Bernoulli **19**(1), 2013, 205–227
DOI: 10.3150/11-BEJ399

Inference for modulated stationary processes

ZHIBIAO ZHAO* and XIAOYE LI**

Department of Statistics, Penn State University, University Park, PA 16802, USA.
*E-mail: *zuz13@stat.psu.edu; **xul117@stat.psu.edu*

We study statistical inferences for a class of modulated stationary processes with time-dependent variances. Due to non-stationarity and the large number of unknown parameters, existing methods for stationary, or locally stationary, time series are not applicable. Based on a self-normalization technique, we address several inference problems, including a self-normalized central limit theorem, a self-normalized cumulative sum test for the change-point problem, a long-run variance estimation through blockwise self-normalization, and a self-normalization-based wild bootstrap. Monte Carlo simulation studies show that the proposed self-normalization-based methods outperform stationarity-based alternatives. We demonstrate the proposed methodology using two real data sets: annual mean precipitation rates in Seoul from 1771–2000, and quarterly U.S. Gross National Product growth rates from 1947–2002.

Keywords: change-point analysis; confidence interval; long-run variance; modulated stationary process; self-normalization; strong invariance principle; wild bootstrap

1. Introduction

In time series analysis, stationarity requires that dependence structure be sustained over time, and thus we can borrow information from one time period to study model dynamics over another period; see Fan and Yao [20] for nonparametric treatments and Lahiri [29] for various resampling and block bootstrap methods. In practice, however, many climatic, economic and financial time series are non-stationary and therefore challenging to analyze. First, since dependence structure varies over time, information is more localized. Second, non-stationary processes often require extra parameters to account for time-varying structure. One way to overcome these issues is to impose certain local stationarity; see, for example, Dahlhaus [15] and Adak [1] for spectral representation frameworks and Dahlhaus and Polonik [16] for a time domain approach.

In this article we study a class of modulated stationary processes (see Adak [1])

$$X_i = \mu + \sigma_i e_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where e_i are stationary time series with zero mean, and $\sigma_i > 0$ are unknown constants adjusting for time-dependent variances. Then X_i oscillates around the constant mean μ, whereas its variance changes over time in an unknown manner. In the special case of $\sigma_i \equiv 1$, (1.1) reduces to stationary case. If $\sigma_i = s(i/n)$ for a Lipschitz continuous function $s(t)$ on [0, 1], then (1.1) is locally stationary. For the general non-stationary case (1.1), the number of unknown parameters is larger than the number of observations, and it is infeasible to estimate σ_i. Due to non-stationarity and the large number of unknown parameters, existing methods that are developed for (locally) stationary processes are not applicable, and our main purpose is to develop new statistical inference techniques.

First, we establish a uniform strong approximation result which can be used to derive a self-normalized central limit theorem (CLT) for the sample mean $\bar{X}$ of (1.1). For stationary case $\sigma_i \equiv 1$, by Fan and Yao [20], under mild mixing conditions,

$$\sqrt{n}(\bar{X}-\mu) \Rightarrow N(0,\tau^2), \qquad \text{where } \tau^2 = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k \text{ and } \gamma_k = \mathrm{Cov}(e_i, e_{i+k}). \tag{1.2}$$

For the modulated stationary case (1.1), it is non-trivial whether $\sqrt{n}(\bar{X}-\mu)$ has a CLT without imposing further assumptions on σ_i and the dependence structure of e_i. Moreover, even when the latter CLT exists, it is difficult to estimate the limiting variance due to the large number of unknown parameters; see De Jong and Davidson [18] for related work assuming a near-epoch dependent mixing framework. Zhao [41] studied confidence interval construction for μ in (1.1) by assuming a block-wise asymptotically equal cumulative variance assumption. The latter assumption is rather restrictive and essentially requires that block averages be asymptotically independent and identically distributed (i.i.d.). In this article, we deal with the more general setting (1.1). Under a strong invariance principle assumption, we establish a self-normalized CLT with the self-normalizing constant adjusting for time-dependent non-stationarity. The obtained CLT is an extension of the classical CLT for i.i.d. data or stationary time series to modulated stationary processes. Furthermore, we extend the idea to linear combinations of means over different time periods, which allows us to address inference regarding mean levels over multiple time periods.

Second, we study the wild bootstrap for modulated stationary processes. Since the seminal work of Efron [19], a great deal of research has been done on the bootstrap under various settings, ranging from bootstrapping for i.i.d. data in Efron [19], wild bootstrapping for independent observations with possibly non-constant variances in Wu [39] and Liu [30], to various block bootstrapping and resampling methods for stationary time series in Künsch [27], Politis and Romano [34], Bühlmann [12] and the monograph Lahiri [29]. With the established self-normalized CLT, we propose a wild bootstrap procedure that is tailored to deal with modulated stationary processes: the dependence is removed through a scaling factor, and the non-constant variance structure of the original data is preserved in the wild bootstrap data-generating mechanism. Our simulation study shows that the wild bootstrap method outperforms the widely used stationarity-based block bootstrap.

Third, we address change-point analysis. The change-point problem has been an active area of research; see Pettitt [32] for proportion changes in binary data, Horváth [25] for mean and variance changes in Gaussian observations, Bai and Perron [8] for coefficient changes in linear models, Aue *et al.* [6] for coefficient changes in polynomial regression with uncorrelated errors, Aue *et al.* [7] for mean change in time series with stationary errors, Shao and Zhang [37] for change-points for stationary time series and the monograph by Csörgő and Horváth [14] for more discussion. Most of these works deal with stationary and/or independent data. Hansen [24] studied tests for constancy of parameters in linear regression models with non-stationary regressors and conditionally homoscedastic martingale difference errors. Here we consider

$$H_0\colon\ X_i = \mu_i + \sigma_i e_i,\ \mu_1 = \cdots = \mu_n, \qquad H_a\colon\ \mu_1 = \cdots = \mu_J \neq \mu_{J+1} = \cdots = \mu_n, \tag{1.3}$$

where J is an unknown change point. The aforementioned works mainly focused on detecting changes in mean while the error variance is constant. On the other hand, researchers have also realized the importance of the variance/covariance structure in change point analysis. For example, Inclán and Tiao [26] studied change in variance for independent data, and Aue *et al.* [5] and Berkes, Gombay and Horváth [10] considered change in covariance for time series data. To our knowledge, there has been almost no attempt to advance change point analysis under the non-constant variances framework in (1.3). Andrews [4] studied change point problem under near-epoch dependence structure that allows for non-stationary processes, but his Assumption 1(c) on page 830 therein essentially implies that the process has constant variance. The popular cumulative sum (CUSUM) test is developed for stationary time series and does not take into account the time-dependent variances. Using the self-normalization idea, we propose a self-normalized CUSUM test and a wild bootstrap method to obtain its critical value. Our empirical studies show that the usual CUSUM tests tend to over-reject the null hypothesis in the presence of non-constant variances. By contrast, the self-normalized CUSUM test yields size close to the nominal level.

Fourth, we estimate the long-run variance τ^2 in (1.2). Long-run variance plays an essential role in statistical inferences involving time series. Most works in the literature deal with stationary processes through various block bootstrap and subsampling approaches; see Carlstein [13], Künsch [27], Politis and Romano [34], Götze and Künsch [21] and the monograph Lahiri [29]. De Jong and Davidson [18] established the consistency of kernel estimators of covariance matrices under a near epoch dependent mixing condition. Recently, Müller [31] studied robust long-run variance estimation for locally stationary process. For model (1.1), the error process $\{e_i\}$ is contaminated with unknown standard deviations $\{\sigma_i\}$, and we apply blockwise self-normalization to remove non-stationarity, resulting in asymptotically stationary blocks.

Fifth, the proposed methods can be extended to deal with the linear regression model

$$X_i = U_i\beta + \sigma_i e_i, \tag{1.4}$$

where $U_i = (u_{i,1}, \dots, u_{i,p})$ are deterministic covariates, and $\beta = (\beta_1, \dots, \beta_p)'$ is the unknown column vector of parameters. For $p = 2$, Hansen [23] established the asymptotic normality of the least-squares estimate of the slope parameter under a fairly general framework of non-stationary errors. While Hansen [23] assumed that the errors form a martingale difference array so that they are uncorrelated, the framework in (1.4) is more general in that it allows for correlations. On the other hand, Hansen [23] allowed the conditional volatilities to follow an autoregressive model, hence introducing stochastic volatilities. Phillips, Sun and Jin [33] considered (1.4) for stationary errors, and their approach is not applicable here due to the unknown non-constant variances σ_i^2. In Section 2.6 we consider self-normalized CLT for the least-squares estimator of β in (1.4). In the polynomial regression case $u_{i,r} = (i/n)^{r-1}$, Aue *et al.* [6] studied a likelihood-based test for constancy of β in (1.4) for uncorrelated errors with constant variance. Due to the presence of correlation and time-varying variances, it is more challenging to study the change point problem for (1.4) and this is beyond the scope of this article.

The rest of this article is organized as follows. We present theoretical results in Section 2. Sections 3–4 contain Monte Carlo studies and applications to two real data sets.

2. Main results

For sequences $\{a_n\}$ and $\{b_n\}$, write $a_n = \mathrm{O}(b_n)$, $a_n = \mathrm{o}(b_n)$ and $a_n \asymp b_n$, respectively, if $|a_n/b_n| < c_1$, $a_n/b_n \to 0$ and $c_2 < |a_n/b_n| < c_3$, for some constants $0 < c_1, c_2, c_3 < \infty$. For $q > 0$ and a random variable e, write $e \in \mathcal{L}^q$ if $\|e\|_q := \{E(|e|^q)\}^{1/q} < \infty$.

