamgaing (2012) and Theorem 3.2.1 of Lin and Lu (1996).

Lemma 1. *(a) Under Assumptions $\mathcal{A}$.1 to $\mathcal{A}$.5, we have $\hat{\boldsymbol{\theta}}_m = \boldsymbol{\theta}_0 + O_p(m^{-1/2})$.*

(b) Under Assumptions $\mathcal{A}$.2 and $\mathcal{A}$.4, we have the weak invariance principle of the partial sum process

$$\frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} L(\mathbf{X}_t, \boldsymbol{\theta}_0) \overset{\mathcal{D}[0,1+T]}{\longrightarrow} \mathbf{W}_{\mathbf{M}}(r)\,,$$

as $m \to \infty$, where $r \in [0, 1+T]$, $\lfloor x \rfloor$ is the largest integer smaller than

x and $\overset{\mathcal{D}[0,1+T]}{\longrightarrow}$ denotes weak convergence in $\mathcal{D}[0, 1+T]$, the space of right-continuous function with left limit on $[0, 1+T]$. Here $\mathbf{W}_{\mathbf{M}}(r)$ is a d-dimensional Gaussian process with mean $\mathbb{E}(\mathbf{W}_{\mathbf{M}}(r)) = 0$ and covariance function $\mathbb{E}(\mathbf{W}_{\mathbf{M}}(u)\mathbf{W}'_{\mathbf{M}}(v)) = \min(u, v)\mathbf{M}(\boldsymbol{\theta}_0)$, where $\mathbf{M}(\boldsymbol{\theta}_0)$ is the long-run covariance matrix defined as

$$\mathbf{M}(\boldsymbol{\theta}_0) = \sum_{i=-\infty}^{\infty} \mathbb{E}[L(\mathbf{X}_1, \boldsymbol{\theta}_0)L(\mathbf{X}_{i+1}, \boldsymbol{\theta}_0)'] .$$

The following theorem shows the weak convergence of $D_m(\hat{\boldsymbol{\theta}}_m)$ of SNSMS. The results can be used to derived the decision boundary $c = c(\alpha, d, T)$ with an asymptotically correct size α such that we can control the Type I error of the procedures.

Theorem 1. *Under Assumptions $\mathcal{A}$.1 to $\mathcal{A}$.5, we have*

(a) $D_m(\hat{\boldsymbol{\theta}}_m) \xrightarrow{\mathcal{D}} \mathbf{M}(\boldsymbol{\theta}_0)^{\frac{1}{2}}\mathbf{V}\left(\mathbf{M}(\boldsymbol{\theta}_0)^{\frac{1}{2}}\right)'$ as $m \to \infty$, where $\mathbf{V} = \int_0^1 (\mathbb{B}_d(r) - r\mathbb{B}_d(1))(\mathbb{B}_d(r) - r\mathbb{B}_d(1))'dr$, $\mathbb{B}_d(r)$ is the standard d-dimensional Wiener process, and $D_m(\hat{\boldsymbol{\theta}}_m)$ is defined in (2.2).

(b) The asymptotic size of the SNSMS with decision boundary c for $T < \infty$ is given by

$$\lim_{m\to\infty} P(T_m \leq mT|H_0) = P\left(\sup_{0\leq s<T} \frac{\mathbb{U}_d(s)'\mathbf{V}^{-1}\mathbb{U}_d(s)}{(1+s)^2} > c\right), \quad (3.1)$$

where $\mathbb{U}_d(s) = \mathbb{B}_d(1+s) - (1+s)\mathbb{B}_d(1)$.

(c) *For $T=\infty$, if $\{\mathbf{X}_t\}$ is geometrically ρ-mixing sequence, i.e., the ρ-mixing coefficient satisfies*

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

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

$$\begin{aligned}\lim_{m\to\infty} P(T_m<\infty|H_0) &= P\left(\sup_{0\le s<\infty} \frac{\mathbb{U}_d(s)'\mathbf{V}^{-1}\mathbb{U}_d(s)}{(1+s)^2} > c\right)\\ &= P\left(\sup_{0\le u<1} \mathbb{B}_d^*(u)'\mathbf{V}^{-1}\mathbb{B}_d^*(u) > c\right), \end{aligned} \tag{3.2}$$

where $\mathbb{U}_d(s) = \mathbb{B}_d(1+s) - (1+s)\mathbb{B}_d(1)$ and $\mathbb{B}_d^*$ is a standard d-dimensional Wiener process independent of $\mathbf{V}$.

Note that $\mathbf{V}$, which is a functional of $\{\mathbb{B}_d(r)\}_{r\in[0,1)}$, is independent of $\{\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)\}_{s\in[0,\infty)}$, and the last equality of (3.2) is due to the rescaling property of the Brownian motion that

$$\{\mathbb{U}_d(s)\} = \{\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)\} \overset{d}{=} \left\{(1+s)\mathbb{B}_d^*\left(\frac{s}{1+s}\right)\right\}\,,$$

see the proof of Theorem 1(c) in the online supplement and Theorem 1 in Hušková and Koubková (2005) for details. The ρ-mixing assumption for the case $T=\infty$ is required to fulfill the conditions for using the ρ-mixing Hájek-Rényi inequality that controls the tail probability of the monitoring statistics over an infinite horizon; for more details, see Wan (2013).

The probabilities in (3.1) and (3.2) can be evaluated through Monte Carlo simulation and hence the decision boundary $c = c(\alpha, d, T)$ can be determined so that the asymptotic size equals to a pre-specified significance level α, i.e.,

$$P\left(\sup_{0\leq s<T} \frac{[\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)]'\mathbf{V}^{-1}[\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)]}{(1+s)^2} > c\right) = \alpha\,, \tag{3.3}$$

or

$$P\left(\sup_{0\leq u<1} \mathbb{B}_d^*(u)'\mathbf{V}^{-1}\mathbb{B}_d^*(u) > c\right) = \alpha\,. \tag{3.4}$$

Table 1 summarizes the decision boundaries c of SNSMS for different values of (α, d, T), which will be used for the simulation studies in Section 4. Given (α, d, T), to find the decision boundaries c of SNSMS for closed-end and open-end schemes, we use simulations to solve c in (3.3) and (3.4), respectively. The standard Brownian motions $\{\mathbb{B}_d(u) : u \in [0, 1+T)\}$ and $\{\mathbb{B}_d^*(v) : v \in [0, 1)\}$ are approximated by partial sum processes of independent normal random variables in a dense grid of length 10^{-4}. The decision boundaries c is taken as the empirical percentile of the test statistic in $5,000,000$ repetitions.

