

## A SELF-NORMALIZED APPROACH TO SEQUENTIAL CHANGE-POINT DETECTION FOR TIME SERIES

Ngai Hang Chan<sup>1,2</sup>, Wai Leong Ng<sup>3</sup> and Chun Yip Yau<sup>2</sup>

<sup>1</sup>*Southwestern University of Finance and Economics*, <sup>2</sup>*The Chinese University of Hong Kong* and <sup>3</sup>*Hang Seng University of Hong Kong*

*Abstract:* We propose a self-normalization sequential change-point detection method for time series. To test for parameter changes, most traditional sequential monitoring tests use a cumulative sum-based test statistic, which involves a long-run variance estimator. However, such estimators require choosing a bandwidth parameter, which may be sensitive to the performance of the test. Moreover, traditional tests usually suffer from severe size distortion as a result of the slow convergence rate to the limit distribution in the early monitoring stage. We propose self-normalization method to address these issues. We establish the null asymptotic and the consistency of the proposed sequential change-point test under general regularity conditions. Simulation experiments and an applications to railway-bearing temperature data illustrate and verify the proposed method.

*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  $\theta \in \Theta$ . We say that there is a change-point at time  $t^*$  if the model for  $\{X_1, X_2, \dots, X_{t^*-1}\}$  has a parameter  $\theta_0$ , but the model for  $\{X_{t^*}, X_{t^*+1}, \dots\}$  has a parameter  $\theta_1 \neq \theta_0$ . 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 online monitoring problem.

Sequential change-point detection received considerable attention with advances in serial data collection in engineering, econometrics, finance, and statistics. Being able to quickly detect of changes in the underlying process structure using a controlled false alarm rate is crucial because it allows practitioners to make immediate decisions and necessary adjustments in a timely manner. Se-

---

Corresponding author: Ngai Hang Chan, Department of Statistics, The Chinese University of Hong Kong, Central Ave, Hong Kong. E-mail: [nhchan@sta.cuhk.edu.hk](mailto:nhchan@sta.cuhk.edu.hk).

quential change-point detection schemes can be classified into two broad 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 a false alarm; see also Lai (1998); Yakir (1997); Polunchenko and Tartakovsky (2010); Han, Tsung and Xian (2017). However, as Lai (1995) points out, although the ARL constraint stipulates a long expected duration to a false alarm, it does not necessarily imply that the probability of having a false alarm before some specified time is small; hence false alarms at the initial stage cannot be controlled. In practice, frequent false alarms will result in a waste of resources in terms of repairing and replacement.

The second category uses a hypothesis testing framework so that the probability of false alarms can be controlled using a Type-I error. To control the Type-I error in a sequential setting, one straightforward approach is to employ retrospective change-point tests repeatedly using a multiple-testing procedure, such as the simulation-based adjustment, Bonferroni adjustment, or false discovery rate (FDR) method; see Hawkins, Qiu and Kang (2003); Hawkins and Zamba (2005); Choi, Ombao and Ray (2008). However, the simulation-based adjustment only works in simple models with an independence assumption, and the Bonferroni and FDR adjustments will lead to conservative results. A more sophisticated approach to controlling the Type-I error to establish the asymptotic distribution of the running maximum of a cumulative sum (CUSUM) process, which is standardized using a long-run variance estimator; see Chu, Stinchcombe and White (1996); Zeileis et al. (2005) in a linear regression setting; Gut and Steinebach (2002) for the 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 the covariance structure of weakly stationary time series; and Na, Lee and Lee (2011); Kirch and Tadjuidje Kamgaing (2015); Leung, Ng and Yau (2017) for general time series models. However, a difficulty with using 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 bandwidth parameter, which is difficult to choose 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 an asymptotically exact

Type-I error, without needing to estimate 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 proved successful in various statistical problems; see, for example, Lobato (2001); Shao (2010); Shao and Zhang (2010); Zhang, Li and Shao (2014); Huang, Volgushev and Shao (2015); Zhang and Lavitas (2018). In particular, the self-normalization method uses a self-normalizer to replace the long-run variance estimator when standardizing the test statistic. Thus, bypass the estimation of the long-run variance and, hence, no user-chosen bandwidth parameter is involved. As a result, the asymptotic null distribution of the SN-based test is pivotal, and the critical values can be tabulated using simulations. Moreover, the SN-based test exhibits “better size, but less power”; see Lobato (2001); Shao and Zhang (2010). This is a desirable feature in terms of resolving the aforementioned size distortion problem. For a detailed review on self-normalization for time series and its recent developments, see Shao (2015). In this study, we derive an SN-based statistic that we use to sequentially monitor the change-points of 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 the test statistic based on the long-run variance estimator; because it accounts for the extra randomness due to the self-normalizer. The test is shown to have an asymptotically zero Type-II error, with a prescribed level of Type-I error. Simulation studies demonstrate that the scheme exhibits a significant improvement in size distortion, while maintaining a good power.