2.1. Uniform approximations for modulated stationary processes

In (1.1), assume without loss of generality that $E(e_i) = 0$ and $E(e_i^2) = 1$ so that $\{e_i\}$ and $\{e_i^2 - 1\}$ are centered stationary processes. With the convention $S_0 = S_0^* = 0$, define

$$S_i = \sum_{j=1}^{i} e_j \quad \text{and} \quad S_i^* = \sum_{j=1}^{i} (e_j^2 - 1), \qquad i = 1, 2, \ldots. \tag{2.1}$$

Assumption 2.1. *There exist standard Brownian motions $\{B_t\}$ and $\{B_t^*\}$ such that*

$$\max_{1 \le i \le n} |S_i - \tau B_i| = \mathrm{o}_{\text{a.s.}}(\Delta_n) \quad \textit{and} \quad \max_{1 \le i \le n} |S_i^* - \tau^* B_i^*| = \mathrm{o}_{\text{a.s.}}(\Delta_n), \tag{2.2}$$

where Δ_n is the approximation error, τ^2 and τ^{*2} are the long-run variances of $\{e_i\}$ and $\{e_i^2 - 1\}$, respectively. Further assume $\tau^2 > 0$ to avoid the degenerate case $\tau^2 = 0$.

The uniform approximations in (2.2) are generally called strong invariance principle. The two Brownian motions $\{B_t\}$ and $\{B_t^*\}$ may be defined on different probability spaces and hence are not jointly distributed, which is not an issue because our argument does not depend on their joint distribution. To see how to use (2.2), under H_0 in (1.3), consider

$$F_j = j(\underline{X}_j - \mu) \quad \text{and} \quad \underline{V}_j^2 = \sum_{i=1}^{j} (X_i - \underline{X}_j)^2, \qquad \text{where } \underline{X}_j = j^{-1} \sum_{i=1}^{j} X_i. \tag{2.3}$$

Theorem 2.1 below presents uniform approximations for F_j and V_j^2. Define

$$r_n = |\sigma_n| + \sum_{i=2}^{n} |\sigma_i - \sigma_{i-1}| \quad \text{and} \quad r_n^* = |\sigma_n^2| + \sum_{i=2}^{n} |\sigma_i^2 - \sigma_{i-1}^2|, \tag{2.4}$$

$$\Sigma_j^2 = \sum_{i=1}^{j} \sigma_i^2 \quad \text{and} \quad \Sigma_j^{*2} = \left(\sum_{i=1}^{j} \sigma_i^4 \right)^{1/2}. \tag{2.5}$$

Theorem 2.1. *Let (2.2) hold. For any $c \in (0, 1]$, the following uniform approximations hold:*

$$\max_{cn \le j \le n} \left| F_j - \tau \sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) \right| = \mathrm{O}_{\text{a.s.}}(r_n \Delta_n), \tag{2.6}$$

$$\max_{cn \le j \le n} |\underline{V}_j^2 - \Sigma_j^2| = \mathrm{O_p}\{(r_n^2 \Delta_n^2 + \Sigma_n^2)/n + \Sigma_n^{*2} + r_n^* \Delta_n\}. \tag{2.7}$$

Theorem 2.1 provides quite general results under (2.2). We now discuss sufficient conditions for (2.2). Shao [36] obtained sufficient mixing conditions for (2.2). In this article, we briefly introduce the framework in Wu [40]. Assume that e_i has the causal representation $e_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$, where ε_i are i.i.d. innovations, and G is a measurable function such that e_i is well defined. Let $\{\varepsilon'_i\}_{i \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$. Assume

$$\sum_{i=1}^{\infty} i \|e_i - e'_i\|_8 < \infty, \qquad \text{where } e'_i = G(\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i). \tag{2.8}$$

Proposition 2.1 below follows from Corollary 4 in Wu [40].

Proposition 2.1. *Assume that* (2.8) *holds. Then* (2.2) *holds with* $\Delta_n = n^{1/4} \log(n)$, *the optimal rate up to a logarithm factor.*

For linear process $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$ with $\varepsilon_i \in \mathcal{L}^8$ and $E(\varepsilon_i) = 0$, $\|e_i - e'_i\|_8 = \|\varepsilon_0 - \varepsilon'_0\|_8 |a_i|$. If $\sum_{i=1}^{\infty} i|a_i| < \infty$, then (2.2) holds with $\Delta_n = n^{1/4} \log(n)$. For many nonlinear time series, $\|e_i - e'_i\|_8$ decays exponentially fast and hence (2.8) holds; see Section 3.1 of Wu [40]. From now on we assume (2.2) holds with $\Delta_n = n^{1/4} \log(n)$.

Remark 2.1. If e_i are i.i.d. with $E(e_i) = 0$ and $e_i \in \mathcal{L}^q$ for some $2 < q \le 4$, the celebrated "Hungarian embedding" asserts that $\sum_{j=1}^{i} e_j$ satisfies a strong invariance principle with the optimal rate $\mathrm{o_{a.s.}}(n^{1/q})$. Thus, it is necessary to have the moment assumption $e_i \in \mathcal{L}^8$ in Proposition 2.1 in order to ensure strong invariance principles for both S_i and S_i^* in (2.1) with approximation rate $n^{1/4} \log(n)$. On the other hand, one can relax the moment assumption by loosening the approximation rate. For example, by Corollary 4 in Wu [40], assume $e_i \in \mathcal{L}^{2q}$ for some $q > 2$ and $\sum_{i=1}^{\infty} i \|e_i - e_i^*\|_{2q} < \infty$, then (2.2) holds with $\Delta_n = n^{1/\min(q,4)} \log(n)$.

As shown in Examples 2.1–2.3 below, r_n and r_n^* in (2.4) often have tractable bounds.

Example 2.1. If σ_i is non-decreasing in i, then $\sigma_n \le r_n \le 2\sigma_n$ and $\sigma_n^2 \le r_n^* \le 2\sigma_n^2$. If σ_i is non-increasing in i, then $r_n = \sigma_1$ and $r_n^* = \sigma_1^2$. If σ_i are piecewise constants with finitely many pieces, then $r_n, r_n^* = \mathrm{O}(1)$.

Example 2.2. Let $\sigma_i = s(i/n^{\gamma})$ for $\gamma \in [0, 1]$ and a Lipschitz continuous function $s(t), t \in [0, \infty)$, $\sup_{t \in [0,\infty)} s(t) < \infty$. Then $r_n, r_n^* = \mathrm{O}(n^{1-\gamma})$. If $\gamma = 1$, we obtain a locally stationary case with the time window $i/n \in [0, 1]$; if $\gamma \in [0, 1)$, we have the infinite time window $[0, \infty)$ as $n/n^{\gamma} \to \infty$, which may be more reasonable for data with a long time horizon.

Example 2.3. If $\sigma_i = i^{\beta} L(i)$ for a slowly varying function $L(\cdot)$ such that $L(cx)/L(x) \to 1$ as $x \to \infty$ for all $c > 0$. Then we can show $r_n = \mathrm{O}\{n^{\beta} L(n)\}$ or $\mathrm{O}(1)$ and $r_n^* = \mathrm{O}\{n^{2\beta} L^2(n)\}$ or $\mathrm{O}(1)$, depending on whether $\beta > 0$ or $\beta < 0$. For the boundary case $\beta = 0$, assume

$L(i+1)/L(i) = 1 + \mathrm{O}(1/i)$ uniformly, then $r_n = L(n) + \mathrm{O}(1)\sum_{i=2}^n L(i)/i = \mathrm{O}\{\log(n) \times \max_{1\le i\le n} L(i)\}$. Similarly, $r_n^* = \mathrm{O}\{\log(n)\max_{1\le i\le n} L^2(i)\}$.

2.2. Self-normalized central limit theorem

In this section we establish a self-normalized CLT for the sample average $\bar{X}$. To understand how non-stationarity makes this problem difficult, elementary calculation shows

$$\mathrm{Var}\big\{\sqrt{n}(\bar{X}-\mu)\big\} = \frac{\gamma_0}{n}\sum_{i=1}^n \sigma_i^2 + \frac{2}{n}\sum_{1\le i<j\le n}\sigma_i\sigma_j\gamma_{j-i} := \tau_n^2, \tag{2.9}$$

where $\gamma_k = \mathrm{Cov}(e_0, e_k)$. In the stationary case $\sigma_i \equiv 1$, under condition $\sum_{k=0}^\infty |\gamma_k| < \infty$, $\tau_n^2 \to \tau^2$, the long-run variance in (1.2). For non-constant variances, it is difficult to deal with τ_n^2 directly, due to the large number of unknown parameters and complicated structure. See De Jong and Davidson [18] for a kernel estimator of τ_n^2 under a near-epoch dependent mixing framework.

To attenuate the aforementioned issue, we apply the uniform approximations in Theorem 2.1. Assume that (2.10) below holds. Note that the increments $B_i - B_{i-1}$ of standard Brownian motions are i.i.d. standard normal random variables. By (2.6), $n(\bar{X}-\mu)$ is equivalent to $N(0, \tau^2\Sigma_n^2)$ in distribution. By (2.7), $\underline{V}_n/\Sigma_n \to 1$ in probability. By Slutsky's theorem, we have Proposition 2.2.

Proposition 2.2. *Let* (2.2) *hold with* $\Delta_n = n^{1/4}\log(n)$. *For* $r_n, r_n^*, \Sigma_n^2, \Sigma_n^{*2}$ *in* (2.4)–(2.5), *assume*

$$\delta_n = r_n\Delta_n/\Sigma_n + (r_n^*\Delta_n + \Sigma_n^{*2})/\Sigma_n^2 \to 0. \tag{2.10}$$

Recall $\underline{V}_n^2$ *in* (2.3). *Then as* $n\to\infty$, $n(\bar{X}-\mu)/\underline{V}_n \Rightarrow N(0,\tau^2)$. *Consequently, a* $(1-\alpha)$ *asymptotic confidence interval for* μ *is* $\bar{X} \pm z_{\alpha/2}\hat{\tau}\underline{V}_n/n$, *where* $\hat{\tau}$ *is a consistent estimate of* τ (*Section* 2.5 *below*), *and* $z_{\alpha/2}$ *is* $(1-\alpha/2)$ *standard normal quantile.*

Proposition 2.2 is an extension of the classical CLT for i.i.d. data or stationary processes to modulated stationary processes. If X_i are i.i.d., then $n(\bar{X}-\mu)/\underline{V}_n \Rightarrow N(0,1)$. In Proposition 2.2, τ^2 can be viewed as the variance inflation factor due to the dependence of $\{e_i\}$. For stationary data, the sample variance $\underline{V}_n^2/n$ is a consistent estimate of the population variance. Here, for non-constant variances case (1.1), by (2.7) in Theorem 2.1, $\underline{V}_n^2/n$ can be viewed as an estimate of the time-average "population variance" Σ_n^2/n. So, we can interpret the CLT in Proposition 2.2 as a self-normalized CLT for modulated stationary processes with the self-normalizing term $\underline{V}_n$, adjusting for non-stationarity due to $\sigma_1,\dots,\sigma_n$ and τ^2, accounting for dependence of $\{e_i\}$. Clearly, parameters $\sigma_1,\dots,\sigma_n$ are canceled out through self-normalization. Finally, condition (2.10) is satisfied in Example 2.2 with $\gamma > 3/4$ and Example 2.3 with $\beta > -1/4$.

