

Inference of trends in time series

Wei Biao Wu and Zhibiao Zhao

University of Chicago, USA

[Received November 2005. Final revision November 2006]

Summary. We consider statistical inference of trends in mean non-stationary models. A test statistic is proposed for the existence of structural breaks in trends. On the basis of a strong invariance principle of stationary processes, we construct simultaneous confidence bands with asymptotically correct nominal coverage probabilities. The results are applied to global warming temperature data and Nile river flow data. Our confidence band of the trend of the global warming temperature series supports the claim that the trend is increasing over the last 150 years.

Keywords: Confidence bands; Global warming; Invariance principle; Non-linear time series; Non-parametric regression

1. Introduction

An important problem in time series analysis is the estimation of trends. Assume that the data X_1, \dots, X_n are observed from the model

$$X_k = \mu(k/n) + e_k, \quad k = 1, \dots, n, \quad (1)$$

where μ is an unknown regression function defined on $[0, 1]$ and e_k is a mean 0 stationary process. The process X_k is mean non-stationary and can be interpreted as a signal μ plus noise e_k model. The paper has two primary goals. The first is to develop statistical procedures to test whether the trend μ in model (1) has structural breaks or jumps. If the curve μ is smooth, our second goal is to construct simultaneous confidence bands (SCBs) for μ . Throughout the paper we consider *a posteriori* or off-line inference, namely the data have already been collected before the analysis.

For model (1), the classical changepoint analysis concerns testing the null hypothesis $\mu_1 = \dots = \mu_n$ against the alternative of one or multiple changepoints,

$$\mu_1 = \dots = \mu_{k_1} \neq \mu_{k_1+1} = \dots = \mu_{k_2} \neq \mu_{k_2+1} = \dots \neq \mu_{k_J+1} = \dots = \mu_n, \quad (2)$$

where $\mu_k = \mu(k/n)$ and k_1, \dots, k_J are called changepoints. The alternative hypothesis says that μ is piecewise constant. Here we shall generalize the classical setting of piecewise constant functions to piecewise Lipschitz continuous functions. The latter setting seems more reasonable in practical situations in which trends are expected to change smoothly instead of staying at the same level between successive abrupt events. Let $\mu(t)$, $t \in [0, 1]$, be a piecewise Lipschitz continuous function. Discontinuous points of μ are called structural breaks. In practice, structural breaks may be caused by sudden events, abrupt policy changes and catastrophes among others.

Address for correspondence: Wei Biao Wu, Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637, USA.
E-mail: wbwu@galton.uchicago.edu

Non-parametric inference of regression functions with jumps has been an active area of research. It would be impossible to have a complete list here and we mention only some representatives: Müller (1992), Wu and Chu (1993), Qiu and Yandell (1998), Spokoiny (1998), Müller and Stadtmüller (1999), Grégoire and Hamrouni (2002), Qiu (2003) and Gijbels and Goderniaux (2004). See also references therein for further information. In the majority of the above-mentioned results, the errors e_k are assumed to be independent. The independence assumption is a serious restriction and it excludes many important applications. The restriction is particularly problematic in time series analysis in which dependence is the rule rather than the exception and is actually one of the main objectives of interest. Tang and MacNeill (1993) argued that serial correlation can seriously affect the distributions of changepoint statistics.

For our second goal of constructing SCBs for μ , we assume that μ is smooth. SCBs can be used to find parametric forms of μ . For example, in the study of global temperature series, an interesting problem is to test whether the trend is linear, quadratic or of other patterns. Under the assumption of independent errors, the construction of SCBs has been discussed by Johnston (1982), Härdle (1989), Knafl *et al.* (1985), Hall and Titterton (1988), Härdle and Marron (1991), Eubank and Speckman (1993), Sun and Loader (1994), Xia (1998), Cummins *et al.* (2001) and Dümbgen (2003) among others. Eubank and Speckman (1993) applied the strong invariance principle of Komlós *et al.* (1975) and constructed SCBs for μ with asymptotically correct nominal values. In the context of kernel density estimation, Bickel and Rosenblatt (1973) obtained SCBs for density functions. The construction of SCBs has been a difficult problem if the errors e_k are dependent. Partial answers are given in Bühlmann (1998). In this paper, by applying the strong invariance principle of stationary processes in Wu (2006), we shall provide a solution to the problem and construct SCBs with asymptotically correct nominal coverage probabilities.

We now introduce some notation. Let I be an interval of \mathbb{R} . A function f is said to be Lipschitz continuous on I , denoted by $f \in L(I)$, if

$$\sup_{x \neq x'} |f(x) - f(x')|/|x - x'| < \infty.$$

Let $\mathcal{C}^m[0, 1]$, $m = 0, 1, \dots$, denote the collection of functions having up to m th-order derivatives. For a function g we say that g has bounded variation if

$$V(g) := \sup \left\{ \sum_i |g(t_i) - g(t_{i-1})| \right\} < \infty,$$

where the supremum is taken over all $\dots < t_{i-1} < t_i < \dots$. Denote by ' \Rightarrow ' convergence in distribution and by $N(m, \sigma^2)$ the normal distribution with mean m and variance σ^2 . For a random variable X write $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$, $p > 0$, and $\|X\| = \|X\|_2$. Write $S_n = \sum_{i=1}^n e_i$ and $S_{-n} = \sum_{i=1}^n e_{-i}$, $n \geq 0$. For two real sequences (a_n) and (b_n) write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} (a_n/b_n) = 1$ and $a_n \asymp b_n$ if $0 < \liminf_{n \rightarrow \infty} |a_n/b_n| \leq \limsup_{n \rightarrow \infty} |a_n/b_n| < \infty$.

The rest of the paper is organized as follows. Structural assumptions on the error sequence e_i are made in Section 2. Section 3 concerns testing the existence of structural breaks in μ . Section 4 discusses the construction of SCBs of μ in the presence of dependent errors e_i . To apply the results in Sections 3 and 4, we need to choose smoothing parameters and to estimate the long run variance of e_i . This problem is discussed in Section 5. A simulation study is carried out in Section 6. Section 7 contains applications in the global warming temperature data and the Nile river data. Proofs are given in Appendix A.

The data and programs that were used to analyse them can be obtained from

<http://www.blackwellpublishing.com/rss>

2. The error structure

We assume that the error process e_i in model (1) is stationary and causal. Let $\varepsilon_i, i \in \mathbb{Z}$, be independent and identically distributed (IID) random variables and G a measurable function such that

$$e_i = G(\dots, \varepsilon_{i-1}, \varepsilon_i) \quad (3)$$

is a proper random variable with mean 0 and finite variance. Let ε'_j be an IID copy of ε_j and $e_i^* = G(\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_i)$. Assume that $\mathbb{E}(|e_i|^p) < \infty$, $p > 2$, and

$$\sum_{i=1}^{\infty} i \|e_i - e_i^*\|_p < \infty. \quad (4)$$

Wu (2006) established the following strong approximation or strong invariance principle. Under inequality (4), there is a standard Brownian motion \mathbb{B} such that, on a richer probability space, S_i can be uniformly approximated by $\sigma \mathbb{B}(i)$:

$$\max_{i \leq n} |S_i - \sigma \mathbb{B}(i)| = o_{AS}\{n^{1/p'} \log(n)\}, \quad p' = \min(4, p), \quad (5)$$

where $\sigma^2 = \sum_{k \in \mathbb{Z}} \mathbb{E}(e_0 e_k)$ is the long run variance. The celebrated strong invariance principle of Komlós *et al.* (1975) asserts that, if e_i are IID, then condition (5) holds with the optimal error bound $o_{AS}(n^{1/p})$ and $\sigma = \|e_i\|$. In our problem the results by Komlós *et al.* (1975) are not applicable owing to the dependence between the e_i .

The strong invariance principle is a very useful tool to access asymptotic properties of S_n and it plays an important role in the related asymptotic inference since Brownian motions have many nice analytical and probabilistic properties. Wu (2005) defined $\|e_i - e_i^*\|_p$ as the physical dependence measure which quantifies the degree of dependence of outputs on inputs. Wu and Shao (2004) showed that, for a variety of non-linear time series models, $\|e_n - e_n^*\|_p = O(r^n)$ for some $r \in (0, 1)$ and hence inequality (4) trivially holds. Consider the autoregressive moving average ARMA process

$$e_n - \sum_{i=1}^l \psi_i e_{n-i} = \sum_{j=0}^q \theta_j \varepsilon_{n-j},$$

where ψ_1, \dots, ψ_l and $\theta_0, \dots, \theta_q$ are real parameters. If the roots of the equation

$$\lambda^l - \sum_{i=1}^l \psi_i \lambda^{l-i} = 0$$

are all inside the unit disc, then $e_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$ with $|a_i| = O(r^i)$ for some $r \in (0, 1)$ and thus inequality (4) also holds.