The remainder of 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 in which we monitor of railway bearing temperatures are provided in Section 4. Section 5 concludes the paper. Technical proofs are provided in the online Supplementary Material.

## 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  $\{X_t\}_{t=1,2,\dots}$  is a stationary and ergodic random process, with joint density  $f_{\boldsymbol{\theta}}$ , where  $\boldsymbol{\theta}$  is the parameter vector in a compact space  $\Theta$ , and  $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$  can be regarded as a class of parametric models with parameter  $\boldsymbol{\theta}$ . Examples include the ARMA( $p,q$ )-GARCH( $r,s$ ) models and the stochastic volatility models SV( $p$ ) with autoregressive order  $p$ , see Section 4.

Before monitoring changes in the sequentially collected data sequence, we assume a training sample of  $m$  historical stationary data points  $\{x_1, x_2, \dots, x_m\}$  has been observed for the initial estimation of the pre-change parameter value, say  $\boldsymbol{\theta}_0$ , because we need to know the initial values of the parameters that are subject 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  $t = m + 1$ , we then observe fresh data  $\{X_t\}_{t=m+1,m+2,\dots}$  sequentially, and monitor whether a change has occurred in  $\boldsymbol{\theta}$  using the null hypothesis:

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

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,  $t^* = m + k^*$ , 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 while  $T$  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  $\{x_1, x_2, \dots, x_m\}$ , 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 the maximum likelihood estimators, M-estimators, least-squares estimators, and generalized moment estimators. The parameter estimates  $\hat{\boldsymbol{\theta}}_m$  then satisfy

the following system of equations:

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

where  $\{\mathbf{X}_t\}_{t=1,\dots,m}$  denotes a 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  $\{\mathbf{X}_t\}_{t=1,\dots,m}$  can be univariate (i.e.,  $\mathbf{X}_t = x_t$ ) or multivariate (i.e.,  $\mathbf{X}_t = (x_{t-r}, x_{t-r+1}, \dots, x_{t-1}, x_t)$ ), depending on the objective function  $L$  and the underlying time series models; for example,  $\mathbf{X}_t = x_t$  for independent and identically distributed (i.i.d.) data, and  $\mathbf{X}_t = (x_{t-p}, x_{t-p+1}, \dots, x_{t-1}, x_t)$  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  $m \rightarrow \infty$ .

**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. Because  $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}_0)) = 0$  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  $\{\mathbf{X}_t\}_{t=m+1,m+2,\dots}$  using the following 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 this is as follows. If the process remains unchanged, then  $\{L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\}_{t=m+1,m+2,\dots}$  should have an expectation close to zero. Hence, under a weak dependence condition on  $\{L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)\}$ , the CUSUM 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 nonzero expectation. Hence, the cumulative sum  $S_m(k, \hat{\boldsymbol{\theta}}_m)$  diverges as  $k \rightarrow \infty$ . This difference can be used to distinguish between the null and alternative hypotheses under an appropriate standardization of  $S_m(k, \hat{\boldsymbol{\theta}}_m)$ . Traditional sequential tests or monitoring schemes require a long-run variance estimator to standardize the CUSUM process. In practice, it is difficult to choose an appropriate choice of bandwidth parameter for the lag-window-type and kernel-type long-run variance estimators. Moreover, large size distortions are commonly found in existing empirical studies; see, for

example, Berkes et al. (2004); Zeileis et al. (2005); Na, Lee and Lee (2011). To address these problems, we employ a self-normalization method, where we replace the long-run variance estimator with the self-normalization factor, which does not require choosing 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

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

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 to avoiding an estimation of the long-run variance.

The case  $T \in (0, \infty)$  is called a closed-end monitoring scheme, where we stop monitoring after a fixed number of observations  $mT$ . 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 the parameter has occurred at some time on or before  $k$ . Otherwise, we proceed to time  $k + 1$  and continue monitoring. For closed-end monitoring, if the condition  $\mathbb{M}_m(k) > c$  has not been met at time  $mT$ , then we set  $T_m = mT + 1$ , and the monitoring scheme terminates by declaring that a change in the parameter did not occur, and  $H_0$  is not rejected. Note that the update at each time point does not involve numerical optimization or 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 occurred at  $k^* \leq mT$ , and the run length is the time until an alarm is given, that is,  $T_m$ .

