Testing for Change Points in Time Series

Xiaofeng SHAO and Xianyang ZHANG

This article considers the CUSUM-based (cumulative sum) test for a change point in a time series. In the case of testing for a mean shift, the traditional Kolmogorov–Smirnov test statistic involves a consistent long-run variance estimator, which is needed to make the limiting null distribution free of nuisance parameters. The commonly used lag-window type long-run variance estimator requires to choose a bandwidth parameter and its selection is a difficult task in practice. The bandwidth that is a fixed function of the sample size (e.g., $n^{1/3}$, where n is sample size) is not adaptive to the magnitude of the dependence in the series, whereas the data-dependent bandwidth could lead to nonmonotonic power as shown in previous studies. In this article, we propose a self-normalization (SN) based Kolmogorov–Smirnov test, where the formation of the self-normalizer takes the change point alternative into account. The resulting test statistic is asymptotically distribution free and its power is monotonic. Furthermore, we extend the SN-based test to test for a change in other parameters associated with a time series, such as marginal median, autocorrelation at lag one, and spectrum at certain frequency bands. The use of the SN idea thus allows a unified treatment and offers a new perspective to the large literature of change point detection in the time series setting. Monte Carlo simulations are conducted to compare the finite sample performance of the new SN-based test with the traditional Kolmogorov–Smirnov test. Illustrations using real data examples are presented.

KEY WORDS: CUSUM; Invariance principle; Self-normalization.

1. INTRODUCTION

In the modeling of time series, structural stability is of prime importance. To assess structural stability, it is of practical interest to test for change points in a time series in view of the often empirical evidence for structural change. There is a huge literature on testing for change points (structural changes) for a sequence of independent and identically distributed (iid) random variables; see Csörgő and Horváth (1997) and Brodsky and Darkhovskay (1993) for accounts of various methods. From a methodological standpoint, the test statistics developed for change point detection in the iid context may not work in the time series setup and suitable modification is needed to account for the temporal dependence in the data; see, for example, Tang and MacNeill (1993), Antoch, Hušková, and Prášková (1997). In the case of testing for a mean shift, a common finding is that one needs to obtain a consistent estimate of the so-called long-run variance, or equivalently the spectral density function at zero frequency. Typically, a bandwidth parameter is involved in consistent estimation and its selection greatly affects the finite sample performance. In the literature, it has been found that the data-dependent bandwidth could lead to the nonmonotonic power problem. In other words, the power can decrease when the alternative gets farther away from the null. Both theoretical and empirical studies show that the nonmonotonic power is due to the data-dependent bandwidth, which is designed under the null and may be severely biased under the alternative. Nonmonotonic power is an undesirable feature of a test statistic, so methods have been proposed to overcome the problem; see Altissimo and Corradi (2003) and Juhl and Xiao (2009). However, the methods proposed in the above-mentioned papers involve a choice of another bandwidth parameter. As mentioned by Perron (2006), "there is no reliable method to appropriately choose this parameter in the context of structural changes." In this sense, the nonmonotonic power problem remains to be solved.

In this article, we propose a new test statistic to test for a change point in the mean. The basic ingredient of our proposal is to extend the self-normalization (SN) idea (see Lobato 2001; Shao 2010) into the change-point detection problem. The extension is very nontrivial as a naive extension is shown to fail in Section 2.2. The SN-based test does not involve any user-chosen number or smoothing parameter. Its asymptotic null distribution is pivotal and the (approximate) critical values are tabulated through simulations. The test is simple to implement yet powerful in the sense that it is consistent and achieves $\sqrt{n}$ local power. A desirable feature of the SN-based test is that its empirical power is seen to be monotonic, although there is a moderate power loss compared to the traditional Kolmogorov–Smirnov test. As a compensation, the SN-based test has better size. The "better size but less power" phenomenon for the SN-based test is consistent with the findings in other testing contexts; see Lobato (2001).

There has been a large amount of work on change-point detection in the time series setting. Here we do not try to provide a complete list of references, but mention some representative works. In the literature, the tests developed for univariate/multivariate time series include Horváth, Kokoszka, and Steinebach (1999), Vogelsang (1998, 1999) for a change in the mean; Inclan and Tiao (1994), Lee and Park (2001), Gombay, Horváth, and Hušková (1996) for a change in the marginal variance; Giraitis, Leipus, and Surgailis (1996) and Inoue (2001) for a change in the marginal distribution function; Picard (1985), Giraitis and Leipus (1992), and Lavielle and Ludena (2000) for a change in the spectrum; Berkes, Gombay, and Horváth (2009), Galeano and Peña (2007) for a change in the autocovariance function at certain lags. For change-point detection in time series models or regression models with dependent errors, see Andrews (1993), Davis, Huang, and Yao (1995), Lee, Ha, and Na (2003), Ling (2007), Qu and Perron (2007), Aue et al. (2008), Gombay (2008), among others. We

Xiaofeng Shao is Assistant Professor (E-mail: *xshao@uiuc.edu*) and Xianyang Zhang is Ph.D. Candidate, Department of Statistics, University of Illinois at Urbana–Champaign, Champaign, IL 61820. We are grateful to the two referees and Zhibiao Zhao for detailed comments, which led to substantial improvements. We thank Wei Biao Wu for providing Argentina rainfall data used in this paper. The research is supported in part by NSF grants DMS-0804937 and DMS-0724752.

Journal of the American Statistical Association
September 2010, Vol. 105, No. 491, Theory and Methods
DOI: 10.1198/jasa.2010.tm10103

refer the interested readers to the excellent review articles by Kokoszka and Leipus (2002) and Perron (2006) for more references. It seems that the techniques developed for change-point detection are specific to the quantity/parameter of interest. For example, the test for a change in variance is quite different from the test for a change in spectrum. This brings considerable difficulty to the practitioners, who want to use these tests routinely to check the stability of certain characteristics of a time series at hand. In this article, we build on the new SN-based test for the mean, adopt the idea of recursive estimation in Shao (2010), and further extend our test to test for a change in other parameters associated with a time series, such as marginal variance, marginal quantile, autocovariance/autocorrelation at certain lags, and spectrum at certain frequency bands. We establish a unified framework to allow the test for a change in the above-mentioned parameters to be treated in the same fashion. As a result, our test can be readily used by the applied researchers in their analysis of time series data.

The rest of the article is organized as follows. Section 2 describes the traditional Kolmogorov–Smirnov test statistic and its practical difficulty in choosing the smoothing parameter, discusses the idea of self-normalization, and introduces our new SN-based test statistic with its power properties. Section 3 extends the SN-based test statistic developed for the mean case to a more general framework. Section 4 presents simulation results to examine the finite sample size and power properties of our new test in comparison with the traditional Kolmogorov–Smirnov test, where consistent estimation of asymptotic variance is involved. Empirical illustrations are provided in Section 5 and conclusions are made in Section 6. We leave the technical details to the Appendix.

2. TESTING FOR A CHANGE IN MEAN

Suppose our interest is to test for a change point in the mean of a univariate time series, that is,

$$H_0: \mathbb{E}(X_1) = \cdots = \mathbb{E}(X_n) = \mu$$

versus

$$H_a: \mathbb{E}(X_1) = \cdots = \mathbb{E}(X_{k^*}) \neq \mathbb{E}(X_{k^*+1}) = \cdots = \mathbb{E}(X_n), \quad 1 \le k^* < n \text{ is unknown.}$$

To facilitate our discussion, we introduce some notation. Let $\mathcal{D}[0,1]$ be the space of functions on $[0,1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley 1968). Denote by "$\Rightarrow$" weak convergence in $\mathcal{D}[0,1]$ or more generally in the $\mathbb{R}^d$-valued function space $\mathcal{D}^d[0,1]$, where $d \in \mathbb{N}$. For a column vector $x = (x_1, \ldots, x_q)' \in \mathbb{R}^q$, let $|x| = (\sum_{j=1}^q x_j^2)^{1/2}$. The symbol $o_p(1)$ signifies convergence to zero in probability. Denote by $\lfloor a \rfloor$ the integer part of $a \in \mathbb{R}$.

Let $\gamma(k) = \text{cov}(X_0, X_k)$, $f(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda}$, and $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$. A class of commonly used test statistics is based on the so-called CUSUM (cumulative sum) process, which is defined as

$$T_n(\lfloor nr \rfloor) = n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} (X_t - \bar{X}_n), \qquad r \in [0,1].$$

Under appropriate moment and weak dependence assumptions on X_t (see Phillips 1987), we have that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \{X_t - \mathbb{E}(X_t)\} \Rightarrow \sigma B(r), \tag{1}$$

where $\sigma^2 = \lim_{n\to\infty} n \,\text{var}(\bar{X}_n) = \sum_{k \in \mathbb{Z}} \gamma(k) > 0$ is the so-called long-run variance and $B(r)$ is the one-dimensional Brownian motion. Under the null hypothesis, we have $T_n(\lfloor nr \rfloor) \Rightarrow \sigma\{B(r) - rB(1)\}$. The well-known Kolmogorov–Smirnov test statistic is defined as

$$KS_n = \sup_{r \in [0,1]} \left| T_n(\lfloor nr \rfloor)/\hat{\sigma}_n \right| = \sup_{k=1,\ldots,n} |T_n(k)/\hat{\sigma}_n|,$$

where $\hat{\sigma}_n^2$ is a consistent estimator of σ^2. A commonly used estimate for σ^2 admits the form

$$\hat{\sigma}_n^2 = \sum_{k=-l_n}^{l_n} \hat{\gamma}(k) K(k/l_n), \tag{2}$$

where $\hat{\gamma}(k) = n^{-1} \sum_{j=1}^{n-|k|} (X_j - \bar{X}_n)(X_{j+|k|} - \bar{X}_n)$ is the sample autocovariance estimate at lag k, $K(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter. Under appropriate regularity conditions, including $1/l_n + l_n/n = o(1)$, $\hat{\sigma}_n^2$ is a consistent estimator of σ^2, so the asymptotic null distribution of KS_n is $\sup_{r \in [0,1]} |B(r) - rB(1)|$, for which critical values have been tabulated in the literature.

A difficult issue in practice is the selection of l. Data-dependent l may yield nonmonotonic power as shown in Vogelsang (1999) and Crainiceanu and Vogelsang (2007). The fixed bandwidth is immune to the nonmonotonic power problem but does not perform well across models with various degree of autocorrelations. The latter authors propose to use a robust estimate of σ^2 based on the ordinary least square regression residuals obtained under the alternative. However, simulation results suggest that there is a large size distortion when strong persistence is present; see Section 4.1. Also see Vogelsang (1999) for similar findings. Theoretical analysis in Vogelsang (1999) and Deng and Perron (2008) show that the decrease in power accompanied with larger shift is due to the fact that the bandwidth l is severely biased upward, which leads to an inflation in the estimate of the scale. To overcome the nonmonotonic power problem, Juhl and Xiao (2009) suggest to estimate the long-run variance σ^2 using the residuals obtained by nonparametric regression. However, the size and power are both sensitive to the bandwidth used in nonparametric regression, and there seems no satisfactory solution provided in their paper.