3.4 Asymptotic properties of SNSMS under H_1

The following theorem shows that under H_1, the asymptotic power of the SNSMS converges to 1 as $m \to \infty$. Hence, when there is a change-point, the proposed monitoring scheme declares a change in parameter with prob-

Table 1: *Decision boundaries c under different α, d and T for T_m in SNSMS.*

		d=1				d=2				d=3			
	T	1	2	10	∞	1	2	10	∞	1	2	10	∞
α	0.05	33.1	44.2	60.5	66.2	69.3	92.3	126.4	138.4	112.0	149.5	204.2	223.6
	0.1	22.6	30.2	41.3	45.2	50.8	67.7	92.7	101.4	85.2	113.8	155.5	170.3

ability approaching 1.

Theorem 2. *Consider the SNSMS with decision boundary c satisfying* (3.3) *or* (3.4) *for a given significance level $\alpha \in (0, 1)$. Under Assumptions $\mathcal{B}$.1 and $\mathcal{B}$.2, the asymptotic power of the SNSMS equals to* 1*, i.e.,*

$$\lim_{m\to\infty} P(T_m \le mT|H_1) = 1 \text{ if } T < \infty\,, \text{ and } \lim_{m\to\infty} P(T_m < \infty|H_1) = 1 \text{ if } T = \infty\,.$$

4. Simulation and empirical studies

In this section, we study the finite sample performance of the SNSMS by considering two models for simulation experiments. We focus on the ARMA(p,q)-GARCH(r,s) models and stochastic volatility models SV(p), and compare the performance with other existing monitoring schemes in terms of their size, power and average run length (ARL), i.e., the average time until an alarm is given. Since an inflated empirical size may lead to a higher power and shorter run length, we report the size-corrected powers and ARLs corresponding to the true size α for a fair comparison.

4.1 Change in mean levels in ARMA(p,q)-GARCH(r,s) models

Consider the ARMA(p,q)-GARCH(r,s) model with mean level at μ_t,

$$\begin{aligned} X_t &= \mu_t + Y_t\,, \qquad Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} \ldots + \theta_q \epsilon_{t-q}\,, \\ \epsilon_t &= \sigma_t \eta_t\,, \qquad \sigma_t^2 = \omega + \sum_{1\le i\le r} \alpha_i X_{t-i}^2 + \sum_{1\le j\le s} \beta_j \sigma_{t-j}^2\,, \end{aligned}$$

where $(\omega, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)$ are parameters of the model, $\{\eta_t\}$ are mean-zero i.i.d. random variables with unit variance, and the mean level $\boldsymbol{\theta} = \mu_t$ is the parameter of interest.

ARMA(p,q)-GARCH(r,s) model is a general time series model which includes the famous GARCH model as a special case. In recent decades, the use of ARMA(p,q)-GARCH(r,s) model in time series modeling and forecasting become popular in engineering, econometrics and finance. For example, Pham and Yang (2010) used ARMA-GARCH models to explain the wear and fault condition of machine, Liu and Shi (2013) applied ARMA-GARCH models in forecasting short-term electricity prices. In financial time series, the GARCH model is commonly used as an alternative to stochastic volatility models, see Tsay (2010, 2012). To monitor the mean level changes in the process, CUSUM monitoring scheme is used, see Chu, Stinchcombe and White (1996) and Na, Lee and Lee (2011) for details. For simplicity, we call it the CUSUM sequential monitoring scheme (CUSMS).

The CUSMS for the mean level of ARMA(p,q)-GARCH(r,s) model is based on the stopping time

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

where the sample mean $\hat{\mu}_m = \frac{1}{m}\sum_{t=1}^{m} X_t$, and the long-run variance estimator

$$\hat{\sigma}_m^2 = \sum_{j=-\lceil m^{\frac{1}{3}} \rceil}^{\lceil m^{\frac{1}{3}} \rceil}\left(1 - \frac{|j|}{\lceil m^{\frac{1}{3}} \rceil}\right)\hat{\gamma}_l(j)$$

where $\hat{\gamma}_l(j) = \frac{1}{m}\sum_{t=j+1}^{m}(X_t - \hat{\mu}_m)(X_{t-j} - \hat{\mu}_m)$. We reject the null hypothesis of no change-point when $C_m < mT + 1$. We also apply the self-normalization approach with $L(X_t, \hat{\boldsymbol{\theta}}_m) = X_t - \hat{\mu}_m$ to obtain the self-normalization CUSUM sequential monitoring scheme (SN-CUSMS):

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

where $D_m(\hat{\mu}_m) = \frac{1}{m^2}\sum_{t=1}^{m}\left\{\left(\sum_{j=1}^{t}(X_t - \hat{\mu}_m)\right)^2\right\}$ is the self-normalizer. In the following subsections, we compare CUSMS and SN-CUSMS on ARMA$(1,1)$-GARCH$(1,1)$ models in terms of their size, size-corrected power and size-corrected ARL. Table 2 summarizes the decision boundary c for CUSMS. Time series plots of some realizations of the models are provided in the online supplement. The decision boundary c of CUSMS for each (α, d, T) is obtained by numerically solving c from the equation

$$1 - \left(\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{\pi^2(2k+1)^2T}{8c^2(1+T)}\right)\right)^d = \alpha\,, \tag{4.1}$$

where d is the dimension of the process; see Lemma 2 and equation (3.8) in Leung, Ng and Yau (2017). The decision boundary c for SN-CUSMS with $d = 1$ is given in Table 1.