### 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 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  $\|A\| = \sup_{\mathbf{x}: \|\mathbf{x}\|=1} \|A\mathbf{x}\|$ . 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 the regularity conditions under which we can derive asymptotic results for the initial estimation of the parameters, monitoring test statistic, and stopping time of the proposed procedure.

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

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

**Assumption 3.**  *$\mathbb{E}(\sup_{\boldsymbol{\theta} \in \Theta} \|L(\mathbf{X}_t, \boldsymbol{\theta})\|) < \infty$  and  $\boldsymbol{\theta}_0$  is the unique zero of  $\mathbb{E}(L(\mathbf{X}_t, \boldsymbol{\theta}))$ ; that is, 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 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*

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

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

**Assumption 5.**  *$L(\mathbf{X}_t, \boldsymbol{\theta})$  is continuously differentiable with respect to  $\boldsymbol{\theta}$  in a neighborhood  $V_{\boldsymbol{\theta}_0}$  of  $\boldsymbol{\theta}_0$ . In addition,  $\mathbb{E}(L'(\mathbf{X}_t, \boldsymbol{\theta}_0))$  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  $m \rightarrow \infty$  if a change occurs.

**Assumption 6.** *The change-point occurs after time  $m$ ; that is,  $t^* = \lfloor m\phi \rfloor$ , 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 7.

(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 6 is a condition on the location of the change-point to ensure that enough post-change observations will be accumulated for monitoring. Assumption 7 on the moment conditions for the post-change process  $\{\mathbf{X}_t^*\}$  is crucial for the consistency of the sequential monitoring scheme, because 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 1 to 5, we have  $\hat{\boldsymbol{\theta}}_m = \boldsymbol{\theta}_0 + O_p(m^{-1/2})$ .*

(b) Under Assumptions 2 and 4, we have the weak invariance principle of the partial sum process

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

as  $m \rightarrow \infty$ , where  $r \in [0, 1 + T]$ ,  $\lfloor x \rfloor$  is the largest integer smaller than  $x$ , and  $\xrightarrow{\mathcal{D}[0,1+T]}$  denotes weak convergence in  $\mathcal{D}[0, 1 + T]$ , the space of right-continuous functions with a 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  $\alpha$ , such that we can control the Type-I error of the procedures.

**Theorem 1.** Under Assumptions 1 to 5, we have

(a)  $D_m(\hat{\boldsymbol{\theta}}_m) \xrightarrow{\mathcal{D}} \mathbf{M}(\boldsymbol{\theta}_0)^{1/2}\mathbf{V}(\mathbf{M}(\boldsymbol{\theta}_0)^{1/2})'$  as  $m \rightarrow \infty$ , 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

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

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 a geometrically  $\rho$ -mixing sequence, that is, the  $\rho$ -mixing coefficient satisfies

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

where  $0 < a < 1$ ,  $\mathcal{A}_0$  and  $\mathcal{B}_k$  are the  $\sigma$ -fields generated by  $\{\mathbf{X}_t; t \leq 0\}$

and  $\{\mathbf{X}_t; t \geq k\}$ , respectively, and  $\mathcal{L}^2(\mathcal{F})$  is the space of square-integrable  $\mathcal{F}$ -measurable random variables, then

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

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  $\{\mathbb{B}_d(1+s) - (1+s)\mathbb{B}_d(1)\}_{s \in [0,\infty)}$ . In addition, the last equality of (3.2) is the result of the rescaling property of the Brownian motion that

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

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

The probabilities in (3.1) and (3.2) can be evaluated using a Monte Carlo simulation and, hence, the decision boundary  $c = c(\alpha, d, T)$  can be determined such that the asymptotic size is equal to a prespecified significance level  $\alpha$ ; that is,

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

or

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

Table 1 summarizes the decision boundaries  $c$  of the SNSMS for different values of  $(\alpha, d, T)$ , which we use in the simulation studies in Section 4. Given  $(\alpha, d, T)$ , to find the decision boundaries  $c$  of the SNSMS for closed- and open-ended schemes, we use simulations to solve for  $c$  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

Table 1. Decision boundaries  $c$  under different  $\alpha$ ,  $d$ , and  $T$  for  $T_m$  in SNSMS.

