

# Testing for jumps in the presence of smooth changes in trends of nonstationary time series

Ting Zhang\*

*Department of Mathematics and Statistics*

*Boston University*

*111 Cummings Mall, Boston, MA 02215*

*e-mail: [tingz@bu.edu](mailto:tingz@bu.edu)*

**Abstract:** Nonparametric smoothing methods have been widely used in trend analysis. However, the inference procedure usually requires the crucial assumption that the underlying trend function is smooth. This paper considers the situation where the trend function has potential jumps in addition to smooth changes. In order to determine the existence of jumps, we propose a nonparametric test that can survive under dependent and nonstationary errors, where existing tests assuming independence or stationarity can fail. When the existence of jumps is affirmative, we further consider the problem of estimating the number, location and size of jumps. The results are illustrated via both Monte Carlo simulations and a real data example.

**Keywords and phrases:** Abrupt and smooth changes, change points, local linear estimation, nonparametric hypothesis testing, nonparametric jump detection, nonstationary processes.

Received August 2015.

## 1. Introduction

Given a sequence of observations collected over time, a problem of significant interest is to understand the pattern of the underlying trend (mean) function. For this, a common approach is to form a small window around each time point and use the (weighted) sample average within the selected window as an estimator of the corresponding local mean. The method is nonparametric as it does not impose any parametric assumption on the underlying trend function, and has been frequently used and widely studied in the literature; see for example Härdle and Mammen (1993), Fan and Zhang (1999), Horowitz and Spokoiny (2001), Zhou and Wu (2010), Chen and Hong (2012) and references therein. Nevertheless, the validity of the methods developed in the aforementioned papers relies critically on the assumption that the trend function is smooth, namely it does not contain any jump. Although the smoothness assumption provides the basic motivation and technical justification for using nonparametric smoothing methods, Müller and Stadtmüller (1999) argued that a small number of jumps can exist in addi-

---

\*Research supported by NSF DMS-1461796

tion to smooth changes for a number of applications. Therefore, an immediate and important problem is to detect whether the trend function has any jump before calling the nonparametric estimation procedure. This motivates us to consider a rigorous statistical test for the existence of jumps in the trend function regardless of whether it has smooth changes. Therefore, smooth changes can exist under both the null and alternative hypotheses, which makes the current problem different from the conventional problem of testing for change-points where one is interested in detecting any change from constancy; see for example Shao and Zhang (2010), Zhou (2013), Vogt and Dette (2015) for recent developments and review papers by Aue and Horváth (2013) and Jandhyala et al. (2013) for further references in this direction. As a result, applying the tests developed in the aforementioned papers to the current problem may result in distorted  $p$ -values, as the current null hypothesis contains smooth changes which however belong to the alternative of the aforementioned tests.

Nonparametric inference of regression functions with jumps has been an active area of research; see for example Hall and Titterton (1992), Müller (1992), Loader (1996), Qiu and Yandell (1998), Spokoiny (1998), Grégoire and Hamrouni (2002a,b), Gijbels et al. (2007) and Joo and Qiu (2009). Nevertheless, most of the existing results focused on estimating the discontinuous regression curve by assuming that the errors are independent and identically distributed (iid). In contrast, the problem of testing for jumps has been studied mainly in the setting that the potential jump location is known to the practitioner, while the difficult problem of establishing an asymptotically valid global test has been much less explored in the literature. Assuming that the errors are iid, Wu and Chu (1993) proposed a test based on the maximal discrepancy between two kernel estimators that use the same bandwidth but different kernels; while Müller and Stadtmüller (1999) proposed a test by estimating the sum of squared jump sizes via simple linear regression. However, as commented by Wu and Zhao (2007), the independence assumption used in Wu and Chu (1993) and Müller and Stadtmüller (1999) can cause serious restriction on their applicability to time series data where dependence is the rule rather than the exception. To accommodate for serial correlation, Wu and Zhao (2007) proposed a dependence-corrected test by assuming that the error process is strictly stationary; see also Gao et al. (2008) for the case with stationary  $\alpha$ -mixing observations. We shall here substantially generalize existing results by allowing dependent and nonstationary errors, which appear commonly in practice (Elsner et al., 2008; Zhou and Wu, 2009; Degras et al., 2012; Guinness and Stein, 2013).

Nevertheless, generalizing the aforementioned tests to allow nonstationary error processes can be a nontrivial task. For example, the method of Müller and Stadtmüller (1999) relies on the relationship in their equation (3.2) to estimate the jump indicator parameter. However, if the error process is nonstationary, then the intercept in their equation (3.2) may also depend on time or the index variable, making it no longer suitable for the aforementioned purpose. Also, the test of Gao et al. (2008) requires a stationary regressor process with a continuous density function to guarantee its convergence to a proper limiting distribution, which is violated in the current setting where the data collection time serves

as the index variable which is deterministic and discrete. Since nonstationary processes with time-varying features have experienced a surge of attention in the recent literature, the current paper makes an effort and consider a nonparametric test that can survive under both dependence and nonstationarity. In addition, to overcome the slow convergence issue of the asymptotic extreme value distribution, we propose to use a pivotalized simulation-assisted (PSA) procedure, and the current test seems to outperform the one by Wu and Zhao (2007) even for stationary error processes for which the aforementioned test is designed; see the simulation comparison in Section 5.1. Moreover, the current test largely relaxes the restrictive bandwidth condition required by Wu and Zhao (2007), which can help bring a great convenience to practitioners. To be more specific, in order for their test statistic to have a proper limiting distribution, Wu and Zhao (2007) requires that the bandwidth  $b_n$  satisfies the condition  $nb_n^3 \log n \rightarrow 0$ ; see also Wu and Chu (1993) for a similar condition on the bandwidth. Although the aforementioned condition is able to regularize the bandwidth sequence to produce a smaller bias, it can cause a larger asymptotic variance of the resulting kernel estimators and in turn bring more randomness or uncertainties into the resulting test statistic making it less powerful. A popular bandwidth choice that is known to balance the bias-variance trade-off is of the form  $b_n = cn^{-1/5}$  for some constant  $0 < c < \infty$ , which has been widely used for estimation and testing purposes in the literature; see the discussion in Zhang and Wu (2011) for further references. Nevertheless, this popular choice is not allowed by the test of Wu and Zhao (2007), which can cause a great inconvenience for practitioners as finding another bandwidth sequence that has the nice interpretation and satisfies the conditions imposed by Wu and Chu (1993) and Wu and Zhao (2007) can be a nontrivial task. Wu and Zhao (2007) did not provide a solution on this and simply used  $b_n = n^{-2/5}$  in their applications. In contrast, the current test relaxes the condition to  $nb_n^7 \log n \rightarrow 0$ , and therefore one can conveniently use the popular bandwidth choice  $b_n = cn^{-1/5}$  in practice to avoid the hassle of finding another appropriate bandwidth subject to the conditions imposed by Wu and Chu (1993) and Wu and Zhao (2007). Furthermore, the current paper makes an effort to further consider the situation when the null hypothesis of no jump point is rejected by the proposed test. In this case, we further propose an algorithm that one can use to estimate the number of jumps, locations of jump points and their jump sizes. The corresponding asymptotic theory is also established under the nonstationary framework introduced in Section 2. Unlike the case for iid normal observations, the current proof uses probabilistic tools including the  $m$ -dependence approximation and martingale decomposition to handle nonstationary error processes, and our asymptotic theory allows a growing number of jump points.

The rest of the paper is organized as follows. Section 2 introduces the nonstationary framework and basic assumptions. Section 3 contains our main results, including a nonparametric test for determining the existence of jumps and an algorithm for estimating the number, location and size of jumps. Section 4 deals with various implementation issues, and Section 5 contains Monte Carlo simulations and a real data application. Technical proofs are deferred to the Appendix.

## 2. Framework and basic assumptions

Assume that the data  $(y_i)_{i=1}^n$  are observed from the model

$$y_i = \mu(i/n) + e_i, \quad i = 1, \dots, n, \tag{1}$$

where  $\mu : [0, 1] \rightarrow \mathbb{R}$  is the unknown trend function and  $(e_i)_{i=1}^n$  is a zero-mean error process which can be nonstationary. If the trend function is smooth, then one can naturally estimate it, along with its derivative, by the local linear estimator (Fan and Gijbels, 1996)

$$\{\tilde{\mu}_n(t; K), \tilde{\mu}'_n(t; K)\} = \operatorname{argmin}_{\eta_0, \eta_1 \in \mathbb{R}} \sum_{i=1}^n \{y_i - \eta_0 - (i/n - t)\eta_1\}^2 K\left(\frac{i/n - t}{b_n}\right) \tag{2}$$

for any  $t \in [0, 1]$ , where  $K(\cdot)$  is the kernel function and  $b_n$  is the bandwidth sequence satisfying  $b_n \rightarrow 0$  and  $nb_n \rightarrow \infty$ . Throughout the paper we assume that the kernel  $K(\cdot)$  is a symmetric nonnegative function in  $\mathcal{C}^1[-1, 1]$  satisfying  $\int_{-1}^1 K(v)dv = 1$  and  $K(0) > 0$ . For example, if  $K(v) = \mathbb{I}_{\{|v| \leq 1\}}/2$  is the rectangle kernel, then the local linear estimator (2) becomes the least squares estimator based on the local data points  $\{y_i : i/n \in [t - b_n, t + b_n]\}$ . Other popular choices include the Epanechnikov kernel  $K(v) = 3 \max(0, 1 - v^2)/4$ , the quartic kernel  $K(v) = 15\{\max(0, 1 - v^2)\}^2/16$  and the triweight kernel  $K(v) = 35\{\max(0, 1 - v^2)\}^3/32$ . For a random variable  $X$ , we write  $X \in \mathcal{L}^q$ ,  $q > 0$ , if  $\|X\|_q = \{E(|X|^q)\}^{1/q} < \infty$  and denote  $\|\cdot\| = \|\cdot\|_2$ .

As commented by Müller and Stadtmüller (1999), in many applications the underlying function is smooth everywhere except for a certain number of points where jumps occur. We shall here consider the situation that the trend function  $\mu(\cdot)$  is piecewise smooth on  $[0, 1]$  with a finite number of jump points. To be more precise, we say that a function  $f \in \text{PS}_M[0, 1]$  if there exists  $0 = t_0 < t_1 < \dots < t_M < t_{M+1} = 1$  such that on each of the intervals  $[t_0, t_1), \dots, [t_{M-1}, t_M)$  and  $[t_M, 1]$ ,  $f$  has bounded third-order derivatives, while  $\lim_{s \uparrow t_k} f(s) \neq f(t_k)$ ,  $k = 1, \dots, M$ . Hence,  $M$  represents the total number of jump points. We want to test the null hypothesis that the trend function does not contain any jump, namely

$$H_0 : \mu \in \text{PS}_0[0, 1]. \tag{3}$$

Let

$$K_-(v) = K(v)\mathbb{I}_{\{v < 0\}}, \quad K_+(v) = K(v)\mathbb{I}_{\{v \geq 0\}}$$

be the two one-sided kernel functions deduced from  $K(\cdot)$ , and

$$\hat{\mu}_{n,-}(t) = \tilde{\mu}_n(t; K_-), \quad \hat{\mu}_{n,+}(t) = \tilde{\mu}_n(t; K_+), \quad t \in [b_n, 1 - b_n],$$

be the corresponding local linear estimators. Then  $\hat{\mu}_{n,-}(t)$  and  $\hat{\mu}_{n,+}(t)$  represent the left and right local linear estimators by using the data points  $\{y_i : i/n \in [t - b_n, t)\}$  and  $\{y_i : i/n \in [t, t + b_n]\}$  respectively. If the function  $\mu(\cdot)$  is smooth, then both  $\hat{\mu}_{n,-}(t)$  and  $\hat{\mu}_{n,+}(t)$  are consistent estimator of  $\mu(t)$ , and thus their difference is expected to be small. On the other hand, if the trend function

has a jump at time  $t$  and  $\mu(t-) = \lim_{s \uparrow t} \mu(s) \neq \mu(t)$ , then  $\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)$  provides an estimator of the jump size  $\mu(t) - \mu(t-)$  which is relatively large. The formulation of a formal test would require an asymptotic theory on the discrepancy between the left and right local linear estimators, which can be quite nontrivial even for stationary error processes. We shall here substantially generalize earlier results by allowing nonstationary and nonlinear error processes so that our asymptotic theory can be widely applicable.