Table 2: *Decision boundaries c for C_m in CUSMS when $d = 1$.*

α	$T = 1$	$T = 2$	$T = 10$	$T = \infty$
0.05	1.585	1.830	2.137	2.241
0.1	1.386	1.600	1.869	1.960

4.2 Change in mean levels in ARMA(p,q)-GARCH(r,s) models: Simulation results under H_0

To investigate the empirical sizes of CUSMS and SN-CUSMS under H_0, we conduct simulations based on the ARMA$(1,1)$-GARCH$(1,1)$ models with parameter $\boldsymbol{\theta} = (\mu_t, \omega, \phi_1, \theta_1, \alpha_1, \beta_1)$, where $\phi_1 < 1$, $\theta_1 < 1$, $\alpha_1 + \beta_1 < 1$ and $\mu_t = 0$ for all t under H_0. For comparison, the following models are considered: Model 1: $(\omega, \phi_1, \theta_1, \alpha_1, \beta_1) = (0.8, 0.5, 0.5, 0.15, 0.2)$; Model 2: $(\omega, \phi_1, \theta_1, \alpha_1, \beta_1) = (0.6, 0.7, 0.8, 0.2, 0.1)$. We considered the combinations of $\alpha = 0.05$, 0.1 and $T = 1$, 2, 10, i.e., the monitoring horizons mT are m, $2m$ and $10m$ respectively. Table 3 reports the proportion of rejection of H_0 when $m = 100$, 300, 500, 1000 and 2000. The row $T = 10^*$ corresponds to the open-end scheme $T = \infty$ with a monitoring horizon of

$10m$, as an infinite monitoring horizon is impossible in practice. Since the decision boundaries for $T = \infty$ is always larger than that for finite T under the same (α, d) as shown in Table 1, the empirical sizes for the row $T = 10$ are expected to be larger than that of the row $T = 10^*$. From Table 3, the size distortion of CUSMS is much more severe than that of SN-CUSMS. The proposed SN-CUSMS generally has an empirical size close to the significance level α. Note that the decay of the covariance structure of the ARMA-GARCH model is slower in Model 2 than that in Model 1. Thus, the long-run variance estimator in Model 2 is more difficult to estimate accurately. This is a possible reason for the greater size distortion of CUSMS in Model 2 than Model 1.

4.3 Change in mean levels in ARMA(p,q)-GARCH(r,s) models: Simulation results under H_1

To investigate the size-corrected empirical powers and ARLs of CUSMS and SN-CUSMS under H_1, we conducted simulations based on the ARMA$(1, 1)$-GARCH$(1, 1)$ models with parameter $(\mu_t, \omega, \phi_1, \theta_1, \alpha_1, \beta_1)$ satisfying $\phi_1 < 1$, $\theta_1 < 1$, $\alpha_1 + \beta_1 < 1$ and change-point at $t^* = m + k^*$.

Model $\mathcal{A}$: $(\omega, \phi_1, \theta_1, \alpha_1, \beta_1) = (0.6, 0.7, 0.8, 0.2, 0.1)$ with mean level changed from $\mu_t = 0$ for $t < m + k^*$ to $\mu_t = \Delta$ for $t \geq m + k^*$.

The parameter values of the Model $\mathcal{A}$ are the same as that of Model 2 in

Table 3: *Empirical sizes for Models 1 and 2 of CUSMS (C) and SN-CUSMS (SN-C) in ARMA(1,1)-GARCH(1,1) models with m = 100, 300, 500, 1000 and 2000. The number of replications for each pair of (α, T) is 2500.*

		Model 1					Model 2				
T	Method	m=100	300	500	1000	2000	m=100	300	500	1000	2000
		Significance level $\alpha = 0.05$									
1	C	0.124	0.087	0.089	0.077	0.076	0.209	0.150	0.132	0.116	0.088
1	SN-C	0.055	0.044	0.052	0.052	0.053	0.06	0.053	0.049	0.051	0.046
2	C	0.145	0.115	0.090	0.082	0.076	0.222	0.154	0.136	0.107	0.095
2	SN-C	0.052	0.05	0.054	0.051	0.053	0.062	0.053	0.054	0.045	0.051
10	C	0.157	0.116	0.098	0.087	0.072	0.249	0.173	0.148	0.121	0.101
10	SN-C	0.053	0.053	0.053	0.053	0.051	0.067	0.058	0.064	0.054	0.05
10^*	C	0.130	0.085	0.081	0.064	0.057	0.234	0.151	0.118	0.107	0.092
10^*	SN-C	0.045	0.047	0.041	0.045	0.044	0.060	0.055	0.045	0.048	0.045
		Significance level $\alpha = 0.1$									
1	C	0.204	0.166	0.156	0.143	0.132	0.278	0.228	0.214	0.17	0.156
1	SN-C	0.107	0.106	0.092	0.097	0.088	0.118	0.097	0.098	0.088	0.092
2	C	0.228	0.156	0.157	0.143	0.13	0.307	0.238	0.225	0.194	0.163
2	SN-C	0.109	0.095	0.102	0.099	0.099	0.122	0.098	0.098	0.101	0.095
10	C	0.242	0.177	0.159	0.151	0.131	0.357	0.255	0.211	0.203	0.151
10	SN-C	0.115	0.11	0.1	0.095	0.096	0.132	0.099	0.1	0.106	0.104
10^*	C	0.192	0.146	0.128	0.119	0.104	0.326	0.226	0.201	0.168	0.143
10^*	SN-C	0.094	0.083	0.084	0.088	0.088	0.119	0.095	0.087	0.088	0.086

Section 4.2 except for the mean level μ_t. Table 4 reports the size-corrected empirical powers and ARLs for various k^* and Δ with $m = 500$. The size-corrected empirical power is the proportion of simulation trials that CUSMS and SN-CUSMS rejects H_0 using a decision boundary calibrated so that the empirical size of the scheme is α under H_0. The ARLs are computed by taking average of the stopping times conditioning on an alarm is given. From Table 4, while SN-CUSMS generally maintains a reasonable power and ARL, CUSMS usually has a better power and a shorter ARL. This "better size but less power" phenomenon is consistent with the findings in the literature in other contexts; see Lobato (2001), and Shao and Zhang (2010). The results from Table 4 also indicates that the detection rate and timing depends on the magnitude of the parameter change Δ and the location of the change-point $t^* = m + k^*$. The smaller the k^* (closer to time m) and larger the magnitude of the change Δ, the higher the detection rate and shorter the detection delay. This phenomenon is mainly due to the fact that the CUSUM statistics $S_m(k, \hat{\boldsymbol{\theta}}_m) = \sum_{t=m+1}^{m+k} L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)$ contains more pre-change data when k^* is large, and hence more post-change data are needed before significance is reached; see, e.g. Chu, Stinchcombe and White (1996) and Na, Lee and Lee (2011) for similar observations.

Also, in general, under the same (T, Δ, k^*), larger m will results in better power since we can estimate the unknown parameters more accurately with a larger training sample. Hence, with a more accurate estimated pre-change model, the monitoring scheme will be more sensitive to the deviations from the incoming data following a post-change model with different parameter values. This phenomenon is also mentioned in Chu, Stinchcombe and White (1996), Berkes *et al.* (2004), Kirch and Tadjuidje Kamgaing (2015) and many others.