With the help of the strong invariance principle (5), we can conduct a systematic study of the asymptotic properties of estimates of μ . In the rest of the paper it is always implicitly assumed that condition (4), and consequently condition (5), holds with $p=4$ and $\sigma > 0$.

Non-parametric inference of μ in model (1) typically involves the quantity

$$Y_n(t) = \sum_{i=1}^n w_n(t, i) e_i, \quad (6)$$

where $w_n(t, i)$ are suitable weights. To see how to apply condition (5) to equation (6), we introduce

$$\Omega_n(t) = |w_n(t, 1)| + \sum_{i=2}^n |w_n(t, i) - w_n(t, i-1)|, \quad \Omega_n = \max_{0 \leq t \leq 1} \{\Omega_n(t)\}, \quad (7)$$

and the Gaussian process

$$Y_n^\diamond(t) = \sum_{i=1}^n w_n(t, i) \sigma \{\mathbb{B}(i) - \mathbb{B}(i-1)\}. \quad (8)$$

Since condition (5) holds with $p=4$, using the summation by parts formula, we have

$$|Y_n(t) - Y_n^\diamond(t)| \leq \Omega_n(t) \max_{i \leq n} |S_i - \sigma \mathbb{B}(i)| = o_{AS}\{\Omega_n(t)n^{1/4} \log(n)\} \quad (9)$$

and the uniform approximation

$$\max_{0 \leq t \leq 1} |Y_n(t) - Y_n^\diamond(t)| = o_{AS}\{\Omega_n n^{1/4} \log(n)\}. \quad (10)$$

If $w_n(t, i)$ is sufficiently smooth in i , then $\Omega_n(t)$ and Ω_n have tractable bounds. For example, for the Priestley–Chao estimate (see equation (18) in Section 4), we have

$$w_n(t, i) = \frac{K\{(t-i/n)/b_n\}}{nb_n}$$

and hence $\Omega_n = O\{(nb_n)^{-1}\}$ if K has bounded variation. For local linear estimates (see Section 4.1 or Fan and Gijbels (1996)), if K is Lipschitz continuous and has bounded support, elementary calculations show that $\Omega_n = O\{(nb_n)^{-1}\}$ also holds. Thus, with properly chosen b_n , the asymptotic properties of $Y_n(t)$ follow from those of $Y_n^\diamond(t)$. In other words, model (1) can be reduced to the conventional model

$$X_k^\diamond = \mu(k/n) + \sigma Z_k, \quad k = 1, \dots, n, \quad (11)$$

where Z_k are IID standard normal distributions and σ and μ are unknown. This idea is implemented in Sections 3 and 4.

3. Inference of structural breaks

Let $\text{PL}[0, 1]$ be the set of piecewise Lipschitz continuous functions on $[0, 1]$ with a finite number of jumps. For a formal definition, $f \in \text{PL}[0, 1]$ if there are $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = 1$ with $k < \infty$ such that f is Lipschitz continuous on the interval $[t_j, t_{j+1})$, $j = 0, \dots, k$, and the jumps $f(t_l) - f(t_l-) \neq 0$, $1 \leq l \leq k$. Here the left-hand limit $f(t-) = \lim_{u \uparrow t} \{f(u)\}$. We generically call jumps structural breaks. Assume that $\mu \in \text{PL}[0, 1]$.

Important problems in the inference of structural breaks include

- (a) testing the hypothesis of no structural change $H_0: \mu \in L[0, 1]$ and
- (b) estimating the locations and sizes of structural breaks.

Various aspects of the second problem have been studied; see Pettitt (1980), Csörgő and Horváth (1997), Lavielle (1999) and Davis *et al.* (2006) among others. In this paper we focus on the first problem. The null hypothesis implies a smooth change in the trend: for some constant $c > 0$, $|\mu_{k+1} - \mu_k| \leq c/n$ holds for all $n \geq 2$ and $1 \leq k < n$. In comparison, in the classical changepoint inference, μ_k does not change.

The formulation $\mu \in \text{PL}[0, 1]$ is more general than that in the classical setting. However, without the piecewise constancy assumption, we can only use local information since a Lipschitz continuous function can be locally approximated by a constant. For example, to test whether a given $t \in (0, 1)$ is a discontinuous point, we can compare the local averages of X_j over

$nt < j < nt + k_n$ and over $nt - k_n < j < nt$, where k_n is the block length satisfying $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$. If the two averages are close, then t is unlikely to be a discontinuous point. A global measure of the discrepancy is

$$D_n^* = \frac{1}{k_n} \max_{k_n \leq i \leq n-k_n} \left| \sum_{j=i+1}^{k_n+i} X_j - \sum_{j=i-k_n+1}^i X_j \right|. \quad (12)$$

A non-overlapping version of D_n^* is given by

$$D_n = \max_{1 \leq i \leq m-1} |A_i - A_{i-1}|, \quad A_i = A_{i,n} = \frac{1}{k_n} \sum_{j=1}^{k_n} X_{j+ik_n}. \quad (13)$$

Here $m = \lfloor n/k_n \rfloor$ is the largest integer not exceeding n/k_n . Let $\omega_i = \mathbb{E}(A_i)$. For $1 \leq i \leq m$ let the interval $I_i = (ik_n/n, (i+1)k_n/n]$. If there is a discontinuous point of μ in I_i , then we expect that either $|A_i - A_{i-1}|$ or $|A_{i+1} - A_i|$ would take large values. So D_n can also be used to test whether μ has discontinuous points. There are certainly other ways to detect discontinuities; see the references that were cited in Section 1.

Theorem 1 concerns the asymptotic distributions of D_n and D_n^* under the null hypothesis of no structural changes $H_0: \mu \in L[0, 1]$. Using condition (5), we show that after proper centering and scaling both D_n and D_n^* have asymptotic extreme value distributions.

Theorem 1. Assume that $\mu \in L[0, 1]$ and

$$k_n^{-1} n^{1/2} \log(n)^3 + n^{-2/3} \log(n)^{1/3} k_n \rightarrow 0. \quad (14)$$

Let $\gamma_m = [4 \log(m) - 2 \log\{\log(m)\}]^{1/2}$. Then we have

$$\sqrt{\log(m)} (k_n^{1/2} \sigma^{-1} D_n - \gamma_m) \Rightarrow V \quad (15)$$

and

$$\sqrt{\log(m)} k_n^{1/2} \sigma^{-1} D_n^* - [2 \log(m) + \frac{1}{2} \log\{\log(m)\}] - \log(3) \Rightarrow V, \quad (16)$$

where V has the extreme value distribution $\mathbb{P}(V \leq x) = \exp\{-\pi^{-1/2} \exp(-x)\}$.

In condition (14), the first part $n^{1/2} \log(n)^3 = o(k_n)$ suggests that the block length k_n should not be too small, thus ensuring the validity of the strong approximation by Brownian motions. In contrast, the second part $k_n = o\{n^{2/3} \log(n)^{-1/3}\}$ suggests that k_n should not be too large so that the maximum difference $\max_{1 \leq i \leq m-1} |\omega_i - \omega_{i-1}|$ can be controlled. If $k_n \asymp n^\beta$ with $\beta \in (\frac{1}{2}, \frac{2}{3})$, then expression (14) holds.

Theorem 1 is not yet directly applicable since the long run variance σ^2 is typically unknown and it needs to be estimated. In Section 5 several estimates of σ are proposed satisfying $\hat{\sigma} - \sigma = O_{\mathbb{P}}(n^{-\gamma})$ for some $\gamma > 0$ (see theorem 3 in Section 5). By Slutsky's theorem, theorem 1 still holds if σ therein is replaced by $\hat{\sigma}$.

For a given level $\alpha \in (0, 1)$, let

$$c_\alpha = -\log\{-\log(1-\alpha)\} - \frac{1}{2} \log(\pi)$$

be the $(1-\alpha)$ th quantile of V in expression (15). By theorem 1, we reject the null hypothesis $\mu \in L[0, 1]$ if

$$D_n > k_n^{-1/2} \sigma_n \{c_\alpha \log(m)^{-1/2} + \gamma_m\}. \quad (17)$$

Now we consider the power of the test. Consider the local alternative in which there is a jump at $\theta \in (0, 1)$ with jump size $\delta_n = \mu(\theta) - \mu(\theta-) \neq 0$. If $\log(n) = o(k_n \delta_n^2)$ and expression (14) holds, since $\mu \in \text{PL}[0, 1]$,

$$\max_{1 \leq i \leq m-1} |\omega_i - \omega_{i-1}| \geq \frac{1}{2} |\delta_n| + O(k_n/n).$$

By theorem 1 it is easily seen that the power goes to 1. A simulation study is carried out in Section 6.1.