The problem of modeling nonstationary processes has been an active area of research; see for example Dahlhaus (1996), Nason et al. (2000), Ombao et al. (2005), Subba Rao (2006) and references therein. A more detailed comparison on existing frameworks can be found in Zhang and Wu (2011); see also Dahlhaus and Subba Rao (2006) and Vogt (2012) for additional discussions. We shall here generalize the framework of Draghicescu et al. (2009) and assume that there exists a zero-mean nonstationary process  $\{G(i/n; \mathcal{F}_i)\}_{i=1}^n$  in the sense of Draghicescu et al. (2009) such that

$$\max_{1 \leq i \leq n} \|e_i - G(i/n; \mathcal{F}_i)\|_p = O(n^{-1}) \quad (4)$$

holds for some  $p \geq 2$ , where  $\mathcal{F}_i = (\dots, \epsilon_{i-1}, \epsilon_i)$  is the shift process of iid random variables  $\epsilon_j$ ,  $j \in \mathbb{Z}$ , and  $G$  is a measurable function. Therefore, one can interpret  $\mathcal{F}_i$  and  $G(i/n; \mathcal{F}_i)$  as the input and output of a time-varying physical system  $G$ , which approximates the underlying data generating mechanism by (4). As discussed in Draghicescu et al. (2009) and Zhang and Wu (2011), this framework covers a wide range of nonstationary processes and naturally extends many existing stationary time series models to their nonstationary counterparts.

We shall now introduce the functional dependence measure that will be useful in our asymptotic theory. Let  $(\epsilon_j^*)_{j \in \mathbb{Z}}$  be an iid copy of  $(\epsilon_j)_{j \in \mathbb{Z}}$  and  $\mathcal{F}_i^* = (\mathcal{F}_{-1}, \epsilon_0^*, \epsilon_1, \dots, \epsilon_i)$  be the coupled shift process, we define the functional dependence measure

$$\theta_{i,q} = \sup_{t \in [0,1]} \|G(t; \mathcal{F}_i) - G(t; \mathcal{F}_i^*)\|_q, \quad (5)$$

which quantifies the dependence of  $G(t; \mathcal{F}_i)$  on the single innovation  $\epsilon_0$  over  $t \in [0, 1]$ . The functional dependence measure (5) enables us to develop an asymptotic theory for complicated statistics of time series data; see Wu (2005) for a comparison with strong mixing conditions and near-epoch dependence conditions. If the short range dependence condition  $\Theta_{0,q} = \sum_{i=0}^{\infty} \theta_{i,q} < \infty$  holds for some  $q \geq 2$ , then the long-run variance function

$$g(t) = \sum_{k \in \mathbb{Z}} E\{G(t; \mathcal{F}_0)G(t; \mathcal{F}_k)\}$$

is uniformly bounded over  $t \in [0, 1]$ . Recall that  $\text{PS}_M[0, 1]$  represents the class of functions on  $[0, 1]$  with  $M$  jump points. We need the following technical assumptions in establishing the main results.

(A1) The trend function  $\mu \in \text{PS}_M[0, 1]$ , where  $M \geq 0$  can either be fixed or grow with the sample size.

(A2) There exists  $c_0 < \infty$  such that  $\|G(s_1; \mathcal{F}_i) - G(s_2; \mathcal{F}_i)\| \leq c_0|s_1 - s_2|$  holds for all  $s_1, s_2 \in [0, 1]$ .

(A3) The long-run variance function  $g(\cdot)$  is bounded away from zero on  $[0, 1]$ .

Condition (A2) suggests that the error process is approximately locally stationary. In particular, consider the process  $\zeta_j(i/n) = G(i/n; \mathcal{F}_j)$ ,  $j \in \mathbb{Z}$ , which is generated by using the same physical system  $G(i/n, \cdot)$  and is thus stationary. Then by condition (A2), in the small neighborhood  $i - k_n \leq l \leq i + k_n$  with  $k_n/n \rightarrow 0$ , we have  $\|e_l - \zeta_l(i/n)\| = O(k_n/n + 1/n) \rightarrow 0$ .

### 3. Main results

#### 3.1. A nonparametric test for jumps

Recall that the left and right local linear estimators  $\hat{\mu}_{n,-}(t)$  and  $\hat{\mu}_{n,+}(t)$  are based on the data points  $\{y_i : i/n \in [t - b_n, t]\}$  and  $\{y_i : i/n \in [t, t + b_n]\}$  respectively, and thus the difference  $\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)$  provides a natural statistic for testing whether there is a jump at time  $t$ . Let

$$\kappa_l = \int_0^1 v^l K(v) dv, \quad l = 0, 1, 2, \dots,$$

and

$$\phi = \int_0^1 \{(\kappa_0 \kappa_2 - \kappa_1^2)^{-1} (\kappa_2 - \kappa_1 v) K(v)\}^2 dv.$$

Theorem 3.1 provides the central limit theorem of  $\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)$  which can be used for testing whether a jump occurs at a certain time point.

**Theorem 3.1.** *Assume (A1), (A2) with jump points  $0 < t_1 < \dots < t_M < 1$ ,  $\Theta_{0,2} < \infty$ ,  $nb_n \rightarrow \infty$  and  $nb_n^7 \rightarrow 0$ . If  $\min_{k \in \{1, \dots, M+1\}} |t_k - t_{k-1}|/b_n \rightarrow \infty$ , then (i) for any continuous point  $t \in (0, 1) \setminus \{t_1, \dots, t_M\}$ ,*

$$(nb_n)^{1/2} \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} \Rightarrow N\{0, 2\phi g(t)\};$$

and (ii) for any jump point  $t_k$  with jump size  $d_k = \mu(t_k) - \lim_{s \uparrow t_k} \mu(s)$ ,

$$\hat{\mu}_{n,+}(t_k) - \hat{\mu}_{n,-}(t_k) = d_k + O_p\{(nb_n)^{-1/2} + b_n^2\}, \quad k = 1, \dots, M.$$

If  $t \in (0, 1) \setminus \{t_1, \dots, t_M\}$  is a continuous point, then the bias of current estimator is  $E\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} = O\{b_n^3 + (nb_n)^{-1}\}$  as can be seen from the proof of Theorem 3.1. For stationary error processes, Wu and Zhao (2007) considered using the difference between the left and right local averages

$$\Delta_n(t) = (nb_n)^{-1} \left( \sum_{i=1}^n y_i \mathbb{1}_{\{i/n \in (t, t+b_n]\}} - \sum_{i=1}^n y_i \mathbb{1}_{\{i/n \in (t-b_n, t]\}} \right), \quad (6)$$

and the corresponding bias is  $E\{\Delta_n(t)\} = b_n \mu'(t) + O\{b_n^2 + (nb_n)^{-1}\}$ . Hence, the finite-sample performance of their method can be greatly affected by the

steepness of the underlying function. Intuitively, the right and left local averages in (6) estimate  $\mu(t + b_n/2)$  and  $\mu(t - b_n/2)$  respectively, and thus the associated bias is of order  $O(b_n)$ , which is larger than its local linear counterpart. Due to the large bias, the method of Wu and Zhao (2007) can be very restrictive on the choice of the bandwidth. In fact, they require that the bandwidth satisfies  $nb_n^3 \log n \rightarrow 0$ , which excludes the popular choice of  $b_n = cn^{-1/5}$ , where  $0 < c < \infty$  is a constant; see also the discussion in Section 1. It is remarkable that using the technique of local linear estimation usually can only bring significant improvements for points in the boundary area (Fan and Gijbels, 1996), which only constitutes a small proportion of the whole region and thus the difference is usually neglected. However, for the specific problem of testing for jumps, left and right local estimators are formed and compared, which makes each time point a boundary point and thus improving the boundary performance by using local linear estimators can bring significant advantages to the current problem but was not taken into account by the test of Wu and Zhao (2007).

If one has the prior knowledge that a certain time point  $t \in (0, 1)$  is a potential jump point due to the occurrence of some sudden events at that time, then Theorem 3.1 is useful in providing a statistical test. In particular, by Theorem 3.1 (i), one rejects the null hypothesis that  $t$  is not a jump point at level  $\alpha \in (0, 1)$  if

$$(nb_n)^{1/2} \left| \frac{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)}{\{2\phi g(t)\}^{1/2}} \right| > z_{1-\alpha/2},$$

where  $z_{1-\alpha/2}$  is the  $(1-\alpha/2)$ -th quantile of the standard normal distribution. By Theorem 3.1 (ii), the above test has unit asymptotic power as  $n \rightarrow \infty$ . Nevertheless, for many applications it is usually the case that potential jump locations are unknown to the practitioner, and the problem of identifying jump locations can require a series of statistical analysis. We shall here provide a global test that can be used for determining the existence of jumps in the trend function without prespecifying potential jump locations. Theorem 3.2 shows that under the null hypothesis (3), after proper centering and scaling the maximal discrepancy between the left and right local linear estimators has the asymptotic extreme value distribution.

**Theorem 3.2.** *Assume (A1)–(A3) with  $M = 0$ ,  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota \geq 2$ , and*

$$nb_n^{5/2}/(\log n)^{15/2} \rightarrow \infty, \quad nb_n^7 \log n \rightarrow 0. \quad (7)$$

*Then as  $n \rightarrow \infty$ ,*

$$\begin{aligned} & \text{pr} \left[ (nb_n)^{1/2} \sup_{t \in \mathcal{T}_n} \left| \frac{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)}{\{2\phi g(t)\}^{1/2}} \right| - B_K(m_n) \leq \frac{u}{(2 \log m_n)^{1/2}} \right] \\ & \rightarrow \exp\{-2 \exp(-u)\}, \end{aligned}$$

*where  $\mathcal{T}_n = [b_n, 1 - b_n]$ ,  $m_n = 1/b_n$ , and*

$$B_K(m_n) = (2 \log m_n)^{1/2} + (2 \log m_n)^{-1/2} [\log \log m_n / 2 + \log\{C_K / (2\pi^{1/2}\phi)\}],$$

with

$$C_K = \{(\kappa_2 - \kappa_1)^2 K(1)^2 + 2\kappa_2^2 K(0)^2\} / (\kappa_0 \kappa_2 - \kappa_1^2)^2.$$

By Theorem 3.2, we reject the null hypothesis (3) at level  $\alpha \in (0, 1)$  if

$$\sup_{t \in \mathcal{T}_n} \left| \frac{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)}{\{2\phi g(t)\}^{1/2}} \right| > (nb_n)^{-1/2} \left[ B_K(m_n) - \frac{\log \log\{(1 - \alpha)^{-1/2}\}}{(2 \log m_n)^{1/2}} \right]. \quad (8)$$

If there is at least one jump point in the trend function, namely  $\mu \in \text{PS}_M[0, 1]$  for some  $M > 0$ , then by Theorem 3.1 (ii), the above test has unit asymptotic power as  $n \rightarrow \infty$ . If the null hypothesis (3) of no jump point is rejected, then one would be interested in knowing the number of jump points along with their locations and sizes, which we shall discuss in Section 3.2.

### 3.2. Estimating the location and size of jumps

#### 3.2.1. The case with a single jump point

If one has prior knowledge about the jump location, say  $t_1 \in (0, 1)$ , then the corresponding jump size can be simply estimated by the difference  $\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)$ , whose consistency is guaranteed by Theorem 3.1 (ii). We shall here consider the situation that the jump location is unknown and needs to be estimated. Recall that the difference  $\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)$  provides a consistent estimator of the jump size that is zero for continuous points, and thus it is natural to estimate the jump location by the maximizer

$$\hat{t}_1 = \operatorname{argmax}_{t \in \mathcal{T}_n} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|.$$

The corresponding jump size can then be estimated by

$$\hat{d}_1 = \hat{\mu}_{n,+}(\hat{t}_1) - \hat{\mu}_{n,-}(\hat{t}_1).$$

Fryzlewicz (2014) considered the problem of estimating the number and locations of change-points in a piecewise-constant mean function with iid observations. We shall here consider the problem of distinguishing jumps from smooth changes with dependent and nonstationary observations. The following theorem states that  $\hat{t}_1$  and  $\hat{d}_1$  consistently estimate the true jump location and the corresponding jump size respectively.

**Theorem 3.3.** *Assume (A1)–(A3) with  $M = 1$  and  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota \geq 2$ . If  $\int_0^{\kappa_2/\kappa_1} (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} (\kappa_2 - \kappa_1 v) K(v) dv < 2$  and there exist  $\nu_1 > \nu_2 \geq 2$  such that  $nb_n^{\nu_2} \rightarrow \infty$  and  $nb_n^{\nu_1} \rightarrow 0$ , then for any  $\varepsilon > 0$  which can be taken arbitrarily small, we have (i)*

$$\hat{t}_1 - t_1 = O_p\{n^{-(1-\varepsilon)}\};$$

and (ii)

$$\hat{d}_1 - d_1 = O_p\{(nb_n)^{-1/2} + b_n^2\}.$$



### 3.2.2. The Case with multiple jump points

In the existence of multiple jump points, by the discussion in Section 3.2.1 one can naturally estimate the corresponding jump locations by local maximizers of the discrepancy function  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$ . To be more specific, we shall in the following present an algorithm that one can use to estimate the number, location and size of jumps.