4.4 Change in parameters in time series models with latent process

Consider the stochastic volatility model SV(p) with autoregressive order p,

$$X_t = Z_t e^{\frac{1}{2}(\alpha_t+\beta)}\,, \alpha_t = \eta_1\alpha_{t-1} + \eta_2\alpha_{t-2} + \ldots + \eta_p\alpha_{t-p} + \epsilon_t\,,$$

where Z_t are i.i.d. $N(0,1)$, ϵ_t are i.i.d. $N(0,\sigma^2)$, $|\eta_i| < 1$ for $i = 1,\ldots,p$ and $\boldsymbol{\theta} = (\eta_1,\ldots,\eta_p,\sigma,\beta)$ are the parameters of interest. For the sequential change-point analysis for SV(p), Leung, Ng and Yau (2017) suggested a sequential monitoring scheme PLSMS using pairwise likelihood for general time series model. Define the stopping time $P_m(l)$ of PLSMS by

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

where c is the corresponding decision boundary of PLSMS, and mT is the pre-specified maximum inspection time, $L_t'(l;\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}L_t(l;\boldsymbol{\theta}) = \sum_{j=1}^{l}\frac{\partial}{\partial\boldsymbol{\theta}}p_t(j;\boldsymbol{\theta})$

Table 4: *Size-corrected empirical powers and average run lengths (ARL) for CUSMS (C) and SN-CUSMS (SN-C) in ARMA(1,1)-GARCH(1,1) Model $\mathcal{A}$ with $m = 500$, $\Delta = 0.5$, 1 and 2, and $k^* = 50, 250$ and 500 with significance levels $\alpha = 0.05$ and 0.1. The number of replications for each pair of (α, T) is 2500.*

			$\Delta = 0.5$					
α	T	Method	$k^* = 50$	ARL	$k^* = 250$	ARL	$k^* = 500$	ARL
0.05	1	C	0.218	337.1	0.091	362.4	-	-
	1	SN-C	0.159	316.5	0.082	334.3	-	-
	2	C	0.302	570.8	0.196	663.5	0.112	698.3
	2	SN-C	0.204	532.1	0.146	593.9	0.098	631.1
	10	C	0.437	1490.8	0.378	1887.3	0.324	2224.5
	10	SN-C	0.284	1514.8	0.252	1713.8	0.239	2102.0
0.1	1	C	0.315	317.6	0.168	348.2	-	-
	1	SN-C	0.254	295.4	0.152	314.5	-	-
	2	C	0.424	525.8	0.288	637.3	0.171	664.5
	2	SN-C	0.32	504.9	0.229	561.6	0.160	588.6
	10	C	0.547	1268.7	0.513	1698.1	0.446	2021.1
	10	SN-C	0.405	1302.3	0.387	1575.2	0.351	1839.2
			$\Delta = 1$					
0.05	1	C	0.677	298.9	0.239	399.2	-	-
	1	SN-C	0.447	294.8	0.178	377.1	-	-
	2	C	0.821	458.0	0.612	653.5	0.293	775.1
	2	SN-C	0.580	466.4	0.418	638.4	0.212	739.3
	10	C	0.939	833.8	0.920	1276.3	0.895	1710.4
	10	SN-C	0.734	1107.9	0.692	1483.3	0.646	1864.3
0.1	1	C	0.759	278.8	0.346	376.0	-	-
	1	SN-C	0.587	274.4	0.285	350.8	-	-
	2	C	0.895	396.8	0.715	603.9	0.432	736.3
	2	SN-C	0.724	431.5	0.55	597.8	0.339	684.9
	10	C	0.968	676.4	0.96	1048.0	0.947	1480.1
	10	SN-C	0.848	898.1	0.813	1292.6	0.784	1625.5
			$\Delta = 2$					
0.05	1	C	0.998	192.7	0.745	394.4	-	-
	1	SN-C	0.910	236.5	0.524	392.9	-	-
	2	C	1	218.3	0.995	484.3	0.858	756.5
	2	SN-C	0.964	315.6	0.889	551.1	0.622	757.6
	10	C	1	262.5	1	551.2	1	894.2
	10	SN-C	0.991	459.7	0.986	785.0	0.978	1184.0
0.1	1	C	1	168.0	0.841	376.4	-	-
	1	SN-C	0.964	201.4	0.678	371.9	-	-
	2	C	1	189.5	1	443.4	0.929	721.3
	2	SN-C	0.99	251.6	0.943	498.1	0.752	725.3
	10	C	1	229.2	1	496.3	1	824.0
	10	SN-C	0.998	343.4	0.996	639.4	0.996	1002.3

is the sum of score functions of pairwise likelihoods at time t up to lag l and $\widehat{M}_m(l)$ is the long-run variance estimator defined as

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

where $\hat{\gamma}_l(j) = \frac{1}{m}\sum_{t=j+1}^{m} L_t'(l;\hat{\boldsymbol{\theta}}_m)(L_{t-j}'(l;\hat{\boldsymbol{\theta}}_m))^T$. On the other hand, we can apply the self-normalization approach on the PLSMS procedure. Following (2.3), the stopping time of the new procedure, SN-PLSMS, is given by $P_m^{(SN)}(l) = \min\left\{\min\left\{k : \mathbb{M}_{m,l}(k) > c\right\}, mT+1\right\}$, where

$$\mathbb{M}_{m,l}(k) = \frac{\left(S_m(k,l;\hat{\boldsymbol{\theta}}_m)\right)' J_m(\hat{\boldsymbol{\theta}}_m)^{-1}\left(S_m(k,l;\hat{\boldsymbol{\theta}}_m)\right)}{m\left(1+\frac{k}{m}\right)^2}\,,$$

with $S_m(k,l;\hat{\boldsymbol{\theta}}_m) = \sum_{t=m+1}^{m+k} L_t'(l;\hat{\boldsymbol{\theta}}_m)$, and

$$J_m(\hat{\boldsymbol{\theta}}_m) = \frac{1}{m^2}\sum_{t=1}^{m}\left\{\left(\sum_{j=1}^{t} L_j'(l;\hat{\boldsymbol{\theta}}_m)\right)\left(\sum_{j=1}^{t} L_j'(l;\hat{\boldsymbol{\theta}}_m)\right)'\right\}.$$

In the following subsections, we will compare the sizes, size-corrected powers and size-corrected ARLs of PLSMS and SN-PLSMS on SV(1) models. Table 5 summarizes the decision boundary c for PLSMS by solving (4.1) with $d = 3$. The decision boundary c for SN-PLSMS with $d = 3$ is given in Table 1.