2.1 Some Background on Self-Normalization

In this article, we propose a new test statistic, that bypasses direct estimation of σ^2. The basic idea is to extend the self-normalization (SN) method (Lobato 2001; Shao 2010) to the change-point testing problem. For the sake of readership, we describe the SN method in the context of inference for the marginal mean of a stationary time series. Under appropriate mixing and moment conditions, $\sqrt{n}(\bar{X}_n - \mu) \to_D N(0, \sigma^2)$, where "$\to_D$" denotes convergence in distribution. In order to construct a confidence interval for μ, the traditional approach replaces the unknown σ^2 by its consistent estimate $\hat{\sigma}_n^2$; see (2).

Since $n(\bar{X}_n - \mu)^2/\hat{\sigma}_n^2 \to_D \chi^2(1)$, the confidence interval for μ is constructed using the critical values from the $\chi^2(1)$ distribution. A major difficulty associated with the traditional approach is the choice of l_n, which is a smoothing parameter whose effect does not appear in the limiting distribution. To avoid the selection of l_n, Lobato (2001) proposed the SN approach as a good alternative to the traditional approach. Let $D_n^2 = n^{-2}\sum_{t=1}^n\{\sum_{j=1}^t(X_j - \bar{X}_n)\}^2$. Assuming (1), then the continuous mapping theorem implies that

$$n(\bar{X}_n - \mu)^2/D_n^2 \to_D \frac{B(1)^2}{\int_0^1\{B(r) - rB(1)\}^2\,dr} =: U_1.$$

The limiting distribution U_1 is pivotal and its critical values have been tabulated by Lobato (2001). The key ingredient is to replace the consistent estimator of σ^2, as used in the traditional approach, with the inconsistent estimator D_n^2. Since the normalization factor D_n^2 is proportional to σ^2, the nuisance parameter σ^2 is canceled out in the limiting distribution of the resulting statistic. Kiefer and Vogelsang (2002) showed that $2D_n^2 = \hat{\sigma}_n^2$ when $K(x) = (1 - |x|)\mathbf{1}(|x| \le 1)$ (i.e., Bartlett kernel) and $b = l_n/n = 1$. Thus the SN method is a special case of the fixed-b paradigm, as advocated by Kiefer and Vogelsang (2005); see Shao (2010) for more discussions.

2.2 SN-Based Test Statistics

Following the description of the SN idea in the previous subsection, a naive extension of the SN idea to the change-point testing problem is to replace $\hat{\sigma}_n$ in KS_n by D_n. In other words, let

$$\widetilde{KS}_n = \sup_{r\in[0,1]}\left|T_n(\lfloor nr\rfloor)/D_n\right| = \sup_{k=1,\ldots,n}|T_n(k)/D_n|.$$

Under H_0, $\widetilde{KS}_n \to_D \sup_{r\in[0,1]}|B(r) - rB(1)|/[\int_0^1\{B(r) - rB(1)\}^2\,dr]^{1/2}$. Figure 1 shows the power (rejection percentage) of $\widetilde{KS}_n$ for the following alternative:

$$X_t = \begin{cases} u_t, & 1 \le t \le n/2 \\ \eta + u_t, & n/2 + 1 \le t \le n = 200, \end{cases} \tag{3}$$

where $u_t = 0.5u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{iid}\,N(0,1)$. It is seen that when the magnitude of change η gets large, the power of $\widetilde{KS}_n$ decreases to zero. The complete loss of power is attributed to the increase of the denominator in $\widetilde{KS}_n$ (i.e., D_n) with respect to η. So a naive extension of the SN idea fails.

The major problem with $\widetilde{KS}_n$ is that the self-normalizer (i.e., denominator) of $\widetilde{KS}_n$ does not take into account the change-point alternative. To circumvent the problem, we propose the following SN-based test with a new self-normalizer, which accounts for the one change point alternative. The SN-based test for the multiple change-point alternative is discussed in Section 2.3. Let $S_{t_1,t_2} = \sum_{j=t_1}^{t_2} X_j$ if $t_1 \le t_2$ and 0 otherwise. We define the normalization process $V_n(\cdot)$ as follows. For $k = 1, \ldots, n-1$, let

$$V_n(k) = n^{-2}\left[\sum_{t=1}^k\{S_{1,t} - (t/k)S_{1,k}\}^2 + \sum_{t=k+1}^n\{S_{t,n} - (n-t+1)/(n-k)S_{k+1,n}\}^2\right].$$

Our test statistic is defined as

$$G_n = \sup_{k=1,\ldots,n-1} T_n(k)'V_n^{-1}(k)T_n(k). \tag{4}$$

Assuming (1), we can derive the limiting null distribution of G_n as

$$G_n \to_D \sup_{r\in[0,1]}\{B(r) - rB(1)\}'V^{-1}(r)\{B(r) - rB(1)\},$$

where $V(r) = \int_0^r\{B(s) - (s/r)B(r)\}^2\,ds + \int_r^1[B(1) - B(s) - (1-s)/(1-r)\{B(1) - B(r)\}]^2\,ds$. Note that the normalization factor $V_n(k)$ in our test depends on k, whereas the normalization factor in $\widetilde{KS}_n$, that is, D_n is the same for all k. This distinction has important implications in their power behaviors. It is not hard to see that the magnitude of $V_n(k^*)$ does not depend on $\Delta_n := \mathbb{E}(X_{k^*+1}) - \mathbb{E}(X_{k^*})$ because the two sums in $V_n(k^*)$, which involve the forward partial sum before k^* and the backward partial sum after k^*, are invariant with respect to Δ_n. On the other hand, the larger the magnitude of Δ_n, the larger $|T_n(k^*)|$ becomes. Heuristically, the magnitudes of $T_n(k^*)'V_n^{-1}(k^*)T_n(k^*)$ and G_n both get larger (in distributional sense) as $|\Delta_n|$ increases, so we get more power. The monotonic power for our test is also confirmed in simulation studies.

Next, we investigate the power of G_n under both local and fixed alternatives. Under H_a, let $k^* = \lfloor\lambda n\rfloor$ for $\lambda \in (0,1)$. The following theorem states the consistency of our test.

Theorem 2.1. Suppose that (1) holds. If $\Delta_n = \Delta \ne 0$ is fixed, then G_n diverges to ∞ in probability. If $\Delta_n = n^{-1/2}L$, $L \ne 0$, then $\lim_{|L|\to\infty}\lim_{n\to\infty} G_n = \infty$ in probability.

We mention in passing that our SN-based test statistic is tailored to the testing problem. To estimate a change point, presumably one can use $\hat{k}$, where

$$\hat{k} = \mathop{\arg\max}_{k=1,\ldots,n-1} T_n(k)'V_n(k)^{-1}T_n(k).$$

However, it seems difficult to obtain the asymptotic distribution of $\hat{k}$ as an estimator of k^*. See Bai (1994, 1997) for early work on the estimation of a change point in the time series setting.

As pointed out by a referee, our test statistic G_n can be made more general by incorporating a weighting scheme and replacing $\sup_{k=1,\ldots,n-1}$ by $n^{-1}\sum_{k=1}^{n-1}$. Specifically, let $\{w(t), t \in [0,1]\}$ be the weight function, which can reflect prior information about the location of a change point. Then we can use $\sup_{1\le k\le n-1} w(k/n)T_n(k)'V_n(k)^{-1}T_n(k)$ or $n^{-1}\sum_{k=1}^{n-1} w(k/n) \times T_n(k)'V_n(k)^{-1}T_n(k)$ to test for a change point. To keep the focus of the article, we shall leave the asymptotic and finite sample investigations of these two test statistics (with a few different weighting schemes) in a separate work.

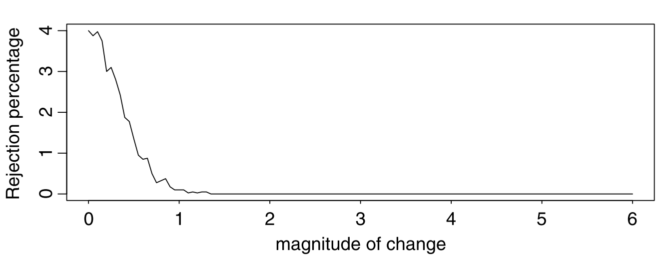

Figure 1. The empirical rejection percentage of the naive SN-based test $\widetilde{KS}_n$ under the alternative model (3).

2.3 Multiple Change-Point Alternatives

Our test statistic G_n is designed for a single change-point alternative and it can be extended to allow for multiple change points. To illustrate the idea, we consider the case of two change points. Let $\bar{X}_{t,t'} = (t'-t+1)^{-1}\sum_{j=t}^{t'} X_j$ for $1 \le t \le t' \le n$. For $1 \le n_1 < n_2 \le n$, we define a CUSUM process on the basis of the subsample $\{X_{n_1+1}, \ldots, X_{n_2}\}$ as

$$\begin{aligned} T_{n_1+1,n_2}(k) &= \frac{1}{\sqrt{n_2-n_1}} \sum_{t=n_1+1}^{k} \left(X_t - \bar{X}_{n_1+1,n_2}\right) \\ &= \frac{(n_2-k)(k-n_1)}{(n_2-n_1)^{3/2}} \left(\overline{X}_{n_1+1,k} - \overline{X}_{k+1,n_2}\right), \\ &\quad n_1+1 \le k \le n_2. \end{aligned}$$

We further extend $V_n(k)$ to its subsampled version as

$$\begin{aligned} V_{n_1+1,n_2}(k) &= \frac{1}{(n_2-n_1)^2} \\ &\quad \times \Bigg[\sum_{t=n_1+1}^{k} \left\{S_{n_1+1,t} - (t-n_1)/(k-n_1) S_{n_1+1,k}\right\}^2 \\ &\quad + \sum_{t=k+1}^{n_2} \left\{S_{t,n_2} - (n_2-t+1)/(n_2-k) S_{k+1,n_2}\right\}^2\Bigg], \\ &\quad n_1+1 \le k \le n_2-1. \end{aligned}$$

Fix $0 < \epsilon < 1/3$ and define $\boldsymbol{\Omega}_n(\epsilon) := \{(k_1,k_2) : \lfloor \epsilon n \rfloor \le k_1 < k_2 \le \lfloor (1-\epsilon)n \rfloor, k_2 - k_1 \ge \lfloor \epsilon n \rfloor\}$. For any $(k_1,k_2) \in \boldsymbol{\Omega}_n(\epsilon)$, define

$$\begin{aligned} H_n(k_1,k_2) &:= T_{1,k_2}(k_1)' V_{1,k_2}^{-1}(k_1) T_{1,k_2}(k_1) \\ &\quad + T_{k_1+1,n}(k_2)' V_{k_1+1,n}^{-1}(k_2) T_{k_1+1,n}(k_2) \end{aligned}$$

and $Q_n(\epsilon) := \sup_{(k_1,k_2)\in\boldsymbol{\Omega}_n(\epsilon)} H_n(k_1,k_2)$. Similar to the one change-point case, we can derive the limiting null distribution of $Q_n(\epsilon)$. The asymptotic critical values can be obtained through simulations, although the computation would be more expensive in this case. The extension to $m \ge 3$ change points is quite straightforward so we omit the details here.