- (a) Find the critical region  $\mathcal{J}_n \subseteq \mathcal{T}_n$  where  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$  exceeds a threshold, say  $\varphi_n > 0$ ;
- (b) Find the location  $\hat{\vartheta}_1$  where  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$  achieves its maximum on  $\mathcal{J}_n$ , and stop the algorithm if  $\mathcal{J}_n$  is empty;
- (c) Exclude the neighborhood  $[\hat{\vartheta}_1 - 2b_n, \hat{\vartheta}_1 + 2b_n]$  from  $\mathcal{J}_n$  and repeat step (b) until all local maximizers are found, denoted by  $\hat{\vartheta}_1, \dots, \hat{\vartheta}_N$ ;
- (d) Estimate the number of jump points by  $\hat{M} = N$  and sort  $\hat{\vartheta}_1, \dots, \hat{\vartheta}_N$  in an increasing order, denoted by  $\hat{\vartheta}_{(1)} < \dots < \hat{\vartheta}_{(N)}$ ;
- (e) Estimate the jump locations and sizes by  $\hat{t}_k = \hat{\vartheta}_{(k)}$  and  $\hat{d}_k = \hat{\mu}_{n,+}(\hat{t}_k) - \hat{\mu}_{n,-}(\hat{t}_k)$ ,  $k = 1, \dots, N$ , respectively.

The following theorem provides the estimation consistency of the above algorithm.

**Theorem 3.4.** *Assume (A1)–(A3) with  $M \geq 0$  and remaining conditions of Theorem 3.3. If  $\min_{k \in \{1, \dots, M+1\}} |t_k - t_{k-1}|/b_n \rightarrow \infty$  and the threshold satisfies  $\varphi_n \rightarrow 0$  and  $\varphi_n/\{(nb_n)^{-1/2}(\log n)^{1/2} + b_n^3\} \rightarrow \infty$ , then (i)*

$$\lim_{n \rightarrow \infty} \text{pr}(\hat{M} = M) = 1;$$

and (ii) the result of Theorem 3.3 continues to hold for  $\hat{t}_k$  and  $\hat{d}_k$ ,  $k = 1, \dots, \hat{M}$ .

## 4. Implementation

### 4.1. Estimation of long-run variance

To apply Theorems 3.1 and 3.2, we need an estimate of the long-run variance function  $g(t)$ ,  $t \in (0, 1)$ . Let  $\tau_n$  and  $\varrho_n$  be bandwidth sequences satisfying  $\tau_n \rightarrow 0$ ,  $\varrho_n \rightarrow 0$  and  $n\tau_n\varrho_n \rightarrow \infty$ , and

$$\lambda_i = \begin{cases} e_i^2 + 2e_i \sum_{j=1}^n e_j \mathbb{I}_{\{0 < j/n - i/n \leq \tau_n \varrho_n\}}, & \text{if } i/n \leq \tau_n \varrho_n; \\ e_i^2 + 2e_i \sum_{j=1}^n e_j \mathbb{I}_{\{0 < i/n - j/n \leq \tau_n \varrho_n\}}, & \text{if } i/n \geq 1 - \tau_n \varrho_n; \\ e_i \sum_{j=1}^n e_j \mathbb{I}_{\{|i/n - j/n| \leq \tau_n \varrho_n\}}, & \text{otherwise.} \end{cases} \quad (9)$$

Following Zhang and Wu (2012), we estimate  $g(t)$ ,  $t \in [0, 1]$ , by

$$\hat{g}(t) = \sum_{i=1}^n \lambda_i w_{i, \tau_n}(t), \quad (10)$$

where  $w_{i,b}(t) = K\{(i/n - t)/b\}\{P_{b,2}(t) - (t - i/n)P_{b,1}(t)\}/\{P_{b,2}(t)P_{b,0}(t) - P_{b,1}(t)^2\}$  are local linear weights with bandwidth  $b$  and  $P_{b,l}(t) = \sum_{j=1}^n (t - j/n)^l K\{(j/n - t)/b\}$ . Theorem 4.1 provides the asymptotic property of our long-run variance estimate (10), where (i) is a result of Proposition 4.1 of Zhang and Wu (2012) and (ii) concerns the uniform consistency which is more desirable for the current problem. Note that uniform results as in Theorem 4.1 (ii) are generally much more difficult to obtain than its counterpart as in (i), and the detailed proof is given in the Appendix.

**Theorem 4.1.** *Assume (A2),  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota > 5/4$  and (4) holds for some  $p \geq 4$ . If  $g \in \mathcal{C}^2[0, 1]$ , then (i) for any  $t \in [0, 1]$ ,*

$$\hat{g}(t) - g(t) = O_p\{\varrho_n^{1/2} + (n\tau_n\varrho_n)^{-(\iota-1)} + (\tau_n\varrho_n)^{(\iota-1)/\iota} + \tau_n^2\}; \quad (11)$$

and (ii)

$$\sup_{t \in [0,1]} |\hat{g}(t) - g(t)| = O_p\{\psi_n + (n\tau_n\varrho_n)^{-(\iota-1)} + (\tau_n\varrho_n)^{(\iota-1)/\iota} + \tau_n^2\}, \quad (12)$$

where  $\psi_n = n^{-1/4}(\tau_n\varrho_n \log n)^{1/2} + n^{1/2}\varrho_n$ .

We shall here briefly discuss the bounds in (11) and (12). For both of them, the optimal bounds are complicated and depend on  $\iota$ , the decay rate of the dependence. If  $\iota \geq 2$  as assumed in Theorem 3.2, then the bound in (11) is minimized and becomes  $O_p\{n^{-2(\iota-1)/(5\iota-3)}\}$  if  $\tau_n \asymp n^{-(\iota-1)/(5\iota-3)}$  and  $\varrho_n \asymp n^{-4(\iota-1)/(5\iota-3)}$ , where for two positive sequences  $(r_n)$  and  $(s_n)$  we write  $r_n \asymp s_n$  if  $r_n/s_n + s_n/r_n$  is bounded for all large  $n$ . On the other hand, the bound in (12) is minimized and becomes  $O_p\{n^{-(\iota-1)/(3\iota-1)}\}$  if  $\tau_n \asymp n^{-(\iota-1)/(6\iota-2)}$  and  $\varrho_n \asymp n^{-(5\iota-3)/(6\iota-2)}$ . Hence if the error process satisfies the geometric moment contraction condition, namely  $\theta_{n,4} = O(\rho^n)$  for some  $0 \leq \rho < 1$  as for finite-order autoregressive processes, then the bounds in (11) and (12) can achieve  $O_p(n^{-2/5+\varepsilon})$  and  $O_p(n^{-1/3+\varepsilon})$  respectively for some arbitrarily small  $\varepsilon > 0$ . In practice, the error process  $(e_i)_{i=1}^n$  is usually not observable, and we shall replace it by the estimated residuals. We suggest using  $\hat{e}_i = y_i - \hat{\mu}_n(i/n)$ ,  $i = 1, \dots, n$ , where for each  $t$ ,  $\hat{\mu}_n(t)$  is one of the one-sided local linear estimates  $\hat{\mu}_{n,\pm}(t)$  that has the smaller weighted residual mean squares

$$\Psi_{n,\pm}(t) = \frac{\sum_{i=1}^n \{y_i - \hat{\mu}_{n,\pm}(t) - (i/n - t)\hat{\mu}'_{n,\pm}(t)\}^2 K_{\pm}\{(i/n - t)/b_n\}}{\sum_{i=1}^n \mathbb{I}_{\{K_{\pm}\{(i/n - t)/b_n\} > 0\}} - 2},$$

where  $\sum_{i=1}^n \mathbb{I}_{\{K_{\pm}\{(i/n - t)/b_n\} > 0\}} - 2$  represents the degree of freedom of the local fit. Let  $\tilde{g}(t)$ ,  $t \in [0, 1]$ , denote the resulting estimate of the long-run variance function, and its asymptotic property is provided in Theorem 4.2.

**Theorem 4.2.** *Assume conditions of Theorems 3.2 and 4.1, then*

$$\sup_{t \in [0,1]} |\tilde{g}(t) - g(t)| = O_p\{\chi_n + \psi_n + (n\tau_n\varrho_n)^{-(\iota-1)} + (\tau_n\varrho_n)^{(\iota-1)/\iota} + \tau_n^2\},$$

where  $\chi_n = \{(nb_n)^{-1/2}(\log n)^{1/2} + b_n^2\}n^{5/4}\tau_n\varrho_n$ .

#### 4.2. A pivotalized simulation-assisted testing procedure

It is well known that the convergence to the extreme value distributions as in Theorem 3.2 can be quite slow, and a very large sample size would be needed for the approximation to be reasonably accurate. We shall here consider a pivotalized simulation-assisted (PSA) procedure that can help improve the finite-sample performance of the proposed test. In particular, by (8), we reject the null hypothesis (3) at level  $\alpha \in (0, 1)$  if

$$\begin{aligned} T_n &= \sup_{t \in \mathcal{T}_n} \left| \frac{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)}{\{\hat{g}(t)\}^{1/2}} \right| \\ &> \{2\phi/(nb_n)\}^{1/2} \left[ B_K(m_n) - \frac{\log \log\{(1-\alpha)^{-1/2}\}}{(2 \log m_n)^{1/2}} \right]. \end{aligned}$$

We then generate iid standard normal random variables  $y_i^\circ$  and compute the corresponding test statistic  $T_n^\circ$ . We repeat this for many times and obtain the empirical quantile  $\hat{q}_{1-\alpha}$  of  $T_n^\circ$ . Since  $T_n$  and  $T_n^\circ$  are both asymptotically pivotal and share the same asymptotic distribution, we reject the null hypothesis (3) at level  $\alpha \in (0, 1)$  if  $T_n > \hat{q}_{1-\alpha}$ . It can be seen from the simulation study in Section 5.1 that the above PSA procedure can help largely improve the finite-sample performance, and the current method can outperform the one by Wu and Zhao (2007) even for stationary error processes for which the aforementioned test is designed.

#### 4.3. Bandwidth selection

Bandwidth selection is a nontrivial problem in the application of nonparametric methods. In the context of nonparametric hypothesis testing, it has been studied by Hall and Hart (1990), Kulasekera and Wang (1997) and Gao and Gijbels (2008) among others. Although many candidates have been proposed in the literature, Wang (2008) commented that usually there is no uniform guidance for an optimal choice. On the positive side, our simulation results in Section 5.1 suggest that the performance of the proposed testing procedure is not very sensitive to the choice of the bandwidth. Therefore, one can simply choose  $b_n = n^{-1/5}$  as suggested by Zhang and Wu (2011) that has the nice interpretation. As an alternative, we consider the generalized cross-validation (GCV) selector, and correct the dependence by estimating the covariance matrix  $\mathbf{\Gamma}_n = \{E(e_i e_j)\}_{1 \leq i, j \leq n}$  as suggested by Wang (1998). In particular, let  $\mathbf{Y} = (y_1, \dots, y_n)^\top$ , where  $^\top$  is the transpose operator. Then for any bandwidth  $b \in (0, 1)$ , recall that  $\hat{\mu}_n(t; b)$  is one of the one-sided local linear estimates  $\hat{\mu}_{n,\pm}(t; b)$  that has the smaller weighted residual mean squares (2), we denote the corresponding fitted values as  $\hat{\mathbf{Y}}(b)$ . We choose the bandwidth  $\tilde{b}_n$  that minimizes

$$\text{GCV}(b) = \frac{n^{-1} \{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}^\top \hat{\mathbf{\Gamma}}_n^{-1} \{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}}{\{1 - K(0)/(nb)\}^2}. \quad (13)$$

An estimate of the covariance matrix  $\Gamma_n$  can be obtained by using the banding technique as in Bickel and Levina (2008) and Wu and Pourahmadi (2009). The GCV selector (13) works reasonably well in our simulation studies. For the choice of  $\tau_n$  and  $\varrho_n$  in estimating the long-run variance, we suggest using the data-driven selector of Zhang and Wu (2012). When implementing the algorithm described in Section 3.2.2 for finite-sample problems, we suggest using the cut-off value from the hypothesis test as the threshold to device the critical region.