4.5 Change in parameters in time series models with latent process: Simulation results under H_0

To investigate the empirical sizes of PLSMS and SN-PLSMS under H_0, we conducted simulations based on the SV(1) models, Model 1: $\eta = 0.7$, $\beta = 1$, $\sigma_\epsilon = 1$, 2 and 3; Model 2: $\eta = 0.5$, $\beta = 2$, $\sigma_\epsilon = 1$, 2 and 3.

Table 5: *Decision boundaries c for $P_m(l)$ in PLSMS when $d = 3$.*

α	$T = 1$	$T = 2$	$T = 10$	$T = \infty$
0.05	1.861	2.149	2.510	2.633
0.1	1.684	1.944	2.270	2.381

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, 10, i.e., the monitoring horizons mT are m, $2m$ and $10m$ respectively. Figure $S.4$ in the online supplement 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. Table 6 reports the proportions of rejection of H_0 for the models when $m = 500$, 1000, and 1500. The row $T = 10^*$ are defined similarly as in Section 4.2. From Table 6, the size distortions of PLSMS is much more severe than that of SN-PLSMS when m is small. The proposed SN-PLSMS generally has an empirical size close to the significance level α.

4.6 Change in parameters in time series models with latent process: Simulation results under H_1

To investigate the size-corrected empirical powers and size-corrected ARLs under H_1, we conducted simulations based on three change-point models with change-point at $t^* = m + k^*$:

Table 6: *Empirical sizes for Models 1 and 2 of PLSMS (P) and SN-PLSMS (SN-P) in SV(1) models with m = 500, 1000 and 1500 when $l = 1$. The number of replications for each pair of (α, T) is 1000.*

		Model 1								
		$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
T	Method	m=500	1000	1500	m=500	1000	1500	m=500	1000	1500
		Significance level $\alpha = 0.05$								
1	P	0.16	0.134	0.132	0.148	0.124	0.121	0.149	0.122	0.121
1	SN-P	0.064	0.067	0.065	0.067	0.057	0.051	0.063	0.054	0.051
2	P	0.164	0.139	0.109	0.15	0.132	0.131	0.167	0.138	0.128
2	SN-P	0.061	0.069	0.054	0.057	0.055	0.056	0.054	0.055	0.053
10	P	0.185	0.156	0.132	0.17	0.133	0.139	0.188	0.147	0.138
10	SN-P	0.071	0.068	0.07	0.067	0.056	0.06	0.072	0.062	0.059
10*	P	0.16	0.127	0.095	0.158	0.121	0.109	0.159	0.115	0.097
10*	SN-P	0.062	0.058	0.047	0.064	0.045	0.046	0.057	0.045	0.049
		Significance level $\alpha = 0.1$								
1	P	0.228	0.181	0.18	0.222	0.196	0.166	0.243	0.205	0.197
1	SN-P	0.122	0.104	0.108	0.118	0.106	0.102	0.128	0.111	0.102
2	P	0.248	0.196	0.186	0.257	0.175	0.224	0.27	0.205	0.175
2	SN-P	0.124	0.115	0.111	0.132	0.108	0.114	0.126	0.106	0.111
10	P	0.275	0.189	0.172	0.228	0.195	0.168	0.281	0.201	0.195
10	SN-P	0.144	0.113	0.109	0.111	0.117	0.113	0.129	0.114	0.106
10*	P	0.223	0.187	0.161	0.192	0.187	0.167	0.225	0.171	0.168
10*	SN-P	0.128	0.092	0.087	0.099	0.103	0.084	0.106	0.081	0.083

		Model 2								
		$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
T	Method	m=500	1000	1500	m=500	1000	1500	m=500	1000	1500
		Significance level $\alpha = 0.05$								
1	P	0.13	0.107	0.097	0.101	0.093	0.082	0.112	0.091	0.079
1	SN-P	0.069	0.058	0.049	0.051	0.046	0.048	0.059	0.047	0.053
2	P	0.13	0.122	0.089	0.144	0.101	0.08	0.131	0.101	0.083
2	SN-P	0.067	0.066	0.05	0.062	0.063	0.044	0.069	0.058	0.057
10	P	0.136	0.113	0.103	0.149	0.112	0.088	0.127	0.106	0.105
10	SN-P	0.06	0.056	0.057	0.068	0.064	0.051	0.067	0.056	0.045
10*	P	0.127	0.098	0.077	0.126	0.094	0.062	0.106	0.089	0.079
10*	SN-P	0.051	0.049	0.043	0.057	0.041	0.038	0.052	0.038	0.032
		Significance level $\alpha = 0.1$								
1	P	0.213	0.171	0.158	0.202	0.158	0.158	0.211	0.155	0.141
1	SN-P	0.123	0.106	0.095	0.118	0.104	0.101	0.12	0.106	0.1
2	P	0.201	0.178	0.166	0.188	0.168	0.153	0.199	0.169	0.157
2	SN-P	0.119	0.116	0.112	0.114	0.099	0.108	0.117	0.102	0.099
10	P	0.238	0.184	0.163	0.194	0.164	0.144	0.206	0.174	0.156
10	SN-P	0.135	0.113	0.106	0.115	0.107	0.096	0.123	0.113	0.095
10*	P	0.193	0.165	0.129	0.178	0.137	0.113	0.176	0.141	0.118
10*	SN-P	0.113	0.102	0.084	0.105	0.088	0.087	0.101	0.09	0.087

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

Model 2: $(\eta_0, \sigma_{\epsilon,0}, \beta_0) = (0.7, 0.2, -0.1)$ changed to $(\eta_A, \sigma_{\epsilon,A}, \beta_A) = (0.5, 0.4, -0.3)$;