Remark 1. Let τ be a positive integer and assume that $f \in \mathcal{C}^\tau[0, 1]$ and that $f^{(\tau)} \in \text{PL}[0, 1]$ has discontinuous points $0 < t_1 < \dots < t_J < 1$. We say that t_1, \dots, t_J are structural breaks of order τ . Theorem 1 can be generalized to test the existence of higher order structural breaks. Let $\mu \in \mathcal{C}^\tau[0, 1]$. As in expression (13), define $\Delta A_i = A_i - A_{i-1}$ and $D_n^{(\tau)} = \max_{\tau+1 \leq i \leq m-1} |\Delta^{\tau+1} A_i|$. Assume that $k_n \asymp n^b$, $\frac{1}{2} < b < (2\tau + 2)/(2\tau + 3)$. Using the argument in the proof of theorem 1, we can show that, under $H_0: \mu^{(\tau)} \in L[0, 1]$, properly normalized $D_n^{(\tau)}$ has the same asymptotic distribution V as in expressions (15) and (16). The details are omitted.

3.1. The convergence issue

It is well known that the convergence to the extreme value distributions in expressions (15), (16) and (22) below is extremely slow and very large values of n are needed for the approximation to be reasonably accurate. To overcome this disadvantage, with the help of the strong invariance principle (5), we resort to the following simulation method:

- (a) generate n IID standard normal random variables Z_1, \dots, Z_n ;
- (b) use formula (12) and obtain $D_{n,Z}^*$ (say).

By condition (9), we can approximate the distribution of D_n^*/σ by $D_{n,Z}^*$ if the block length k_n is sufficiently large. (This idea is also implemented in the proof of theorem 1; see Appendix A). Clearly the distribution of $D_{n,Z}^*$ can be obtained by repeating steps (a) and (b) for many times. The distribution of D_n/σ can be similarly dealt with.

4. Simultaneous confidence bands

Consider the Priestley–Chao estimate (Priestley and Chao, 1972):

$$\mu_{b_n}(t) = \sum_{i=1}^n w_n(t, i) X_i, \quad w_n(t, i) = \frac{K\{(t - i/n)/b_n\}}{nb_n}. \quad (18)$$

Here the bandwidth b_n satisfies $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$ and K is a kernel with $\int K(s) ds = 1$. Other methods include the Gasser–Müller estimate (Gasser and Müller, 1979), local linear estimate, splines and wavelets. Proposition 1 below can be used to construct pointwise confidence intervals. It is a simple consequence of expressions (6)–(10) and the details are omitted.

Proposition 1. Assume that K has bounded variation, $b_n \rightarrow 0$ and $\log(n)^2 = o\{\sqrt{nb_n}\}$. Then for fixed $0 < t_1 < \dots < t_J < 1$, $\sqrt{nb_n}[\mu_{b_n}(t_j) - \mathbb{E}\{\mu_{b_n}(t_j)\}]$, $1 \leq j \leq J$, are asymptotically IID normals $N\{0, \sigma^2 \int_{\mathbb{R}} K^2(u) du\}$.

In practical situations, however, it is often not very useful to provide only pointwise confidence intervals and an SCB is more desirable. At a given level $\alpha \in (0, 1)$, to construct a $100(1 - \alpha)\%$ asymptotic SCB for μ , we need to find two functions l and u depending on the data $(X_k)_{k=1}^n$ such that

$$\lim_{n \rightarrow \infty} [\mathbb{P}\{l(t) \leq \mu(t) \leq u(t) \text{ for all } t \in (0, 1)\}] = 1 - \alpha. \quad (19)$$

If the trend is of certain parametric forms, then methods such as least squares can be applied to estimate μ and SCBs can be constructed by Scheffé's (1959) procedure in conjunction with asymptotic normal theory. However, in many cases such parametric forms are unknown and we

need to resort to non-parametric techniques since they make very few structural assumptions. However, non-parametric estimates may suggest appropriate parametric models. With SCB (19), it is possible to test the validity of parametric models. We now state some regularity conditions.

Definition 1. Let $\mathcal{H}(\alpha)$, $1 \leq \alpha \leq 2$, be the set of bounded functions H with bounded support satisfying

- (a) $\int_{\mathbb{R}} \Psi_H(u; \delta) du = O(\delta)$ as $\delta \rightarrow 0$, where $\Psi_H(u; \delta) = \sup\{|H(y) - H(y')| : y, y' \in [u - \delta, u + \delta]\}$, and
- (b) the limit $D_{H,\alpha} = \lim_{\delta \rightarrow 0} [|\delta|^{-\alpha} \int_{\mathbb{R}} \{H(x + \delta) - H(x)\}^2 dx]$ exists and $D_{H,\alpha} \neq 0$.

Let $\kappa_H^2 = \int_{\mathbb{R}} H^2(s) ds$. For $m \geq 3$ define

$$B_{H,\alpha}(m) = \sqrt{\{2 \log(m)\}} + \frac{1}{\sqrt{\{2 \log(m)\}}} \left[\frac{2-\alpha}{2\alpha} \log\{\log(m)\} + \log\left(\frac{C_{H,\alpha}^{1/\alpha} h_\alpha 2^{1/\alpha}}{2\sqrt{\pi}}\right) \right], \quad (20)$$

where $C_{H,\alpha} = D_{H,\alpha}/2\kappa_H^2$ and h_α is the Pickands constant (see theorem A1 in Bickel and Rosenblatt (1973)). Two values of h_α are known: $h_1 = 1$ and $h_2 = \pi^{-1/2}$.

Theorem 2. Assume that $K \in \mathcal{H}(\alpha)$ is a symmetric kernel with support $[-\omega, \omega]$. Let $\beta = \int K(u)u^2 du/2$. Further assume that $\mu \in C^3[0, 1]$ and

$$\frac{\log(n)^3}{b_n \sqrt{n}} + nb_n^7 \log(n) \rightarrow 0. \quad (21)$$

Let $m = 1/b_n$ and the interval $\mathcal{T} = [\omega b_n, 1 - \omega b_n]$. Then for every $u \in \mathbb{R}$, as $n \rightarrow \infty$,

$$\mathbb{P} \left[\frac{\sqrt{(nb_n)}}{\sigma \kappa_K} \sup_{t \in \mathcal{T}} |\mu_{b_n}(t) - \mu(t) - b_n^2 \beta \mu''(t)| - B_{K,\alpha}(m) \leq \frac{u}{\sqrt{\{2 \log(m)\}}} \right] \rightarrow \exp\{-2 \exp(-u)\}. \quad (22)$$

Condition (b) in definition 1 is useful in the extremal value theory of Gaussian processes (see theorem A1 in Bickel and Rosenblatt (1973)). Elementary calculations show that we have $K \in \mathcal{H}(\alpha)$ with $\alpha = 2$ for the triangle, quartic, Epanechnikov and Parzen kernels and $\alpha = 1$ for the rectangle kernel. For a kernel with unbounded support, under suitable conditions on its tail, theorem 2 is still applicable; see remark 3 in Appendix A.

As in condition (14), the first part in condition (21) ensures the validity of the strong approximation and the second part controls the bias (see lemmas 2 and 3 in Appendix A). Condition (21) is mild and it is satisfied if $b_n \asymp n^{-\gamma}$, $1/7 < \gamma < 1/2$. In particular, it holds if $\gamma = 1/5$, which corresponds to the optimal bandwidth under the mean-squared error (MSE) criterion.

Remark 2. In the literature strong invariance principles obtained for dependent random variables typically have rates of the form $o_{AS}(n^{1/2-\delta})$ for some arbitrarily small $\delta > 0$ (Philipp and Stout, 1975; Eberlein, 1986). As can be seen from condition (9) and the proof of lemma 2, in our problem such error bounds are too crude to be useful. It requires $b_n \rightarrow 0$ and $n^{2\delta} b_n \rightarrow \infty$. The latter condition is prohibitively restrictive if δ is close to 0. Recall that in the independence case Eubank and Speckman (1993) have applied the strong invariance principle of Komlós *et al.* (1975), which has the optimal rate. In comparison with the result of Komlós *et al.* (1975), our bound in condition (5) is optimal up to a multiplicative logarithmic factor if $2 < p \leq 4$ and it is sufficiently sharp for asymptotic inference of non-parametric estimates.