## 5. Numerical experiments

### 5.1. Monte Carlo simulations

We shall in this section carry out Monte Carlo simulations to examine the finite-sample performance of the proposed test and compare it with the existing method. The problem of testing for jumps in regression functions has attracted much attention in the literature; see for example Hall and Titterton (1992), Qiu and Yandell (1998), Grégoire and Hamrouni (2002b) and Joo and Qiu (2009) for local tests designed specifically for a given potential jump point, and Wu and Chu (1993) and Müller and Stadtmüller (1999) for global tests with unknown jump locations. However, as commented by Wu and Zhao (2007), existing results usually assume that the errors are iid, which can cause a serious limitation on their applicability to time series data where dependence is the rule rather than the exception. The aforementioned paper proposed a dependence-corrected test by assuming that the error process is strictly stationary. We shall here provide a brief review of their test and compare it with the proposed method. In particular, Wu and Zhao (2007) proposed to use the test statistic

$$T_{\text{WZ}} = \max_{k_n \leq i \leq n - k_n} |\Delta_n(i/n)|, \quad (14)$$

where  $\Delta_n(\cdot)$  is the difference between left and right local averages as defined in (6), and  $k_n = \lfloor nb_n \rfloor$  is the largest integer not exceeding  $nb_n$ . To obtain the cut-off values, Wu and Zhao (2007) suggested a simulation scheme by using quantiles of  $T_{\text{WZ}}^\circ$ , test statistics obtained from independent standard normal observations. Nevertheless, due to the lack of pivotalization, asymptotic distributions of  $T_{\text{WZ}}$  and  $T_{\text{WZ}}^\circ$  can be different by a scale, which needs to be estimated by

$$\hat{r} = \left( \frac{k_n}{2q_{\Phi,0.75}} \right)^{1/2} \text{median}(|A_{i,n} - A_{i-1,n}|, 1 \leq i \leq \lfloor n/k_n \rfloor - 1),$$

where  $q_{\Phi,0.75} = 0.674$  is the third quartile of the standard normal distribution and  $A_{i,n} = k_n^{-1} \sum_{j=1}^{k_n} y_{ik_n+j}$ . The cut-off value is then set to be the corresponding quantile of  $\hat{r}T_{\text{WZ}}^\circ$ . We shall here compare it with the proposed test for both stationary and nonstationary error processes. Let  $n = 500$  and the trend function  $\mu(t) = 2 \sin(2\pi t)$ , we consider the problem of testing the null hypothesis (3). Throughout our numerical experiments, the rectangle kernel  $K(v) = \mathbb{1}_{\{|v| \leq 1\}}/2$

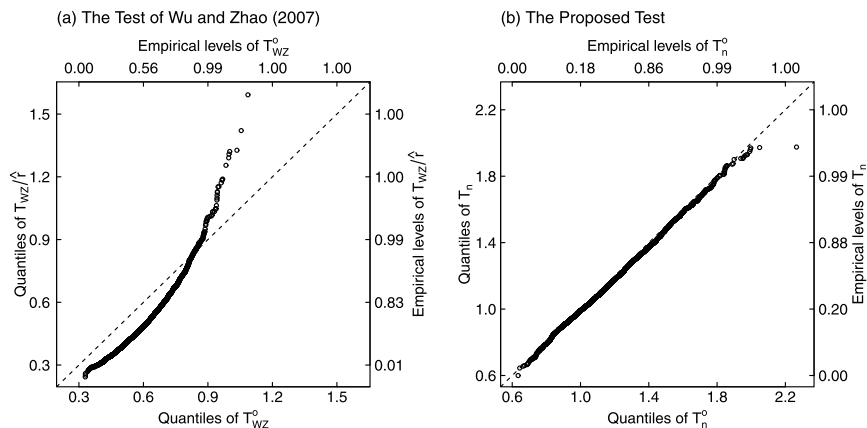


FIG 1. A comparison of the test by Wu and Zhao (2007) in the left panel with the proposed test in the right panel for stationary autoregressive error processes. Q-Q plots of (a)  $T_{WZ}/\hat{r}$  against  $T_{WZ}^{\circ}$ ; and (b)  $T_n$  against  $T_n^{\circ}$ . The dashed line in both plots has unit slope and zero intercept.

is used. Since the test of Wu and Zhao (2007) was developed specifically for stationary error processes, we first consider a stationary autoregressive error process  $e_i = 0.3e_{i-1} + \epsilon_i$ , where  $\epsilon_i$ ,  $i \in \mathbb{Z}$ , are iid standard normal random variables, and we use the bandwidth suggested by Wu and Zhao (2007). The results are summarized in Figure 1, from which we can see that the performance of the simulation method of Wu and Zhao (2007) may still not be decent enough as quantiles of  $T_{WZ}$  are not well approximated by those of  $\hat{r}T_{WZ}^{\circ}$ . In contrast, the proposed test seems to perform reasonably well as can be seen from Figure 1 (b).

We shall now consider the situation with nonstationary error processes. In this case, the asymptotic distribution derived by Wu and Zhao (2007) for their test statistic may no longer hold, as  $\Delta_n(t)$  in (14) at different time points can behave differently for nonstationary processes. Write  $\zeta_i(t) = G(t; \mathcal{F}_i)$  where  $\mathcal{F}_i = (\dots, \epsilon_{i-1}, \epsilon_i)$ , we consider

- (a) a time-varying autoregressive error process:  $\zeta_i(t) = \rho(t)\zeta_{i-1}(t) + \epsilon_i$ , where  $\epsilon_i$ ,  $i \in \mathbb{Z}$ , are iid random variables with  $\text{pr}(\epsilon_i = -1) = \text{pr}(\epsilon_i = 1) = 1/2$ ;
- (b) a time-varying nonlinear error process:  $\zeta_i(t) = \epsilon_i - \rho(t)\{|\epsilon_{i-1}| - (2/\pi)^{1/2}\}$ , where  $\epsilon_i$ ,  $i \in \mathbb{Z}$ , are iid standard normal random variables.

Model (a) generalizes the traditional autoregressive model by allowing the regression coefficient to change over time, and has been widely used in modeling nonstationary signals. By the moving-average representation of autoregressive processes, condition (A2) is satisfied if the coefficient function  $\rho(\cdot)$  is Lipschitz continuous on  $[0, 1]$  and satisfies  $\sup_{t \in [0, 1]} |\rho(t)| < 1$ . For model (b), by elementary calculation  $\|\zeta_i(t_1) - \zeta_i(t_2)\| = |\rho(t_1) - \rho(t_2)| \cdot \|\epsilon_{i-1}| - (2/\pi)^{1/2}\|$ , and thus condition (A2) is satisfied if  $\rho(\cdot)$  is Lipschitz continuous on  $[0, 1]$ . Let

TABLE 1  
 Empirical acceptance probabilities (in percentage) of the proposed method based on  $T_n$  and the test based on  $T_{WZ}$  of Wu and Zhao (2007) for different choices of sample size  $n$  and bandwidth  $b$ .

b	Model (a)						Model (b)					
	$T_n$			$T_{WZ}$			$T_n$			$T_{WZ}$		
	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
0.05	93.8	96.5	99.1	88.0	92.0	96.3	88.3	94.2	98.9	91.5	94.6	98.2
0.06	92.8	96.1	99.1	94.2	96.3	98.6	89.8	95.4	99.0	95.9	97.8	99.2
0.07	92.8	96.1	99.1	98.1	98.9	99.7	89.9	95.1	99.3	98.5	99.1	99.8
0.08	93.8	97.4	99.4	98.7	99.2	99.6	90.3	94.4	98.8	99.3	99.6	99.9
0.09	93.2	96.7	99.5	99.8	99.9	100.0	91.0	95.5	99.3	99.8	100.0	100.0
0.10	93.0	96.0	99.3	99.9	100.0	100.0	90.2	94.9	98.9	100.0	100.0	100.0
0.11	93.7	96.7	99.3	99.9	100.0	100.0	90.7	95.1	98.9	100.0	100.0	100.0
0.12	94.2	97.5	99.5	99.9	100.0	100.0	90.3	95.3	98.9	100.0	100.0	100.0
0.13	93.8	96.7	99.5	100.0	100.0	100.0	90.8	94.9	98.7	100.0	100.0	100.0
0.14	93.1	96.4	99.2	100.0	100.0	100.0	90.7	95.3	98.9	100.0	100.0	100.0
0.15	92.1	95.8	99.1	99.9	100.0	100.0	90.9	95.1	98.6	99.8	99.9	100.0
0.16	92.4	96.1	99.1	100.0	100.0	100.0	90.2	95.0	98.8	99.9	100.0	100.0
0.17	92.1	96.1	99.2	100.0	100.0	100.0	88.8	93.7	98.8	100.0	100.0	100.0
0.18	91.3	95.3	99.0	100.0	100.0	100.0	88.4	93.8	98.6	100.0	100.0	100.0
0.19	89.6	94.7	99.0	100.0	100.0	100.0	85.7	91.7	97.8	100.0	100.0	100.0
0.20	88.3	94.0	98.4	100.0	100.0	100.0	82.5	90.5	97.6	100.0	100.0	100.0

$\rho(t) = 0.5t - 0.2$ , we consider the problem of testing the null hypothesis (3). The results are summarized in Table 1 for different choices of bandwidths, from which we can see that the empirical acceptance probabilities of the proposed test based on  $T_n$  are fairly close to their nominal levels (90%, 95% and 99%) and the results are not very sensitive to the choice of bandwidths. As a comparison, empirical acceptance probabilities of the test by Wu and Zhao (2007) generally deviate from their nominal levels.

## 5.2. Application to global temperature data

The data set at <http://cdiac.ornl.gov/ftp/trends/temp/jonescru/global.txt> contains global monthly temperature anomalies in Celsius from 1850 to 2012. The data set has been widely studied in the literature on understanding the trend pattern, and a time series plot is provided in Figure 2. In particular, Wu et al. (2001) modeled the trend function as piecewise constant, namely changes in the mean can only occur by means of pure jumps. On the other hand, Wu and Zhao (2007) proposed to model the trend as a smooth function as their test suggested that there is no jump in the trend function with a  $p$ -value of 0.22. Nevertheless, the aforementioned results relied on the assumption of stationarity, while Zhou and Wu (2009) argued that the series should be treated as nonstationary. We shall here model the error process by (4) and apply the proposed method to test whether the trend function contains any jump. The selected bandwidth is  $\hat{b}_n = 0.18$ , and the corresponding test statistic is  $T_n = 0.537$ . By the PSA procedure described in Section 4.2 with 5000 simulated  $T_n^\circ$ , the  $p$ -value is 0.011 thus we reject the null hypothesis of no jump at

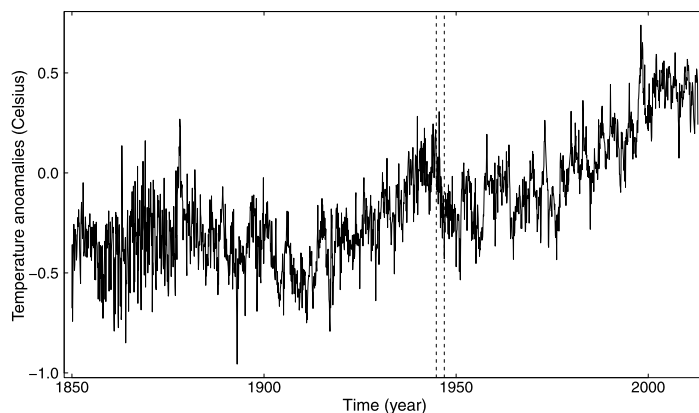


FIG 2. Monthly global temperature anomalies in Celsius from 01/1850 to 12/2012. The period between the two dashed lines corresponds to 11/1944–12/1946.

the 5% significance level. We provide in Figure 2 the region (between the two dashed lines) where the local discrepancy exceeds the critical value. It can be seen from Figure 2 that there is a huge decrease in the mean during the small period between the two dashed lines, which should be treated as a jump as suggested by our analysis. By the algorithm in Section 3.2.2, the jump location is estimated at March 1946 with a jump size  $\hat{d} = -0.26$  in Celsius, indicating a sudden cooling after the Second World War during which atomic bombs were dropped at Japanese cities of Hiroshima and Nagasaki. Also, the end of World War II could lead to significant changes in naval activities at and under the sea surface, national and foreign policies of various countries and other human activities that could potentially affect temperature measurements.

## Appendix

### A.1. Technical proofs

Define the projection operator

$$\mathcal{P}_k(\cdot) = E(\cdot | \mathcal{F}_k) - E(\cdot | \mathcal{F}_{k-1}), \quad k \in \mathbb{Z}.$$

Let  $e_i^{\natural} = G(i/n; \mathcal{F}_i)$ ,  $i = 1, \dots, n$ , and for  $t \in (0, 1)$  let

$$R_{n,\pm}(t) = (nb_n)^{-1} \sum_{i=1}^n (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \{\kappa_2 \mp \kappa_1(i/n - t)/b_n\} K_{\pm}\{(i/n - t)/b_n\} e_i$$

and

$$R_{n,\pm}^{\natural}(t) = (nb_n)^{-1} \sum_{i=1}^n (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \{\kappa_2 \mp \kappa_1(i/n - t)/b_n\} K_{\pm}\{(i/n - t)/b_n\} e_i^{\natural}.$$

The following lemma provides the asymptotic properties of  $R_{n,\pm}(t)$  and  $R_{n,\pm}^{\sharp}(t)$ ,  $t \in (0, 1)$ , and would be useful in proving Theorems 3.1 and 3.2. The proof of Lemma A.1 is given in Section A.2.