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

Model 2 has two parameters changed with small magnitude, while only η is changed in Model 1. In Model 3, the parameters are restricted to change in a way such that the variance of the observed process $\{X_t\}$ remains unchanged. Table 7 reports the size-corrected empirical powers and ARLs for various k^* with $m = 750$. The size-corrected empirical power is the proportion of simulation trials in which PLSMS or SN-PLSMS rejects H_0 using a decision boundary calibrated so that the empirical size of the scheme is α under H_0. The ARLs are computed by taking average of the stopping times conditioning on an alarm is given. From Table 7, while SN-PLSMS generally maintain a reasonable power, the power of PLSMS is usually higher than that of SN-PLSMS in Model 1 and 3. The ARL of PLSMS is also shorter than that of SN-PLSMS in general. For Model 2, the size-corrected power of SN-PLSMS is higher than that of PLSMS. A possible reason is that the size distortion of PLSMS in stationary process following Model 2 (with parameter values $(\eta_0, \sigma_{\epsilon,0}, \beta_0) = (0.7, 0.2, -0.1)$) is much more severe than that of other two. Unreported simulations show that the size distortion of PLSMS in stationary process following Model 2 is about 0.16 above the significance level, compared with that of Model 1 which is about 0.07 above the significance level. Indeed, the size-corrected decision boundary for Model 2 is much higher than that for Models 1 and 3. For example, under $(\alpha, T, k^*) = (0.05, 2, 50)$, the size-corrected decision boundary c for Model $1, 2$, and 3 are $2.4, 4.98$ and 2.46, respectively. Hence, the size-corrected power of PLSMS for Model 2 is heavily affected.

4.7 Empirical studies on railway bearing temperature data

In this section, an application to railway bearing temperature data is studied. Figure 1 depicts the temperature (TDFA) of two railway bearings, namely AxlePos 1 and AxlePos 7, from 12 August 2016 to 7 January 2017. Since most of the abnormality of the bearing condition, for example, grease hardening, will induce an increase in the bearing temperature, detecting temperature changes in the bearing helps to reflect potential abnormality in the bearings. On the other hand, punctuality of the service and the cost of maintenance are crucial for the operation of a railway company, and will be adversely affected by false alarms. Hence, a controlled Type I error of the monitoring scheme is important.

To monitor the mean level of the railway bearing temperature, we applied the CUSMS and SN-CUSMS procedures described in Section 4.1 and compared their performance. The data from 12 August 2016 to 30 August

Table 7: *Size-corrected empirical powers and average run lengths (ARL) for Models 1 to 3 of PLSMS (P) and SN-PLSMS (SN-P) with m =750 and k^* =50, 250 and 500. The number of replications for each pair of* (α, T) *is 1000.*

			Model 1					
α	T	Method	$k^* = 50$	ARL	$k^* = 250$	ARL	$k^* = 500$	ARL
0.05	1	P	0.843	374.8	0.611	527.5	0.193	618.6
	1	SN-P	0.685	402.8	0.459	527.9	0.17	622.9
	2	P	0.948	517.2	0.875	777.1	0.744	1003.5
	2	SN-P	0.801	606.0	0.665	784.6	0.559	1005.8
	10	P	0.998	711.2	0.991	1135.0	0.99	1605.8
	10	SN-P	0.911	1196.4	0.875	1529.4	0.871	2046.6
0.1	1	P	0.934	328.0	0.682	503.8	0.327	594.6
	1	SN-P	0.808	339.7	0.561	504.8	0.26	561.3
	2	P	0.985	452.2	0.928	704.3	0.819	938.1
	2	SN-P	0.87	550.4	0.769	729.5	0.652	931.3
	10	P	1	605.0	0.997	948.3	0.997	1433.1
	10	SN-P	0.952	922.4	0.954	1248.9	0.912	1816.2
			Model 2					
α	T	Method	$k^* = 50$	ARL	$k^* = 250$	ARL	$k^* = 500$	ARL
0.05	1	P	0.074	465.0	0.051	502.2	0.047	555.9
	1	SN-P	0.14	446.2	0.084	493.7	0.074	464.3
	2	P	0.066	901.1	0.065	984.8	0.046	1012.2
	2	SN-P	0.188	853.1	0.145	875.8	0.122	892.7
	10	P	0.103	1415.3	0.058	2227.6	0.062	2620.5
	10	SN-P	0.296	2286.0	0.223	2890.6	0.223	3024.4
0.1	1	P	0.157	440.4	0.138	396.0	0.114	420.7
	1	SN-P	0.254	447.9	0.194	458.2	0.133	427.4
	2	P	0.15	748.3	0.15	701.7	0.105	694.3
	2	SN-P	0.314	780.8	0.254	825.0	0.177	896.8
	10	P	0.232	2517.7	0.223	2615.3	0.201	2529.6
	10	SN-P	0.422	2088.0	0.41	2332.8	0.368	2710.3
			Model 3					
α	T	Method	$k^* = 50$	ARL	$k^* = 250$	ARL	$k^* = 500$	ARL
0.05	1	P	0.357	464.6	0.195	529.4	0.073	581.1
	1	SN-P	0.299	447.9	0.137	526.0	0.074	528.9
	2	P	0.485	756.8	0.374	909.6	0.236	1052.8
	2	SN-P	0.322	751.6	0.251	913.7	0.166	987.0
	10	P	0.629	1934.6	0.607	2421.4	0.52	2741.9
	10	SN-P	0.435	2027.2	0.408	2488.3	0.376	2747.7
0.1	1	P	0.461	440.8	0.305	518.4	0.17	524.9
	1	SN-P	0.373	421.9	0.265	488.7	0.171	521.0
	2	P	0.61	755.7	0.486	868.2	0.357	1010.0
	2	SN-P	0.476	725.1	0.39	857.1	0.272	935.5
	10	P	0.76	1719.8	0.721	2134.8	0.661	2553.9
	10	SN-P	0.607	1844.4	0.574	2109.9	0.511	2508.6

2016, which involve 350 observations ($m = 350$), are used as the training dataset. From the time series plot of the temperature of railway bearings in Figure 1, the 350 training data points appear to be stationary and free from structural break. We apply the retrospective change-point test proposed by Shao and Zhang (2010) for mean change on the training data set. The test results suggest that there is no change in mean level in the training data set under the significance level $\alpha = 5\%$. Using the training dataset, we estimate the mean and the self-normalization factor by

$$\hat{\mu}_m = \frac{1}{m}\sum_{t=1}^{m} X_t\,, \quad D_m(\hat{\mu}_m) = \frac{1}{m^2}\sum_{t=1}^{m}\left\{\left(\sum_{j=1}^{t}(X_t - \hat{\mu}_m)\right)\left(\sum_{j=1}^{t}(X_t - \hat{\mu}_m)\right)'\right\}.$$

The data from 30 August 2016 to 7 January 2017, which involve 2530 observations ($mT = 2530$), are used as the monitoring dataset. Therefore, we set $T = 2530/350 = 7.228571$. We performed the detection schemes under $\alpha = 0.05$ and 0.1 by monitoring the CUSUM statistics, $S(k) = \sum_{t=m+1}^{m+k}(X_t - \hat{\mu}_m)$ for time $m+k$, $k = 1, \ldots, 2530$. If the deviation from the mean is significant such that $m^{-1}\left(1 + \frac{k}{m}\right)^{-2} S(k)' D_m(\hat{\mu}_m)^{-1} S(k) > c$, then we declare that a mean change of the bearing temperature is occurred before time $m + k$.