In practice, the number of change points may be unknown and needs to be estimated. One approach is to treat the change-point estimation and testing as a model selection problem and adopt a suitably chosen information criterion; see Yao (1988) and Zhang and Siegmund (2007) for the use of BIC and a modified version. Davis, Lee, and Rodriguez-Yam (2006) use minimum description length criterion to perform automatic segmentation of a nonstationary time series and find the best combination of the number and location of change points, as well as the order and the parameter estimates for the piecewise AR processes. Alternatively, one can use a sequential testing procedure to decide the number of change points; see Bai and Perron (1998) and Qu and Perron (2007) and the references therein.

3. TESTING FOR A CHANGE POINT IN A GENERAL FRAMEWORK

In this section, we shall extend the SN-based change point test statistic from the mean case to more general settings. We adopt the framework in Shao (2010) and let $\mathbf{F}^m$ denote the mth marginal distribution of X_t, where the dimension m is fixed but arbitrary. Let $\mathbf{Y}_t = (X_t, \ldots, X_{t+m-1})'$, $t = 1, \ldots, N = n-m+1$, and $\mathbf{F}_t^m$ denotes the distribution of $\mathbf{Y}_t$. Let $\boldsymbol{\theta}_t = \mathbf{T}(\mathbf{F}_t^m) \in \mathbb{R}^q$, $t = 1, \ldots, N$, be the quantity of interest, where $\mathbf{T}$ is a functional that takes values in $\mathbb{R}^q$. We are interested in testing

$$H_0 : \boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_N$$

versus

$$H_a : \boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_{k^*} \ne \boldsymbol{\theta}_{k^*+1} = \cdots = \boldsymbol{\theta}_N \quad \text{for some unknown } k^*,\ 1 \le k^* < N.$$

Important examples that fall into the above framework include: (i) marginal mean of X_t, that is, $q = m = 1$, $T(F^1) = \int_{\mathbb{R}} x\, dF^1(x)$; (ii) marginal variance of X_t, that is, $q = m = 1$, $T(F^1) = \int_{\mathbb{R}} x^2\, dF^1(x) - \{\int_{\mathbb{R}} x\, dF^1(x)\}^2$; (iii) autocorrelation function at lags $(1, \ldots, k)$, that is, $q = k$, $m = k+1$, and $\mathbf{T}(\mathbf{F}^m) = (\rho(1), \ldots, \rho(k))'$, where $\rho(k) = \gamma(k)/\gamma(0)$; (iv) pth quantile of the distribution F^1, where $p \in (0,1)$. In this case, $T(F^1) = (F^1)^{-1}(p)$.

Let $\boldsymbol{\rho}_{N_1,N_2}$ be the empirical distribution based on $\{\mathbf{Y}_j\}_{j=N_1}^{N_2}$ and $\hat{\boldsymbol{\theta}}_{N_1,N_2} = \mathbf{T}(\boldsymbol{\rho}_{N_1,N_2})$. For any $k = 1, \ldots, N-1$, we define $\mathbf{T}_n(k) = k/\sqrt{N}(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N})$ and

$$\begin{aligned} \mathbf{V}_n(k) = N^{-2} \Bigg\{ &\sum_{t=1}^{k} t^2 (\hat{\boldsymbol{\theta}}_{1,t} - \hat{\boldsymbol{\theta}}_{1,k})(\hat{\boldsymbol{\theta}}_{1,t} - \hat{\boldsymbol{\theta}}_{1,k})' \\ &+ \sum_{t=k+1}^{N} (N-t+1)^2 (\hat{\boldsymbol{\theta}}_{t,N} - \hat{\boldsymbol{\theta}}_{k+1,N})(\hat{\boldsymbol{\theta}}_{t,N} - \hat{\boldsymbol{\theta}}_{k+1,N})' \Bigg\}. \end{aligned}$$

Then our test statistic is $G_n := \sup_{k=1,\ldots,N-1} \mathbf{T}_n(k)' \mathbf{V}_n(k)^{-1} \times \mathbf{T}_n(k)$. In the foregoing expressions, we use both forward and backward recursive estimates to mimic the forward and backward sums in the mean case. In the case of the mean, $t(\hat{\theta}_{1,t} - \hat{\theta}_{1,k}) = S_{1,t} - (t/k) S_{1,k}$ and our statistics $\mathbf{T}_n(k)$, $\mathbf{V}_n(k)$, and G_n reduce to those defined in Section 2.2.

Following Shao (2010), we restrict our attention to the so-called approximately linear statistic $\mathbf{T}(\boldsymbol{\rho}_{1,N})$, that is,

$$\mathbf{T}(\boldsymbol{\rho}_{1,N}) = \mathbf{T}(\mathbf{F}^m) + N^{-1} \sum_{t=1}^{N} \mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m) + \mathbf{R}_{1,N},$$

where $\mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m)$ is the influence function of $\mathbf{T}$ (Hampel et al. 1986) defined by

$$\mathbf{IF}(\mathbf{y}; \mathbf{F}^m) = \lim_{\epsilon \downarrow 0} \frac{\mathbf{T}\{(1-\epsilon)\mathbf{F}^m + \epsilon \delta_{\mathbf{y}}\} - \mathbf{T}(\mathbf{F}^m)}{\epsilon}$$

and $\mathbf{R}_{1,N}$ is the reminder term. Similarly, we have that for any $1 \le N_1 \le N_2 \le N$,

$$\mathbf{T}\left(\boldsymbol{\rho}_{N_1,N_2}\right) = \mathbf{T}(\mathbf{F}^m) + (N_2 - N_1 + 1)^{-1} \sum_{t=N_1}^{N_2} \mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m) + \mathbf{R}_{N_1,N_2}. \quad (5)$$

To obtain the asymptotic null distribution of G_n, we impose the following two assumptions:

Assumption 3.1. Assume $\mathbb{E}\{\mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m)\} = \mathbf{0}$ and

$$N^{-1/2} \sum_{t=1}^{\lfloor rN \rfloor} \mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m) \Rightarrow \boldsymbol{\Delta}\mathbf{B}_q(r),$$

where $\boldsymbol{\Delta}$ is a $q \times q$ lower triangular matrix with nonnegative diagonal entries and $\mathbf{B}_q(\cdot)$ is a q-dimensional vector of independent Brownian motions. Assume that $\boldsymbol{\Delta}\boldsymbol{\Delta}' = \boldsymbol{\Sigma}(\mathbf{F}^m) = \sum_{k=-\infty}^{\infty} \operatorname{cov}\{\mathbf{IF}(\mathbf{Y}_0; \mathbf{F}^m), \mathbf{IF}(\mathbf{Y}_k; \mathbf{F}^m)\}$ is positive definite.

Assumption 3.2. Assume that $\sup_{k=1,\ldots,N} |k\mathbf{R}_{1,k}| = o_p(N^{1/2})$ and $\sup_{k=1,\ldots,N} |k\mathbf{R}_{N-k+1,N}| = o_p(N^{1/2})$.

Assumption 3.1 holds under suitable moment condition on $\mathbf{IF}(\mathbf{Y}_t; \mathbf{F}^m)$ and mixing condition on X_t; see Assumption 2.1 and Lemma 2.2 of Phillips (1987). The verification of Assumption 3.2 is highly nontrivial and is left for future research. Let $\mathbf{V}_q(r) = \int_0^r \mathbf{W}_{1,q}(s,r)\mathbf{W}_{1,q}(s,r)'\,ds + \int_r^1 \mathbf{W}_{2,q}(s,r)\mathbf{W}_{2,q}(s,r)'\,ds$, where $\mathbf{W}_{1,q}(s,r) = \mathbf{B}_q(s) - (s/r)\mathbf{B}_q(r)$ for $s \in [0,r]$ and $\mathbf{W}_{2,q}(s,r) = \{\mathbf{B}_q(1) - \mathbf{B}_q(s)\} - (1-s)/(1-r)\{\mathbf{B}_q(1) - \mathbf{B}_q(r)\}$ for $s \in [r,1]$. The following theorem states the asymptotic null distribution of G_n.

Theorem 3.1. Under the null hypothesis, suppose that Assumptions 3.1 and 3.2 hold. Then the limiting null distribution for G_n is $G(q) := \sup_{r\in[0,1]} \{\mathbf{B}_q(r) - r\mathbf{B}_q(1)\}'\mathbf{V}_q(r)^{-1}\{\mathbf{B}_q(r) - r\mathbf{B}_q(1)\}$.

Table 1 presents the simulated critical values for $G(q)$ based on $n = 5000$ and 10,000 replications for $q = 1, \ldots, 10$. In what follows, we provide a discussion on the power of our test. Under the alternative, there is a change point for the quantity $\boldsymbol{\theta}$ at time k^*. Since $\mathbf{Y}_t = (X_t, \ldots, X_{t+m-1})'$, it is reasonable to assume that for X_t, the change point occurs at time $k^* + m$. In particular, we assume that the observations $(X_1, \ldots, X_{k^*+m-1})$ come from a stationary process $\{X_t^{(1)}\}_{t\in\mathbb{Z}}$, whereas the observations $(X_{k^*+m}, \ldots, X_n)$ come from another stationary process $\{X_t^{(2)}\}_{t\in\mathbb{Z}}$. For stationary processes $X_t^{(j)}, j = 1, 2$, we can similarly define $\mathbf{Y}_t^{(1)}$ and $\mathbf{Y}_t^{(2)}$. For $j = 1, 2$, we assume that (a) the expansion (5) holds for the process $\mathbf{Y}_t^{(j)}$ with influence functions $\mathbf{IF}^{(j)}(\mathbf{Y}_t^{(j)}; \mathbf{F}^m)$ and remainder terms $\mathbf{R}_{N_1,N_2}^{(j)}$ for $j = 1, 2$; (b) $N^{-1/2} \sum_{t=1}^{\lfloor rN \rfloor} \{\mathbf{IF}^{(1)}(\mathbf{Y}_t^{(1)}; \mathbf{F}^m)', \mathbf{IF}^{(2)}(\mathbf{Y}_t^{(2)}; \mathbf{F}^m)'\}' \Rightarrow \tilde{\boldsymbol{\Delta}}\mathbf{B}_{2q}(r)$, where $\tilde{\boldsymbol{\Delta}}$ is a $2q \times 2q$ lower triangular matrix with nonnegative diagonal entries and $\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}}'$ is positive definite; (c) The remainder terms $(\mathbf{R}_{1,k}^{(j)}, \mathbf{R}_{N-k+1,N}^{(j)})_{k=1}^N$ satisfy Assumption 3.2 for $j = 1, 2$. Then following the arguments in the proofs of Theorem 2.1 and Theorem 3.1, we can show that G_n is consistent and it has nontrivial power against local alternatives or order $N^{-1/2}$.