**Lemma A.1.** *Assume (A2),  $\Theta_{0,2} < \infty$ ,  $b_n \rightarrow 0$  and  $nb_n \rightarrow \infty$ . Then (i)*

$$\sup_{t \in (0,1)} \|R_{n,\pm}^{\sharp}(t)\| = O\{(nb_n)^{-1/2}\};$$

(ii) for each  $t \in (0, 1)$ ,

$$(nb_n)^{1/2}\{R_{n,+}^{\sharp}(t) - R_{n,-}^{\sharp}(t)\} \Rightarrow N\{0, 2\phi g(t)\};$$

and (iii) the same results hold with  $R_{n,\pm}^{\sharp}(t)$  in (i) and (ii) replaced by  $R_{n,\pm}(t)$ .

*Proof.* (Theorem 3.1) Observe that the left and right local linear estimates have the closed form solution

$$\begin{bmatrix} \hat{\mu}_{n,\pm}(t) \\ b_n \hat{\mu}'_{n,\pm}(t) \end{bmatrix} = \begin{bmatrix} U_{n,0,\pm}(t) & U_{n,1,\pm}(t) \\ U_{n,1,\pm}(t) & U_{n,2,\pm}(t) \end{bmatrix}^{-1} \begin{bmatrix} V_{n,0,\pm}(t) \\ V_{n,1,\pm}(t) \end{bmatrix}, \quad (15)$$

where

$$U_{n,l,\pm}(t) = (nb_n)^{-1} \sum_{i=1}^n \{(i/n - t)/b_n\}^l K_{\pm}\{(i/n - t)/b_n\},$$

with the convention that  $0^0 = 1$  and

$$V_{n,l,\pm}(t) = (nb_n)^{-1} \sum_{i=1}^n \{(i/n - t)/b_n\}^l K_{\pm}\{(i/n - t)/b_n\} y_i.$$

Note that  $\{U_{n,0,\pm}(t), U_{n,1,\pm}(t), U_{n,2,\pm}(t)\} = (\kappa_0, \pm\kappa_1, \kappa_2) + O\{(nb_n)^{-1}\}$  uniformly over  $t \in [b_n, 1 - b_n]$ . By Lemma A.1 (iii), we have

$$\hat{\mu}_{n,\pm}(t) - E\{\hat{\mu}_{n,\pm}(t)\} = R_{n,\pm}(t) + O_p\{(nb_n)^{-1}\{(nb_n)^{-1/2} + b_n^2\}\}. \quad (16)$$

If  $t \in (0, 1) \setminus \{t_1, \dots, t_M\}$  is a continuous point, then by (15),

$$E\{\hat{\mu}_{n,\pm}(t)\} - \mu(t) = b_n^2(\kappa_2^2 - \kappa_1\kappa_3)\mu''(t)/\{2(\kappa_0\kappa_2 - \kappa_1^2)\} + O\{b_n^3 + (nb_n)^{-1}\}. \quad (17)$$

Therefore,  $E\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} = O\{b_n^3 + (nb_n)^{-1}\}$  and (i) follows by (16), (17) and Lemma A.1 (iii). On the other hand, for any jump point  $t_k$ ,  $k = 1, \dots, M$ , by properties of local linear estimates, we have  $E\{\hat{\mu}_{n,+}(t_k) - \hat{\mu}_{n,-}(t_k)\} = d_k + O\{b_n^2 + (nb_n)^{-1}\}$ . Then (ii) follows by (16) and Lemma A.1 (iii).  $\square$

Recall that  $D_{\pm}(v) = (\kappa_0\kappa_2 - \kappa_1^2)^{-1}(\kappa_2 \mp \kappa_1 v)K_{\pm}(v)$ ,  $v \in \mathbb{R}$ , from the proof of Lemma A.1. The following lemmas are useful in proving Theorem 3.2, and their proofs are given in Section A.2.



**Lemma A.2.** Let  $Z_1, \dots, Z_n$  be iid standard normal random variables and

$$Q_n(t) = (2\phi nb_n)^{-1/2} \sum_{i=1}^n [D_+\{(i/n-t)/b_n\} - D_-\{(i/n-t)/b_n\}] Z_i, \quad t \in (0, 1).$$

If  $b_n \rightarrow 0$  and  $nb_n^2 \rightarrow \infty$ , then

$$\text{pr} \left\{ \sup_{t \in \mathcal{T}_n} |Q_n(t)| - B_K(m_n) \leq \frac{u}{(2 \log m_n)^{1/2}} \right\} \rightarrow \exp\{-2 \exp(-u)\}.$$

**Lemma A.3.** Assume (A2), (A3) and  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota \geq 2$ . If  $b_n \rightarrow 0$  and  $nb_n^{5/2}/(\log n)^{15/2} \rightarrow \infty$ , then (i)

$$\begin{aligned} & \text{pr} \left\{ (nb_n)^{1/2} \sup_{t \in \mathcal{T}_n} \left| \frac{R_{n,+}^h(t) - R_{n,-}^h(t)}{\{2\phi g(t)\}^{1/2}} \right| - B_K(m_n) \leq \frac{u}{(2 \log m_n)^{1/2}} \right\} \\ & \rightarrow \exp\{-2 \exp(-u)\}; \end{aligned}$$

and (ii) the same result holds for  $R_{n,+}(t) - R_{n,-}(t)$ .

*Proof.* (Theorem 3.2) Let  $U_{n,l,\pm}(t)$  and  $V_{n,l,\pm}(t)$  be defined as in (15), then by (4) and Lemma A.3 of Zhang and Wu (2012),

$$\sup_{t \in \mathcal{T}_n} |V_{n,l,\pm}(t) - E\{V_{n,l,\pm}(t)\}| = O_p\{(nb_n)^{-1/2}(\log n)^{1/2} + (nb_n)^{-1}n^{1/4}\}$$

for  $l = 0, 1$ , and thus

$$\sup_{t \in [b_n, 1-b_n]} |\hat{\mu}_{n,\pm}(t) - E\{\hat{\mu}_{n,\pm}(t)\}| = O_p\{(nb_n)^{-1/2}(\log n)^{1/2} + (nb_n)^{-1}n^{1/4}\}.$$

Note that  $nb_n^2 \rightarrow \infty$  from (7), we have  $n^{1/4} = o\{(nb_n)^{1/2}\}$  and thus by (4) and (15),

$$\sup_{t \in [b_n, 1-b_n]} |\hat{\mu}_{n,\pm}(t) - E\{\hat{\mu}_{n,\pm}(t)\} - R_{n,\pm}(t)| = O_p\{(nb_n)^{-3/2}(\log n)^{1/2} + (nb_n)^{-1}\}.$$

Since  $\mu \in \text{PS}_0[0, 1]$  that has bounded third-order derivatives, by (15) and (17) we have  $E\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} = O(b_n^3)$  uniformly over  $t \in [b_n, 1 - b_n]$ , and thus

$$\begin{aligned} & \sup_{t \in [b_n, 1-b_n]} |\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} - \{R_{n,+}(t) - R_{n,-}(t)\}| \\ & = O_p\{(nb_n)^{-3/2}(\log n)^{1/2} + (nb_n)^{-1} + b_n^3\}. \end{aligned}$$

Since  $(nb_n \log n)^{1/2}\{(nb_n)^{-3/2}(\log n)^{1/2} + (nb_n)^{-1} + b_n^3\} \rightarrow 0$  from (7), Theorem 3.2 follows by Lemma A.3 (ii).  $\square$

For any  $r > 0$ , let  $\mathcal{N}(t_1, r) = \{t \in [0, 1] : |t - t_1| \leq r\}$  be the neighborhood of  $t_1$  with radius  $r$ . The following lemmas are useful in proving Theorem 3.3.

**Lemma A.4.** *Assume conditions of Theorem 3.3. Then for any positive non-increasing sequence  $\xi_n$  satisfying  $\xi_n/(b_n^2 + n^{-1/2}b_n^{1/2}) \rightarrow \infty$ , we have*

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ |\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)| > \sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n)} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)| \right\} = 1.$$

*Proof.* We shall first deal with the region where  $t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, b_n)$ . Since  $M = 1$  is assumed, the mean function does not have any jump other than  $t_1$ , and thus by the proof of Theorem 3.2,

$$\sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, b_n)} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)| = O_p\{(nb_n)^{-1/2}(\log n)^{1/2} + b_n^3\}.$$

On the other hand, by Theorem 3.1 (ii),

$$|\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)| = |d_1| + O_p\{(nb_n)^{-1/2} + b_n^2\},$$

and thus

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ |\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)| > \sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, b_n)} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)| \right\} = 1.$$

Therefore, it suffices to deal with the remaining region where  $t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n)$ , for which we need the following preparation. Let  $\zeta_k(t) = G(t; \mathcal{F}_k)$  and write  $\Xi_{k,m,n} = E\{\zeta_k(k/n) \mid \mathcal{F}_{k-m,k}\} - E\{\zeta_k(k/n) \mid \mathcal{F}_{k-m+1,k}\}$ , then  $e_k^{\sharp} = \sum_{m=0}^{\infty} \Xi_{k,m,n}$  and  $\Xi_{k,m,n}$ ,  $k = 1, \dots, n$ , form a sequence of martingale differences. Note that  $\max_{1 \leq i \leq n} \|e_i - e_i^{\sharp}\| = O(n^{-1})$  and

$$\|\Xi_{k,m,n}\| = \|E\{\zeta_m(k/n) \mid \mathcal{F}_{0,m}\} - E\{\zeta_m(k/n) \mid \mathcal{F}_{1,m}\}\| \leq \theta_{m,2},$$

by Doob's inequality and the summation by parts formula, we have

$$\sup_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n} |V_{n,l,\pm}(t) - E\{V_{n,l,\pm}(t)\}| = O_p\{(nb_n)^{-1/2}\}, \quad l = 0, 1,$$

and thus

$$\sup_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n} |\hat{\mu}_{n,\pm}(t) - E\{\hat{\mu}_{n,\pm}(t)\}| = O_p\{(nb_n)^{-1/2}\}.$$

Without loss of generality, assume that the jump size  $d_1 > 0$ . Let  $\omega_n = b_n + (nb_n)^{-1}$ , then by (15),

$$\sup_{t \in [t_1 - b_n, t_1] \cap \mathcal{T}_n} \left| E\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} - d_1 \int_{(t_1-t)/b_n}^1 D_+(v)dv \right| = O(\omega_n),$$

and

$$\sup_{t \in [t_1, t_1 + b_n] \cap \mathcal{T}_n} \left| E\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} - d_1 \int_{-1}^{-(t-t_1)/b_n} D_-(v)dv \right| = O(\omega_n).$$

Let  $P_K(s) = \int_0^s D_+(v)dv = \int_0^s (\kappa_0\kappa_2 - \kappa_1^2)^{-1}(\kappa_2 - \kappa_1v)K(v)dv$ ,  $s \in [0, 1]$ , then  $P_K(\cdot)$  is a continuous function on  $[0, 1]$  satisfying  $P_K(0) = 0$ ,  $P_K(1) = 1$  and  $0 < P_K(s) < 2$  if  $s \in (0, 1)$ . Since the kernel function  $K(\cdot)$  is symmetric on  $[-1, 1]$ , we have

$$\sup_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n} |\{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\} - d_1 \{1 - P_K(|t_1 - t|/b_n)\}| = O\{(nb_n)^{-1/2} + b_n\},$$

and thus

$$\lim_{n \rightarrow \infty} \text{pr} \left( \inf_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n} [\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\} + \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\}] > 0 \right) = 1. \tag{18}$$

If  $|t - t_1| > \xi_n$ , then by the condition on  $\xi_n$  we have  $(|t - t_1|/b_n)/\{b_n + (nb_n)^{-1/2}\} \rightarrow \infty$  and

$$\text{pr} \left( \inf_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n)} [\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\} - \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\}] > 0 \right)$$

tends to one as  $n \rightarrow \infty$ . Therefore,

$$\text{pr} \left( \inf_{t \in \mathcal{N}(t_1, b_n) \cap \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n)} [\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\}^2 - \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\}^2] > 0 \right)$$

also tends to one as  $n \rightarrow \infty$ , which entails the desired result. If the jump size  $d_1 < 0$ , then one can apply the above argument to the transformed data  $-y_i$ ,  $i = 1, \dots, n$ , in which case  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$  remains at the same but the direction of jump is flipped.  $\square$

**Lemma A.5.** *Assume conditions of Theorem 3.3. Let  $\xi_n$  be as in Lemma A.4, then for any positive nonincreasing sequence  $\xi_n^\#$  satisfying  $\xi_n^\# / (\xi_n b_n + n^{-1/2} \xi_n^{1/2}) \rightarrow \infty$ , we have*

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ |\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)| > \sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n^\#)} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)| \right\} = 1.$$