In classical statistics, the width of confidence intervals usually shrinks as sample size increases. By Proposition 2.2 and Theorem 2.1, the width of the constructed confidence interval for μ is proportional to $\underline{V}_n/n$ or, equivalently, Σ_n/n. Thus, a necessary and sufficient condition for shrinking confidence interval is $\sum_{i=1}^n \sigma_i^2/n^2 \to 0$, which is satisfied if $\sigma_i = \mathrm{o}(\sqrt{i})$. An intuitive explanation is as follows. For i.i.d. data, sample mean converges at a rate of $\mathrm{O}(\sqrt{n})$. In (1.1), if σ_i grows faster than $\mathrm{O}(\sqrt{i})$, the contribution of a new observation is negligible relative to its noise level.

Example 2.4. If $\sigma_i \asymp i^\beta$ with $\beta \in [0, 1/2)$, the length of confidence interval is proportional to $\Sigma_n/n \asymp n^{\beta-1/2}$. In particular, if $c_1 < \sigma_i < c_2$ for some positive constants c_1 and c_2, then Σ_n/n achieves the optimal rate $\mathrm{O}(n^{-1/2})$. If $\sigma_i \asymp \log(i)$, then $\Sigma_n/n \asymp \log(n)/\sqrt{n}$.

The same idea can be extended to linear combinations of means over multiple time periods. Suppose we have observations from k consecutive time periods $\mathcal{T}_1, \ldots, \mathcal{T}_k$, each of the form (1.1) with different means, denoted by $\mu_1, \ldots, \mu_k$, and each having time-dependent variances. Let $\nu = \beta_1\mu_1 + \cdots + \beta_k\mu_k$ for given coefficients $\beta_1, \ldots, \beta_k$. For example, if we are interested in mean change from $\mathcal{T}_1$ to $\mathcal{T}_2$, we can take $\nu = \mu_2 - \mu_1$; if we are interested in whether the increase from $\mathcal{T}_3$ to $\mathcal{T}_4$ is larger than that from $\mathcal{T}_1$ to $\mathcal{T}_2$, we can let $\nu = (\mu_4 - \mu_3) - (\mu_2 - \mu_1)$. Proposition 2.3 below extends Proposition 2.2 to multiple means.

Proposition 2.3. *Let $\nu = \beta_1\mu_1 + \cdots + \beta_k\mu_k$. For $\mathcal{T}_j$, denote its sample size n_j and its sample average $\bar{X}(j)$. Assume that (2.10) holds for each individual time period $\mathcal{T}_j$ and, for simplicity, that $n_1, \ldots, n_k$ are of the same order. Then*

$$\frac{\sum_{j=1}^k \beta_j \bar{X}(j) - \nu}{\Lambda_n} \Rightarrow N(0, \tau^2), \qquad \text{where } \Lambda_n^2 = \sum_{j=1}^k \left\{ \frac{\beta_j^2}{n_j^2} \sum_{i \in \mathcal{T}_j} [X_i - \bar{X}(j)]^2 \right\}.$$

2.3. Wild bootstrap for self-normalized statistic

Recall $\sigma_i e_i$ in (1.1). Suppose we are interested in the self-normalized statistic

$$H_n = \frac{\sum_{i=1}^n \sigma_i e_i}{\sqrt{\sum_{i=1}^n \sigma_i^2 e_i^2}}.$$

For problems with small sample sizes, it is natural to use bootstrap distribution instead of the convergence $H_n \Rightarrow N(0, \tau^2)$ in Proposition 2.2. Wu [39] and Liu [30] have pioneered the work on the wild bootstrap for independent data with non-identical distributions. We shall extend their wild bootstrap procedure to the modulated stationary process (1.1).

Let $\{\alpha_i\}$ be i.i.d. random variables independent of $\{e_i\}$ satisfying $\alpha_i \in \mathcal{L}^3$, $E(\alpha_i) = 0$, $E(\alpha_i^2) = 1$. Define the self-normalized statistic based on the following new data:

$$H_n^* = \frac{\sum_{i=1}^n \xi_i}{\sqrt{\sum_{i=1}^n (\xi_i - \bar{\xi})^2}}, \qquad \text{where } \xi_i = \sigma_i e_i \alpha_i \text{ and } \bar{\xi} = \frac{\xi_1 + \cdots + \xi_n}{n}.$$

Clearly, ξ_i inherits the non-stationarity structure of $\sigma_i e_i$ by writing $\xi_i = \sigma_i e_i^*$ with $e_i^* = e_i\alpha_i$. On the other hand, for the new error process $\{e_i^*\}$, $E(e_i^{*2}) = E(e_i^2) = 1$ and $\mathrm{Cov}(e_i^*, e_j^*) = 0$ for

$i \neq j$. Thus, $\{e_i^*\}$ is a white noise sequence with long-run variance one. By Proposition 2.2, the scaled version $H_n/\tau \Rightarrow N(0, 1)$ is robust against the dependence structure of $\{e_i\}$, so we expect that H_n^* should be close to H_n/τ in distribution.

Theorem 2.2. *Let the conditions in Proposition 2.2 hold. Further assume*

$$\left(\sum_{i=1}^{n} \sigma_i^3\right)^2 \left(\sum_{i=1}^{n} \sigma_i^2\right)^{-3} \to 0. \tag{2.11}$$

Let $\hat{\tau}$ be a consistent estimate of τ. Denote by $\mathbb{P}^*$ the conditional law given $\{e_i\}$. Then

$$\sup_{x \in \mathbb{R}} |\mathbb{P}^*(H_n^* \leq x) - \mathbb{P}(H_n/\hat{\tau} \leq x)| \to 0, \qquad \textit{in probability}. \tag{2.12}$$

Theorem 2.2 asserts that, H_n^* behaves like the scaled version $H_n/\hat{\tau}$, with the scaling factor $\hat{\tau}$ coming from the dependence of $\{e_i\}$. Here we use the sample mean $\bar{X}$ in (1.1) to illustrate a wild bootstrap procedure to obtain the distribution of $n(\bar{X} - \mu)/(\tau \underline{V}_n)$ in Proposition 2.2.

(i) Apply the method in Section 2.5 to $X_1, \ldots, X_n$ to obtain a consistent estimate $\hat{\tau}$ of τ.
(ii) Subtract the sample mean $\bar{X}$ from data to obtain $\epsilon_i = X_i - \bar{X}$, $i = 1, \ldots, n$.
(iii) Generate i.i.d. random variables $\alpha_1, \ldots, \alpha_n$ satisfying $E(\alpha_i) = 0$, $E(\alpha_i^2) = 1$.
(iv) Based on ϵ_i in (ii) and α_i in (iii), generate bootstrap data $\xi_i^b = \epsilon_i \alpha_i$, and compute

$$H_n^b = \frac{\sum_{i=1}^{n} \xi_i^b}{\hat{\tau}^b \sqrt{\sum_{i=1}^{n} (\xi_i^b - \bar{\xi}^b)^2}},$$

where $\hat{\tau}^b$ is a long-run variance estimate (see Section 2.5) for bootstrap data ξ_i^b.
(v) Repeat (iii)–(iv) many times and use the empirical distribution of those realizations of H_n^b as the distribution of $n(\bar{X} - \mu)/(\tau \underline{V}_n)$.

The proposed wild bootstrap is an extension of that in Liu [30] for independent data to modulated stationary case, and it has two appealing features. First, the scaling factor $\hat{\tau}$ makes the statistic independent of the dependence structure. Second, the bootstrap data-generating mechanism is adaptive to unknown time-dependent variances $\{\sigma_i^2\}$. For the distribution of α_i in step (iii), we use $\mathbb{P}(\alpha_i = -1) = \mathbb{P}(\alpha_i = 1) = 1/2$, which has some desirable properties. For example, it preserves the magnitude and range of the data. As shown by Davidson and Flachaire [17], for certain hypothesis testing problems in linear regression models with symmetrically distributed errors, the bootstrap distribution is exactly equal to that of the test statistic; see Theorem 1 therein.

For the purpose of comparison, we briefly introduce the widely used block bootstrap for a stationary time series $\{X_i\}$ with mean μ. By (1.2), $\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, \tau^2)$. Suppose that we want to bootstrap the distribution of $\sqrt{n}(\bar{X} - \mu)$. Let $k_n, \ell_n, \mathcal{I}_1, \ldots, \mathcal{I}_{\ell_n}$ be defined as in Section 2.5 below. The non-overlapping block bootstrap works in the following way:

(i) Take a simple random sample of size ℓ_n with replacement from the blocks $\mathcal{I}_1, \ldots, \mathcal{I}_{\ell_n}$, and form the bootstrap data $X_1^b, \ldots, X_{n'}^b$, $n' = k_n \ell_n$, by pooling together X_is for which the index i is within those selected blocks.

(ii) Let $\bar{X}^b$ be the sample average of $X_1^b, \dots, X_{n'}^b$. Compute $\Xi_n = \sqrt{n'}\{\bar{X}^b - E^*(\bar{X}^b)\}$, where $E^*(\bar{X}^b) = \sum_{i=1}^{n'} X_i/n'$ is the conditional expectation of $\bar{X}^b$ given $\{X_i\}$.

(iii) Repeat (i)–(ii) many times and use the empirical distribution of Ξ_n's as the distribution of $\sqrt{n}(\bar{X} - \mu)$.