4.1. Implementation

Let $\hat{\sigma}$ and $\hat{\mu}''$ be estimates of σ and μ'' respectively. On the basis of theorem 2, an asymptotic $100(1 - \alpha)\%$ confidence band for μ can be constructed as

$$\mu_{b_n}(t) - b_n^2 \beta \hat{\mu}''(t) \pm l_{u_\alpha}, \quad l_u = \frac{\hat{\sigma} \kappa_K}{\sqrt{(nb_n)}} \left[B_{K, \alpha}(b_n^{-1}) + \frac{u}{\sqrt{\{2 \log(b_n^{-1})\}}} \right] \quad (23)$$

and $u_\alpha = -\log[\log\{(1 - \alpha)^{-1/2}\}]$. Eubank and Speckman (1993) proposed expression (23) for the independent error case. The estimation of μ'' is not easy when data points are not abundant.

To circumvent the difficulty, we adopt a jackknife-type bias correction scheme. Assume that the bias

$$\mathbb{E}\{\mu_{b_n}(t)\} - \mu(t) = b_n^2 \beta \mu''(t) + O(\chi_n)$$

(see lemma 3 in Appendix A), where $\chi_n = b_n^3 + n^{-1} b_n^{-1}$. Consider the simple estimate of the form

$$\tilde{\mu}_b(t) = 2 \mu_b(t) - \mu_{b/2}(t). \quad (24)$$

Then $\mathbb{E}\{\tilde{\mu}_{b_n}(t)\} = \mu(t) + O(\chi_n)$. So we do not need to estimate the unpleasant term μ'' . For similar forms see Härdle (1986). As pointed out by a referee, using equation (24) is equivalent to using the higher order kernel

$$K^*(s) = 2K(s) - \frac{K(s/\sqrt{2})}{\sqrt{2}}.$$

For this K^* theorem 2 is still applicable if the conditions therein are satisfied. We do not recommend using kernels with too high orders when data points are not abundant since the goal of bias reduction is achieved at the cost of a significant increase in variance. For related discussions see Fan and Gijbels (1996), section 3.3, and Fan and Zhang (2003). For our K^* , the increase in variance is not severe. For example, for the Epanechnikov kernel, $\kappa_{K^*}/\kappa_K \approx 1.53$.

Because of the dependence, the optimal MSE bandwidth b_n is different from the bandwidth under independence (Herrmann *et al.*, 1992). By the variance estimates in Herrmann *et al.* (1992) and Ruppert *et al.* (1995), a simple choice is $\tilde{b}_n = \rho^{1/5} b_n^*$, where b_n^* is the optimal bandwidth calculated as if the data were independent, and $\rho = \sigma^2 / \|e_0\|^2$ is the *variance correction factor*. The Priestley–Chao estimate (18) suffers the boundary problem. We shall use the local linear estimate (Fan and Gijbels, 1996) with

$$w_n(t, i) = K\left(\frac{t - i/n}{b_n}\right) \frac{S_2(t) - (t - i/n) S_1(t)}{S_2(t) S_0(t) - S_1^2(t)},$$

where

$$S_j(t) = \sum_{i=1}^n (t - i/n)^j K\left(\frac{t - i/n}{b_n}\right),$$

and the automatic bandwidth selector of Ruppert *et al.* (1995). Then $\tilde{b}_n \sim cn^{-1/5}$ for some constant c . The bias correction (24) allows larger b'_n such that $b_n'^3 + (nb'_n)^{-1} \sim (nb'_n)^{-1/2}$, or $b'_n \sim cn^{-1/7}$. However, it is non-trivial to find an optimal c that has good practical performance. On the basis of \tilde{b}_n , in our applications we let $b'_n = \phi \rho^{1/5} b_n^*$, $1 \leq \phi \leq 4$.

Theorem 3 shows that, to estimate σ^2 , the MSE optimal block length $k_n \sim cn^{1/3}$. It is unclear how to choose the optimal c . In practice we choose $k_n \in (n^{1/3}, n^{1/2})$.

Since the convergence in expression (22) is slow, we shall, as in Section 3.1, apply a similar simulation-based approach. For ease of application, we combine procedures mentioned above and list the details below.

- Choose an appropriate $k_n \in (n^{1/3}, n^{1/2})$ and obtain an estimate $\hat{\sigma}$ of σ .
- Let $b = 2\hat{\rho}^{1/5}b^*$, where b^* is calculated from Ruppert *et al.* (1995), and $\hat{\rho} = \hat{\sigma}^2/\hat{\nu}$, $\hat{\nu} = n^{-1} \sum_{i=1}^n \hat{e}_i^2$, $\hat{e}_i = X_i - \{2\mu_{b^*}(i/n) - \mu_{\sqrt{2}b^*}(i/n)\}$.
- Generate IID standard normals Z_1, \dots, Z_n and calculate $\sup_{0 \leq t \leq 1} |\tilde{\mu}_b^\diamond(t)|$, where $\tilde{\mu}_b^\diamond(t) = 2\mu_b^\diamond(t) - \mu_{b\sqrt{2}}^\diamond(t)$ and $\mu_b^\diamond(t) = \sum_{i=1}^n w_n(t, i)Z_i$.
- Repeat step (c) and obtain the estimated quantile $\hat{q}_{0.95}$ of $\sup_{0 \leq t \leq 1} |\tilde{\mu}_b^\diamond(t)|$.
- The 95% SCB is $\tilde{\mu}_b \pm \hat{\sigma}\hat{q}_{0.95}$, where $\tilde{\mu}_b(t) = 2\mu_b(t) - \mu_{b\sqrt{2}}(t)$.

5. Estimating σ

To apply theorems 1 and 2, we should deal with the crucial issue of estimating the long run variance σ^2 . If μ is a constant, since $\sigma^2/2\pi$ is the spectral density function of the process e_i at 0, there is a variety of ways to estimate σ^2 , such as lag window estimates and smoothed periodogram estimates. See Bühlmann (2002) and Politis *et al.* (1999) among others. The situation is slightly more complicated owing to the presence of a non-constant mean trend, which could possibly be discontinuous. Assuming that (e_i) are IID and μ is continuous, Hall *et al.* (1990) considered difference-based estimation of variance. See also Herrmann *et al.* (1992), which assumed very strong moment conditions.

Recall expression (13) for the definition of A_i . For a real sequence a_1, \dots, a_k , denote its median by $\text{median}(a_1, \dots, a_n)$. Here we consider three asymptotically consistent estimates:

$$\begin{aligned}\hat{\sigma}_1 &= \frac{\sqrt{(\pi k_n)}}{2(m-1)} \sum_{i=1}^{m-1} |A_i - A_{i-1}|, \\ \hat{\sigma}_2 &= \frac{\sqrt{k_n}}{\sqrt{(2u_{1/4})}} \text{median}(|A_i - A_{i-1}|, 1 \leq i \leq m-1), \\ \hat{\sigma}_3 &= \frac{\sqrt{k_n}}{\sqrt{\{2(m-1)\}}} \left(\sum_{i=1}^{m-1} |A_i - A_{i-1}|^2 \right)^{1/2}.\end{aligned}\quad (25)$$

In $\hat{\sigma}_2$, $u_{1/4} = 0.674 \dots$ is the third quartile of the standard normal distribution. Carlstein (1986) considered strong mixing processes by using non-overlapping blocks. Our $\hat{\sigma}_3$ is closely related to Carlstein's subseries variance estimate.

Theorem 3. Assume that $\mu \in L[0, 1]$.

- Let $k_n \asymp n^{5/8}$. Then $\hat{\sigma}_1, \hat{\sigma}_2 = \sigma + O_{\mathbb{P}}\{n^{-1/16} \log(n)\}$.
- Let $k_n \asymp n^{1/3}$. Then $\mathbb{E}(|\hat{\sigma}_3^2 - \sigma^2|^2) = O(n^{-2/3})$.

We conjecture that, like $\hat{\sigma}_3$, the other two estimates also satisfy $\hat{\sigma}_1, \hat{\sigma}_2 = \sigma + O_{\mathbb{P}}(n^{-1/3})$ if $k_n \asymp n^{1/3}$. Our simulation study (which is not reported here) shows that $\hat{\sigma}_2$ is more robust whereas $\hat{\sigma}_1$ and $\hat{\sigma}_3$ are vulnerable to large jumps in μ . For autoregressive AR(1) models with IID normal innovations, Carlstein (1986) argued that his subseries variance estimate has the optimal MSE $O(n^{-2/3})$ if the block length is of the order $n^{1/3}$. A result of similar vein based on the block-wise bootstrap is given in Künsch (1989). It is interesting to note that our underlying condition (4) plays two important roles at the same time: one is to ensure the strong invariance

principle (5) whereas the other is to achieve the MSE optimal variance estimate as in theorem 3, part (b).