*Proof.* In view of Lemma A.4, it suffices to consider the case where  $\xi_n < b_n$  and  $\xi_n/b_n \rightarrow 0$ . Let  $\ell_n = n\xi_n$ , then by the technique of martingale decomposition as in the proof of Lemma A.4, one can obtain that for  $l \in \{0, 1\}$ ,

$$\begin{aligned} & \sup_{t \in \mathcal{N}(t_1, \xi_n) \cap \mathcal{T}_n} |V_{n,l,\pm}(t_1) - V_{n,l,\pm}(t) - E\{V_{n,l,\pm}(t_1) - V_{n,l,\pm}(t)\}| \\ &= O_p \left\{ \frac{(nb_n)^{-1/2} \ell_n + \ell_n^{1/2}}{nb_n} \right\}, \end{aligned}$$

by splitting the sum in  $V_{n,l,\pm}(t_1) - V_{n,l,\pm}(t)$  into common and non-common indices. Since  $\ell_n/(nb_n) \rightarrow 0$ , we have

$$\sup_{t \in \mathcal{N}(t_1, \xi_n) \cap \mathcal{T}_n} |\hat{\mu}_{n,\pm}(t_1) - \hat{\mu}_{n,\pm}(t) - E\{\hat{\mu}_{n,\pm}(t_1) - \hat{\mu}_{n,\pm}(t)\}| = O_p\{(nb_n)^{-1} \ell_n^{1/2}\}. \tag{19}$$

Recall the definition of  $P_K(s)$ ,  $s \in [0, 1]$ , from the proof of Lemma A.4, and without loss of generality assume that the jump size  $d_1 > 0$ . The case with  $d_1 < 0$  can be similarly handled as in the proof of Lemma A.4. Since the trend function  $\mu(\cdot)$  is smooth on  $[0, t_1]$ , we have  $\mu(t) - \mu(t_1) = O(\xi_n)$  uniformly over  $t \in [t_1 - \xi_n, t_1] \cap \mathcal{T}_n$ , and thus by (15),

$$\sup_{t \in [t_1 - \xi_n, t_1] \cap \mathcal{T}_n} |E\{\hat{\mu}_{n,-}(t_1) - \hat{\mu}_{n,-}(t)\}| = O(n^{-1}\ell_n). \tag{20}$$

On the other hand, by the smoothness of the trend function on  $[t_1, 1]$ , one can group nonzero summands in  $E\{\hat{\mu}_{n,+}(t_1)\}$  and  $E\{\hat{\mu}_{n,+}(t)\}$  according to the closest weight or order and obtain that

$$\sup_{t \in [t_1 - \xi_n, t_1] \cap \mathcal{T}_n} |E\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,+}(t)\} - d_1 P_K(|t_1 - t|/b_n)| = O\{(nb_n)^{-1} + n^{-1}\ell_n\}. \tag{21}$$

Let  $\Omega_n(t_1, t) = \{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,+}(t)\} - \{\hat{\mu}_{n,-}(t_1) - \hat{\mu}_{n,-}(t)\}$ , then by (19)–(21) and a similar argument on  $t \in [t_1, t_1 + \xi_n] \cap \mathcal{T}_n$ , we have

$$\sup_{t \in \mathcal{N}(t_1, \xi_n) \cap \mathcal{T}_n} |\Omega_n(t_1, t) - d_1 P_K(|t_1 - t|/b_n)| = O_p\{(nb_n)^{-1}\ell_n^{1/2} + n^{-1}\ell_n\}.$$

If  $t \notin \mathcal{N}(t_1, \xi_n^\#)$ , then  $(|t - t_1|/b_n)/\{\xi_n + (nb_n)^{-1/2}(\xi_n/b_n)^{1/2}\} \rightarrow \infty$ . Since  $\xi_n = \ell_n/n$  and  $(nb_n)^{-1/2}(\xi_n/b_n)^{1/2} = (nb_n)^{-1}\ell_n^{1/2}$ , we have

$$\text{pr} \left( \inf_{t \in \mathcal{N}(t_1, \xi_n) \cap \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n^\#)} [\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\} - \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\}] > 0 \right)$$

tends to one as  $n \rightarrow \infty$ . Therefore, by (18) we have

$$\text{pr} \left( \inf_{t \in \mathcal{N}(t_1, \xi_n) \cap \mathcal{T}_n \setminus \mathcal{N}(t_1, \xi_n^\#)} [\{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\}^2 - \{\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)\}^2] > 0 \right)$$

also tends to one as  $n \rightarrow \infty$ , and the result follows in view of Lemma A.4.  $\square$

*Proof.* (Theorem 3.3) Let  $\ell_{n,0} = nb_n$  and construct  $\ell_{n,k}$  recursively by

$$\ell_{n,k} = \ell_{n,k-1}b_n + \ell_{n,k-1}^{1/2}, \quad k > 0.$$

Under the stated bandwidth condition,  $b_n = o(n^{-1/\nu_1})$  for some  $\nu_1 > 1$ , and thus  $\ell_{n,0} = o(n^{1-1/\nu_1})$ . Let  $\varsigma_0 = 1 - 1/\nu_1$  and construct  $\varsigma_k$  recursively by

$$\varsigma_k = \max(\varsigma_{k-1} - 1/\nu_1, \varsigma_{k-1}/2), \quad k > 0,$$

then  $\varsigma_k$ ,  $k \geq 0$ , form a monotonic sequence on  $(0, 1)$  and zero is its limit point. Therefore, for any given  $\varepsilon > 0$ , there exists a  $K_{\varepsilon, \nu_1} < \infty$  such that  $\varsigma_k < \varepsilon/2$  holds for all  $k \geq K_{\varepsilon, \nu_1}$ . By induction, we have  $\ell_{n,k} = o(n^{\varsigma_k})$  and thus  $\ell_{n,k} = o(n^{\varepsilon/2})$

for all  $k \geq K_{\varepsilon, \nu_1}$ . We now define  $\ell_{n,k}^\sharp = \ell_{n,k}(\log n)^k$ ,  $k \geq 0$ , then for each  $k > 0$  as  $n \rightarrow \infty$ ,

$$\ell_{n,k}^\sharp / \{\ell_{n,k-1}^\sharp b_n + (\ell_{n,k-1}^\sharp)^{1/2}\} \rightarrow \infty,$$

and  $\ell_{n,K_{\varepsilon, \nu_1}}^\sharp = o(n^\varepsilon)$ . Since  $\ell_{n,0}^\sharp = \ell_{n,0} = nb_n$ , by Lemma A.4,

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ \left| \hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1) \right| > \sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, \ell_{n,1}^\sharp/n)} \left| \hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t) \right| \right\} = 1,$$

and thus by induction and the proof of Lemma A.5 we have

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ \left| \hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1) \right| > \sup_{t \in \mathcal{T}_n \setminus \mathcal{N}(t_1, \ell_{n,K_{\varepsilon, \nu_1}}^\sharp/n)} \left| \hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t) \right| \right\} = 1.$$

Since  $\ell_{n,K_{\varepsilon, \nu_1}}^\sharp/n = o(n^{\varepsilon-1})$  and  $\hat{t}_1$  by definition is the maximizer of  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$ , we have  $\hat{t}_1 \in \mathcal{N}(t_1, n^{\varepsilon-1})$  with probability tending to one as  $n \rightarrow \infty$  and thus (i) follows. For (ii), by the proof of Lemma A.5 we have

$$\sup_{t \in \mathcal{N}(t_1, n^{\varepsilon-1}) \cap \mathcal{T}_n} \left| \{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,+}(t)\} - \{\hat{\mu}_{n,-}(t_1) - \hat{\mu}_{n,-}(t)\} \right| = O_p\{(nb_n)^{-1}n^\varepsilon\}.$$

Recall from the proof of Lemma A.5 that  $\Omega_n(t_1, t) = \{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,+}(t)\} - \{\hat{\mu}_{n,-}(t_1) - \hat{\mu}_{n,-}(t)\}$ , then for any positive sequence  $N_n \rightarrow \infty$  we have

$$\begin{aligned} & \text{pr}\{n^{1-\varepsilon}b_n|\hat{d}_1 - \{\hat{\mu}_{n,+}(t_1) - \hat{\mu}_{n,-}(t_1)\}| \geq N_n\} \\ & \leq \text{pr} \left[ n^{1-\varepsilon}b_n \sup_{t \in \mathcal{N}(t_1, n^{\varepsilon-1}) \cap \mathcal{T}_n} |\Omega_n(t_1, t)| \geq N_n \right] + \text{pr}(\hat{t}_1 - t_1 > n^{\varepsilon-1}), \end{aligned}$$

which goes to zero as  $n \rightarrow \infty$ . Since  $\varepsilon > 0$  can be taken arbitrarily small, the result follows by Theorem 3.1 (ii). □

*Proof.* (Theorem 3.4) Let  $\mathcal{J}_n^\sharp = \cup_{k=1}^M \mathcal{N}(t_k, b_n)$  be the collection of jump neighborhoods, then by the proof of Theorem 3.2,

$$\sup_{t \in \mathcal{T}_n \setminus \mathcal{J}_n^\sharp} |\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)| = O_p\{(nb_n)^{-1/2}(\log n)^{1/2} + b_n^3\},$$

and thus

$$\lim_{n \rightarrow \infty} \text{pr} \left\{ \sup_{t \in \mathcal{T}_n \setminus \mathcal{J}_n^\sharp} \left| \hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t) \right| \geq \varphi_n \right\} = 0.$$

On the other hand, by the properties of local linear estimators and the proof of Theorem 3.1 (ii), we have

$$\lim_{n \rightarrow \infty} \text{pr} \left[ \min_{k \in \{1, \dots, M\}} \left| \hat{\mu}_{n,+}(t_k) - \hat{\mu}_{n,-}(t_k) \right| \geq \varphi_n \right] = 1.$$

Therefore, with probability tending to one, there will be  $M$  separate regions where  $|\hat{\mu}_{n,+}(t) - \hat{\mu}_{n,-}(t)|$  exceeds the threshold  $\varphi_n$ . Since the window size  $b_n \rightarrow 0$  as  $n \rightarrow \infty$ , (i) follows. The second claim follows by applying the argument of Theorem 3.3 on  $\mathcal{N}(t_k, b_n)$ ,  $k = 1, \dots, M$ , respectively.  $\square$

*Proof.* (Theorem 4.1) Let  $\lambda_i^{\natural}, i = 1, \dots, n$ , be defined as in (9) but with  $(e_j)$  therein replaced by  $(e_j^{\natural})$ , then

$$\begin{aligned} |\lambda_i - \lambda_i^{\natural}| &\leq \left( |e_i^{\natural}| + \left| \sum_{j=1}^n e_j^{\natural} \mathbb{1}_{\{0 < i/n-j/n \leq \tau_n \varrho_n\}} \right| + \left| \sum_{j=1}^n e_j^{\natural} \mathbb{1}_{\{0 > i/n-j/n > -\tau_n \varrho_n\}} \right| \right) \\ &\quad \times 2|e_i - e_i^{\natural}| + 2|e_i| \sum_{j=1}^n |e_j - e_j^{\natural}| \mathbb{1}_{\{|i/n-j/n| \leq \tau_n \varrho_n\}}. \end{aligned}$$

Hence, by (4) and the proof of Lemma A.1 (i), we have

$$\max_{1 \leq i \leq n} \|\lambda_i - \lambda_i^{\natural}\|_1 = O\{n^{-1}(n\tau_n \varrho_n)^{1/2} + \tau_n \varrho_n\},$$

and (i) follows by Proposition 4.1 of Zhang and Wu (2012). For (ii), note that by (4),

$$\max_{1 \leq i \leq n} |e_i - e_i^{\natural}| \leq \left\{ \sum_{i=1}^n (e_i - e_i^{\natural})^4 \right\}^{1/4} = O_p(n^{-3/4}),$$

and thus  $\max_{1 \leq i \leq n} |e_i| \leq \max_{1 \leq i \leq n} |e_i^{\natural}| + \max_{1 \leq i \leq n} |e_i - e_i^{\natural}| = O_p(n^{1/4})$  and

$$\begin{aligned} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n |e_j - e_j^{\natural}| \mathbb{1}_{\{|i/n-j/n| \leq \tau_n \varrho_n\}} \right| &\leq \left( \max_{1 \leq k \leq n} |e_k - e_k^{\natural}| \right) (2n\tau_n \varrho_n + 1) \\ &= O_p(n^{1/4} \tau_n \varrho_n). \end{aligned}$$

In addition, by the proof of Theorem 3.2, we have

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n e_j^{\natural} \mathbb{1}_{\{0 < i/n-j/n \leq \tau_n \varrho_n\}} \right| = O_p\{(n\tau_n \varrho_n \log n)^{1/2} + n^{1/4}\},$$

and thus

$$\max_{1 \leq i \leq n} |\lambda_i - \lambda_i^{\natural}| = O_p\{n^{-1/4}(\tau_n \varrho_n \log n)^{1/2} + n^{-1/2} + n^{1/2} \tau_n \varrho_n\}.$$