In step (ii), another choice is the studentized version $\tilde{\Xi}_n = \sqrt{n'}\{\bar{X}^b - E^*(\bar{X}^b)\}/\hat{\tau}^b$, where $\hat{\tau}^b$ is a consistent estimate of τ based on bootstrap data. Assuming stationarity and $k_n \to \infty$, the blocks are asymptotically independent and share the same model dynamics as the whole data, which validates the above block bootstrap. We refer the reader to Lahiri [29] for detailed discussions. For a non-stationary process, block bootstrap is no longer valid, because individual blocks are not representative of the whole data. By contrast, the proposed wild bootstrap is adaptive to unknown dependence and the non-constant variances structure.

2.4. Change point analysis: Self-normalized CUSUM test

To test a change point in the mean of a process $\{X_i\}$, two popular CUSUM-type tests (see Section 3 of Robbins *et al.* [35] for a review and related references) are

$$T_n^1 = \max_{cn \le j \le (1-c)n} \frac{\hat{\tau}^{-1}|S_X(j)|}{\sqrt{j(1-j/n)}} \quad \text{and} \quad T_n^2 = \max_{cn \le j \le (1-c)n} \hat{\tau}^{-1}|S_X(j)|, \tag{2.13}$$

where $\hat{\tau}^2$ is a consistent estimate of the long-run variance τ^2 of $\{X_i\}$, and

$$S_X(j) = \left(1 - \frac{j}{n}\right)\sum_{i=1}^{j} X_i - \frac{j}{n}\sum_{i=j+1}^{n} X_i. \tag{2.14}$$

Here $c > 0$ ($c = 0.1$ in our simulation studies) is a small number to avoid the boundary issue. For i.i.d. data, $j(1 - j/n)$ is proportional to the variance of $S_X(j)$, so T_n^1 is a studentized version of T_n^2. For i.i.d. Gaussian data, T_n^1 is equivalent to likelihood ratio test; see Csörgő and Horváth [14]. Assume that, under null hypothesis,

$$\left\{n^{-1/2}\sum_{i=1}^{\lfloor nt \rfloor}[X_i - E(X_i)]\right\}_{0\le t\le 1} \Rightarrow \tau\{B_t\}_{0\le t\le 1}, \qquad \text{in the Skorohod space} \tag{2.15}$$

for a standard Brownian motion $\{B_t\}_{t\ge 0}$. The above convergence requires finite-dimensional convergence and tightness; see Billingsley [11]. By the continuous mapping theorem, $T_n^1 \Rightarrow \max_{c\le t\le 1-c} |B_t - tB_1|/\sqrt{t(1-t)}$ and $T_n^2/\sqrt{n} \Rightarrow \max_{c\le t\le 1-c} |B_t - tB_1|$.

For the modulated stationary case (1.3), (2.15) is no longer valid. Moreover, since T_n^1 and T_n^2 do not take into account the time-dependent variances σ_i^2, an abrupt change in variances may lead to a false rejection of H_0 when the mean remains constant. For example, our simulation study in Section 3.3 shows that the empirical false rejection probability for T_n^1 and T_n^2 is about 10% for nominal level 5%. To alleviate the issue of non-constant variances, we adopt the self-normalization approach as in previous sections. Recall F_j and $\underline{V}_j$ in (2.3). For each fixed

$cn \le j \le (1-c)n$, by Theorem 2.1 and Slutsky's theorem, $F_j/\underline{V}_j \Rightarrow N(0, \tau^2)$ in distribution, assuming the negligibility of the approximation errors. Therefore, the self-normalization term $\underline{V}_j$ can remove the time-dependent variances. In light of this, we can simultaneously self-normalize the two terms $\sum_{i=1}^{j} X_i$ and $\sum_{i=j+1}^{n} X_i$ in (2.14) and propose the self-normalized test statistic

$$T_n^{\mathrm{SN}} = \max_{cn \le j \le (1-c)n} \hat{\tau}^{-1} |T_n(j)|, \quad \text{where } T_n(j) = \frac{S_X(j)}{\sqrt{(1-j/n)^2 \underline{V}_j^2 + (j/n)^2 \overline{V}_j^2}}. \tag{2.16}$$

Here, $\underline{V}_j^2$ is defined as in (2.3), $\overline{V}_j^2 = \sum_{i=j+1}^{n} (X_i - \overline{X}_j)^2$ with $\overline{X}_j = (n-j)^{-1} \sum_{i=j+1}^{n} X_i$.

Theorem 2.3. *Assume* (2.2) *holds. Let* $\delta_n \to 0$ *be as in* (2.10). *Under* H_0, *we have*

$$\max_{cn \le j \le (1-c)n} |T_n(j) - \tau \widetilde{T}_n(j)| = \mathrm{O}_{\mathrm{p}}(\delta_n),$$

where

$$\widetilde{T}_n(j) = \frac{(1-j/n)\sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) - j/n \sum_{i=j+1}^{n} \sigma_i (B_i - B_{i-1})}{\sqrt{(1-j/n)^2 \sum_{i=1}^{j} \sigma_i^2 + (j/n)^2 \sum_{i=j+1}^{n} \sigma_i^2}}.$$

By Theorem 2.3, under H_0, T_n^{SN} is asymptotically equivalent to $\max_{cn \le j \le (1-c)n} |\widetilde{T}_n(j)|$. Due to the self-normalization, for each j, the time-dependent variances are removed and $\widetilde{T}_n(j) \sim N(0,1)$ has a standard normal distribution. However, $\widetilde{T}_n(j)$ and $\widetilde{T}_n(j')$ are correlated for $j \ne j'$. Therefore, $\{\widetilde{T}_n(j)\}$ is a non-stationary Gaussian process with a standard normal marginal density. Due to the large number of unknown parameters σ_i, it is infeasible to obtain the null distribution directly. On the other hand, Theorem 2.3 establishes the fact that, asymptotically, the distribution of T_n^{SN} in (2.16) depends only on $\sigma_1, \dots, \sigma_n$ and is robust against the dependence structure of $\{e_i\}$, which motivates us to use the wild bootstrap method in Section 2.3 to find the critical value of T_n^{SN}.

(i) Compute $T_n(j)$ and find $\hat{J} = \operatorname{argmax}_{cn \le j \le (1-c)n} |T_n(j)|$.
(ii) Divide the data into two blocks $X_1, \dots, X_{\hat{J}}$ and $X_{\hat{J}+1}, \dots, X_n$. Within each block, subtract the sample mean from the observations therein to obtain centered data. Pool all centered data together and denote them by $\epsilon_1, \dots, \epsilon_n$.
(iii) Based on $\epsilon_1, \dots, \epsilon_n$, obtain an estimate $\hat{\tau}$ of τ. See Section 2.5 below.
(iv) Compute the test statistic T_n^{SN} in (2.16).
(v) Based on ϵ_i in (ii), use the wild bootstrap method in Section 2.3 to generate synthetic data $\xi_1, \dots, \xi_n$, and use (i)–(iv) to compute the bootstrap test statistic T_n^b based on the bootstrap data $\xi_1, \dots, \xi_n$.
(vi) Repeat (v) many times and find $(1-\alpha)$ quantile of those T_n^bs.

As argued in Section 2.3, the synthetic data-generating scheme (v) inherits the time-varying non-stationarity structure of the original data. Also, the statistic T_n^{SN} is robust against the dependence structure, which justifies the proposed bootstrap method. If H_0 is rejected, the change point is then estimated by $\hat{J} = \operatorname{argmax}_{cn \le j \le (1-c)n} |T_n(j)|$.

If there is no evidence to reject H_0, we briefly discuss how to apply the same methodology to test $\tilde{H}_0$: $\sigma_1 = \cdots = \sigma_J \neq \sigma_{J+1} = \cdots = \sigma_n$, that is, whether there is a change point in the variances σ_i^2. By (1.1), we have $(X_i - \mu)^2 = \sigma_i^2 + \sigma_i^2 \zeta_i$, where $\zeta_i = e_i^2 - 1$ has mean zero. Therefore, testing a change point in the variances σ_i^2 of X_i is equivalent to testing a change point in the mean of the new data $\tilde{X}_i = (X_i - \bar{X})^2$.

2.5. Long-run variance estimation

To apply the results in Sections 2.2–2.4, we need a consistent estimate of the long-run variance τ^2. Most existing works deal with stationary time series through various block bootstrap and subsampling approaches; see Lahiri [29] and references therein. Assuming a near-epoch dependent mixing condition, De Jong and Davidson [18] established the consistency of a kernel estimator of $\mathrm{Var}(\sum_{i=1}^n X_i)$, and their result can be used to estimate τ_n^2 in (2.9) for the CLT of $\sqrt{n}(\bar{X} - \mu)$. However, for the change point problem in Section 2.4, we need an estimator of the long-run variance τ^2 of the unobservable process $\{e_i\}$, so the method in De Jong and Davidson [18] is not directly applicable.