6. A simulation study

In this section we shall present a simulation study for the performance of our test for structural breaks in Section 3 and the nominal levels (coverage probabilities) of our SCB in Section 4. Let ε_i be IID standard normals and $|\theta| < 1$. Consider the process

$$e_i = \theta|e_{i-1}| + \sqrt{(1 - \theta^2)}\varepsilon_i. \quad (26)$$

If $\theta = 0$, then $e_i = \varepsilon_i$ are IID. Otherwise e_i forms a non-linear autoregressive process. Since $|\theta| < 1$, by theorem 2 in Wu and Shao (2004), equation (26) has a stationary distribution and $\|e_n - e_n^*\|_4 = O(|\theta|^n)$. Hence condition (4) holds with $p = 4$. Interestingly, $F(u) = \mathbb{P}(e_i \leq u)$ has a skew normal density of the form $f(u) = 2\phi(u)\Phi(\delta u)$, where ϕ and Φ are respectively the standard normal density and normal distribution function (Andel *et al.*, 1984). The extra parameter $\delta = \theta/\sqrt{(1 - \theta^2)}$ regulates the skewness. Skew normal time series have been widely used to model processes with asymmetric and/or non-normal distributions. Simple calculations show that

$$\mathbb{E}(e_i) = \int_{\mathbb{R}} u f(u) du = \theta\sqrt{(2/\pi)}$$

and

$$\text{var}(e_i) = 1 - 2\theta^2/\pi.$$

Let $\sigma^2(\theta)$ be the long run variance of e_i in equation (26). For each level of $\theta = 0.0, 0.1, \dots, 0.9$, we apply $\hat{\sigma}_3$ of equation (25) with length 10^5 and $k_n = 47$. The 10 values of $\hat{\sigma}(\theta)$ are reported in Table 1. The other two estimates $\hat{\sigma}_1$ and $\hat{\sigma}_2$ yield very similar results.

6.1. Power curves for the test of structural breaks

We choose the mean function $\mu_\delta(t) = \delta \cos(2\pi t) \mathbf{1}_{t > 0.5}$ with a jump of size δ at location $t = 0.5$ and consider the model $X_k = \mu_\delta(k/n) + e'_k$, where

$$e'_k = \frac{e_k - \theta\sqrt{(2/\pi)}}{\hat{\sigma}(\theta)}$$

and e_k is defined by equation (26). Fig. 1(a) shows the plot of $\mu_\delta(t)$ with $\delta = 1$. Let the sample size $n = 200$ and $k_n = \lfloor n^{0.6} \rfloor = 24$. We use the simulation procedure that is listed in Section 3.1 with 4×10^5 repetitions to obtain the 0.95-quantile of $D_{n,Z}^*$. Then we simulate 4×10^4 sets of samples and the power is calculated by the proportion of samples for which the null hypothesis $H_0: \mu_\delta \in L[0, 1]$ is rejected when μ_δ has a jump of size δ . Power curves are plotted in Fig. 1(b). For $\theta = 0, 0.3, 0.6$, the p -values at $\delta = 0$ are 0.049, 0.047 and 0.048 respectively. They are close to the nominal level 0.05. However, for $\theta = 0.9$, the test performs poorly. It has less power and the

Table 1. Estimated long run standard deviations $\hat{\sigma}(\theta)$

θ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\sigma}(\theta)$	1.00	1.01	1.02	1.04	1.07	1.11	1.17	1.28	1.46	1.87

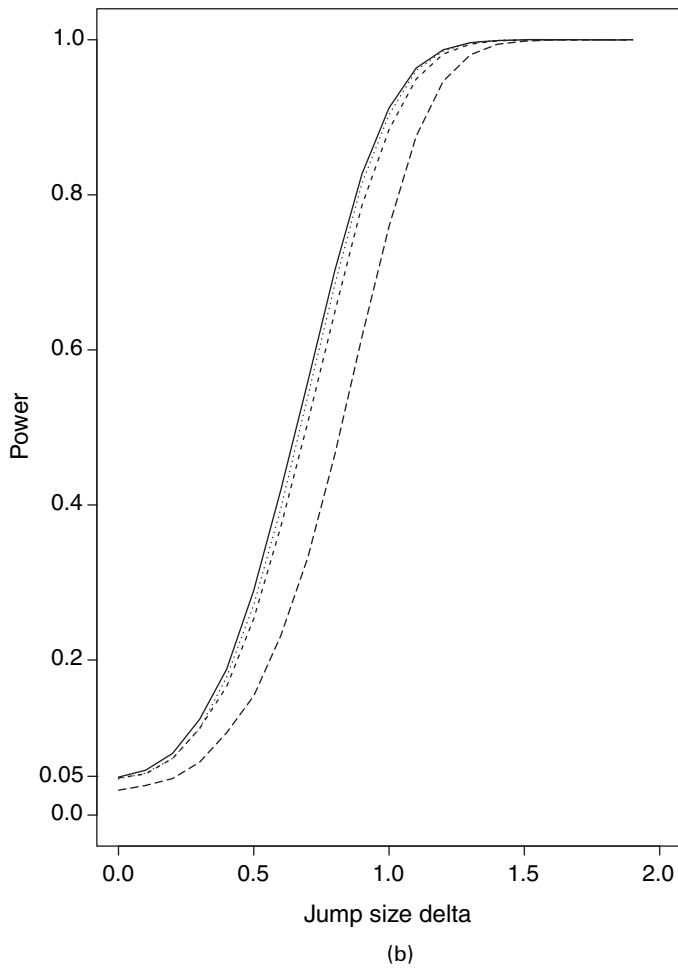
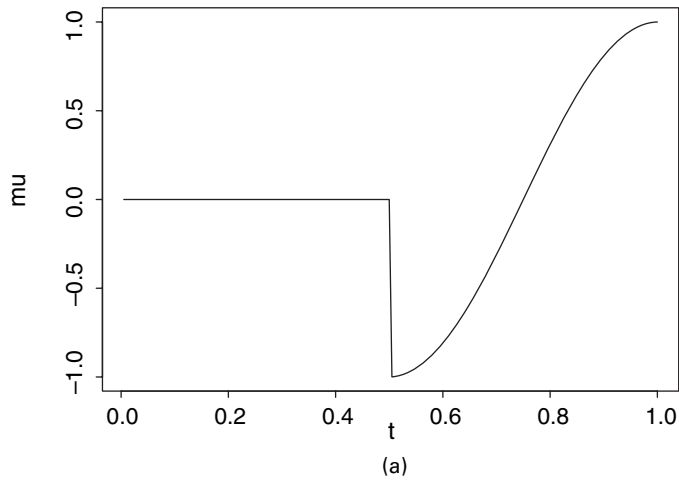


Fig. 1. (a) Mean curve $\mu(t) = \cos(2\pi t) \mathbf{1}_{t>0.5}$ and (b) power curves for the test statistic D_n^* that is proposed in Section 3 under $\mu_\delta(t) = \delta \mu(t)$, $0 \leq \delta \leq 2$: —, $\theta = 0$; ·····, $\theta = 0.3$; ----, $\theta = 0.6$; ---, $\theta = 0.9$

p -value is 0.032. See Section 6.2 for more discussion on the relationship between bandwidths and dependence in the context of SCBs.

6.2. Coverage probabilities of simultaneous confidence bands

Consider the model $X_k = \mu(k/n) + e'_k$, where

$$e'_i := \frac{e_i - \theta\sqrt{(2/\pi)}}{\hat{\sigma}(\theta)}$$

and the mean function $\mu(t) = \cos(2\pi t)$, $0 \leq t \leq 1$. Let $n = 200$. To study how bandwidths and dependence affect the coverage probabilities, we choose 10 levels of θ ($\theta = 0.0, 0.1, \dots, 0.9$) and 15 levels of b ($b = 0.01, 0.02, \dots, 0.15$). For each level of b , following steps (c) and (d) in Section 4, we use local linear regression program `locpoly` in the R package `KernSmooth` and estimate $q_{0.95} = q_{0.95}(b)$ with $N = 10^4$ repetitions. The estimated quantiles are shown in the second column of Table 2. For each of the $10 \times 15 = 150$ combinations of θ and b , we generate 10^4 realizations of $\tilde{\mu}_b(t) = 2\mu_b(t) - \mu_{\sqrt{2b}}(t)$. The SCB is constructed as $\tilde{\mu}_b \pm \hat{\sigma}q_{0.95}$. If μ lies within this band, namely $\sup_{0 \leq t \leq 1} |\mu(t) - \tilde{\mu}_b(t)| \leq \hat{\sigma}q_{0.95}$, then we say that the SCB covers μ . The fourth to 13th columns show the simulated coverage probabilities with the error process e'_i for $\theta = 0.0, 0.1, \dots, 0.9$. The third column shows the bias $\max_{0 \leq t \leq 1} |\mathbb{E}\{\tilde{\mu}_b(t)\} - \mu(t)|$.