Table 8 reports the corresponding time points where change-points are declared for the significance level $\alpha = 0.05$ and 0.1 under closed-end scheme with $T = 7.228571$ and open-end scheme with $T = \infty$ respectively. In order to compare the two monitoring procedures, we find the first change-point in the whole monitoring dataset by an offline change-point estimation. We apply the PELT algorithm for multiple change-points detection proposed by Killick, Fearnhead and Eckley (2012) on the data set. The estimation result shows that the first possible change-point is at time 908 which corresponding to 29 September 2016.

Figure 1 depicts the temperature of two railway bearings data from 12 August 2016 to 7 January 2017, and the results of CUSMS and SN-CUSMS under the open-end scheme. The observations on the left-hand side of the thin solid line are training data of size $m = 350$. The dotted and dashed lines represent the time points at which SN-CUSMS declare a change at $\alpha = 0.05$ and $\alpha = 0.1$, respectively. The dotdashed and longdashed lines represent the time points at which CUSMS declare a change at $\alpha = 0.05$ and $\alpha = 0.1$, respectively. The thick solid line represents the possible first change-point estimated by the PELT algorithm using the whole monitoring data set. Figure 1 suggests that SN-CUSMS successfully detects a change in the bearing temperature. On the other hand, the change-point detected by CUSMS appears to occur before the estimated change-point and thus could be a false alarm. This is in line with the findings in Section 4.2 that CUSMS tends to suffer from a large size distortions and rejects the null hypothesis more frequent than the nominal level.

Table 8: *Performances of the CUSMS and SN-CUSMS for the temperature of two railway bearing with $d = 2$ and $m = 350$ using closed-end monitoring scheme with $T = 7.228571$ and open-end monitoring scheme with $T = \infty$.*

	Closed-end monitoring scheme	Open-end monitoring scheme
	CUSMS	
α	Change-point declared (Decision boundary c)	Change-point declared
0.05	439 (2.337)	446
0.1	428 (2.091)	433
	SN-CUSMS	
α	Change-point declared (Decision boundary c)	Change-point declared
0.05	1029 (122.1)	1038
0.1	993 (89.5)	1008

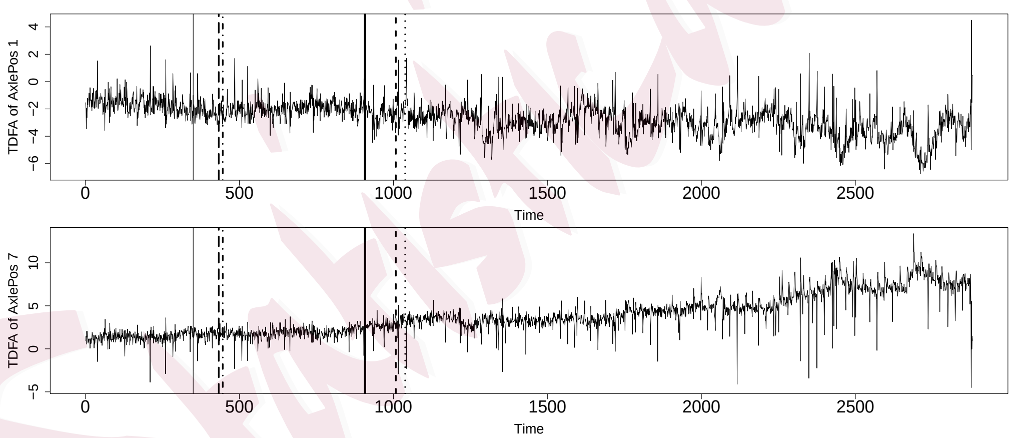

Figure 1: Plots of temperature of the bearing AxlePos 1 and AxlePos 7 from August 2016 to January 2017, the observations on the left-hand side of the thin solid line are the training data. The thick solid line is the estimated first change-point by PELT algorithm using the whole monitoring data set. The dotted and dashed lines represent the time points at which open-end SN-CUSMS declare a change at $\alpha = 0.05$ and $\alpha = 0.1$, respectively. The dotdashed and longdashed lines represent the time points at which open-end CUSMS declare a change at $\alpha = 0.05$ and $\alpha = 0.1$, respectively.

5. Discussions and Conclusions

This paper proposes a self-normalization (SN)-based sequential change-point detection method for detecting changes in parameter values in time series models. The monitoring scheme is shown to have asymptotically zero Type II error for any prescribed level of Type I error. By incorporating the self-normalization method, estimation of long-run variance and arbitrary choices of the bandwidth for kernel estimators, whose effect does not appear in the limit distribution, are bypassed. Simulation and empirical studies show that, the proposed method substantially improves the large size distortions occurred in traditional methods and at the same time, maintains a reasonable power.

The proposed sequential monitoring procedure is closely related to the theory of sequential tests with power one, which is a problem of determining a stopping rule τ such that $P(\tau < \infty | H_0) \leq \alpha$ and $P(\tau < \infty | H_1) = 1$ for a given significance level α. Under this framework, the false alarm rate of the monitoring scheme is controlled. Hence, the scheme is best in addressing applications in which 1) the costs associated with false alarm are higher than the costs associated with detection delay; or 2) the system requires a huge cost to be reset after a false alarm. As argued in Chu, Stinchcombe and White (1996) and Berkes *et al.* (2004), this framework is particularly useful in the sequential analysis of economic and financial data in which the sampling is costless under the null hypothesis of no change-point, and no action is required if the observed processes is “in control”, i.e., there is no change in the parameters of data generating mechanisms. See Berkes *et al.* (2004) for more financial applications in portfolio risk measure and option pricing. Also, in some engineering applications, punctuality of the service and the cost of maintenance are crucial for the operation of a company, and will be adversely affected by false alarms. Frequent false alarms will also result in a waste of resources in manual checking, repairing and replacement. Hence, a controlled Type I error of the monitoring scheme is important.