To attenuate the non-stationarity issue, we extend the idea in Section 2.2 to blockwise self-normalization. Let k_n be the block length. Denote by $\ell_n = \lfloor n/k_n \rfloor$ the largest integer not exceeding n/k_n. Ignore the boundary and divide $1, \ldots, n$ into ℓ_n blocks

$$\mathcal{I}_j = \{(j-1)k_n + 1, \ldots, jk_n\}, \qquad j = 1, \ldots, \ell_n. \tag{2.17}$$

Recall the overall sample mean $\bar{X}$. For each block j, define the self-normalized statistic

$$D_j = \frac{k_n[\bar{X}(j) - \bar{X}]}{V(j)}, \qquad \text{where } \bar{X}(j) = \frac{1}{k_n} \sum_{i \in \mathcal{I}_j} X_i, V^2(j) = \sum_{i \in \mathcal{I}_j} [X_i - \bar{X}(j)]^2. \tag{2.18}$$

By Proposition 2.2, the self-normalized statistics $D_1, \ldots, D_{\ell_n} \sim N(0, \tau^2)$ are asymptotically i.i.d. Thus, we propose estimating τ^2 by

$$\hat{\tau}^2 = \frac{1}{\ell_n} \sum_{j=1}^{\ell_n} D_j^2. \tag{2.19}$$

As in (2.4)–(2.5), we define the quantities on block j

$$r(j) = |\sigma_{jk_n}| + \sum_{i \in \mathcal{I}_j} |\sigma_i - \sigma_{i-1}| \quad \text{and} \quad r^*(j) = |\sigma_{jk_n}^2| + \sum_{i \in \mathcal{I}_j} |\sigma_i^2 - \sigma_{i-1}^2|, \tag{2.20}$$

$$\Sigma^2(j) = \sum_{i \in \mathcal{I}_j} \sigma_i^2 \quad \text{and} \quad \Sigma^{*2}(j) = \left(\sum_{i \in \mathcal{I}_j} \sigma_i^4 \right)^{1/2}. \tag{2.21}$$

Theorem 2.4. *Let (2.2) hold with $\Delta_n = n^{1/4}\log(n)$. Recall r_n, Σ_n in (2.4)–(2.5). Define*

$$M_n = \frac{1}{k_n} + \max_{1\le j\le \ell_n} \frac{\Sigma^{*2}(j) + r^*(j)\Delta_n}{\Sigma^2(j)} + \max_{1\le j\le \ell_n} \frac{r(j)\Delta_n}{\Sigma(j)}. \tag{2.22}$$

Assume that $r_n\Delta_n/\Sigma_n \to 0$ and

$$\chi_n = \ell_n^{-1/2} + \log(n)M_n + \sqrt{\log(n)}\frac{\Sigma_n}{\ell_n^2}\sum_{j=1}^{\ell_n}\frac{1}{\Sigma(j)} + \frac{\Sigma_n^2}{\ell_n^3}\sum_{j=1}^{\ell_n}\frac{1}{\Sigma^2(j)} \to 0. \tag{2.23}$$

Then $\hat\tau^2 - \tau^2 = \mathrm{O_p}(\chi_n)$. Consequently, $\hat\tau$ is a consistent estimate of τ.

Consider Example 2.2 with $\gamma \in [0, 1)$. Then $\chi_n \asymp \sqrt{\log(n)/\ell_n} + \log^2(n)(n^{1/4}/\sqrt{k_n} + n^{5/4-\gamma}/k_n + \sqrt{k_n}n^{1/4-\gamma})$. For $\gamma \in (3/4, 1)$, it can be shown that the optimal rate is $\chi_n \asymp n^{-1/8}\log^{5/4}(n)$ when $k_n \asymp n^{3/4}\log^{3/2}(n)$. In Example 2.3 with $\sigma_i = i^\beta$ for some $\beta \in [0, 1)$, elementary but tedious calculations show that the optimal rate is

$$\chi_n \asymp \begin{cases} n^{-1/8}\log^{5/4}(n), & k_n \asymp n^{3/4}\log^{3/2}(n), \\ \quad \beta \in [0, 3/4], & \\ n^{(\beta-1)/(5-4\beta)}\{\log(n)\}^{(8(1-\beta))/(5-4\beta)}, & k_n \asymp n^{(4.5-4\beta)/(5-4\beta)}\{\log(n)\}^{4/(5-4\beta)}, \\ \quad \beta \in (3/4, 1). & \end{cases}$$

2.6. Some possible extensions

The self-normalization approaches in Sections 2.2–2.5 can be extended to linear regression model (1.4) with modulated stationary time series errors. The approach in Phillips, Sun and Jin [33] is not applicable here due to non-stationarity. For simplicity, we consider the simple case that $p = 2$, $U_i = (1, i/n)$, and $\beta = (\beta_0, \beta_1)'$. Hansen [23] studied a similar setting for martingale difference errors. Denote by $\hat\beta_0$ and $\hat\beta_1$ the simple linear regression estimates of β_0 and β_1 given by

$$\hat\beta_1 = \frac{n\sum_{i=1}^n iX_i - \sum_{i=1}^n i\sum_{i=1}^n X_i}{\sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2/n} \quad \text{and} \quad \hat\beta_0 = \bar X - \hat\beta_1(n+1)/(2n). \tag{2.24}$$

Then simple algebra shows that

$$\hat\beta_0 - \beta_0 = \frac{2}{n^2 - n}\sum_{i=1}^n (2n - 3i + 1)\sigma_i e_i, \qquad \hat\beta_1 - \beta_1 = \frac{6}{n^2 - 1}\sum_{i=1}^n (2i - n - 1)\sigma_i e_i.$$

The latter expressions are linear combinations of $\{e_i\}$. Thus, by the same argument in Proposition 2.2 and Theorem 2.1, we have self-normalized CLTs for $\hat\beta_0$ and $\hat\beta_1$.

Theorem 2.5. *Let $s_{i,0} = (2n - 3i + 1)\sigma_i$ and $s_{i,1} = (2i - n - 1)\sigma_i$. Assume that $\{s_{i,0}\}_{1\le i\le n}$ and $\{s_{i,1}\}_{1\le i\le n}$ satisfy condition* (2.10). *Then as $n \to \infty$,*

$$\frac{n^2(\hat{\beta}_0 - \beta_0)}{2V_{n,0}} \Rightarrow N(0, \tau^2), \qquad \text{where } V_{n,0}^2 = \sum_{i=1}^{n}(2n - 3i + 1)^2(X_i - \hat{\beta}_0 - \hat{\beta}_1 i/n)^2,$$

$$\frac{n^2(\hat{\beta}_1 - \beta_1)}{6V_{n,1}} \Rightarrow N(0, \tau^2), \qquad \text{where } V_{n,1}^2 = \sum_{i=1}^{n}(2i - n - 1)^2(X_i - \hat{\beta}_0 - \hat{\beta}_1 i/n)^2.$$

The long-run variance τ^2 can be estimated using the idea of blockwise self-normalization in Section 2.5. Let k_n, ℓ_n and $\mathcal{I}_j$ be defined as in Section 2.5. Then we propose

$$\hat{\tau}^2 = \frac{1}{\ell_n}\sum_{j=1}^{\ell_n} D_j^2, \qquad \text{where } D_j = \frac{\sum_{i\in\mathcal{I}_j}(X_i - \hat{\beta}_0 - \hat{\beta}_1 i/n)}{\sqrt{\sum_{i\in\mathcal{I}_j}(X_i - \hat{\beta}_0 - \hat{\beta}_1 i/n)^2}}. \tag{2.25}$$

Here, $D_1, \dots, D_{\ell_n}$ are asymptotically i.i.d. normal random variables with mean zero and variance τ^2. Consistency can be established under similar conditions as in Theorem 2.4.

For the general linear regression model (1.4), the linearly weighted average structure of linear regression estimates allows us to obtain self-normalized CLTs as in Theorem 2.5 under more complicated conditions. Also, it is possible to extend the proposed method to the nonparametric regression model with time-varying variances

$$X_i = f(i/n) + \sigma_i e_i, \tag{2.26}$$

where $f(\cdot)$ is a nonparametric time trend of interest. Nonparametric estimates, for example, the Nadaraya–Watson estimate, are usually based on locally weighted observations. The latter feature allows us to derive similar self-normalized CLT. However, the change point problem for (1.4) and (2.26) will be more challenging, and Aue *et al.* [6] studied (1.4) for uncorrelated errors with constant variance. Also, it is more difficult to address the bandwidth selection issues; see Altman [2] for related contribution when $\sigma_i \equiv 1$. It remains a direction of future research to investigate (1.4) and (2.26).

3. Simulation study

3.1. Selection of block length k_n for $\hat{\tau}$

Recall that $D_1, \dots, D_{\ell_n}$ in (2.25) are asymptotically i.i.d. normal random variables. To get a sensible choice of the block length parameter k_n, we propose a simulation-based method by minimizing the empirical mean squared error (MSE):

(i) Simulate n i.i.d. standard normal random variables $Z_1, \dots, Z_n$.
(ii) Based on $Z_1, \dots, Z_n$, obtain $\hat{\tau}$ with block length k.

(iii) Repeat (i)–(ii) many times, compute empirical MSE(k) as the average of realizations of $(\hat{\tau}-1)^2$, and find the optimal k by minimizing MSE(k).

We find that the optimal block length k is about 12 for $n=120$, about 15 for $n=240$, about 20 for $n=360, 600$ and about 25 for $n=1200$.

3.2. Empirical coverage probabilities

Let sample size $n=120$. Recall e_i and σ_i in (1.1). For σ_i, consider four choices:

$$\begin{aligned}
\text{A1:}\quad & \sigma_i = 0.2\mathbf{1}_{i\le n/2} + 0.6\mathbf{1}_{i>n/2},\\
\text{A2:}\quad & \sigma_i = 0.2\{1+\cos^2(i/n^{4/5})\},\\
\text{A3:}\quad & \sigma_i = 0.2 + 0.1\log(1+|i-n/2|),\\
\text{A4:}\quad & \sigma_i = 0.3 + \phi(i/60),
\end{aligned}$$

where ϕ is the standard normal density, and $\mathbf{1}$ is the indicator function. The sequences A1–A4 exhibit different patterns, with a piecewise constancy for A1, a cosine shape for A2, a sharp change around time $n/2$ for A3 and a gradual downtrend for A4. Let ε_i be i.i.d. N(0, 1). For e_i, we consider both linear and nonlinear processes.

$$\text{B1:}\quad e_i = \{\eta_i - E(\eta_i)\}/\sqrt{\mathrm{Var}(\eta_i)}, \qquad \text{where } \eta_i = \theta|\eta_{i-1}| + \sqrt{1-\theta^2}\varepsilon_i,\ |\theta|<1.$$

$$\text{B2:}\quad e_i = \sum_{j=0}^{\infty} a_j\varepsilon_{i-j}, \qquad \text{where } a_j = \frac{(j+1)^{-\beta}}{\sqrt{\sum_{j=0}^{\infty}(j+1)^{-2\beta}}},\ \beta > 1/2.$$

For B1, by Wu [40], (2.8) holds. By Andel, Netuka and Svara [3], $E(\eta_i) = \theta\sqrt{2/\pi}$ and $\mathrm{Var}(\eta_i) = 1 - 2\theta^2/\pi$. To examine how the strength of dependence affects the performance, we consider $\theta = 0, 0.4, 0.8$, representing independence, intermediate and strong dependence, respectively. For B2 with $\beta > 2$, (2.2) holds with $\Delta_n = n^{1/4}\log(n)$, and we consider three cases $\beta = 2.1, 3, 4$. To assess the effect of block length k_n, three choices $k_n = 8, 10, 12$ are used. Thus, we consider all 72 combinations of $\{\text{A1, A2, A3, A4}\} \times \{\text{B1}, \theta = 0, 0.4, 0.8; \text{B2}, \beta = 2.1, 3, 4\} \times \{k_n = 8, 10, 12\}$.