Table 2 shows that the coverage probabilities of our SCB are reasonably close to the set nominal level 95%, especially when the bandwidth $b \leq 0.11$ and θ is not very large. For the model $X_k = \cos(2\pi k/n) + \varepsilon_k$, $1 \leq k \leq n$, where $n = 200$ and ε_k are IID standard normal random variables, the automatic bandwidth selection procedure of Ruppert *et al.* (1995) shows that the optimal bandwidth b for the local linear regression is around 0.07. In this case, for $\theta = 0.0, 0.1, \dots, 0.8$, the coverage probabilities range from 0.950 to 0.958. They are quite close to the nominal level 95%. However, if θ is close to 1, then the dependence is strong and we need to choose a relatively large bandwidth to ensure the validity of the strong approximation. The last column of Table 2 supports this claim. However, too large b increases the bias; hence the coverage probabilities decrease.

Table 2. Coverage probabilities of simultaneous confidence bands

b	$q_{0.95}$	bias	Coverage probabilities for the following values of θ :									
			0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.01	2.020	0.001	0.951	0.953	0.953	0.955	0.966	0.969	0.976	0.985	0.995	1.00
0.02	1.611	0.004	0.954	0.954	0.954	0.954	0.962	0.969	0.970	0.975	0.986	0.999
0.03	1.366	0.009	0.954	0.954	0.954	0.952	0.957	0.959	0.963	0.967	0.978	0.993
0.04	1.210	0.015	0.955	0.951	0.951	0.955	0.955	0.959	0.960	0.964	0.970	0.989
0.05	1.101	0.022	0.957	0.956	0.954	0.952	0.954	0.955	0.958	0.962	0.966	0.980
0.06	1.003	0.030	0.948	0.951	0.950	0.952	0.952	0.955	0.950	0.954	0.959	0.973
0.07	0.940	0.039	0.950	0.952	0.952	0.953	0.953	0.957	0.954	0.957	0.958	0.971
0.08	0.878	0.046	0.954	0.949	0.953	0.950	0.950	0.956	0.949	0.952	0.952	0.963
0.09	0.838	0.053	0.948	0.951	0.952	0.951	0.952	0.949	0.955	0.955	0.948	0.958
0.10	0.790	0.059	0.946	0.948	0.947	0.944	0.948	0.945	0.947	0.948	0.942	0.949
0.11	0.769	0.071	0.947	0.950	0.948	0.945	0.950	0.944	0.943	0.943	0.941	0.950
0.12	0.721	0.095	0.936	0.935	0.933	0.934	0.935	0.937	0.933	0.935	0.932	0.940
0.13	0.692	0.123	0.925	0.922	0.923	0.921	0.917	0.917	0.915	0.920	0.913	0.910
0.14	0.680	0.154	0.918	0.911	0.908	0.907	0.907	0.907	0.911	0.914	0.906	0.904
0.15	0.655	0.187	0.896	0.886	0.886	0.886	0.882	0.892	0.893	0.895	0.893	0.893

7. Applications

7.1. Nile river data

The data X_i , $1 \leq i \leq 100$, consist of measurements of the annual flow of the River Nile at Aswan from 1871 to 1970. Since Cobb (1978), the Nile river data have been extensively studied. It is believed that there is a jump (decrease) in year 1899 which may be due to the construction of a new dam at Aswan. However, it seems that in the literature most analyses of these data assume that the observations are independent. Here we shall apply theorem 1 and test the existence of jumps without assuming independence.

Since the convergence in expression (16) is very slow, we shall use the simulation method in Section 3.1 to obtain cut-off values. Let $n = 100$, $k = 15$ and $m = 1 + \lfloor n/k \rfloor = 7$. We repeat the following process for 10^4 times: generate n IID normals $N(0, 1)$ and calculate expression (12). The 95% and 99% simulated quantiles are 1.07 and 1.24 respectively. For the Nile river data, with $k = \lfloor n^{0.6} \rfloor$ we obtain $D_n^* = 254.06$, which needs to be rescaled by the long run standard deviation $\hat{\sigma}$. Assuming that the observations are independent, Cobb (1978) suggested $\hat{\sigma} = 125$. Here we shall calculate it by expression (25). Examining the plot of the three estimates, we choose lag 9 and obtain $\hat{\sigma}_1 = 176$, $\hat{\sigma}_2 = 162$ and $\hat{\sigma}_3 = 194$. As mentioned in Section 5, $\hat{\sigma}_2 = 162$ is preferred for robustness. Therefore the value of the test statistic $D_n^*/\hat{\sigma}_2 = 254.06/162 = 1.57$. Since this value is larger than the 99%-quantile 1.24, we conclude that the jump does exist at the 1% level (it is significant even if $\hat{\sigma}_3$ is used: $D_n^*/\hat{\sigma}_3 = 254.06/194 = 1.31$).

The evidence would be more substantial if Cobb's estimate $\hat{\sigma} = 125$ were used, in which case the test statistic $D_n^*/\hat{\sigma} = 254.06/125 = 2.03$. The variance correction factor $\rho = 162^2/125^2 = 1.68$. A simple way to relax his independence assumption is to use autoregressive moving average modelling. The mean levels before and after year 1899 are 1097.75 and 849.97 respectively. Then the estimated noise processes are $\hat{e}_i = X_i - 1097.75$ if $i \leq 28$ and $\hat{e}_i = X_i - 849.97$ if $i > 28$. Using the Akaike information criterion, the estimate noise processes \hat{e}_i can be modelled as an AR(1) process: $\hat{e}_i = 0.16\hat{e}_{i-1} + \varepsilon_i$, where ε_i are IID with mean 0 and standard deviation $15859^{1/2} \approx 126$. The AR(1) model implies that the long run standard deviation for e_i is $126/(1 - 0.16) \approx 150$, which in a certain sense justifies our estimate $\hat{\sigma}_2 = 162$.

7.2. Global warming data

Global temperature series have been extensively studied in the statistics community; see for example Bloomfield and Nychka (1992), Vogelsang (1998) and Wu *et al.* (2001) among others. Here we consider the series that was compiled by Jones *et al.* (2006) (see <http://cdiac.esd.ornl.gov/ftp/trends/temp/jonescru/>.) It contains global monthly temperature anomalies from 1856 to 2000 (Fig. 2(a)). Assuming that the trend is non-decreasing, Wu *et al.* (2001) fitted an isotonic regression for the annual temperature sequence. However, Wu *et al.* (2001) did not address the key issue of how to test the monotonicity assumption.

We first test whether jumps exist. There are 145 years and the length of the series is $n = 12 \times 145 = 1740$. In the three estimates in expression (25), we choose $k = 36$. Then $\hat{\sigma}_1 = 0.45$, $\hat{\sigma}_2 = 0.45$ and $\hat{\sigma}_3 = 0.44$. They are consistent and we choose $\hat{\sigma} = 0.44$. To calculate D_n^* of expression (12), as in Section 7.1, let $k = \lfloor n^{0.6} \rfloor$ such that condition (14) holds. Then $D_n^* = 0.218$ and $D_n^*/\hat{\sigma} = 0.495$. On the basis of the simulation method that was outlined in Section 3.1, we obtain the simulated p -value 22%. Therefore we are pleased to conclude that there is no evidence for jumps in the mean trend. In an interesting paper, Müller and Stadtmüller (1999) analysed infant growth data and argued that the growth of children occurs in jumps in the sense that there is a short period of fast growth, since it is unlikely that a mathematical jump discontinuity exists