Since  $\tau_n \rightarrow 0$  and  $n\tau_n \varrho_n \rightarrow \infty$ , it suffices to prove that the same result holds for

$$\hat{g}^{\natural}(t) = \sum_{i=1}^n \lambda_i^{\natural} w_{i, \tau_n}(t).$$

For this, note that

$$\begin{aligned} \|\mathcal{P}_{i-l}(e_i^{\natural}e_j^{\natural})\| &= \|E(e_i^{\natural}e_j^{\natural} \mid \mathcal{F}_{i-l}) - E(e_i^{\natural}e_j^{\natural} \mid \mathcal{F}_{i-l-1})\| \\ &\leq (\theta_{l,4} + \theta_{l+j-i,4}) \sup_{t \in [0,1]} \|G(t; \mathcal{F}_0)\|_4, \end{aligned}$$

we have

$$\begin{aligned} \|\mathcal{P}_{i-l}\lambda_i^{\natural}\| &\leq 2 \sum_{j=1}^n \|\mathcal{P}_{i-l}(e_i^{\natural}e_j^{\natural})\| \mathbb{1}_{\{|i-j| \leq n\tau_n \varrho_n\}} \\ &\leq 2\{(2n\tau_n \varrho_n + 1)\theta_{l,4} + \Theta_{l+[n\tau_n \varrho_n],4} - \Theta_{l-[n\tau_n \varrho_n],4}\} \\ &\quad \times \sup_{t \in [0,1]} \|G(t; \mathcal{F}_0)\|_4. \end{aligned} \tag{22}$$

For any fixed  $l \in \mathbb{Z}$ ,  $(\mathcal{P}_{i-l}\lambda_i^{\natural})_{i=1}^n$  forms a sequence of martingale differences, and by Doob’s inequality,

$$\left\| \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \mathcal{P}_{i-l}\lambda_i^{\natural} \right| \right\| \leq 2 \left\| \sum_{i=1}^n \mathcal{P}_{i-l}\lambda_i^{\natural} \right\| = 2 \left( \sum_{i=1}^n \|\mathcal{P}_{i-l}\lambda_i^{\natural}\|^2 \right)^{1/2}. \tag{23}$$

Let  $C$  denote an absolute constant whose value may vary from place to place. Note that  $\hat{g}^{\natural}(t) - E\{\hat{g}^{\natural}(t)\} = \sum_{l \in \mathbb{Z}} \sum_{i=1}^n \mathcal{P}_{i-l}\lambda_i^{\natural} w_{i,\tau_n}(t)$  and

$$\sum_{l \in \mathbb{Z}} \{\Theta_{l+[n\tau_n \varrho_n],4} - \Theta_{l-[n\tau_n \varrho_n],4}\} \leq (2n\tau_n \varrho_n + 1)\Theta_{0,4},$$

by (22), (23) and the summation by parts formula,

$$\left\| \sup_{t \in [0,1]} |\hat{g}^{\natural}(t) - E\{\hat{g}^{\natural}(t)\}| \right\| \leq \frac{C}{n\tau_n} \sum_{l \in \mathbb{Z}} \left\| \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \mathcal{P}_{i-l}\lambda_i^{\natural} \right| \right\| \leq Cn^{1/2} \varrho_n \Theta_{0,4}.$$

By the proof of Proposition 4.1 of Zhang and Wu (2012),

$$\max_{1 \leq i \leq n} |E(\lambda_i^{\natural}) - g(i/n)| = O\{(n\tau_n \varrho_n)^{-(\iota-1)} + (\tau_n \varrho_n)^{(\iota-1)/\iota}\},$$

(ii) follows by properties of local linear estimates. □

*Proof.* (Theorem 4.2) Let  $\hat{\lambda}_i$ ,  $i = 1, \dots, n$ , be defined as in (9) but with  $(e_j)$  therein replaced by  $(\hat{e}_j)$ , then

$$\begin{aligned} |\lambda_i - \hat{\lambda}_i| &\leq \left( |e_i| + \left| \sum_{j=1}^n e_j \mathbb{1}_{\{0 < i/n-j/n \leq \tau_n \varrho_n\}} \right| + \left| \sum_{j=1}^n e_j \mathbb{1}_{\{0 > i/n-j/n > -\tau_n \varrho_n\}} \right| \right) \\ &\quad \times 2|e_i - \hat{e}_i| + 2|\hat{e}_i| \sum_{j=1}^n |e_j - \hat{e}_j| \mathbb{1}_{\{|i/n-j/n| \leq \tau_n \varrho_n\}}. \end{aligned}$$

By the proof of Theorem 3.2, we have

$$\max_{1 \leq i \leq n} |e_i - \hat{e}_i| = O_p\{(nb_n)^{-1/2}(\log n)^{1/2} + b_n^2\},$$

and thus  $\max_{1 \leq i \leq n} |\hat{e}_i| \leq \max_{1 \leq i \leq n} |e_i| + \max_{1 \leq i \leq n} |e_i - \hat{e}_i| = O_p(n^{1/4})$  and

$$\sum_{j=1}^n |e_j - \hat{e}_j| \mathbb{1}_{\{|i/n-j/n| \leq \tau_n \varrho_n\}} \leq \left( \max_{1 \leq j \leq n} |e_j - \hat{e}_j| \right) (2n\tau_n \varrho_n + 1) = O_p(n^{-1/4} \chi_n).$$

In addition, by the proof of Theorem 4.1, we have

$$\begin{aligned} & \max_{1 \leq i \leq n} \left| \sum_{j=1}^n e_j \mathbb{1}_{\{0 < i/n-j/n \leq \tau_n \varrho_n\}} \right| \\ & \leq \max_{1 \leq i \leq n} \left| \sum_{j=1}^n e_j^{\sharp} \mathbb{1}_{\{0 < i/n-j/n \leq \tau_n \varrho_n\}} \right| + O_p(n^{1/4} \tau_n \varrho_n) \\ & = O_p\{(n\tau_n \varrho_n \log n)^{1/2} + n^{1/4}\}, \end{aligned}$$

and thus

$$\max_{1 \leq i \leq n} |\lambda_i - \hat{\lambda}_i| = O_p\{[(nb_n)^{-1/2}(\log n)^{1/2} + b_n^2]\{(n\tau_n \varrho_n \log n)^{1/2} + n^{1/4}\} + \chi_n\}.$$

Since  $n\tau_n \varrho_n \rightarrow \infty$ , the result follows by Theorem 4.1. □

### A.2. Additional Technical Proofs

We shall here provide the proofs for Lemmas A.1–A.3.

*Proof.* (Lemma A.1) For any fixed  $l \geq 0$ , the projections  $\mathcal{P}_{i-l} e_i^{\sharp}$ ,  $i = 1, \dots, n$ , form a sequence of martingale differences. Let  $\zeta_k(t) = G(t; \mathcal{F}_k)$ ,  $D_{\pm}(v) = (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} (\kappa_2 \mp \kappa_1 v) K_{\pm}(v)$ ,  $v \in \mathbb{R}$ , and

$$M_{n,\pm}(t) = (nb_n)^{-1} \sum_{i=1}^n D_{\pm}\{(i/n - t)/b_n\} \zeta_i(t).$$

Since  $\|\mathcal{P}_{i-l} \zeta_i(t)\| = \|\mathcal{P}_0 \zeta_l(t)\| \leq \theta_{l,2}$  and  $\|\mathcal{P}_{i-l}\{e_i^{\sharp} - \zeta_i(t)\}\| \leq \|e_i^{\sharp} - \zeta_i(t)\| \leq c_0|i/n - t|$  by (A2), we have

$$\begin{aligned} & \|R_{n,\pm}^{\sharp}(t) - M_{n,\pm}(t)\| \\ & \leq (nb_n)^{-1} \sum_{l=0}^{\infty} \left[ \sum_{i=1}^n |D_{\pm}\{(i/n - t)/b_n\}|^2 \|\mathcal{P}_{i-l}\{e_i^{\sharp} - \zeta_i(t)\}\|^2 \right]^{1/2} \\ & \leq (nb_n)^{-1/2} (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \left\{ 2 \sup_{v \in [-1,1]} |K(v)| \right\}^2 \sum_{l=0}^{\infty} \min(\theta_{l,2}, b_n). \end{aligned}$$



Since  $b_n \rightarrow \infty$  and  $\sum_{l=0}^\infty \theta_{l,2} = \Theta_{0,2} < \infty$ , by the dominated convergence theorem we have

$$\sup_{t \in (0,1)} \|R_{n,\pm}^h(t) - M_{n,\pm}(t)\| = o\{(nb_n)^{-1/2}\}. \tag{24}$$

Note that

$$\begin{aligned} \|M_{n,\pm}(t)\| &\leq (nb_n)^{-1} \sum_{l=0}^\infty \left[ \sum_{i=1}^n |D_\pm\{(i/n - t)/b_n\}|^2 \|\mathcal{P}_{i-l}\zeta_i(t)\|^2 \right]^{1/2} \\ &\leq (nb_n)^{-1/2} \Theta_{0,2} (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \left\{ 2 \sup_{v \in [-1,1]} |K(v)| \right\}^2, \end{aligned}$$

(i) follows. For (ii), let  $\tilde{\zeta}_k(t) = E\{\zeta_k(t) | \mathcal{F}_{k-m,k}\}$  where  $\mathcal{F}_{k-m,k} = (\epsilon_{k-m}, \dots, \epsilon_k)$ , then

$$\tilde{\zeta}_k(t) - \zeta_k(t) = \sum_{l=m}^\infty [E\{\zeta_k(t) | \mathcal{F}_{k-l,k}\} - E\{\zeta_k(t) | \mathcal{F}_{k-l-1,k}\}]$$

and  $\|E\{\zeta_k(t) | \mathcal{F}_{k-l,k}\} - E\{\zeta_k(t) | \mathcal{F}_{k-l-1,k}\}\| \leq \theta_{l,2}$ . Let

$$\tilde{M}_{n,\pm}(t) = (nb_n)^{-1} \sum_{i=1}^n D_\pm\{(i/n - t)/b_n\} \tilde{\zeta}_i(t),$$

then we have

$$(nb_n)^{1/2} \|M_{n,\pm}(t) - \tilde{M}_{n,\pm}(t)\| \leq (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \left\{ 2 \sup_{v \in [-1,1]} |K(v)| \right\}^2 \sum_{l=m}^\infty \theta_{l,2}, \tag{25}$$

which goes to zero as  $m \rightarrow \infty$ . Note that  $\int_{-1}^1 D_+(v)^2 dv = \int_{-1}^1 D_-(v)^2 dv = \phi$  and for each  $t \in (0, 1)$ ,  $\{\tilde{\zeta}_k(t)\}_{k \in \mathbb{Z}}$  forms an  $m$ -dependent stationary process, (ii) follows by (24), (25) and the central limit theorem for  $m$ -dependent random variables; see for example Hoeffding and Robbins (1948). Note that  $\max_{1 \leq i \leq n} \|e_i - e_i^h\| = O(n^{-1})$  by (4), we have  $\sup_{t \in (0,1)} \|R_{n,\pm}^h(t) - R_{n,\pm}(t)\| = O(n^{-1})$ , and (iii) follows.  $\square$

*Proof.* (Lemma A.2) Let  $\mathbb{B}(v)$ ,  $v \in \mathbb{R}$ , be a standard Brownian motion, and

$$\Upsilon(u) = (2\phi)^{-1/2} \int_{\mathbb{R}} \{D_+(v - u) - D_-(v - u)\} d\mathbb{B}(v), \quad u \in \mathbb{R}.$$

Then  $\Upsilon(u)$ ,  $u \in \mathbb{R}$ , is a stationary Gaussian process with mean zero and marginal variance  $(2\phi)^{-1} \int_{\mathbb{R}} \{D_+(v - u) - D_-(v - u)\}^2 dv \equiv 1$ . In addition, let  $D_s(v) = \{D_+(v) - D_-(v)\} - \{D_+(v - s) - D_-(v - s)\}$ ,  $v \in \mathbb{R}$ , then the covariance satisfies

$$\text{cov}\{\Upsilon(u), \Upsilon(u + s)\} = 1 - (4\phi)^{-1} \int_{\mathbb{R}} \{D_s(v - u)\}^2 dv = 1 - (2\phi)^{-1} C_K s + o(s).$$

Since the process  $\{\Upsilon(u)\}_{u \in [1, b_n^{-1} - 1]}$  is identically distributed as  $\{\Upsilon(u)\}_{u \in [0, b_n^{-1} - 2]}$  from the stationarity of  $\Upsilon(u)$ ,  $u \in \mathbb{R}$ , by Corollary A1 of Bickel and Rosenblatt (1973),

$$\begin{aligned} & \text{pr} \left\{ \sup_{u \in [1, b_n^{-1} - 1]} |\Upsilon(u)| - B_K(b_n^{-1} - 2) \leq \frac{u}{\{2 \log(b_n^{-1} - 2)\}^{1/2}} \right\} \\ & \rightarrow \exp\{-2 \exp(-u)\}. \end{aligned}$$