In practice, practitioners have to specify the training sample size m and monitoring horizon ratio T. In simple location model, m can be as small as 100 for a good empirical size, as demonstrated in the simulation results in Table 3. On the other hand, large m of size at least 500 is usually required for complicated model such as stochastic volatility models which involves latent processes, in order to have accurate parameter estimates and good empirical sizes; see Table 6. From the simulation studies in Section 4.3, for simple location model, $m \geq 500$ generally give a good power when the signal-to-noise ratio Δ/σ is around 1; see Table 4 under $\Delta = 2$. In practice, it is difficult to determine which m achieves a power close to 1 since it depends on the unknown break size Δ and the location of the change-point k^*. One way to address this issue is to choose a small T, i.e., $T \in [1, 10)$, and update the parameter estimates in (2.1) and (2.2) using the available data after monitoring mT observations when no change occurs. Consequently, a sufficiently large m can be accumulated in case of no changes, and possible change-point will be closer to the new starting point and will be easier to be detected. Another way to address this issue is to conduct extensive simulations using various assumed or estimated models, projected monitoring horizons mT, projected break sizes Δ and locations of the change-point k^*, and find an appropriate m satisfying the need of the practitioner.

For the choice of T, if the practitioners have a plan in advance that exactly n_0 incoming data are going to be observed and monitored, then T should be equal to $\frac{n_0}{m}$. If the practitioners do not have the exact number of n_0 but have a possible range $n_0 \in [n_a, n_b]$, then using a larger T, i.e., $T = \frac{n_b}{m}$, to determine the decision boundary c will give a more conservative monitoring scheme, which leads to a smaller empirical size, compared to that using $T = \frac{n_a}{m}$. It is because the decision boundary c is increasing with T under the same (α, d). Hence, using the decision boundary c for $T = \infty$ will always give a conservative results, and is only appropriate when the practitioners have little information about n_0 and speculate that a very long monitoring horizon is inspected. From the simulation results in Section 4, $T \in [1, 10]$ generally gives good empirical sizes, reasonable powers and ARLs.

Supplementary Materials

The online supplementary material contains figures for time series plots of realizations of the models in Section 4 and the proofs for the main results in the paper.

Acknowledgments

We would like to thank the Editor, an Associate Editor and two anonymous referees for their thoughtful and useful comments, which led to an improved version of this paper. Research supported in part by HKSAR-RGC-GRF Nos 14300514 and 14325216, HKSAR-RGC-TRF No. T32-101/15-R (Chan); HKSAR-RGC-GRF Nos 405113, 14305517 and 14601015 (Yau).

References

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH (p,q) models. *Econometric Theory* **20**, 1140–1167.

Choi, H., Ombao, H. and Ray, B. (2008). Sequential change-point detection methods for non-stationary time series. *Technometrics* **50**, 40–52.

Chu, C.-S. J., Stinchcombe, M. and White, H. (1996). Monitoring structural change. *Econometrica* **64**, 1045–1065.

Fuh, C.-D. (2006). Optimal change point detection in state space models. *Institute of Statistical Science, Academia Sinica, Tech. Rep, Taiwan.*

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**, 2872–2883.

Gombay, E. and Serban, D. (2009). Monitoring parameter change in AR(p) time series models. *Journal of Multivariate Analysis* **100**, 715–725.

Gut, A. and Steinebach, J. (2002). Truncated Sequential Change-point Detection based on Renewal Counting Processes. *Scandinavian Journal of Statistics* **29**, 693–719.

Han, D., and Tsung, F. and Xian, J. (2017). On the optimality of Bayesian change-point detection. *The Annals of Statistics* **45**, 1375–1402.

Hawkins, D. M., Qiu, P. and Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of Quality Technology* **35**, 355–366.

Hawkins, D. M. and Zamba, K. D. (2005). A change-point model for a shift in variance. *Journal of Quality Technology* **37**, 21–31.

Huang, Y., Volgushev, S. and Shao, X. (2015). On Self-Normalization for Censored Dependent Data. *Journal of Time Series Analysis* **36**, 109–124.

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *Journal of Statistical Research* **39**, 51–70.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* **107**, 1590–1598.

Kirch, C. and Tadjuidje Kamgaing, J. (2012). Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis* **33**, 365–385.

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.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 613–658.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory* **44**, 2917–2929.

Leung, S. H., Ng, W. L. and Yau, C. Y. (2017). Sequential change-point detection in time series models based on pairwise likelihood. *Statistica Sinica* **27**, 575–605.

Lin, Z. and Lu, C. (1996). *Limit Theory for Mixing Dependent Random Variables.* Science Press, Beijing, and Kluwer Academic Publishers, Boston.

Liu, H. and Shi, J. (2013). Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Economics* **37**, 152–166.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association* **96**, 1066–1076.

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

Pham, H. T. and Yang, B.-S. (2010). Estimation and forecasting of machine health condition using ARMA/GARCH model. *Mechanical Systems and Signal Processing* **24**, 546–558.

Polunchenko A. S. and Tartakovsky A. G. (2010). On optimality of the Shiryaev Roberts procedure for detecting a change in distribution. *The Annals of Statistics* **38**, 3445–3457.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 343–366.

Shao, X. (2015). Self-Normalization for Time Series: A Review of Recent Developments. *Journal of the American Statistical Association* **110**, 1797–1817.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105**, 1228–1240.

Tsay, R. S. (2010). *Analysis of Financial Time Series, 3rd eds.* John Wiley & Sons.

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

Wan, Y. (2013). On the ρ-mixing Hájek-Rényi inequality. *Journal of Jianghan University Natural Science Edition* **41**, 43–46.

Yakir, B. (1997). A note on optimal detection of a change in distribution. *The Annals of Statistics* **25**, 2117–2126.

Zeileis, A., Leisch, F., Kleiber, C. and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* **20**, 99–121.

Zhang, T. and Lavitas, L. (2018). Unsupervised Self-Normalized Change-Point Testing for Time Series. *Journal of the American Statistical Association* **113**, 637–648.

Zhang, X., Li, B. and Shao, X. (2014). Self-normalization for Spatial Data. *Scandinavian Journal of Statistics* **41**, 311–324.

Department of Statistics, The Chinese University of Hong Kong

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

Department of Mathematics and Statistics, Hang Seng University of Hong Kong

E-mail: wlng@hsu.edu.hk

Department of Statistics, The Chinese University of Hong Kong

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