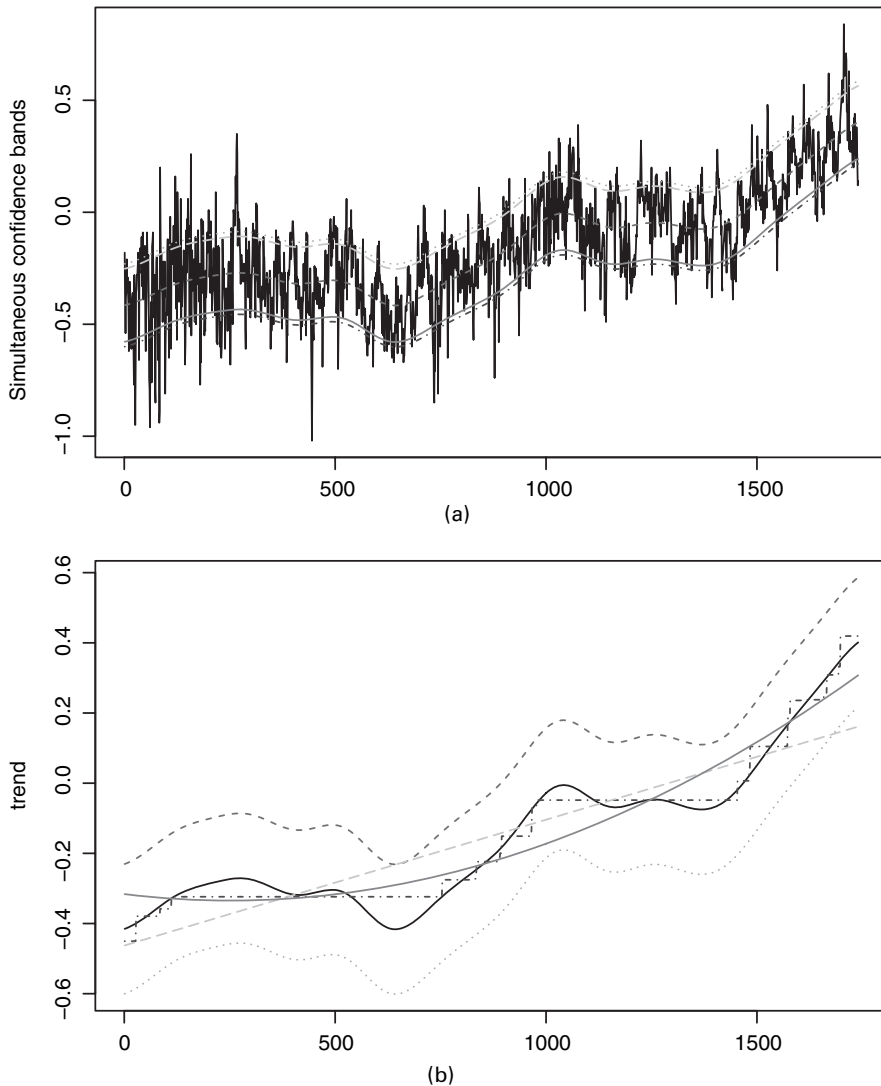


Fig. 2. SCBs for the global temperature data: (a) global monthly temperature anomalies from 1856 to 2000, local linear estimate of the trend curve and its 0.95 and 0.99 SCBs; (b) fitted linear (—), quadratic (---) and isotonic (·····) trends, local linear estimate of the trend (—) and its 0.95 SCB (-----, upper band; ·····, lower band)

in reality. There is no jump in the derivative since the p -value is 0.414.

The automatic bandwidth selection of Ruppert *et al.* (1995) gives $b_n^* = 0.0124$. The variance correction factor $\hat{\rho} = \hat{\sigma}^2 / \hat{\nu} = 0.229 / 0.0208 \approx 11$. For the 95% SCB (Fig. 2(b)), we choose the bandwidth $b_n = 2 \times 0.0124 \times 11^{0.2} \approx 0.04$. We are testing three null hypotheses separately: non-decreasing trend H_{isotonic} , linear trend H_{linear} and quadratic trend $H_{\text{quadratic}}$. The approximate p -values for them are 0.55, 0.008 and 0.15 respectively. Therefore we reject H_{linear} at the 1% level. Woodward and Gray (1993) fitted a linear trend model. The 95% SCB suggests that we accept H_{isotonic} and, surprisingly, $H_{\text{quadratic}}$. The regression equation is $y_i = -0.316 - 0.255(i/n) + 0.879(i/n)^2$. Rust (2003) fitted a quadratic trend and argued that a linear trend is inadequate.

Acknowledgements

We are grateful to the referees, an Associate Editor and the Joint Editor for their many helpful comments. We also thank Steve Lalley, Michael Stein and George Tiao for valuable suggestions.

Appendix A: Proofs

In the proofs below, the symbol C denotes a generic constant which may vary from place to place. Let Φ and $\phi = \Phi'$ be the standard normal distribution and density functions. To prove theorem 1, the following lemma is needed. Recall expression (15) for V .

Lemma 1. Let $Z_i, i \in \mathbb{Z}$, be IID standard normal distributions, $Y_i = |Z_{i+1} - Z_i|$ and $\gamma_n = [4 \log(n) - 2 \log\{\log(n)\}]^{1/2}$. Then, as $n \rightarrow \infty$, $\sqrt{\log(n)}\{\max_{1 \leq i \leq n-1} (Y_i) - \gamma_n\} \Rightarrow V$.

Proof. For $t \geq 0$ let $\Psi(t) = \mathbb{P}(|Z_1| \geq t) = 2\Phi(-t)$. Then $\Psi(t) = \{2 + o(1)\} \phi(t)/t$ as $t \rightarrow \infty$. Since $Y_i/\sqrt{2} = {}^D|Z_1|$, for fixed x , $n \mathbb{P}\{Y_i \geq \gamma_n + \log(n)^{-1/2}x\} \rightarrow \exp(-x)/\sqrt{\pi}$ as $n \rightarrow \infty$. Let $0 < \lambda < 1 - 2^{-1/2}$. Then $\mathbb{P}(Y_1 \geq t, |Z_1| < \lambda t) \leq \mathbb{P}\{|Z_2| \geq (1 - \lambda)t\} = o\{\Psi(t/\sqrt{2})\}$. Since Y_2 and Z_1 are independent, as $t \rightarrow \infty$, we have

$$\mathbb{P}(Y_1 \geq t, Y_2 \geq t) \leq \mathbb{P}(Y_2 \geq t, |Z_1| \geq \lambda t) + \mathbb{P}(Y_1 \geq t, |Z_1| < \lambda t) = o\{\mathbb{P}(Y_1 \geq t)\}.$$

By theorem 3.7.1 in Galambos (1987), the lemma follows.

Lemma 2. Assume that $H \in \mathcal{H}(\alpha)$, $\alpha \in [1, 2]$, $\int_{\mathbb{R}} H^2(u) du = 1$ and H has finite support $[-\omega, \omega]$. Let $b_n \rightarrow 0$ satisfy $\sqrt{(nb_n)}/\log(n)^3 \rightarrow \infty$. For $0 \leq t \leq 1$ define

$$U_n(t) = \frac{1}{\sqrt{(nb_n)}} \sum_{j=1}^n H\left\{m\left(t - \frac{j}{n}\right)\right\} \frac{e_j}{\sigma},$$

where $m = 1/b_n$. Let $B_H(m) = B_{H,\alpha}(m)$ be defined as in equation (20). Then, for $u \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \left(\mathbb{P} \left[\max_{t \in [\omega b_n, 1 - \omega b_n]} |U_n(t) - B_H(m)| \leq \frac{u}{\sqrt{2 \log(m)}} \right] \right) = \exp\{-2 \exp(-u)\}. \quad (27)$$

Proof. Let \mathbb{B} be the Brownian motion in condition (5), $Y(s) = \int_{\mathbb{R}} H(s-u) d\mathbb{B}(u)$, $\tilde{Y}(s) = \int_0^m H(s - [1 + k_n v]/k_n) d\mathbb{B}(v)$, where $k_n = nb_n$, and $Z_j = \mathbb{B}(j) - \mathbb{B}(j-1)$. Then \tilde{Y} is a stationary Gaussian process. Since $H \in \mathcal{H}(\alpha)$, $\alpha \in [1, 2]$, and $\int_{\mathbb{R}} H^2(u) du = 1$, then

$$\gamma(\delta) := \int_{\mathbb{R}} H(u) H(u + \delta) du = 1 - |\delta|^\alpha D_{H,\alpha}/2 + o(|\delta|^\alpha).$$