Remark 3.1. As mentioned above, we assume that the change of the mth marginal distribution of X_t is due to the change in its one-dimensional marginal distribution. It was pointed out by a referee that it seems hard to come up with a situation that some characteristic of the m-dimensional distribution ($m \ge 3$) is constant before the change point and then become another constant afterward, unless the characteristic essentially depends on the bivariate distribution. This is in fact possible as seen from the following example:

Example 3.1. Consider the following model:

$$X_t = \begin{cases} Z_t, & 1 \le t \le k^* \\ Z_t + \eta, & k^* + 1 \le t \le n, \end{cases}$$

where Z_t is a strictly stationary process with $\operatorname{cov}(Z_1, Z_2) \ne 0$ and $\eta \ne 0$. Suppose that $m = 3$ and the quantity of interest is $\theta = T(\mathbf{F}^3) = \mathbb{E}[\{X_1 - \mathbb{E}(X_1)\}\{X_2 - \mathbb{E}(X_2)\}X_3]$. Then $\theta_i = \mathbb{E}[\{X_i - \mathbb{E}(X_i)\}\{X_{i+1} - \mathbb{E}(X_{i+1})\}X_{i+2}]$. Straightforward calculation shows that

$$\theta_1 = \theta_2 = \cdots = \theta_{k^*-2} = \mathbb{E}\big[\{Z_1 - \mathbb{E}(Z_1)\}\{Z_2 - \mathbb{E}(Z_2)\}Z_3\big]$$

and

$$\begin{aligned}\theta_{k^*-1} &= \theta_{k^*} = \cdots = \theta_n \\ &= \mathbb{E}\big[\{Z_1 - \mathbb{E}(Z_1)\}\{Z_2 - \mathbb{E}(Z_2)\}Z_3\big] + \eta \operatorname{cov}(Z_1, Z_2),\end{aligned}$$

which suggest that there is a change point in θ_t at time $k^* - 2$. For general $m \ge 4$, we can define $\theta = \mathbb{E}[\{X_1 - E(X_1)\} \times \{X_{m-1} - E(X_{m-1})\}X_m]$. By a similar argument, we can see that there is a change point in θ_t at time $k^* - m + 1$ provided that $\mathbb{E}[\{Z_1 - \mathbb{E}(Z_1)\} \times \cdots \times \{Z_{m-1} - \mathbb{E}(Z_{m-1})\}] \ne 0$.

Despite the above example, it is in general more natural to consider the following alternative

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_{k^*} \ne \boldsymbol{\theta}_{k^*+m} = \boldsymbol{\theta}_{k^*+m+1} = \cdots = \boldsymbol{\theta}_N,$$

which contains our abrupt change alternative H_a as a special case. It is silent about the quantities θ_t, when t lies in the transition period $[k^* + 1, k^* + m - 1]$. Since m is finite, the contribution of the observations in the transition period is asymptotically negligible and the consistency of our test still holds.

Table 1. Simulated critical values for $G(q)$, $q = 1, \ldots, 10$, based on $n = 5000$ and 10,000 replications

	q									
α%	1	2	3	4	5	6	7	8	9	10
90%	29.6	56.5	81.5	114.7	150.0	183.8	223.5	267.1	308.5	360.0
95%	40.1	73.7	103.6	141.5	182.7	218.8	267.3	317.9	360.7	420.5
97.5%	52.2	92.2	128.9	171.9	218.7	255.0	313.4	367.9	416.3	483.0
99%	68.6	117.7	160.0	209.7	265.8	318.3	368.0	432.5	483.6	567.2
99.5%	84.6	135.3	182.9	246.6	291.7	367.7	410.5	498.1	544.9	621.6
99.9%	121.9	192.5	246.8	319.2	358.1	464.9	530.6	614.1	649.0	751.1

3.1 Testing for a Change Point in Spectrum

For a stationary time series, the spectral density function $\{f(\lambda), \lambda \in (-\pi, \pi]\}$ or the spectral distribution function $\{F(\lambda) = \int_0^\lambda f(w)\,dw, \lambda \in [0, \pi]\}$ fully characterizes its second order properties. Therefore, if the goal is to check the structural stability of second order properties, it is natural to test for a change in spectrum, and see at which frequency band the change occurs. Testing for a change in spectrum has been considered by Picard (1985), Giraitis and Leipus (1992), and Lavielle and Ludena (2000). Picard (1985) developed a Kolmogorov–Smirnov test for spectrum change under the Gaussian assumption, which was later relaxed by Giraitis and Leipus (1992). Note that the limiting null distribution of the test statistic of Giraitis and Leipus (1992) depends on the fourth order cumulants of the process, and it is not clear how to implement their test in practice. Laville and Ludena (2000) allowed multiple change points but assumed a parametric form for the spectral density in each segment. Here we propose a fully nonparametric SN-based test for a change in spectrum. This extension is made possible since the validity of the SN approach has been extended to cover the quantity that is a functional of $\mathbf{F}^\infty$ [i.e., the joint distribution of $(X_t)_{t\in\mathbb{Z}}$] in Shao (2010). In particular, this includes $F(\lambda_1)$ and $F(\lambda_2) - F(\lambda_1)$ for $0 < \lambda_1 < \lambda_2 \le \pi$. Suppose that we want to test if there is a change in $\boldsymbol{\theta} = \{F(\lambda_1), F(\lambda_2) - F(\lambda_1), \ldots, F(\lambda_q) - F(\lambda_{q-1})\}$ (or for each element) for any prespecified frequencies $0 < \lambda_1 < \cdots < \lambda_q = \pi$. Note that $F(\lambda_j) - F(\lambda_{j-1}) = \int_{\lambda_{j-1}}^{\lambda_j} f(\omega)\,d\omega$ measures the total spectral power within the band $[\lambda_{j-1}, \lambda_j]$. A natural estimator for $F(\lambda)$ is $F_n(\lambda) = \int_0^\lambda I_n(w)\,dw$, where

$$I_n(w) = (2\pi n)^{-1}\left|\sum_{j=1}^n (X_j - \bar{X}_n)\exp(ijw)\right|^2$$

is the periodogram. For any subsample $(X_t, \ldots, X_{t'})$, $1 \le t < t' \le n$, we define $I_{t,t'}(w) = \{2\pi(t' - t + 1)\}^{-1}|\sum_{j=t}^{t'}(X_j - \bar{X}_{t,t'})\exp(ijw)|^2$ and $F_{t,t'}(\lambda) = \int_0^\lambda I_{t,t'}(w)\,dw$.

Let $\hat{\boldsymbol{\theta}}_{t,t'} = \{F_{t,t'}(\lambda_1), F_{t,t'}(\lambda_2) - F_{t,t'}(\lambda_1), \ldots, F_{t,t'}(\lambda_q) - F_{t,t'}(\lambda_{q-1})\}$ be the estimator of $\boldsymbol{\theta}$ on the basis of the subsample $(X_t, \ldots, X_{t'})$. Further let $\mathbf{T}_n(k) := k/\sqrt{n}(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,n})$ and

$$\mathbf{V}_n(k) := n^{-2}\left\{\sum_{t=1}^k t^2(\hat{\boldsymbol{\theta}}_{1,t} - \hat{\boldsymbol{\theta}}_{1,k})(\hat{\boldsymbol{\theta}}_{1,t} - \hat{\boldsymbol{\theta}}_{1,k})' + \sum_{t=k+1}^n (n-t+1)^2(\hat{\boldsymbol{\theta}}_{t,n} - \hat{\boldsymbol{\theta}}_{k+1,n})(\hat{\boldsymbol{\theta}}_{t,n} - \hat{\boldsymbol{\theta}}_{k+1,n})'\right\}$$

for $k = 1, \ldots, n-1$. Then our test statistic

$$G_n = \sup_{k=1,2,\ldots,n-1} \mathbf{T}_n(k)'\mathbf{V}_n^{-1}(k)\mathbf{T}_n(k).$$

Under appropriate moment and weak dependence assumptions, we expect to show that the asymptotic null distribution of G_n is again $G(q)$; see Theorem 3.1.

A quantity that is closely related to $F(\lambda)$ is its ratio counterpart $\tilde{F}(\lambda) = F(\lambda)/F(\pi)$. If one is only interested in the pattern of dependence described in terms of autocorrelations, then $\tilde{F}(\lambda)$ is of more practical relevance. Suppose the interest is to test for a change in $\tilde{\boldsymbol{\theta}} = \{\tilde{F}(\lambda_1), \tilde{F}(\lambda_2) - \tilde{F}(\lambda_1), \ldots, \tilde{F}(\lambda_q) - \tilde{F}(\lambda_{q-1})\}$, then one can estimate $\tilde{F}(\lambda)$ by $F_n(\lambda)/F_n(\pi)$ and calculate forward and backward recursive estimates in a similar fashion. The resulting test statistic admits the same form as G_n and its limiting null distribution is also $G(q)$. We shall investigate its finite sample performance in Section 4.3.

4. SIMULATION STUDIES

Through Monte Carlo simulations, we investigate the size and power properties of the new SN-based test statistics for the mean change in Section 4.1, for the median change in Section 4.2, and for the change in the second order property of a time series in Section 4.3. Throughout our simulations, we use 5000 replications.

4.1 Change Point in Mean

In this subsection, we examine the finite sample size and power properties of our SN-based test statistic G_n in detecting a shift in mean, and compare the results with those delivered by KS_n test statistic. Specifically, we compare the following five methods:

(i) FB: we used the Bartlett kernel for $K(\cdot)$ and a fixed bandwidth $l_n = \lfloor n^{1/3}\rfloor$ in the calculation of KS_n.

(ii) DDB1: we used Andrews' AR(1) plug-in bandwidth selection rule to choose l_n in calculating KS_n. In particular, $K(\cdot)$ is the Bartlett kernel and $l(n) = \lfloor 1.1447\{\frac{4\hat{\rho}^2 n}{(1-\hat{\rho}^2)^2}\}^{1/3}\rfloor$, where $\hat{\rho} = \sum_{t=2}^n \hat{u}_t\hat{u}_{t-1}/\sum_{t=2}^n \hat{u}_{t-1}^2$ and $\hat{u}_t = X_t - \bar{X}_n$. This data dependent bandwidth was recommended by Andrews (1991), who showed that it minimizes the approximate mean square error of $\hat{\sigma}_n^2$ if the process u_t admits an AR(1) model.