$T$	$d=1$				$d=2$				$d=3$				
	1	2	10	$\infty$	1	2	10	$\infty$	1	2	10	$\infty$	
$\alpha$	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

variables in a dense grid of length  $10^{-4}$ . The decision boundary  $c$  is taken as the empirical percentile of the test statistic in 5,000,000 repetitions.

### 3.4. Asymptotic properties of the SNSMS under $H_1$

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

**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 6 and 7, the asymptotic power of the SNSMS is equal to 1; that is,*

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

## 4. Simulation and Empirical Studies

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

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

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

The ARMA( $p, q$ )-GARCH( $r, s$ ) model is a general time series model, and includes the well-known GARCH model as a special case. The ARMA( $p, q$ )-GARCH( $r, s$ ) model has become increasingly popular in time series modeling and forecasting in engineering, econometrics, and finance. For example, Pham and Yang (2010) used ARMA-GARCH models to explain the wear and fault conditions of machines, and Liu and Shi (2013) applied ARMA-GARCH models to forecast 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, a CUSUM monitoring scheme is used; see Chu, Stinchcombe and White (1996); Na, Lee and Lee (2011) for details. For simplicity, we call this the CUSUM sequential monitoring scheme (CUSMS).

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

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

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

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

where  $\hat{\gamma}_l(j) = m^{-1} \sum_{t=j+1}^m (X_t - \hat{\mu}_m)(X_{t-j} - \hat{\mu}_m)$ . We reject the null hypothesis of no change-point when  $C_m < mT + 1$ . We also apply the self-normalization approach with  $L(X_t, \hat{\theta}_m) = X_t - \hat{\mu}_m$  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  $D_m(\hat{\mu}_m) = m^{-2} \sum_{t=1}^m \{ (\sum_{j=1}^t (X_t - \hat{\mu}_m))^2 \}$  is the self-normalizer. In the following subsections, we compare the CUSMS and SN-CUSMS on ARMA(1, 1)-

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

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

GARCH(1,1) models in terms of their size, size-corrected power, and size-corrected ARL. Table 2 summarizes the decision boundary  $c$  for the CUSMS. Time series plots of some realizations of the models are provided in the online Supplementary Material. The decision boundary  $c$  of CUSMS for each  $(\alpha, d, T)$  is obtained by numerically solving for  $c$  in the equation

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

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 the SN-CUSMS with  $d = 1$  is given in Table 1.

**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 the CUSMS and SN-CUSMS under  $H_0$ , we performed simulations based on the ARMA(1,1)-GARCH(1,1) models with parameter  $\theta = (\mu_t, \omega, \phi_1, \theta_1, \alpha_1, \beta_1)$ , 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:  $(\omega, \phi_1, \theta_1, \alpha_1, \beta_1) = (0.8, 0.5, 0.5, 0.15, 0.2)$ ; and Model 2:  $(\omega, \phi_1, \theta_1, \alpha_1, \beta_1) = (0.6, 0.7, 0.8, 0.2, 0.1)$ . We considered combinations of  $\alpha = 0.05, 0.1$  and  $T = 1, 2, 10$ ; that is, the monitoring horizons  $mT$  are  $m, 2m$ , and  $10m$  respectively. Table 3 reports the proportion of rejection of  $H_0$  when  $m = 100, 300, 500, 1,000$ , and  $2,000$ . The row  $T = 10^*$  corresponds to the open-end scheme  $T = \infty$  with a monitoring horizon of  $10m$ , because an infinite monitoring horizon is impossible in practice. The decision boundaries for  $T = \infty$  are always larger than those for finite  $T$  under the same  $(\alpha, d)$ , as shown in Table 1. Thus, the empirical sizes for the row  $T = 10$  are expected to be larger than those for the row  $T = 10^*$ . From Table 3, the size distortion of the CUSMS is much more severe than that of the SN-CUSMS. In general, the proposed SN-CUSMS has an empirical size close to the significance level  $\alpha$ . 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

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  $m = 100, 300, 500, 1,000,$  and  $2,000$ . The number of replications for each pair  $(\alpha, T)$  is 2,500.

$T$	Method	Model 1					Model 2				
		$m=100$	300	500	1,000	2,000	$m=100$	300	500	1,000	2,000
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

long-run variance estimator in Model 2 is more difficult to estimate accurately. This is a possible reason for the greater size distortion of the CUSMS in Model 2 than that in 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 power and ARLs of the CUSMS and SN-CUSMS under  $H_1$ , we performed 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 a 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 the mean level changed from  $\mu_t = 0$ , for  $t < m + k^*$ , to  $\mu_t = \Delta$ , for  $t \geq m + k^*$ .