Without loss of generality we examine coverage probabilities based on 10^3 realized confidence intervals for $\mu = 0$ in (1.1). We compare our self-normalization-based confidence intervals to some stationarity-based methods. For (1.1), if we pretend that the error process $\{\tilde{e}_i = \sigma_i e_i\}$ is stationary, then we can use (1.2) to construct an asymptotic confidence interval for μ. Under stationarity, the long-run variance τ^2 of $\{\tilde{e}_i\}$ can be similarly estimated through the block method in Section 2.5 by using the non-normalized version $D_j = \sqrt{k_n}[\bar{X}(j) - \bar{X}]$ in (2.25); see Lahiri [29]. Thus, we compare two self-normalization-based methods and three stationarity-based alternatives: self-normalization-based confidence intervals through the asymptotic theory in Proposition 2.2 (SN) and the wild bootstrap (WB) in Section 2.3; stationarity-based confidence intervals through the asymptotic theory (1.2) (ST), non-overlapping block bootstrap (BB) and studentized non-overlapping block bootstrap (SBB) in Section 2.3. From the results in Table 1, we see

Table 1. Coverage probabilities (in percentage) for μ in (1.1) with e_i from B1 [(a)] and B2 [(b)]. Nominal level is 95%. SN and WB denote self-normalization-based confidence intervals using asymptotic theory in Proposition 2.2 and the wild bootstrap procedure, respectively; ST, BB, SBB denote stationarity-based confidence intervals using asymptotic theory in (1.2), non-overlapping block bootstrap and studentized non-overlapping block bootstrap, respectively

θ	k_n	σ_i	SN	WB	ST	BB	SBB	σ_i	SN	WB	ST	BB	SBB
(a) Model B1													
0.0	8	A1	98.0	94.7	93.1	92.2	92.8	A2	96.6	95.2	92.3	92.5	92.5
	10		98.2	95.0	92.6	92.4	92.2		94.6	94.6	90.0	89.5	89.4
	12		98.1	95.6	91.7	91.4	91.1		92.1	93.7	89.7	89.5	89.6
	8	A3	96.4	95.0	92.5	92.3	92.0	A4	96.6	95.6	93.1	92.6	93.0
	10		94.7	94.7	90.8	90.6	90.6		95.1	95.1	91.4	91.3	91.3
	12		93.7	94.8	90.8	90.4	90.5		92.9	93.7	89.8	89.7	89.5
0.4	8	A1	98.7	95.9	92.7	92.6	92.9	A2	96.6	95.3	92.5	92.4	92.0
	10		98.5	95.7	92.8	92.7	92.3		95.4	95.4	91.6	91.1	91.6
	12		98.0	95.0	90.8	90.8	90.2		92.5	94.0	89.4	89.1	89.4
	8	A3	96.6	95.2	91.7	91.7	91.6	A4	95.4	94.1	90.8	90.9	90.6
	10		95.3	95.5	91.5	91.3	91.5		95.0	94.8	91.2	90.7	90.8
	12		93.1	94.6	90.2	89.9	89.9		94.1	95.1	90.3	89.8	90.1
0.8	8	A1	97.9	94.6	87.8	86.8	87.3	A2	96.1	94.7	87.2	87.3	87.0
	10		97.6	95.5	87.3	87.0	86.7		93.3	92.9	86.4	86.8	86.1
	12		97.3	94.0	85.8	85.5	85.1		92.6	93.4	86.5	86.4	86.4
	8	A3	94.8	93.5	85.7	85.7	86.0	A4	95.5	94.7	86.3	86.1	86.1
	10		93.5	93.8	85.7	85.5	85.2		95.3	95.1	88.5	88.3	88.5
	12		92.4	93.3	87.2	86.7	86.9		92.6	94.2	86.3	85.8	85.7
β													
(b) Model B2													
4.0	8	A1	97.6	94.9	91.8	91.4	91.9	A2	95.9	94.2	91.9	92.0	91.1
	10		97.7	93.2	88.9	88.1	88.3		95.7	95.7	92.1	91.8	92.1
	12		97.9	95.5	90.7	90.2	90.0		93.3	94.6	90.0	89.9	89.7
	8	A3	94.6	93.3	89.8	89.5	89.5	A4	95.6	94.7	91.3	91.7	91.0
	10		95.1	95.2	91.6	91.4	91.5		95.4	95.9	92.8	92.2	93.0
	12		93.8	95.4	90.8	90.6	90.2		93.9	94.9	88.9	88.5	88.6
3.0	8	A1	99.1	95.7	91.1	91.0	91.2	A2	95.8	94.6	90.4	89.8	90.1
	10		98.5	96.4	91.6	90.9	91.1		95.6	95.2	92.1	91.9	91.5
	12		97.9	94.6	89.6	89.3	89.0		94.1	95.0	90.5	90.2	90.4
	8	A3	95.9	94.6	92.0	91.9	91.7	A4	96.0	94.5	90.6	90.4	90.3
	10		94.3	94.4	90.0	89.9	89.8		94.3	94.4	89.2	89.3	88.9
	12		93.2	94.5	88.9	88.6	88.7		93.1	94.1	89.6	88.9	88.8
2.1	8	A1	97.1	92.5	86.2	86.2	85.5	A2	95.7	93.8	88.9	89.0	88.7
	10		97.6	94.7	89.2	88.9	88.6		93.5	93.6	88.8	88.8	88.4
	12		97.2	95.1	87.9	87.5	87.7		92.6	93.9	88.0	87.6	87.7
	8	A3	94.0	93.7	88.5	88.4	88.3	A4	95.0	93.1	88.8	88.7	88.6
	10		93.3	93.8	88.1	87.9	87.8		94.1	94.2	89.1	88.8	89.1
	12		92.9	94.7	89.1	88.4	88.4		91.5	92.6	87.7	87.5	87.5

that the coverage probabilities of the proposed self-normalization-based methods (columns SN and WB) are close to the nominal level 95% for almost all cases considered. By contrast, the stationarity-based methods (columns ST, BB and SBB) suffer from substantial undercoverage, especially when dependence is strong ($\theta = 0.8$ in Table 1(a) and $\beta = 2.1$ in Table 1(b)). For the two self-normalization-based methods, WB slightly outperforms SN.

3.3. Size and power study

In (1.3), we use the same setting for σ_i and e_i as in Section 3.2. For mean μ_i, we consider $\mu_i = \lambda \mathbf{1}_{i>40}, \lambda \geq 0$, and compare the test statistics T_n^1, T_n^2 in (2.13) and T_n^{SN} in (2.16). First, we compare their size under the null with $\lambda = 0$. The critical value of T_n^{SN} is obtained using the wild bootstrap in Section 2.4; for T_n^1 and T_n^2, their critical values are based on the block bootstrap in Section 2.3. In each case, we use 10^3 bootstrap samples, nominal level 5%, and block length $k_n = 10$, and summarize the empirical sizes (under the null $\lambda = 0$) in Table 2 based on 10^3 realizations. While T_n^{SN} has size close to 5%, T_n^1 and T_n^2 tend to over-reject the null, and the false rejection probabilities can be three times the nominal level of 5%. Next, we compare the size-adjusted power. Instead of using the bootstrap methods to obtain critical values, we use 95% quantiles of 10^4 realizations of the test statistics when data are simulated directly from the null model so that the empirical size is exactly 5%. Figure 1 presents the power curves for combinations $\{$A1–A4$\} \times \{$B1 with $\theta = 0.4$; B2 with $\beta = 3.0\}$ with 10^3 realizations each. For A1, T_n^{SN} outperforms T_n^1 and T_n^2; for A2–A4, there is a moderate loss of power for T_n^{SN}. Overall, T_n^{SN} has power comparable to other two tests. In practice, however, the null

Table 2. Size (in percentage) comparison of T_n^1 and T_n^2 in (2.13) and T_n^{SN} in (2.16), with sample size $n = 120$, nominal level 5%, and block length $k_n = 10$

	Model B1				Model B2			
σ_i	θ	T_n^{SN}	T_n^1	T_n^2	β	T_n^{SN}	T_n^1	T_n^2
A1	0.0	4.9	9.1	8.4	2.1	7.3	12.2	13.4
	0.4	4.7	9.4	9.6	3.0	4.7	8.6	9.2
	0.8	6.0	15.1	14.7	4.0	5.6	9.9	7.7
A2	0.0	5.7	8.2	6.1	2.1	5.8	9.5	8.6
	0.4	6.1	8.9	6.8	3.0	5.3	9.6	6.8
	0.8	7.3	12.6	9.3	4.0	4.2	7.5	4.2
A3	0.0	5.0	5.7	4.8	2.1	5.5	7.7	6.7
	0.4	5.3	6.9	5.4	3.0	5.8	6.1	4.9
	0.8	7.0	9.8	10.0	4.0	5.0	6.5	4.2
A4	0.0	5.4	8.4	6.0	2.1	6.9	8.8	7.1
	0.4	5.7	7.9	5.2	3.0	4.8	6.6	6.3
	0.8	7.2	11.1	9.2	4.0	5.3	6.2	5.8