(iii) DDB2: we also tried a robust long run variance estimator, as used in Crainiceanu and Vogelsang (2007). The idea is to estimate the break point and then use the ordinary regression residuals obtained from the alternative one-break model to construct the long-run variance estimate. In particular, we estimated the break point by $\hat{k}$, which maximizes

$$\left\{\frac{k(n-k)}{n^2}\right\}^{1/2}\left|\frac{1}{k}\sum_{t=1}^k X_t - \frac{1}{n-k}\sum_{t=k+1}^n X_t\right|$$

over $k = 1, \ldots, n-1$. Bai (1994) obtained the consistency and the convergence rate of $\hat{k}$ to the true break point. We again used the data dependent bandwidth as in (ii) by replacing $\hat{u}_t$ by $\tilde{u}_t$, where

$$\tilde{u}_t = \begin{cases} X_t - \hat{k}^{-1}\sum_{t=1}^{\hat{k}} X_t, & \text{if } t = 1, \ldots, \hat{k} \\ X_t - (n-\hat{k})^{-1}\sum_{t=\hat{k}+1}^n X_t, & \text{if } t = \hat{k}+1, \ldots, n. \end{cases}$$

(iv) DDB-JX: Juhl and Xiao (2009) recently proposed to use the residuals from nonparametric regression to calculate the long-run variance estimate to alleviate the nonmonotonic power problem. Specifically, they obtained nonparametric residuals as $\bar{u}_t = X_t - (nh)^{-1}\sum_{s=1}^n \tilde{K}(\frac{t-s}{nh})X_s$, where $\tilde{K}(\cdot)$ is a kernel function and h is a bandwidth parameter. Following Juhl and Xiao (2009), we take $\tilde{K}$ to be the Epanechnikov kernel [i.e., $\tilde{K}(x) = 3/4(1-x^2)\mathbf{1}(|x| \le 1)$] and $h = 2n^{-1/5}$, which delivered

Table 2. Empirical sizes (in percentage) for the five methods used in testing for a change in mean

n	ρ	FB	DDB1	DDB2	DDB-JX	SN
200	0	3.5	2.5	4.1	4.2	4.9
	0.5	6.9	4.9	12.8	6.0	6.1
	0.8	20.2	2.4	22.8	5.8	8.6
500	0	3.6	2.7	3.4	4.1	5.2
	0.5	6.2	4.7	8.8	5.1	5.3
	0.8	18.4	4.5	14.1	6.0	6.5

NOTE: Nominal level is 5%. The largest standard error is 0.59%.

the best results in their simulation studies. For the calculation of long-run variance estimate, we use the quadratic spectral kernel [see Equation (2.3) of Juhl and Xiao 2009] with the corresponding data-dependent bandwidth [see Equation (2.5) of Juhl and Xiao 2009].

(v) SN: SN-based test statistic G_n; see (4).

Consider the model

$$X_t = \eta\mathbf{1}(t > 0.5n) + u_t, \qquad t = 1, \ldots, n,$$

where $u_t = \rho u_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \text{iid}\, N(0, 1)$. To examine the size, we let $\eta = 0$, $\rho = 0, 0.5, 0.8$ and $n = 200, 500$. As seen from Table 2, the size distortion for FB and DDB2 is very large when $\rho = 0.8$, which is consistent with the findings reported in Crainiceanu and Vogelsang (2007). For DDB1, it is undersized but its size appears to be quite close to the nominal level when $\rho = 0.5, 0.8$ for $n = 500$. The size performance for Juhl and Xiao's method is quite satisfactory. For the SN-based test, the size distortion corresponding to $\rho = 0.8$ is noticeable when $n = 200$, and it improves at $n = 500$. Figure 2 shows the size-adjusted power for the five methods at $n = 200$ and 500. As we expected, the nonmonotonic power phenomenon occurs for DDB1; see the plots corresponding to $n = 200$, $\rho = 0.5, 0.8$ and $n = 500$, $\rho = 0.8$. It is interesting to note that Juhl and Xiao's method also exhibits nonmonotonic power in the case of $n = 200$ and $\rho = 0.8$. This finding is new, and it suggests that the use of nonparametric regression residuals in the calculation of a long-run variance estimate can alleviate but not eliminate the nonmontonic power problem. In contrast, the SN-based test always delivers monotonic power. Compared to the fixed bandwidth scheme, which delivers the highest power among all the methods for all the scenarios under consideration, the SN-based test is less powerful, but the power loss is fairly moderate. Overall, the size and power performance for the SN-based test in detecting a change point in mean is quite encouraging.

4.2 Change Point in Median

In this subsection, we investigate the size and power performance of our SN-based test G_n in detecting a shift in median. To our knowledge, there has been little work on change-point detection for the median in the time series setup. Following the discussion on the CUSUM process for the mean, we can similarly define a CUSUM-like process for the median, that is, $T_n(k) = k/\sqrt{n}(\hat{\theta}_{1,k} - \hat{\theta}_{1,n})$, $k = 1, \ldots, n$, where $\hat{\theta}_{n_1,n_2}$ is the sample median based on the observations $\{X_{n_1}, \ldots, X_{n_2}\}$. Under the null and appropriate weakly dependent conditions, the weak convergence $T_n(\lfloor nr \rfloor) \Rightarrow \sigma_{med}\{B(r) - rB(1)\}$ is expected to hold, where

$$\sigma_{med}^2 = \{4g^2(\theta)\}^{-1} \sum_{k=-\infty}^{\infty} \operatorname{cov}\{1 - 2\mathbf{1}(X_0 \le \theta), 1 - 2\mathbf{1}(X_k \le \theta)\}.$$

Here θ is the true median of the distribution of X_1 and $g(\cdot)$ is the density function of X_1. To estimate σ_{med}^2, we use the nonoverlapping subsampling method (Carlstein 1986) for simplicity. The use of the overlapping subsampling method (Politis, Romano, and Wolf 1999) is possible but it is computationally more expensive and would not change the results much. Let l be the subsampling width and $\hat{\theta}_i$, $i = 1, \ldots, s_n(l) = \lceil n/l \rceil$, be the sample median for the ith nonoverlapping subsample, where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Then the subsampling-based variance estimator of σ_{med}^2 is defined as

$$\hat{\sigma}_{med}^2 = \frac{l}{s_n(l)} \sum_{i=1}^{s_n(l)} \left(\hat{\theta}_i - s_n(l)^{-1} \sum_{i=1}^{s_n(l)} \hat{\theta}_i \right)^2.$$

The consistency of $\hat{\sigma}_{med}^2$ can be established under suitable weakly dependent conditions on X_t; see Carlstein (1986). Therefore, the asymptotic null distribution of the statistic

$$KS_{n,med} = \sup_{k=1,2,\ldots,n-1} |T_n(k)/\hat{\sigma}_{med}|$$

is $\sup_{r\in[0,1]} |B(r) - rB(1)|$. In practice, the selection of l_n seems to be quite difficult and we are not aware of any data-dependent bandwidth rule in this setting. So we tried $l = cn^{1/3}$, where $c =$ 0.5, 1, 2, 4, and 8.

Let ϵ_{1t}, ϵ_{2t}, and ϵ_{3t} be iid with $N(0, 1)$, $t(5)$ and Cauchy$(0, 1)$ distribution respectively. Consider the following three models:

$$M_1 : X_{1t} = \eta\mathbf{1}\{t > 0.5n\} + u_{1t};$$

$$M_2 : X_{2t} = \eta\mathbf{1}\{t > 0.5n\} + u_{2t}; \qquad X_{3t} = \eta\mathbf{1}\{t > 0.5n\} + u_{3t},$$

where $u_{1t} = 0.7u_{1(t-1)} + \epsilon_{1t}$, $u_{2t} = 0.7u_{2(t-1)} + 0.6^{1/2}\epsilon_{2t}$, and $u_{3t} = 0.7u_{3(t-1)} + \epsilon_{3t}$. The empirical sizes for the SN-based test statistic and $KS_{n,med}$ are presented in Table 3 for $n = 200$ and 500. For the models M_1 and M_2, the size distortion for the subsampling-based test SS_c (i.e., subsampling width is equal to $cn^{1/3}$) at $c = 0.5, 1$ is very severe. The size becomes closer to the nominal level for SS_2, SS_4, and SS_8, where the latter two outperform the SN-based test in size. For the model M_3, the size increases as $c \ge 1$ gets larger. The test statistics $SS_{0.5}$, SS_1, and SS_2 are undersized and the size for SS_4 appears to be the best among all the methods. The opposite patterns corresponding to the models M_1, M_2, and M_3 suggest that the optimal subsampling size could very much depend on the moment property of the underlying process. For the SN-based test, it is oversized and the size distortion diminishes as sample size increases. Compared with the models M_1 and M_2, the model M_3 corresponds to slightly more size distortion for the SN-based test, but the difference is not as drastic as the subsampling-based test.

Figure 3 examines the size-corrected power for the models M_1 and M_2 with $\eta \in [0, 6]$, and for the model M_3 with $\eta \in [0, 30]$. When $n = 200$, for the models M_1 and M_2, the power for $SS_{0.5}$ is the highest, and the SN-based test has a modest power loss. By contrast, SS_4 and SS_8, which deliver accurate size, have a severe power loss. For the model M_3, it is interest-