The parameter values for Model  $\mathcal{A}$  are the same as those for Model 2 in Section 4.2, except for the mean level  $\mu_t$ . Table 4 reports the size-corrected empirical power and ARLs for various  $k^*$  and  $\Delta$ , with  $m = 500$ . The size-corrected empirical power is the proportion of simulation trials in which the CUSMS and SN-CUSMS reject  $H_0$ , using a decision boundary calibrated such that the empirical size of the scheme is  $\alpha$  under  $H_0$ . The ARLs are computed as the average stopping times, conditioning on an alarm being given. From Table 4, whereas the SN-CUSMS generally maintains a reasonable power and ARL, the

CUSMS usually exhibits a better power and a shorter ARL. This “better size, but less power” phenomenon is consistent with the findings reported in the literature in other contexts; see Lobato (2001); Shao and Zhang (2010). The results from Table 4 also show that the detection rate and timing depend on the magnitude of the parameter change  $\Delta$  and the location of the change-point  $t^* = m + k^*$ . Here, a smaller  $k^*$  (closer to time  $m$ ) and a larger magnitude of the change  $\Delta$  indicate a higher detection rate and a shorter detection delay. This phenomenon occurs mainly because the CUSUM statistic  $S_m(k, \hat{\boldsymbol{\theta}}_m) = \sum_{t=m+1}^{m+k} L(\mathbf{X}_t, \hat{\boldsymbol{\theta}}_m)$  contains more pre-change data when  $k^*$  is large, and hence more post-change data are needed before significance is reached; see, for example, Chu, Stinchcombe and White (1996); Na, Lee and Lee (2011) for similar observations.

Furthermore, in general, under the same  $(T, \Delta, k^*)$ , a larger  $m$  will result in better power because we can estimate the unknown parameters more accurately using a larger training sample. Hence, with a more accurate estimated pre-change model, the monitoring scheme will be more sensitive to deviations of incoming data that follow 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 a latent process**

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

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

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

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

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

Table 4. Size-corrected empirical power and average run length (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  $(\alpha, T)$  is 2,500.

$\alpha$	$T$	Method	$\Delta = 0.5$					
			$k^* = 50$		$k^* = 250$		$k^* = 500$	
			$k^*$	ARL	$k^*$	ARL	$k^*$	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	1,490.8	0.378	1,887.3	0.324	2,224.5
	10	SN-C	0.284	1,514.8	0.252	1,713.8	0.239	2,102.0
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	1,268.7	0.513	1,698.1	0.446	2,021.1
	10	SN-C	0.405	1,302.3	0.387	1,575.2	0.351	1,839.2
$\Delta = 1$								
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	1,276.3	0.895	1,710.4
	10	SN-C	0.734	1,107.9	0.692	1,483.3	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	1,048.0	0.947	1,480.1
	10	SN-C	0.848	898.1	0.813	1,292.6	0.784	1,625.5
$\Delta = 2$								
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	1,184.0
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	1,002.3

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

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

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

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

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

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

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

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

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

**4.5. Change in parameters in time series models with a latent process:  
Simulation results under  $H_0$**

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

The models with different values of  $\eta$  represent different degrees of correlation in the latent autoregressive process. Within each model, the values of  $\sigma_\epsilon$  represent

volatilities of the latent autoregressive process. We also considered combinations of  $\alpha = 0.05, 0.1$  and  $T = 1, 2, 10$ , that is, the monitoring horizons  $mT$  are  $m, 2m$ , and  $10m$ , respectively. Figure *S.4* in the online Supplementary Material provides time series plots of some realizations of Model 2 with different  $\sigma_\epsilon$  values. Note that spikes occur more frequently under a larger variance in the latent process.

Table 6 reports the proportions of rejection of  $H_0$  for the models when  $m = 500, 1,000$ , and  $1,500$ . The row  $T = 10^*$  is defined similarly to that in Section 4.2. From Table 6, the size distortions of the PLSMS are much more severe than those of the SN-PLSMS when  $m$  is small. In general, the proposed SN-PLSMS has an empirical size close to the significance level  $\alpha$ .