Note that  $B_K(b_n^{-1} - 2) = B_K(m_n) + o(b_n)$  and  $\log(b_n^{-1} - 2) = \log(m_n) + O(b_n)$ , one obtains that

$$\text{pr} \left\{ \sup_{u \in [1, b_n^{-1} - 1]} |\Upsilon(u)| - B_K(m_n) \leq \frac{u}{(2 \log m_n)^{1/2}} \right\} \rightarrow \exp\{-2 \exp(-u)\}. \quad (26)$$

For  $u \in \mathbb{R}$ , let

$$\tilde{\Upsilon}(u) = (2\phi)^{-1/2} \int_{\mathbb{R}} \left\{ D_+ \left( \frac{\lfloor 1 + nb_n v \rfloor}{nb_n} - u \right) - D_- \left( \frac{\lfloor 1 + nb_n v \rfloor}{nb_n} - u \right) \right\} d\mathbb{B}(v),$$

then by the chaining argument in the proof of Lemma 2 of Wu and Zhao (2007), we have  $\sup_{u \in [1, b_n^{-1} - 1]} |\Upsilon(u) - \tilde{\Upsilon}(u)| = O_p\{(nb_n)^{-1/2} \log n\}$  and thus (26) holds with  $\Upsilon(u)$  therein replaced by  $\tilde{\Upsilon}(u)$ . By the scaling property of Brownian motion, the processes  $\{Q_n(t)\}_{t \in \mathcal{T}_n}$  and  $\{\tilde{\Upsilon}(t/b_n)\}_{t \in \mathcal{T}_n}$  have the same joint distribution, and Lemma A.2 follows.  $\square$

*Proof.* (Lemma A.3) Recall from the proof of Lemma A.1 that  $\zeta_k(t) = G(t; \mathcal{F}_k)$ . By the orthogonality of martingale differences,

$$\begin{aligned} |E\{\zeta_0(t)\zeta_k(t)\}| &= \left| \sum_{l \in \mathbb{Z}} E\{\mathcal{P}_l \zeta_0(t) \mathcal{P}_l \zeta_k(t)\} \right| \\ &\leq \sum_{l \in \mathbb{Z}} \|\mathcal{P}_l \zeta_0(t)\| \|\mathcal{P}_l \zeta_k(t)\| \leq \sum_{i=0}^{\infty} \theta_{i,2} \theta_{i+|k|,2}. \end{aligned}$$

Therefore, for any  $t_1, t_2 \in [0, 1]$  we have

$$|E\{\zeta_0(t_1)\zeta_k(t_1)\} - E\{\zeta_0(t_2)\zeta_k(t_2)\}| \leq 2 \sum_{i=0}^{\infty} \theta_{i,2} \theta_{i+|k|,2}. \quad (27)$$

On the other hand, let  $c_1 = \sup_{t \in [0,1]} \|\zeta_0(t)\|$ , then by the stochastic Lipschitz continuous condition (A2),

$$|E\{\zeta_0(t_1)\zeta_k(t_1)\} - E\{\zeta_0(t_2)\zeta_k(t_2)\}| \leq 2c_0 c_1 |t_1 - t_2|. \quad (28)$$

Let  $\varpi_n = \lfloor b_n^{-1/3} \rfloor$ , then by (27) and (28),

$$\begin{aligned} \sup_{|t_1 - t_2| \leq b_n} |g(t_1) - g(t_2)| &\leq 4 \sum_{k=0}^{\infty} \min \left( c_0 c_1 |t_1 - t_2|, \sum_{i=0}^{\infty} \theta_{i,2} \theta_{i+|k|,2} \right) \\ &\leq 4\varpi_n c_0 c_1 |t_1 - t_2| + 4\Theta_{0,2} \Theta_{\varpi_n+1,2} = O(b_n^{1/3}). \quad (29) \end{aligned}$$

Let  $Q_n(t)$ ,  $t \in (0, 1)$ , be defined as in Lemma A.2 and

$$\check{Q}_n(t) = \{2\phi nb_n g(t)\}^{-1/2} \sum_{i=1}^n [D_+\{(i/n-t)/b_n\} - D_-\{(i/n-t)/b_n\}]\{g(i/n)\}^{1/2} Z_i,$$

where  $Z_1, \dots, Z_n$  are iid standard normal random variables. Then by (29),  $\sup_{t \in \mathcal{T}_n} |\check{Q}_n(t) - Q_n(t)| = O_p\{b_n^{1/3}(\log n)^{1/2}\}$  and thus by Lemma A.2,

$$\text{pr} \left\{ \sup_{t \in \mathcal{T}_n} |\check{Q}_n(t)| - B_K(m_n) \leq \frac{u}{(2 \log m_n)^{1/2}} \right\} \rightarrow \exp\{-2 \exp(-u)\}. \quad (30)$$

Let  $S_0^\ddagger = 0$  and  $S_i^\ddagger = \sum_{j=1}^i e_j^\ddagger$ ,  $i = 1, \dots, n$ . By Theorem 2 in Zhou and Wu (2010), on a richer probability space, there exists a process  $(S_i^\dagger)_{i=0}^n$  and iid standard normal random variables  $Z_1, \dots, Z_n$  such that  $(S_i^\dagger)_{i=0}^n$  has the same joint distribution as  $(S_i^\ddagger)_{i=0}^n$  and

$$\max_{1 \leq i \leq n} \left| S_i^\dagger - \sum_{j=1}^i \{g(j/n)\}^{1/2} Z_j \right| = o_p(n^{3/10} \log n).$$

By the technique of summation by parts, we have

$$\begin{aligned} & \sup_{t \in \mathcal{T}_n} \left| (nb_n)^{-1} \sum_{i=1}^n D_\pm\{(i/n-t)/b_n\} [S_i^\dagger - S_{i-1}^\dagger - \{g(j/n)\}^{1/2} Z_i] \right| \\ &= o_p(n^{-7/10} b_n^{-1} \log n). \end{aligned}$$

Let  $e_j^\dagger = S_j^\dagger - S_{j-1}^\dagger$ ,  $j = 1, \dots, n$ , and

$$R_{n,\pm}^\dagger(t) = (nb_n)^{-1} \sum_{i=1}^n (\kappa_0 \kappa_2 - \kappa_1^2)^{-1} \{\kappa_2 \mp \kappa_1(i/n-t)/b_n\} K_\pm\{(i/n-t)/b_n\} e_i^\dagger,$$

then we have

$$\sup_{t \in \mathcal{T}_n} \left| \frac{(nb_n)^{1/2} \{R_{n,+}^\dagger(t) - R_{n,-}^\dagger(t)\}}{\{2\phi g(t)\}^{1/2}} - \check{Q}_n(t) \right| = o_p(n^{-1/5} b_n^{-1/2} \log n).$$

Since  $(\log n)^{1/2} n^{-1/5} b_n^{-1/2} \log n \rightarrow 0$  from (7), (i) follows by (30). For (ii), let  $S_0 = 0$  and  $S_i = \sum_{j=1}^i e_j$ ,  $i = 1, \dots, n$ , then by (4),

$$\left\| \max_{1 \leq i \leq n} |S_i - S_i^\ddagger| \right\| \leq \sum_{i=1}^n \|e_i - e_i^\ddagger\| = O(1).$$

Therefore, by the technique of summation by parts, we have

$$\sup_{t \in [0,1]} |R_{n,\pm}(t) - R_{n,\pm}^\ddagger(t)| = O_p\{(nb_n)^{-1}\},$$

and (ii) follows. □

## Acknowledgments

I am grateful to the Editor, the Associate Editor, and an anonymous referee for their helpful comments and suggestions.

## References

- AUE, A. AND HORVÁTH, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, **34**, 1–16. [MR3008012](#)
- BICKEL, P. J. AND LEVINA, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, **36**, 199–227. [MR2387969](#)
- BICKEL, P. J. AND ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, **1**, 1071–1095. [MR0348906](#)
- CHEN, B. AND HONG, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, **80**, 1157–1183. [MR2963885](#)
- DAHLHAUS, R. (1996). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications*, **62**, 139–168. [MR1388767](#)
- DAHLHAUS, R. AND SUBBA RAO, S. (2006). Statistical inference for time-varying arch processes. *The Annals of Statistics*, **34**, 1075–1114. [MR2278352](#)
- DEGRAS, D., XU, Z., ZHANG, T. AND WU, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing*, **60**, 1087–1097. [MR2858979](#)
- DRAGHICESCU, D., GUILLAS, S. AND WU, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *Journal of Computational and Graphical Statistics*, **18**, 1–20. [MR2511058](#)
- ELSNER, J. B., KOSSIN, J. P. AND JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, **455**, 92–95.
- FAN, J. AND GIJBELS, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman & Hall, London. [MR1383587](#)
- FAN, J. AND ZHANG, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, **27**, 1491–1518. [MR1742497](#)
- FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, **42**, 2243–2281. [MR3269979](#)
- GAO, J. AND GIJBELS, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, **103**, 1584–1594. [MR2504206](#)
- GAO, J., GIJBELS, I. AND VAN BELLEGEM, S. (2008). Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics*, **143**, 123–142. [MR2384436](#)
- GIJBELS, I., LAMBERT, A. AND QIU, P. (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics*, **59**, 235–272. [MR2394168](#)

- GRÉGOIRE, G. AND HAMROUNI, Z. (2002a). Change point estimation by local linear smoothing. *Journal of Multivariate Analysis*, **83**, 56–83. [MR1934975](#)
- GRÉGOIRE, G. AND HAMROUNI, Z. (2002b). Two non-parametric tests for change-point problems. *Journal of Nonparametric Statistics*, **14**, 87–112. [MR1905587](#)
- GUINNESS, J. AND STEIN, M. L. (2013). Interpolation of nonstationary high frequency spatial-temporal temperature data. *The Annals of Applied Statistics*, **7**, 1684–1708. [MR3127964](#)
- HALL, P. AND HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, **85**, 1039–1049. [MR1134500](#)
- HALL, P. AND TITTERINGTON, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics*, **34**, 429–440. [MR1190262](#)
- HÄRDLE, W. AND MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, **21**, 1926–1947. [MR1245774](#)
- HOEFFDING, W. AND ROBBINS, H. (1948). The central limit theorem for dependent random variables. *Duke Mathematical Journal*, **15**, 773–780. [MR0026771](#)
- HOROWITZ, J. L. AND SPOKOINY, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599–631. [MR1828537](#)
- JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. AND LIU, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, **34**, 423–446. [MR3070866](#)
- JOO, J.-H. AND QIU, P. (2009). Jump detection in a regression curve and its derivative. *Technometrics*, **51**, 289–305. [MR2562274](#)
- KULASEKERA, K. B. AND WANG, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *Journal of the American Statistical Association*, **92**, 500–511. [MR1467844](#)
- LOADER, C. R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics*, **24**, 1667–1678. [MR1416655](#)
- MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics*, **20**, 737–761. [MR1165590](#)
- MÜLLER, H.-G. AND STADTMÜLLER, U. (1999). Discontinuous versus smooth regression. *The Annals of Statistics*, **27**, 299–337. [MR1701113](#)
- NASON, G. P., VON SACHS, R. AND KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 271–292. [MR1749539](#)
- OMBAO, H., VON SACHS, R. AND GUO, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, **100**, 519–531. [MR2160556](#)
- QIU, P. AND YANDELL, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics*, **40**, 141–152. [MR1626927](#)
- SHAO, X. AND ZHANG, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, **105**, 1228–1240. [MR2752617](#)
- SPOKOINY, V. G. (1998). Estimation of a function with discontinuities via local

- polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26**, 1356–1378. [MR1647669](#)
- SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability*, **38**, 1155–1172. [MR2285698](#)
- VOGT, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, **40**, 2601–2633. [MR3097614](#)
- VOGT, M. AND DETTE, H. (2015). Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, **43**, 713–740. [MR3319141](#)
- WANG, L. (2008). Nonparametric test for checking lack-of-fit of quantile regression model under random censoring. *Canadian Journal of Statistics*, **36**, 321–336. [MR2431684](#)
- WANG, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, **93**, 341–348.
- WU, J. S. AND CHU, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics*, **21**, 1545–1566. [MR1241278](#)
- WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14150–14154. [MR2172215](#)
- WU, W. B. AND POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, **19**, 1755–1768. [MR2589209](#)
- WU, W. B., WOODROOFE, M. AND MENTZ, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika*, **88**, 793–804. [MR1859410](#)
- WU, W. B. AND ZHAO, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, 391–410. [MR2323759](#)
- ZHANG, T. AND WU, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika*, **98**, 599–614. [MR2836409](#)
- ZHANG, T. AND WU, W. B. (2012). Inference of time-varying regression models. *The Annals of Statistics*, **40**, 1376–1402. [MR3015029](#)
- ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, **108**, 726–740. [MR3174655](#)
- ZHOU, Z. AND WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics*, **37**, 2696–2729. [MR2541444](#)
- ZHOU, Z. AND WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 513–531. [MR2758526](#)