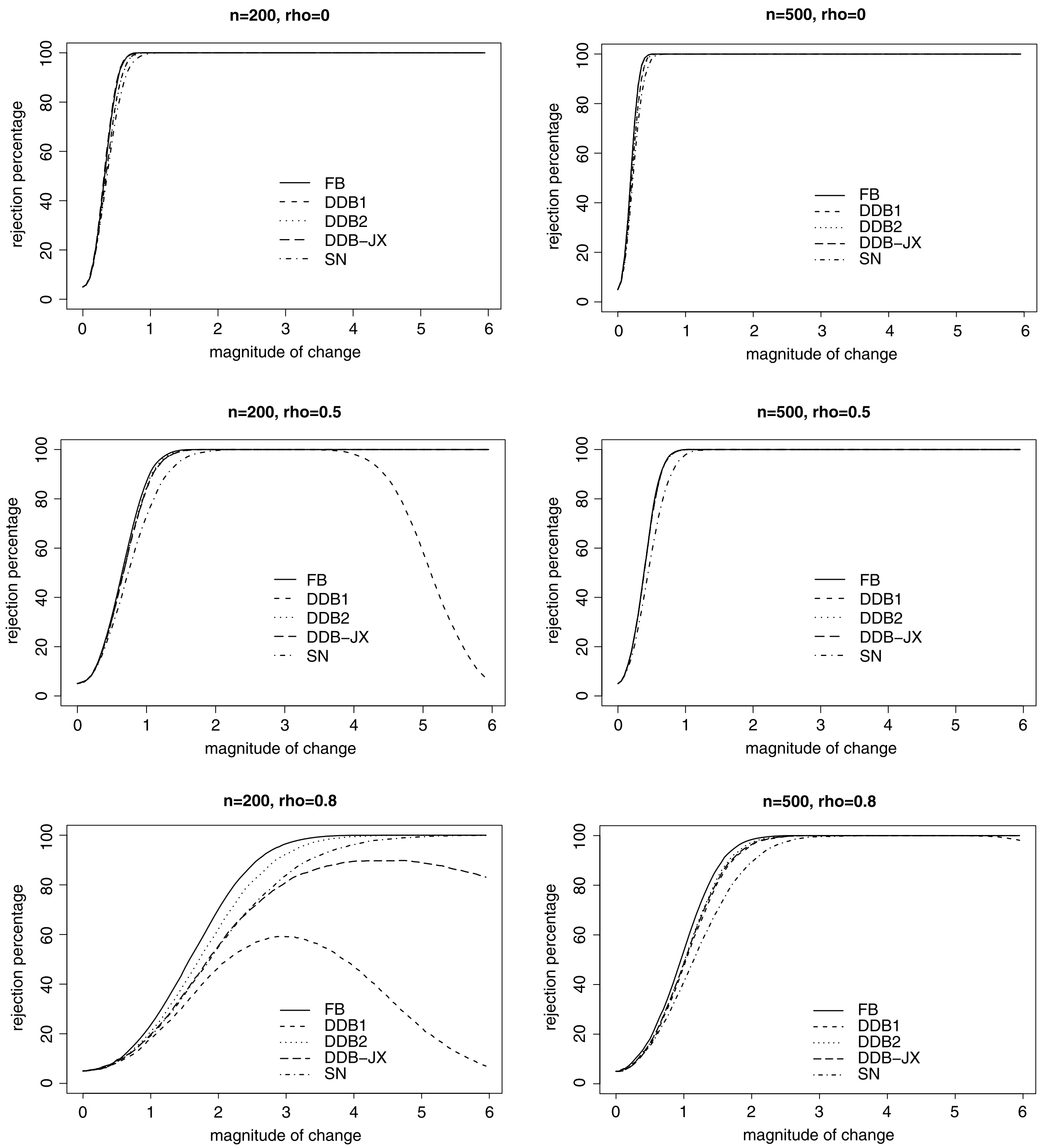

Figure 2. Size-adjusted power curve for the five methods used in detecting a change point in mean for the AR(1) models with $\rho = 0, 0.5, 0.8$. Sample size $n = 200$ (left panel) and 500 (right panel).

ing to observe that the SN-based test outperforms all the other methods in power when $\eta \in [5, 30]$, which suggests that the SN-based test is robust to the heavy tail innovations in the data generating process. When $n = 500$, the power for SS_4 appreciates a lot but SS_8 is still substantially inferior to other tests in power; other patterns are qualitatively similar to the case $n = 200$. Overall, the SN-based test has monotonic power and reasonable size and power performance in testing for a change point in median.

4.3 Change Point in the Second Order Property

In this subsection, we examine the size and power for tests that detect a change in the second order property of a time series. Consider the AR(1) model with a shift in the AR(1) coefficient:

$$X_t = \begin{cases} \rho_1 X_{t-1} + \epsilon_t, & 1 \le t \le n/2 \\ \rho_2 X_{t-1} + \epsilon_t, & n/2 + 1 \le t \le n, \end{cases}$$

where $\epsilon_t \sim \text{iid} N(0, 1)$. Under the null, $\rho_1 = \rho_2 = 0, 0.5, 0.8$. Under the alternatives, $(\rho_1, \rho_2) = (0, 0.5)$, $(0.5, 0.8)$, and $(0, 0.8)$. We apply three types of test statistics: (a) test statistic that aims to detect a change in $\rho(1)$; (b) test statistic that targets a change in $F(\pi/2)$; (c) test statistic that detects a change in $F(\pi/2)/F(\pi)$. Note that under the alternatives, there is a change in all three quantities. Table 4 shows the empirical sizes for the SN-based test and the traditional Kolmogorov–Smirnov

Table 3. Empirical sizes (in percentage) for the SN-based test statistic and $KS_{n,med}$ in testing for a change in median

n	Model	$\alpha\%$	SN	$SS_{0.5}$	SS_1	SS_2	SS_4	SS_8
200	M_1	10%	14.8	69.9	34.1	16.2	12.3	11.9
		5%	9.0	59.4	22.9	9.1	6.4	7.0
	M_2	10%	14.9	65.4	29.7	15.8	12.6	11.9
		5%	9.6	52.6	19.9	8.9	6.2	7.2
	M_3	10%	15.7	4.1	1.4	3.7	11.4	18.9
		5%	10.5	2.2	0.7	2.0	8.0	13.9
500	M_1	10%	12.9	56.7	27.4	15.4	9.3	12.0
		5%	7.4	43.1	17.9	8.4	5.2	5.9
	M_2	10%	12.5	48.8	23.6	14.4	8.9	12.8
		5%	7.5	35.9	15.0	7.4	4.2	6.8
	M_3	10%	13.9	0.2	0.3	2.1	6.2	14.9
		5%	8.7	0.0	0.2	1.4	4.3	10.0

NOTE: In the table, SN denotes the SN-based test and SS_c stands for the size of the subsampling-based test statistic with the subsampling width being $cn^{1/3}$. The largest standard error is 0.71%.

test, where the asymptotic variance is consistently estimated by the nonoverlapping subsampling method. For all three quantities, the subsampling-based test has a noticeable size distortion for all $\rho_1 = \rho_2$ and both sample sizes. The general pattern is that the size can decrease all the way from $c = 1$ to $c = 8$, or it can decrease as c gets larger and then increase after it reaches the lowest value. The size for the SN-based test is quite satisfactory, especially for $\rho_1 = \rho_2 = 0, 0.5$. When $\rho_1 = \rho_2 = 0.8$, the size distortion is apparent at $n = 200$, but it improves at a larger sample size $n = 500$. Table 5 presents the size-corrected power for the three types of tests. As seen from the power for the SN-based test, the test for a change in $\rho(1)$ $[F(\pi/2)]$ is most (least) powerful among the three. A comparison of the SN-based test and the subsampling-based test again suggests that there is a loss of power associated with the SN-based test. It is also worth noting that the subsampling width that yields the highest power corresponds to the largest size distortion. In other words, there is a tradeoff between size distortion and power loss. The SN-based test delivers better size but less power, a finding that has also been reported in other testing contexts; see Lobato (2001) and Shao (2010) for more discussions.

5. EMPIRICAL ILLUSTRATIONS

In this section, we apply our new tests to two real datasets. We first consider the GNP (Gross National Product) dataset, as analyzed in Shumway and Stoffer (2006), page 144. The data are U.S. quarterly U.S. GNP in billions of chained 1996 dollars from 1947(1) to 2002(3) and they have been seasonally adjusted. Following Shumway and Stoffer (2006), we look at the difference of the logarithm of the GNP, which is naturally interpreted as the growth rate of GNP. Although it was stated in Shumway and Stoffer (2006) that "the growth rate appears to be a stable process," Figure 4 shows that there might be a structural break in the variability of the data, with less variability for the data after year 1985. Applying our SN-based test to test for a possible change in the marginal variance, 75% quantile and 25% quantile, respectively, we tabulate the values of our test statistics and their corresponding p-values in Table 6. The test results suggest that there is a change point in the 75% quantile of the series, and thus provide significant evidence against the hypothesis that the series is (strictly) stationary. In Shumway and Stoffer (2006), stationary time series models, such as AR(1) and MA(2), have been used to fit the data. Although both model fits pass the diagnostic checking tests, our results indicate that it might be beneficial to consider a one change point model. Further modeling is beyond the scope of this paper.

Second, we apply our test to detect a change in the mean of Argentina rainfall data, as used in Wu, Woodroofe, and Mentz (2001); see Figure 5. The rainfall data contains yearly rainfall in millimeters in Argentina from 1884 to 1996. The latter authors proposed a test statistic based on isotonic regression. Note that a consistent long-run variance estimate is involved in their procedure and the choice of the truncation lag was based on a visual inspection of the autocorrelation plot of the residuals. Here we apply the SN-based test to test for a change point in mean and our test statistic takes the value 30.5, corresponding to a p-value of about 0.1. Thus it provides some evidence against the constant mean hypothesis, although not significant at the usual 5% significance level. As mentioned in Wu, Woodroofe, and Mentz (2001), the data provider believes that there is a change in the mean, which corresponds to the construction of a dam during 1952–1962. If one has prior beliefs about possible location of the change point, then we can incorporate the beliefs into our SN-based CUSUM test to enhance the power. Specifically, we define

$$G_n(\tau_1, \tau_2) = \sup_{k=\lfloor \tau_1 n\rfloor,\ldots,\lfloor \tau_2 n\rfloor} T_n(k)' V_n^{-1}(k) T_n(k), \quad \text{where } 0 \le \tau_1 < \tau_2 \le 1,$$

and the limiting null distribution for $G_n(\tau_1, \tau_2)$ is

$$G(1; \tau_1, \tau_2) = \sup_{r \in [\tau_1, \tau_2]} \{B(r) - rB(1)\}' V^{-1}(r) \{B(r) - rB(1)\}.$$

We choose $(\tau_1, \tau_2) = (0.6, 0.7)$ as this corresponds to the period 1952–1962. The critical values for $G(1; \tau_1, \tau_2)$ are tabulated in Table 7. In this case, the p-value is in the range $(0.025, 0.05)$, thus we reject the constant mean hypothesis at the 5% significance level. Our conclusion is consistent with that reached in Wu, Woodroofe, and Mentz (2001).

In addition, we also apply the SN-based test to see if there is a change in median. It turns out that our test statistic takes value 155.4 and the p-value is smaller than 0.001, thus providing strong evidence for a change in the marginal median. The finding is quite interesting as median and mean are both location parameters. Since the test statistic for a change in mean may be susceptible to outliers in the data, the test for a change in median could be used as a useful alternative for the change of the center of the marginal distribution.