#### 4.6. Change in parameters in time series models with a latent process: Simulation results under $H_1$

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

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 that with a small magnitude, whereas only  $\eta$  changed in Model 1. In Model 3, the parameters are restricted to change in such a way that the variance of the observed process  $\{X_t\}$  remains unchanged. Table 7 reports the size-corrected empirical power and ARLs for various  $k^*$ , with  $m = 750$ . The size-corrected empirical power is the proportion of the simulation trials in which the PLSMS or SN-PLSMS rejects  $H_0$ , using a decision boundary calibrated such that the empirical size of the scheme is  $\alpha$  under  $H_0$ . The ARLs are computed as the average stopping times, conditioning on an alarm being given. From Table 7, although the SN-PLSMS maintains reasonable power, in general, the power of the PLSMS is usually higher than that of the SN-PLSMS in Models 1 and 3. The ARL of the PLSMS is also shorter than that of the SN-PLSMS, in general. For Model 2, the size-corrected power of the SN-PLSMS is higher than that of the PLSMS. A possible reason for this is that the size distortion of the PLSMS in the stationary process that follows 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 the PLSMS in the stationary process that follows Model 2 is about 0.16 above the significance

Table 6. Empirical sizes for Models 1 and 2 of PLSMS (P) and SN-PLSMS (SN-P) in SV(1) models, with  $m = 500, 1,000,$  and  $1,500,$  when  $l = 1.$  The number of replications for each pair of  $(\alpha, T)$  is 1,000.

		Model 1								
$T$	Method	$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
		$m=500$	1,000	1,500	$m=500$	1,000	1,500	$m=500$	1,000	1,500
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								
$T$	Method	$\sigma_\epsilon = 1$			$\sigma_\epsilon = 2$			$\sigma_\epsilon = 3$		
		$m=500$	1,000	1,500	$m=500$	1,000	1,500	$m=500$	1,000	1,500
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

level; that of Model 1 is about 0.07 above the significance level. Indeed, the size-corrected decision boundary for Model 2 is much higher than those for Models 1 and 3. For example, under  $(\alpha, T, k^*) = (0.05, 2, 50),$  the size-corrected decision boundaries  $c$  for Models 1, 2, and 3 are 2.4, 4.98, and 2.46, respectively. Hence,

the size-corrected power of the PLSMS for Model 2 is affected significantly.

#### 4.7. Empirical studies on railway bearing temperature data

In this section, we apply the proposed method to railway bearing temperature data. Figure 1 depicts the temperatures (TDFA) of two railway bearings, AxlePos 1 and AxlePos 7, for the period August 12, 2016 to January 7, 2017. Because most abnormalities of the bearing condition (e.g., grease hardening) will induce an increase in the bearing temperature, detecting temperature changes helps to identify potential abnormalities in the bearings. On the other hand, the punctuality of its service and the cost of maintenance are crucial to operating a railway company, and both will be adversely affected by false alarms. Hence, a controlled Type-I error of the monitoring scheme is important.

To monitor the mean bearing temperature, we applied the CUSMS and SN-CUSMS procedures described in Section 4.1 and then compared their performance. The data from August 12, 2016, to August 30, 2016, which include 350 observations ( $m = 350$ ), are used as the training data set. From the time series plot of the bearing temperatures in Figure 1, the 350 training data points appear to be stationary and free of structural breaks. We apply the retrospective change-point test proposed by Shao and Zhang (2010) for the mean change on the training data set. The test results suggest there is no change in the mean level in the training data set under the significance level  $\alpha = 5\%$ . Using the training data set, we estimate the mean and the self-normalization factor by

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

respectively. The data from August 30, 2016 to January 7, 2017, which involve 2,530 observations ( $mT = 2,530$ ), are used as the monitoring data set. Therefore, we set  $T = 2,530/350 = 7.228571$ . We performed the detection schemes under  $\alpha = 0.05$  and  $0.1$  by monitoring the CUSUM statistic,  $S(k) = \sum_{t=m+1}^{m+k} (X_t - \hat{\mu}_m)$ , for time  $m+k$ , for  $k = 1, \dots, 2,530$ . If the deviation from the mean is significant, such that  $m^{-1} (1 + k/m)^{-2} S(k)' D_m(\hat{\mu}_m)^{-1} S(k) > c$ , then we declare that a change in the mean of the bearing temperature has occurred before time  $m+k$ .

Table 8 reports the corresponding time points when change-points are declared for significance levels  $\alpha = 0.05$  and  $0.1$  under a closed-end scheme with  $T = 7.228571$  and an open-end scheme with  $T = \infty$ , respectively. In order to compare the two monitoring procedures, we find the first change-point in the

Table 7. Size-corrected empirical power 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  $(\alpha, T)$  is 1,000.