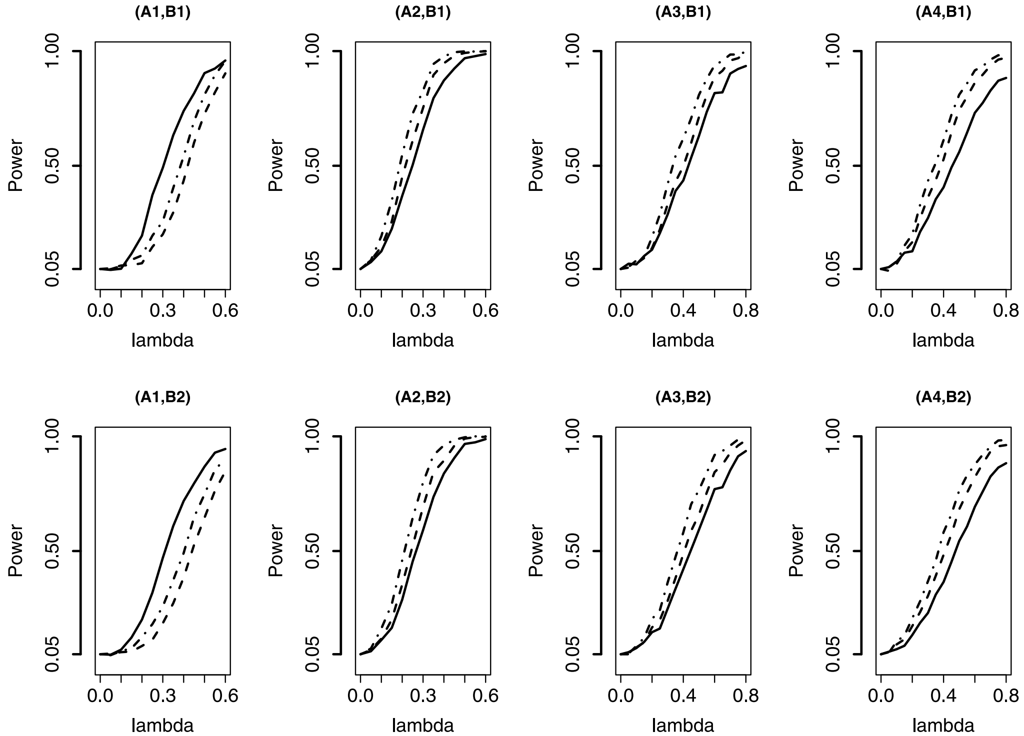

Figure 1. Size-adjusted power curves for T_n^1 (dashed curve) and T_n^2 (dotdash curve) in (2.13) and T_n^{SN} (solid curve) in (2.16) as functions of change size λ (horizontal axis) with sample size $n = 120$ and block length $k_n = 10$. For (A1, B1)–(A4, B1), the error process $\{e_i\}$ is from B1 with $\theta = 0.4$; for (A1, B2)–(A4, B2), the error process $\{e_i\}$ is from B2 with $\beta = 3.0$.

model is unknown, and when one turns to the bootstrap method to obtain the critical values, the usual CUSUM tests T_n^1 and T_n^2 will likely over-reject the null as shown in Table 2. In summary, with such small sample size and complicated time-varying variances structure, T_n^{SN} along with the wild bootstrap method delivers reasonably good power and the size is close to nominal level.

Finally, we point out that the proposed self-normalization-based methods are not robust to models with time-varying correlation structures. For example, consider the model $e_i = 0.3e_{i-1} + \varepsilon_i$ for $1 \le i \le 60$ and $e_i = 0.8e_{i-1} + \varepsilon_i$ for $61 \le i \le n$, where ε_i are i.i.d. N(0, 1). With $k_n = 10$, the sizes (nominal level 5%) for the three tests T_n^{SN}, T_n^1, T_n^2 are 0.154, 0.196, 0.223 for A1. Future research directions include (i) developing tests for change in the variance or covariance structure for (1.1) (See Inclán and Tiao [26], Aue *et al.* [5] and Berkes, Gombay and Horváth [10] for related contributions); and (ii) developing methods that are robust to changes in correlations.

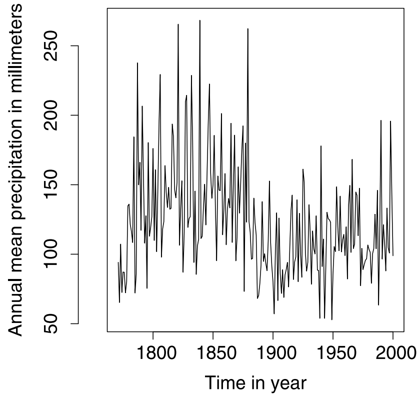

Figure 2. Annual mean precipitation in Seoul from 1771–2000.

4. Applications to two real data sets

4.1. Annual mean precipitation in Seoul during 1771–2000

The data set consists of annual mean precipitation rates in Seoul during 1771–2000; see Figure 2 for a plot. The mean levels seem to be different for the two time periods 1771–1880 and 1881–2000. Ha and Ha [22] assumed the observations are i.i.d. under the null hypothesis. As shown in Figure 2, the variations change over time. Also, the auto-correlation function plot (not reported here) indicates strong dependence up to lag 18. Therefore, it is more reasonable to apply our self-normalization-based test that is tailored to deal with modulated stationary processes. With sample size $n = 230$, by the method in Section 3.1, the optimal block length is about 15. Based on 10^5 bootstrap samples as described in Section 2.4, we obtain the corresponding p-values 0.016, 0.005, 0.045, 0.007, with block length $k_n = 12, 14, 16, 18$, respectively. For all choices of k_n, there is compelling evidence that a change point occurred at year 1880. While our result is consistent with that of Ha and Ha [22], our modulated stationary time series framework seems to be more reasonable. Denote by μ_1 and μ_2 the mean levels over pre-change and post-change time periods 1771–1880 and 1881–2000. For the two sub-periods with sample sizes 110 and 120, the optimal block length is about 12. With $k_n = 12$, applying the wild bootstrap in Section 2.3 with 10^5 bootstrap samples, we obtain 95% confidence intervals $[121.7, 161.3]$ for μ_1, $[100.9, 114.3]$ for μ_2. For the difference $\mu_1 - \mu_2$, with optimal block length $k_n = 15$, the 95% wild bootstrap confidence interval is $[19.6, 48.2]$. Note that the latter confidence interval for $\mu_1 - \mu_2$ does not cover zero, which provides further evidence for $\mu_1 \neq \mu_2$ and the existence of a change point at year 1880.

4.2. Quarterly U.S. GNP growth rates during 1947–2002

The data set consists of quarterly U.S. Gross National Product (GNP) growth rates from the first quarter of 1947 to the third quarter of 2002; see Section 3.8 in Shumway and Stoffer [38]

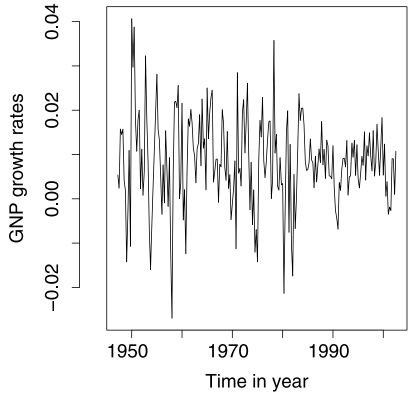

Figure 3. Quarterly U.S. GNP growth rates from 1947–2002.

for a stationary autoregressive model approach. However, the plot in Figure 3 suggests a non-stationary pattern: the variation becomes smaller after year 1985 whereas the mean level remains constant. Moreover, the stationarity test in Kwiatkowski *et al.* [28] provides fairly strong evidence for non-stationarity with a p-value of 0.088. With the block length $k_n = 12, 14, 16, 18$, we obtain the corresponding p-values $0.853, 0.922, 0.903, 0.782$, and hence there is no evidence to reject the null hypothesis of a constant mean μ. Based on $k_n = 15$, the 95% wild bootstrap confidence interval for μ is $[0.66\%, 1.00\%]$. To test whether there is a change point in the variance, by the discussion in the last paragraph of Section 2.4, we consider $\tilde{X}_i = (X_i - \underline{X}_n)^2$. With $k_n = 12, 14, 16, 18$, the corresponding p-values are $0.001, 0.006, 0.001, 0.010$, indicating strong evidence for a change point in the variance at year 1984. In summary, we conclude that there is no change point in the mean level, but there is a change point in the variance at year 1984.

Appendix: Proofs

Proof of Theorem 2.1. Let $r_j = |\sigma_j| + \sum_{i=2}^{j} |\sigma_i - \sigma_{i-1}|$. By the triangle inequality, we have $r_j \le r_n$. Recall S_i in (2.2). By the summation by parts formula, (2.6) follows via

$$\begin{aligned} F_j &= \sum_{i=1}^{j} \sigma_i (S_i - S_{i-1}) = \sigma_j S_j + \sum_{i=1}^{j-1} (\sigma_i - \sigma_{i+1}) S_i \\ &= \sigma_j \tau B_j + \sum_{i=1}^{j-1} (\sigma_i - \sigma_{i+1}) \tau B_i + \mathrm{O}_{\mathrm{a.s.}}(r_n \Delta_n) \qquad \text{(A.1)} \\ &= \tau \sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) + \mathrm{O}_{\mathrm{a.s.}}(r_n \Delta_n). \end{aligned}$$

By Kolmogorov's maximal inequality for independent random variables, for $\delta > 0$,

$$P\left\{\max_{1\le j\le n}\left|\sum_{i=1}^{j}\sigma_i(B_i - B_{i-1})\right| \ge \delta\Sigma_n\right\} \le (\delta\Sigma_n)^{-2}E\left[\left\{\sum_{i=1}^{n}\sigma_i(B_i - B_{i-1})\right\}^2\right] = \delta^{-2}. \qquad \text{(A.2)}$$

Thus, by (A.1), $\max_{1\le j\le n}|F_j| = \mathrm{O_p}(\Sigma_n + r_n\Delta_n)$. Observe that

$$\underline{V}_j^2 - \Sigma_j^2 = W_j - F_j^2/j, \qquad \text{where } W_j = \sum_{i=1}^{j}\sigma_i^2(e_i^2 - 1). \qquad \text{(A.3)}$$

By (2.2), the same argument in (A.1) and (A.2) shows $W_j = \mathrm{O_p}(\Sigma_n^{*2} + r_n^*\Delta_n)$, uniformly. The desired result then follows via (A.3). □