6. CONCLUSIONS

In this article, we propose a new class of test statistics to test for a change point in time series. The appealing features of our SN-based test can be summarized as follows: (a) The test statistic does not involve any user-chosen number or smoothing parameter, and the asymptotic null distribution is nuisance parameter free. (b) The implementation is rather straightforward as the test statistic involves only forward and backward

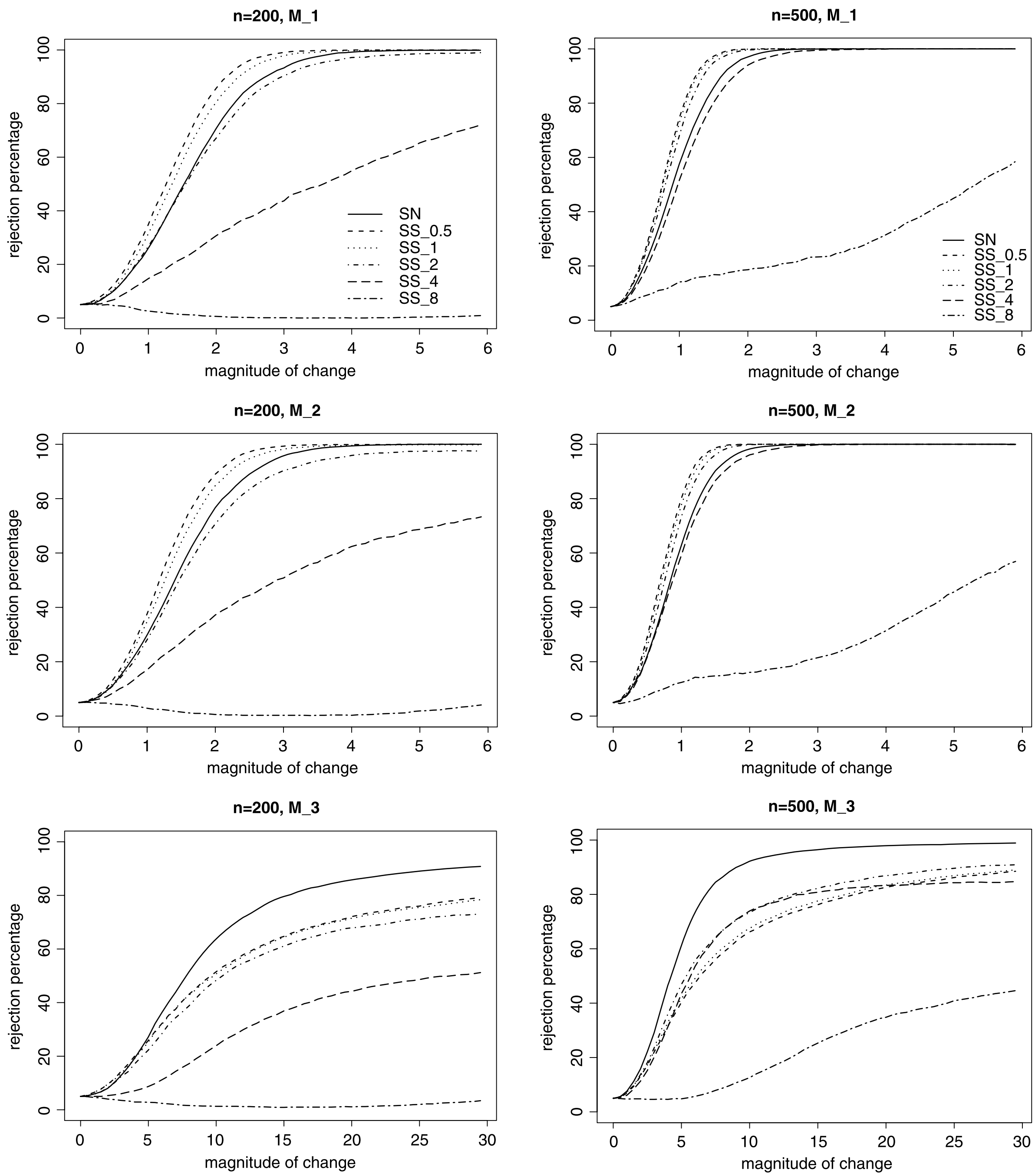

Figure 3. Size-adjusted power curve in detecting a change point in median for the models M_1, M_2, and M_3. Here SN denotes the SN-based method and SS_c denotes the subsampling method with the subsampling width being $cn^{1/3}$. Sample size $n = 200$ (left panel) and 500 (right panel).

recursive estimates. Approximate upper critical values for the limiting null distribution $G(q)$, $q = 1, \ldots, 10$, are provided in Table 1 by means of simulations. The SN-based test has wide applicability because it can be used to test for a change point in the marginal mean, marginal quantile, autocorrelation at certain lags, and second order spectrum etc. Additionally, we are able to treat the change-point detection problem in the above-mentioned quantities in a unified fashion. This feature brings convenience to the user, who intends to check the stability of their data from various aspects. (c) The finite sample performance is encouraging. Compared to the existing approaches, the SN-based test has better size but less power, which is consistent with early findings by Lobato (2001) and Shao (2010) in other contexts. The power loss is moderate and the power is monotonic as demonstrated in simulation studies. On the basis of the above nice characteristics, our test can be recommended to practitioners as a useful inference tool for routine use.

In summary, our SN-based test provides a unified treatment and a new perspective to the large literature of change-point detection in time series. The treatment here is restricted to univariate time series, and we expect that an extension to multivariate setting is possible; see Aue et al. (2009) for a recent work. Fur-

Table 4. Empirical sizes (in percentage) for the SN-based test statistics and the subsampling-based test statistics in testing for a change point in (a) $\rho(1)$, (b) $F(\pi/2)$, and (c) $F(\pi/2)/F(\pi)$

n	$\rho_1 = \rho_2$	$\alpha\%$		SN	SS_1	SS_2	SS_4	SS_8
200	0	10%	(a)	11.3	44.3	16.7	13.6	12.9
		5%	(a)	6.4	33.0	8.8	7.2	8.1
	0.5	10%	(a)	11.7	38.7	9.8	19.4	19.4
		5%	(a)	6.9	29.0	5.9	12.4	13.9
	0.8	10%	(a)	15.1	39.3	8.2	28.3	28.1
		5%	(a)	9.6	30.0	4.7	21.5	21.7
500	0	10%	(a)	10.6	30.6	15.0	8.4	12.2
		5%	(a)	6.0	19.8	8.5	4.1	5.8
	0.5	10%	(a)	11.9	20.4	11.3	4.5	16.3
		5%	(a)	6.7	13.2	6.8	2.3	9.3
	0.8	10%	(a)	13.5	13.9	8.5	2.0	26.6
		5%	(a)	8.3	8.8	5.0	0.9	18.3
200	0	10%	(b)	9.6	43.7	18.2	13.2	10.9
		5%	(b)	5.1	32.2	10.0	7.0	6.4
	0.5	10%	(b)	10.1	82.9	35.2	18.6	13.0
		5%	(b)	5.4	75.2	25.2	11.9	8.4
	0.8	10%	(b)	13.6	99.7	81.5	45.1	22.4
		5%	(b)	8.3	99.4	74.1	36.3	17.2
500	0	10%	(b)	10.3	30.8	15.1	7.9	12.4
		5%	(b)	5.2	20.2	8.4	2.9	5.2
	0.5	10%	(b)	10.0	66.2	28.2	10.6	13.1
		5%	(b)	5.0	53.6	18.1	5.4	6.9
	0.8	10%	(b)	10.7	99.4	75.5	33.6	21.6
		5%	(b)	6.1	98.8	66.6	24.0	13.2
200	0	10%	(c)	11.4	42.0	16.5	13.8	12.6
		5%	(c)	6.3	29.8	9.6	7.7	7.8
	0.5	10%	(c)	11.6	19.7	5.6	16.6	19.7
		5%	(c)	6.9	13.1	2.8	11.3	13.9
	0.8	10%	(c)	16.4	9.1	1.5	25.4	32.1
		5%	(c)	10.1	4.9	0.6	18.7	26.5
500	0	10%	(c)	11.5	29.2	15.8	9.0	12.2
		5%	(c)	6.2	19.1	8.5	4.1	5.9
	0.5	10%	(c)	11.8	8.1	6.2	4.1	15.6
		5%	(c)	7.0	4.4	3.4	2.2	9.0
	0.8	10%	(c)	15.5	1.4	2.6	1.7	28.5
		5%	(c)	9.4	0.7	1.3	0.9	20.5

NOTE: In the table, SN stands for the SN-based test and SS_c denotes the subsampling-based method with the subsampling window width being $cn^{1/3}$. The largest standard error is 0.71%.

thermore, we anticipate that the SN-based change point test can be extended to test for a change point in the parameter vector of a time series regression model or regression model with dependent errors. Further research along these directions are well underway.

APPENDIX

Proof of Theorem 2.1

Note that

$$\frac{k(n-k)}{n^2}\left\{k^{-1}\sum_{t=1}^{k}X_t-(n-k)^{-1}\sum_{t=k+1}^{n}X_t\right\}=\frac{1}{n}\left(\sum_{t=1}^{k}X_t-\frac{k}{n}\sum_{t=1}^{n}X_t\right).$$

Table 5. Size-corrected power (in percentage) for the SN-based test statistics and the subsampling-based test statistics in testing for a change point in (a) $\rho(1)$, (b) $F(\pi/2)$, and (c) $F(\pi/2)/F(\pi)$

n	(ρ_1, ρ_2)		SN	SS_1	SS_2	SS_4	SS_8
200	(0, 0.5)	(a)	78.4	91.9	87.0	74.7	10.1
		(b)	40.6	73.8	60.4	34.2	7.9
		(c)	73.2	80.6	74.8	63.1	11.0
	(0.5, 0.8)	(a)	57.7	69.3	66.0	61.3	22.3
		(b)	32.2	78.2	68.5	45.1	19.8
		(c)	51.6	47.2	47.1	47.0	23.0
	(0, 0.8)	(a)	99.0	100	100	99.7	63.6
		(b)	60.3	99.8	98.2	76.6	22.4
		(c)	98.7	99.8	99.6	98.8	61.7
500	(0, 0.5)	(a)	98.5	100	99.9	99.1	84.5
		(b)	74.6	97.2	93.7	86.9	27.7
		(c)	96.8	99.5	99.5	98.1	76.5
	(0.5, 0.8)	(a)	91.8	97.9	97.3	90.1	84.1
		(b)	61.0	97.5	94.1	86.5	36.5
		(c)	87.6	90.5	91.4	75.8	76.2
	(0, 0.8)	(a)	100	100	100	100	100
		(b)	86.0	100	100	99.6	55.3
		(c)	100	100	100	100	100

NOTE: In the table, SN stands for the SN-based test and SS_c denotes the subsampling-based method with the subsampling window width being $cn^{1/3}$. Nominal level is 5%. The largest standard error is 0.71%.

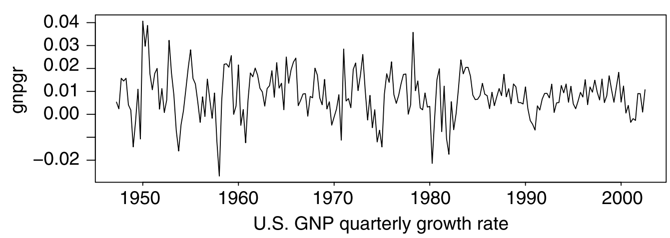

Figure 4. Quarterly U.S. GNP growth rate from 1947(1) to 2002(3).

Table 6. Test statistics and their p-values for quarterly U.S. GNP data

Parameter	Variance	75% quantile	25% quantile	(25% quantile, 75% quantile)
Test statistic	28.7	248.1	14.5	322.4
Range of p-value	(0.1, 1)	(0, 0.001)	(0.1, 1)	(0, 0.001)

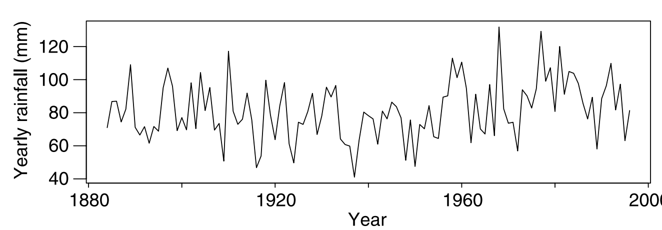

Figure 5. Argentina rainfall data: yearly rainfall (milimeters) in Argentina from 1884 to 1996.