Model 1								
$\alpha$	$T$	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	1,003.5
	2	SN-P	0.801	606.0	0.665	784.6	0.559	1,005.8
	10	P	0.998	711.2	0.991	1,135.0	0.99	1,605.8
	10	SN-P	0.911	1,196.4	0.875	1,529.4	0.871	2,046.6
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	1,433.1
	10	SN-P	0.952	922.4	0.954	1,248.9	0.912	1,816.2
Model 2								
$\alpha$	$T$	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	1,012.2
	2	SN-P	0.188	853.1	0.145	875.8	0.122	892.7
	10	P	0.103	1,415.3	0.058	2,227.6	0.062	2,620.5
	10	SN-P	0.296	2,286.0	0.223	2,890.6	0.223	3,024.4
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	2,517.7	0.223	2,615.3	0.201	2,529.6
	10	SN-P	0.422	2,088.0	0.41	2,332.8	0.368	2,710.3
Model 3								
$\alpha$	$T$	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	1,052.8
	2	SN-P	0.322	751.6	0.251	913.7	0.166	987.0
	10	P	0.629	1,934.6	0.607	2,421.4	0.52	2,741.9
	10	SN-P	0.435	2,027.2	0.408	2,488.3	0.376	2,747.7
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	1,010.0
	2	SN-P	0.476	725.1	0.39	857.1	0.272	935.5
	10	P	0.76	1,719.8	0.721	2,134.8	0.661	2,553.9
	10	SN-P	0.607	1,844.4	0.574	2,109.9	0.511	2,508.6

Table 8. Performance of the CUSMS and SN-CUSMS for the temperatures of two railway bearings with  $d = 2$  and  $m = 350$ , using a closed-end monitoring scheme with  $T = 7.228571$ , and an open-end monitoring scheme with  $T = \infty$ .

Closed-end monitoring scheme		Open-end monitoring scheme
CUSMS		
$\alpha$	Change-point declared (Decision boundary $c$ )	Change-point declared
0.05	439 (2.337)	446
0.1	428 (2.091)	433
SN-CUSMS		
$\alpha$	Change-point declared (Decision boundary $c$ )	Change-point declared
0.05	1,029 (122.1)	1,038
0.1	993 ( 89.5)	1,008

overall monitoring data set using an offline change-point estimation. We apply the PELT algorithm for detecting multiple change-points proposed by Killick, Fearnhead and Eckley (2012) to the data set. The estimation results show that the first possible change-point is at time 908, which corresponds to September 29, 2016.

Figure 1 depicts the temperatures of two railway bearings from August 12, 2016 to January 7, 2017, and the CUSMS and SN-CUSMS results 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 the SN-CUSMS declares changes at  $\alpha = 0.05$  and  $\alpha = 0.1$ , respectively. The dotdashed and longdashed lines represent the time points at which the CUSMS declare a change at  $\alpha = 0.05$  and  $\alpha = 0.1$ , respectively. The thick solid line represents the first possible change-point estimated by the PELT algorithm using the whole monitoring data set. Figure 1 suggests that the SN-CUSMS successfully detects a change in the bearing temperature. On the other hand, the change-point detected by the 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 the CUSMS tends to suffer from large size distortions, and rejects the null hypothesis more frequently than the nominal level.

## 5. Conclusion

We have proposed 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 an asymptotically zero Type-II error for any prescribed level of Type-I error. By incorporating the self-normalization