Proof of Theorem 2.2. Denote by $\Phi(x)$ the standard normal distribution function. By Proposition 2.2 and Slutsky's theorem, $\mathbb{P}(H_n/\hat{\tau} \le x) \to \Phi(x)$ for each fixed $x \in \mathbb{R}$. Since $\Phi(x)$ is a continuous distribution, $\sup_{x\in\mathbb{R}}|\mathbb{P}(H_n/\hat{\tau} \le x) - \Phi(x)| = 0$. It remains to prove $\sup_{x\in\mathbb{R}}|\mathbb{P}^*(H_n^* \le x) - \Phi(x)| \to 0$, in probability. Notice that, conditioning on $\{e_i\}$, $\{\xi_i\}$ are independent random variables with zero mean. By the Berry–Esséen bound in Bentkus, Bloznelis and Götze [9], there exists a finite constant c such that

$$\sup_{x\in\mathbb{R}}|\mathbb{P}^*(H_n^* \le x) - \Phi(x)| \le c\sum_{i=1}^{n}E^*(|\xi_i|^3)\left\{\sum_{i=1}^{n}E^*(|\xi_i|^2)\right\}^{-3/2}, \qquad \text{(A.4)}$$

where E^* denotes conditional expectations given $\{e_i\}$. Clearly, $E^*(|\xi_i|^2) = \sigma_i^2 e_i^2 E(\alpha_1^2)$ and $E(|\xi_i|^3) = \sigma_i^3|e_i^3|E(|\alpha_1^3|)$. Thus, under the assumption $e_i \in \mathcal{L}^3$, we have $\sum_{i=1}^n E^*(|\xi_i|^3) = \mathrm{O_p}(\sum_{i=1}^n \sigma_i^3)$. Meanwhile, by the proof of Theorem 2.1, $\sum_{i=1}^n E^*(|\xi_i|^2) = \sum_{i=1}^n \sigma_i^2 e_i^2 = \{1 + \mathrm{o_p}(1)\}\sum_{i=1}^n \sigma_i^2$. Therefore, the desired result follows from (A.4) in view of (2.11). □

Proof of Theorem 2.3. For $cn \le j \le (1-c)n$, $c \le (1 - j/n)$, $j/n \le 1 - c$. For $S_X(j)$ in (2.14), by (2.6), we have $\max_{cn\le j\le(1-c)n}|S_X(j) - \tau\widetilde{S}_X(j)| = \mathrm{O_{a.s.}}(r_n\Delta_n)$, where

$$\widetilde{S}_X(j) = \left(1 - \frac{j}{n}\right)\sum_{i=1}^{j}\sigma_i(B_i - B_{i-1}) - \frac{j}{n}\sum_{i=j+1}^{n}\sigma_i(B_i - B_{i-1}).$$

By (2.7), $\max_{cn\le j\le(1-c)n}|(1 - j/n)^2\underline{V}_j^2 + (j/n)^2\overline{V}_j^2 - V_j^2| = \mathrm{O_p}(\varpi_n)$, where

$$V_j^2 = (1 - j/n)^2\sum_{i=1}^{j}\sigma_i^2 + (j/n)^2\sum_{i=j+1}^{n}\sigma_i^2 \quad \text{and} \quad \varpi_n = (r_n^2\Delta_n^2 + \Sigma_n^2)/n + \Sigma_n^{*2} + r_n^*\Delta_n.$$

For $cn \le j \le (1-c)n$, $V_j^2 \ge c^2\Sigma_n^2$. Thus, condition (2.10) implies $\varpi_n = \mathrm{o}(V_j^2)$ and $\{V_j^2 + \mathrm{O}_\mathrm{p}(\varpi_n)\}^{1/2} = V_j + \mathrm{O}_\mathrm{p}(\varpi_n/V_j)$. Therefore, uniformly over $cn \le j \le (1-c)n$,

$$T_n(j) - \tau\widetilde{T}_n(j) = \frac{\tau\widetilde{S}_X(j) + \mathrm{O}_{\mathrm{a.s.}}(r_n\Delta_n)}{V_j + \mathrm{O}_\mathrm{p}(\varpi_n/V_j)} - \frac{\tau\widetilde{S}_X(j)}{V_j} = \mathrm{O}_\mathrm{p}\left\{\frac{r_n\Delta_n}{V_j} + \frac{\varpi_n\widetilde{S}_X(j)}{V_j^3}\right\}.$$

By (A.2), $\max_j |\widetilde{S}_X(j)| = \mathrm{O}_\mathrm{p}(\Sigma_n)$. Thus, the result follows in view of $V_j \ge c\Sigma_n$. □

Proof of Theorem 2.4. Condition $M_n \to 0$ implies $\max_{1\le j\le \ell_n} r(j)\Delta_n/\Sigma(j) \to 0$. By (2.7),

$$\omega_j := \frac{V^2(j)}{\Sigma^2(j)} - 1 = \mathrm{O}_\mathrm{p}\left\{\frac{\Sigma^{*2}(j) + r^*(j)\Delta_n}{\Sigma^2(j)} + \frac{1}{k_n}\right\} = \mathrm{O}_\mathrm{p}(M_n) \to 0. \tag{A.5}$$

Define $U_j = \Sigma^{-1}(j)\sum_{i\in\mathcal{I}_j}\sigma_i(B_i - B_{i-1})$. Clearly, $U_1,\dots,U_{\ell_n}$ are independent standard normal random variables. Thus, $\max_{1\le j\le\ell_n}|U_j| = \mathrm{O}_\mathrm{p}\{\sqrt{\log(\ell_n)}\} = \mathrm{O}_\mathrm{p}\{\sqrt{\log(n)}\}$. By (2.6), $\underline{X}_n - \mu = \mathrm{O}_\mathrm{p}\{(\Sigma_n + r_n\Delta_n)/n\} = \mathrm{O}_\mathrm{p}(\Sigma_n/n)$. Recall the definition of D_j in (2.18). By the same argument in (2.6), using $\sqrt{1+x} = 1 + \mathrm{O}(x)$ as $x \to 0$, we have

$$\begin{aligned}
D_j &= \frac{k_n\{\bar{X}(j) - \mu\}}{\Sigma(j)}\frac{1}{\sqrt{1+\omega_j}} + \frac{k_n(\mu - \underline{X}_n)}{\Sigma(j)}\frac{1}{\sqrt{1+\omega_j}} \\
&= \left[\tau U_j + \mathrm{O}_{\mathrm{a.s.}}\left\{\frac{r(j)\Delta_n}{\Sigma(j)}\right\}\right]\{1 + \mathrm{O}(\omega_j)\} + \mathrm{O}_\mathrm{p}\left\{\frac{k_n\Sigma_n}{n\Sigma(j)}\right\} \\
&= \tau U_j + \mathrm{O}_\mathrm{p}\left\{\sqrt{\log(n)}M_n + \frac{\Sigma_n}{\ell_n\Sigma(j)}\right\}.
\end{aligned}$$

By the latter expression and $\log(n)M_n \to 0$, we can easily verify $\hat{\tau}^2 - \tau^2 = \mathrm{O}_\mathrm{p}(\chi_n)$. □

Acknowledgements

We are grateful to the associate editor and three anonymous referees for their insightful comments that have significantly improved this paper. We also thank Amanda Applegate for help on improving the presentation and Kyung-Ja Ha for providing us the Seoul precipitation data. Zhao's research was partially supported by NIDA Grant P50-DA10075-15. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

References

[1] Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643

[2] Altman, N.S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.* **85** 749–759. MR1138355
[3] Anděl, J., Netuka, I. and Zvźra, K. (1984). On threshold autoregressive processes. *Kybernetika (Prague)* **20** 89–106. MR0747062
[4] Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678
[5] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452
[6] Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2008). Testing for changes in polynomial regression. *Bernoulli* **14** 637–660. MR2537806
[7] Aue, A., Horváth, L., Kokoszka, P. and Steinebach, J. (2008). Monitoring shifts in mean: Asymptotic normality of stopping times. *TEST* **17** 515–530. MR2470096
[8] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121
[9] Bentkus, V., Bloznelis, M. and Götze, F. (1996). A Berry–Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.* **9** 765–796. MR1400598
[10] Berkes, I., Gombay, E. and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *J. Statist. Plann. Inference* **139** 2044–2063. MR2497559
[11] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[12] Bühlmann, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074
[13] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813
[14] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2743035
[15] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916
[16] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797
[17] Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *J. Econometrics* **146** 162–169. MR2459651
[18] de Jong, R.M. and Davidson, J. (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* **68** 407–423. MR1748008
[19] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681
[20] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series*: *Nonparametric and Parametric Methods*. *Springer Series in Statistics*. New York: Springer. MR1964455
[21] Götze, F. and Künsch, H.R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24** 1914–1933. MR1421154
[22] Ha, K.J. and Ha, E. (2006). Climatic change and interannual fluctuations in the long-term record of monthly precipitation for Seoul. *Int. J. Climatol.* **26** 607–618.
[23] Hansen, B.E. (1995). Regression with nonstationary volatility. *Econometrica* **63** 1113–1132. MR1348515
[24] Hansen, B.E. (2000). Testing for structural change in conditional models. *J. Econometrics* **97** 93–115. MR1788819
[25] Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* **21** 671–680. MR1232511
[26] Inclán, C. and Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* **89** 913–923. MR1294735

[27] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147

[28] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics* **54** 159–178.

[29] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. New York: Springer. MR2001447

[30] Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR0964947

[31] Müller, U.K. (2007). A theory of robust long-run variance estimation. *J. Econometrics* **141** 1331–1352. MR2413504

[32] Pettitt, A.N. (1980). A simple cumulative sum type statistic for the change-point problem with zero–one observations. *Biometrika* **67** 79–84. MR0570508

[33] Phillips, P.C.B., Sun, Y. and Jin, S. (2007). Long run variance estimation and robust regression testing using sharp origin kernels with no truncation. *J. Statist. Plann. Inference* **137** 985–1023. MR2301731

[34] Politis, D.N. and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181

[35] Robbins, M.W., Lund, R.B., Gallagher, C.M. and Lu, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *J. Amer. Statist. Assoc.* **106** 89–99. MR2816704

[36] Shao, Q.M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48** 319–334. MR1244549

[37] Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617

[38] Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications: With R Examples*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR2228626

[39] Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1295. MR0868303

[40] Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

[41] Zhao, Z. (2011). A self-normalized confidence interval for the mean of a class of non-stationary processes. *Biometrika* **98** 81–90.

Received January 2011 and revised August 2011