Table 7. Simulated critical values of $G(1; \tau_1, \tau_2)$ for $(\tau_1, \tau_2) = (0.6, 0.7)$ based on $n = 5000$ and 10,000 replications

$\alpha\%$	90%	95%	97.5%	99%	99.5%	99.9%
Critical values	16.2	23.7	32.2	45.1	55.9	84.2

So $T_n(k) = \frac{k(n-k)}{n^{3/2}}\{k^{-1}\sum_{t=1}^{k} X_t - (n-k)^{-1}\sum_{t=k+1}^{n} X_t\}$. Under the alternative, we have

$$T_n(k^*) = \frac{k^*(n-k^*)}{n^{3/2}}\left\{(k^*)^{-1}\sum_{t=1}^{k^*}\{X_t - \mathbb{E}(X_t)\} - (n-k^*)^{-1}\sum_{t=k^*+1}^{n}\{X_t - \mathbb{E}(X_t)\} - \Delta_n\right\}.$$

On the other hand, it is not hard to see that $V_n(k^*) \to_D V(\lambda)$. Therefore, if Δ_n is a nonzero fixed constant, we have that

$$G_n \geq T_n(k^*)' V_n(k^*)^{-1} T_n(k^*) \to \infty \quad \text{in probability.}$$

When $\Delta_n = n^{-1/2}L$, we have $G_n \geq T_n(k^*)' V_n(k^*)^{-1} T_n(k^*)$, which converges in distribution to

$$\{-L\lambda(1-\lambda) + B(\lambda) - \lambda B(1)\}' V(\lambda)^{-1}\{-L\lambda(1-\lambda) + B(\lambda) - \lambda B(1)\}.$$

So as $|L| \to \infty$, the above limit diverges to ∞. The conclusion follows.

Proof of Theorem 3.1

With (5), we can derive that for $t = 1, \ldots, k$,

$$t(\hat{\boldsymbol{\theta}}_{1,t} - \hat{\boldsymbol{\theta}}_{1,k}) = \left\{\sum_{j=1}^{t} \mathbf{IF}(\mathbf{Y}_j; \mathbf{F}^m) - (t/k)\sum_{j=1}^{k} \mathbf{IF}(\mathbf{Y}_j; \mathbf{F}^m)\right\} + \left\{t\mathbf{R}_{1,t} - \frac{t}{k}k\mathbf{R}_{1,k}\right\} \quad (6)$$

and for $t = k+1, \ldots, N$,

$$(N-t+1)(\hat{\boldsymbol{\theta}}_{t,N} - \hat{\boldsymbol{\theta}}_{k+1,N}) = \left\{\sum_{j=t}^{N} \mathbf{IF}(\mathbf{Y}_j; \mathbf{F}^m) - \frac{(N-t+1)}{(N-k)}\sum_{j=k+1}^{N} \mathbf{IF}(\mathbf{Y}_j; \mathbf{F}^m)\right\} + \left\{(N-t+1)\mathbf{R}_{t,N} - \frac{(N-t+1)}{(N-k)}(N-k)\mathbf{R}_{k+1,N}\right\}. \quad (7)$$

Under Assumption 3.2, we can show that the terms in the second curly brackets of the Equations (6) and (7) are uniformly negligible, which, along with Assumption 3.1, implies the joint convergence of $\mathbf{T}_n(\lfloor rN\rfloor) \Rightarrow \boldsymbol{\Delta}\{\mathbf{B}_q(r) - r\mathbf{B}_q(1)\}$ and $\mathbf{V}_n(\lfloor rN\rfloor) \Rightarrow \boldsymbol{\Delta}\mathbf{V}_q(r)\boldsymbol{\Delta}'$. Then the conclusion is a direct consequence of the continuous mapping theorem.

[Received February 2010. Revised May 2010.]

REFERENCES

Altissimo, F., and Corradi, V. (2003), "Strong Rules for Detecting the Number of Breaks in a Time Series," *Journal of Econometrics*, 117, 207–244. [1228]

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [1233]

——— (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [1228]

Antoch, J., Hušková, M., and Prášková, Z. (1997), "Effect of Dependence on Statistics for Determination of Change," *Journal of Statistical Planning and Inference*, 60, 291–310. [1228]

Aue, A., Hörmann, L., Horváth, L., and Reimherr, M. (2009), "Break Detection in the Covariance Structure of Multivariate Time Series Models," *The Annals of Statistics*, 37, 4046–4087. [1237]

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2008), "Testing for Changes in Polynomial Regression," *Bernoulli*, 14, 637–660. [1228]

Bai, J. (1994), "Least Square Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472. [1230,1233]

——— (1997), "Estimation of a Change Point in Multiple Regressions," *Review of Economics and Statistics*, 79, 551–563. [1230]

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78. [1231]

Berkes, I., Gombay, E., and Horváth, L. (2009), "Testing for Changes in the Covariance Structure of Linear Processes," *Journal of Statistical Planning and Inference*, 139, 2044–2063. [1228]

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley. [1229]

Brodsky, B. E., and Darkhovsky, B. S. (1993), *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications*, Vol. 243, Dordrecht: Kluwer. [1228]

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Sequence," *The Annals of Statistics*, 14, 1171–1179. [1234]

Crainiceanu, C. M., and Vogelsang, T. J. (2007), "Spectral Density Bandwidth Choice: Source of Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77, 457–476. [1229,1233,1234]

Csörgő, M., and Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, New York: Wiley. [1228]

Davis, R. A., Huang, D., and Yao, Y.-C. (1995), "Testing for a Change in the Parameter Values and Order of an Autoregressive Model," *The Annals of Statistics*, 23, 282–304. [1228]

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006), "Structural Break Estimation for Nonstationary Time Series Models," *Journal of the American Statistical Association*, 101, 223–239. [1231]

Deng, A., and Perron, P. (2008), "A Non-Local Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change," *Journal of Econometrics*, 142, 212–240. [1229]

Galeano, P., and Peña, D. (2007), "Covariance Changes Detection in Multivariate Time Series," *Journal of Statistical Planning and Inference*, 137, 194–211. [1228]

Giraitis, L., and Leipus, R. (1992), "Testing and Estimating in the Change-Point Problem for the Spectral Function," *Lithuanian Mathematical Journal*, 32, 20–38. [1228,1233]

Giraitis, L., Leipus, R., and Surgailis, D. (1996), "The Change-Point Problem for Dependent Observations," *Journal of Statistical Planning and Inference*, 53, 297–310. [1228]

Gombay, E. (2008), "Change Detection in Autoregressive Time Series," *Journal of Multivariate Analysis*, 99, 451–464. [1228]

Gombay, E., Horváth, L., and Huskova, M. (1996), "Estimators and Tests for Change in Variances," *Statistics and Decisions*, 14, 145–159. [1228]

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.

Horváth, L., Kokoszka, P., and Steinebach, J. (1999), "Testing for Changes in Multivariate Dependent Observations With an Application to Temperature Changes," *Journal of Multivariate Analysis*, 68, 96–119. [1228]

Inclán, C., and Tiao, G. C. (1994), "Use of Cumulative Sums of Squares for Retrospective Detection of Change of Variance," *Journal of the American Statistical Association*, 89, 913–923. [1228]

Inoue, A. (2001), "Testing for Distributional Change in Time Series," *Econometric Theory*, 17, 156–187. [1228]

Juhl, T., and Xiao, Z. (2009), "Testing for Changing Mean Monotonic Power," *Journal of Econometrics*, 148, 14–24. [1228,1229,1233,1234]

Kiefer, N. M., and Vogelsang, T. J. (2002), "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093–2095. [1230]

——— (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164. [1230]

Kokoszka, P., and Leipus, R. (2002), "Detection and Estimation of Changes in Regime," in *Long-Range Dependence: Theory and Applications*, eds. P. Doukhan, G. Oppenheim, and M. S. Taqqu, Boston: Birkhauser, pp. 325–337. [1229]

Lavielle, M., and Ludena, C. (2000), "The Multiple Change-Points Problem for the Spectral Distribution," *Bernoulli*, 6, 845–869. [1228,1233]

Lee, S., and Park, S. (2001), "The Cusum of Squares Test for Scale Changes in Infinite Order Moving Average Processes," *Scandinavian Journal of Statistics*, 28, 625–644. [1228]

Lee, S., Ha, J., and Na, O. (2003), "The Cusum Test for Parameter Change in Time Series Models," *Scandinavian Journal of Statistics*, 30, 781–796. [1228]

Ling, S. (2007), "Testing for Change Points in Time Series Models and Limiting Theorems for NED Sequences," *The Annals of Statistics*, 35, 1213–1237. [1228]

Lobato, I. N. (2001), "Testing That a Dependent Process Is Uncorrelated," *Journal of the American Statistical Association*, 96, 1066–1076. [1228-1230, 1236,1237]

Perron, P. (2006), "Dealing With Structural Breaks," in *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, eds. K. Patterson and T. C. Mills, Great Britain: Palgrave Macmillan, pp. 278–352. [1228,1229]

Phillips, P. C. B. (1987), "Time Series Regression With Unit Roots," *Econometrica*, 55, 277–301. [1229,1232]

Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 17, 841–867. [1228,1233]

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, New York: Springer-Verlag. [1234]

Qu, Z., and Perron, P. (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, 75, 459–502. [1228,1231]

Shao, X. (2010), "A Self-Normalized Approach to Confidence Interval Construction in Time Series," *Journal of the Royal Statistical Society, Ser. B*, 72, 343–366. [1228-1231,1233,1236,1237]

Shumway, R. H., and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications: With R Examples*, New York: Springer. [1236]

Tang, S. M., and MacNeill, I. B. (1993), "The Effect of Serial Correlation on Tests for Parameter Change at Unknown Time," *The Annals of Statistics*, 21, 552–575. [1228]

Vogelsang, T. J. (1998), "Testing for a Shift in Mean Without Having to Estimate Serial-Correlation Parameters," *Journal of Business & Economic Statistics*, 16, 73–80. [1228]

——— (1999), "Sources of Nonmonotonic Power When Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88, 283–299. [1228,1229]

Wu, W. B., Woodroofe, M., and Mentz, G. (2001), "Isotonic Regression: Another Look at the Changepoint Problem," *Biometrika*, 88, 793–804. [1236]

Yao, Y.-C. (1988), "Estimating the Number of Change-Points via Schwarz' Criterion," *Statistics & Probability Letters*, 6, 181–189. [1231]

Zhang, N. R., and Siegmund, D. O. (2007), "A Modified Bayes Information Criterion With Applications to the Analysis of Comparative Genomic Hybridization Data," *Biometrics*, 63, 22–32. [1231]