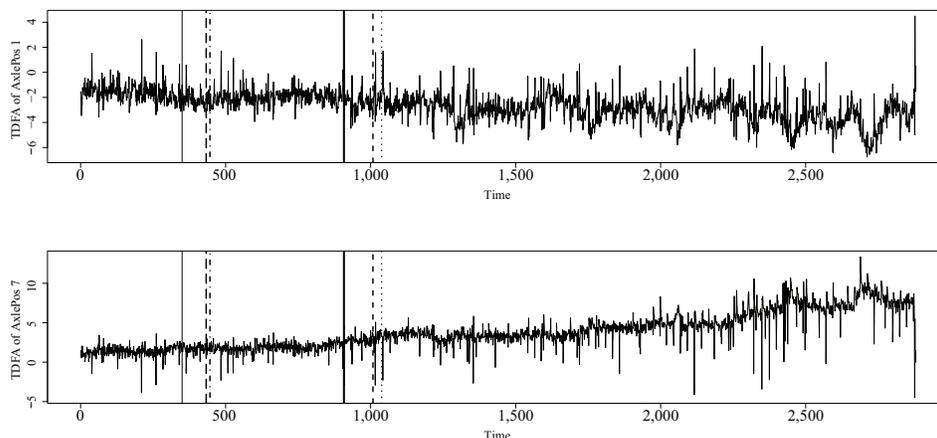


Figure 1. Plots of temperatures of the bearings AxlePos 1 and AxlePos 7 from August 2016 to January 2017. The observations on the left-hand side of the thin solid line are the training data. The thick solid line is the estimated first change-point by the PELT algorithm using the whole monitoring data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

method, we bypass the estimation of the long-run variance and the arbitrary choices of the bandwidth for the kernel estimators, the effect of which do not appear in the limit distribution. Simulation and empirical studies show that the proposed method substantially improves the large size distortions that occur in traditional methods, while maintaining 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  $\tau$ , such that  $P(\tau < \infty | H_0) \leq \alpha$  and  $P(\tau < \infty | H_1) = 1$ , for a given significance level  $\alpha$ . Under this framework, the false alarm rate of the monitoring scheme is controlled. Hence, the scheme performs best in applications in which 1) the costs associated with false alarms are higher than those associated with detection delays; or 2) the system requires a significant cost to reset after a false alarm. As argued in Chu, Stinchcombe and White (1996); 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”; that is, there is no change in the parameters of the data-generating mechanisms.

See Berkes et al. (2004) for additional financial applications in measuring portfolio risk and pricing options. In addition, in some engineering applications, the punctuality of the service and the cost of maintenance are crucial to the operation of a company, and both are adversely affected by false alarms. Frequent false alarms also result in wasted resources, owing to 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 a 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, a large  $m$  of at least 500 is usually required for complicated models, such as stochastic volatility models, which involve latent processes, in order to achieve accurate parameter estimates and good empirical sizes; see Table 6. From the simulation studies in Section 4.3, for a simple location model,  $m \geq 500$  generally yields good power when the signal-to-noise ratio  $\Delta/\sigma$  is around one; see Table 4 under  $\Delta = 2$ . In practice, it is difficult to determine which  $m$  achieves a power close to one because this depends on the unknown break size  $\Delta$  and the location of the change-point  $k^*$ . One way to address this issue is to choose a small  $T$ , such as  $T \in [1, 10)$ , and then 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 the case of no changes, and the possible change-point will be closer to the new starting point, and thus easier to detect. Another way to address this issue is to conduct extensive simulations using various assumed or estimated models, projected monitoring horizons  $mT$ , projected break sizes  $\Delta$ , and locations of the change-point  $k^*$ , and then identify an appropriate  $m$  satisfying the needs 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  $n_0/m$ . 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  $T$ , such as  $T = n_b/m$ , to determine the decision boundary  $c$  will give a more conservative monitoring scheme, which leads to a smaller empirical size than when using  $T = n_a/m$ . This is because the decision boundary  $c$  is increasing with  $T$  under the same  $(\alpha, d)$ . Hence, using the decision boundary  $c$  for  $T = \infty$  will always give conservative results, and is only appropriate when practitioners have little information about  $n_0$  and suspect there is a very long monitoring horizon. From the simulation

results in Section 4, in general,  $T \in [1, 10]$  gives good empirical sizes, reasonable power, and reasonable ARLs.

## Supplementary Material

The online Supplementary Material contains figures for the time series plots of the 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. This research was supported in part by HKSAR-RGC-GRF Nos 14308218 and 14325216, and HKSAR-RGC-TRF No. T32-101/15-R (Chan); and 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., 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 (Statistical Methodology)* **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, 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 Edition. John Wiley & Sons.
- Tsay, R. S. (2012). *An Introduction to Analysis of Financial Data with R*. John Wiley & Sons.
- Wan, Y. (2013). On the  $\rho$ -mixing Hájek-Rényi inequality. *Journal of Jiangnan 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.

Ngai Hang Chan

Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.

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

Wai Leong Ng

Department of Mathematics and Statistics, Hang Seng University of Hong Kong, Hang Shin Link, Siu Lek Yuen, Hong Kong.

E-mail: wlng@hsu.edu.hk

Chun Yip Yau

Department of Statistics, The Chinese University of Hong Kong, Central Ave, Hong Kong.

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

(Received July 2018; accepted April 2019)