Note that $\mathbb{E}\{Y(s) Y(s + \delta)\} = \gamma(\delta)$. By corollary A1 of Bickel and Rosenblatt (1973), for $u \in \mathbb{R}$,

$$\lim_{m \rightarrow \infty} \left(\mathbb{P} \left[\max_{s \in \mathcal{I}} |Y(s) - B_H(m')| \leq \{2 \log(m')\}^{-1/2} u \right] \right) = \exp\{-2 \exp(-u)\}, \quad (28)$$

where $\mathcal{I} = [\omega, m - \omega]$ and $m' = m - 2\omega$. Since $B_H(m') = B_H(m) + o(m^{-1})$ and $\log(m')^{-1/2} = \log(m)^{-1/2} + o(m^{-1})$, by Slutsky's theorem, condition (28) also holds with m' therein replaced by m . Let $R_n = \max_{1 \leq j \leq n} |S_j - \sigma \mathbb{B}(j)|/\sigma \sqrt{(nb_n)}$ and

$$\begin{aligned} W_n(t) &= \sum_{j=1}^n \frac{H\{m(t - j/n)\}}{\sqrt{(nb_n)}} Z_j \\ &= \int_0^n \frac{H(mt - m[1 + u]/n)}{\sqrt{(nb_n)}} d\mathbb{B}(u). \end{aligned}$$

By condition (5) with $p=4$, since $\sqrt{(nb_n)}/\log(n)^3 \rightarrow \infty$, we have $R_n = o_{AS}\{\log(n)^{-1/2}\}$. Using the summation by parts formula, since $\int_{\mathbb{R}} \Psi_H(u; \delta) du = O(\delta)$, we have

$$\begin{aligned} |U_n(t) - W_n(t)| &= O(R_n) \left[1 + \int_0^n \left| H\left\{m\left(t - \frac{\lfloor 1+u \rfloor}{n}\right)\right\} - H\left\{m\left(t - \frac{\lfloor u \rfloor}{n}\right)\right\} \right| du \right] \\ &= O(R_n) \left[1 + \int_{\mathbb{R}} \Psi_H\left\{m\left(t - \frac{u}{n}\right); \frac{m}{n}\right\} du \right] \\ &= o_{AS}\{\log(n)^{-1/2}\} \end{aligned} \quad (29)$$

uniformly over $t \in [0, 1]$. By the scaling property of Brownian motion,

$$(W_n(s/m), s \in \mathcal{I}) \stackrel{\mathcal{D}}{=} (\tilde{Y}(s), s \in \mathcal{I}).$$

So equation (27) follows from equations (28) and (29) if $\Theta = O_{\mathbb{P}}(r_n)$, where $\Theta = \max_{s \in \mathcal{I}} |\Delta(s)|$, $\Delta(s) = Y(s) - \tilde{Y}(s)$ and $r_n = \sqrt{\log(n)}/\sqrt{k_n} = o\{\log(n)^{-1/2}\}$.

We now show that $\Theta = O_{\mathbb{P}}(r_n)$. Note that $\|\Delta(s)\|^2 \leq \int_{\mathbb{R}} \Psi_H^2(s-v; k_n^{-1}) dv \leq Ck_n^{-1}$ holds for some constant C . Similarly, if $|s-s'| \leq \frac{1}{2}$, we have $\|Y(s) - Y(s')\|^2 \leq C|s-s'|$ and $\|\tilde{Y}(s) - \tilde{Y}(s')\|^2 \leq C|s-s'|$. So $\|\Delta(s) - \Delta(s')\|^2 \leq 4C|s-s'|$. Let $\varepsilon = 2^{-1}$, $\varepsilon_j = (j+3)^{-2}$, $\delta_j = (2^j n^2)^{-1}$ and $\mathcal{X}_j = \{k\delta_j, k \in \mathbb{Z}\} \cap \mathcal{I}$, $j \geq 0$. Then the cardinality $|\mathcal{X}_j| \leq m/\delta_j$ and $\varepsilon + \sum_{j=0}^{\infty} \varepsilon_j < 1$. Let $\lambda = 8\sqrt{C}$. By the chaining lemma (see lemma 4.1 in Cranston *et al.* (2000)), we have

$$\begin{aligned} \mathbb{P}(\Theta > \lambda r_n) &\leq \mathbb{P}\{|\Delta(w)| > \lambda r_n \varepsilon\} + \sum_{j=0}^{\infty} |\mathcal{X}_j| \sup_{|s-s'| \leq \delta_j} \mathbb{P}\{|\Delta(s) - \Delta(s')| > \lambda r_n \varepsilon_j\} \\ &\leq 2 \left[1 - \Phi \left\{ \frac{\lambda r_n \varepsilon}{\sqrt{(Ck_n^{-1})}} \right\} \right] + \sum_{j=0}^{\infty} 2 \frac{m}{\delta_j} \left[1 - \Phi \left\{ \frac{\lambda r_n \varepsilon_j}{\sqrt{(C\delta_j)}} \right\} \right]. \end{aligned} \quad (30)$$

Since $1 - \Phi(t) \sim \phi(t)/t$ as $t \rightarrow \infty$, elementary calculations show that $\mathbb{P}(\Theta > \lambda r_n) = O(n^{-2})$. Then $\Theta = O_{AS}(r_n)$ and the lemma follows.

Lemma 3. Let $K \in \mathcal{H}(\alpha)$ be a symmetric kernel with support $[-\omega, \omega]$ and $\mu \in C^3[0, 1]$. Then $\mathbb{E}\{\mu_{b_n}(t)\} - \mu(t) = b_n^2 \mu''(t)\beta + O(b_n^3 + n^{-1}b_n^{-1})$ uniformly over $t \in \mathcal{T} = [\omega b_n, 1 - \omega b_n]$.

Proof. Let $k_n = nb_n$ and $K_j(v) = K(v)v^j$. Since $K \in \mathcal{H}(\alpha)$, by property (a) in definition 1, elementary calculations show that, for $j=0, 1, 2$, we have

$$\sup_{t \in \mathcal{T}} \left\{ \int_0^n \left| K_j \left(\frac{\lfloor 1+v \rfloor - nt}{k_n} \right) - K_j \left(\frac{v-nt}{k_n} \right) \right| dv \right\} = O(1).$$

Since $k_n^{-1} \int_0^n K_j \{(v-nt)/k_n\} dv = \int_{\mathbb{R}} K_j(u) du$ when $t \in \mathcal{T}$, by Taylor's expansion $\mu(t+\delta) = \mu(t) + \delta \mu'(t) + \delta^2 \mu''(t)/2 + O(\delta^3)$ as $\delta \rightarrow 0$, the lemma follows.

Remark 3. In lemmas 2 and 3, the kernel K is assumed to have bounded support. Assume that $b_n \log(n) \rightarrow 0$. Similar but lengthy calculations show that, if K is the normal density ϕ , then these two lemmas are still valid with the interval $\mathcal{T} = [\omega b_n, 1 - \omega b_n]$ therein replaced by $[b_n \log(n), 1 - b_n \log(n)]$. The details of the derivation are omitted.

A.1. Proof of theorem 1

We shall first prove expression (15). Let \mathbb{B} be the Brownian motion in the strong invariance principle (5). Then $Z_{i,n} = k_n^{-1/2}[\mathbb{B}\{(i+1)k_n\} - \mathbb{B}(ik_n)]$, $i=0, \dots, m-1$, are IID standard normal distributions. Let $R_{i,n} = S_{(i+1)k_n} - \sigma \mathbb{B}\{(i+1)k_n\} - S_{ik_n} + \sigma \mathbb{B}(ik_n)$ and write

$$A_i = \sigma k_n^{-1/2} Z_{i,n} + k_n^{-1} \sum_{j=1}^{k_n} \mu \left(\frac{j+ik_n}{n} \right) + k_n^{-1} R_{i,n}, \quad (31)$$

By condition (5), $\max_{i < m} |R_{i,n}| = o_{AS}\{n^{1/4} \log(n)\}$. Since $\mu \in L[0, 1]$, $\mu\{j/n + ik_n/n\} - \mu\{j/n + (i-1)k_n/n\} = O(k_n/n)$ uniformly over i and j . Recall that $m = \lfloor n/k_n \rfloor$. By expression (14),

$$\begin{aligned} k_n^{1/2} \sigma^{-1} (A_i - A_{i-1}) &= Z_{i,n} - Z_{i-1,n} + O_{AS}\{k_n^{3/2}/n + k_n^{-1/2} n^{1/4} \log(n)\} \\ &= Z_{i,n} - Z_{i-1,n} + o_{AS}\{\log(m)^{-1/2}\}. \end{aligned} \quad (32)$$

So expression (15) follows from equation (32) and lemma 1.

The relation (16) can be similarly proved. The function $H(u) = (\mathbf{1}_{0 \leq u < 1} - \mathbf{1}_{-1 < u < 0})/\sqrt{2}$ satisfies the conditions in lemma 2 with $\alpha = 1$ and $D_{H,1} = 3$. Like equation (32),

$$\begin{aligned} \frac{k_n^{1/2} D_n^*}{\sigma \sqrt{2}} &= \frac{1}{\sqrt{(2k_n)}} \max_{k_n \leq i \leq n-k_n} |\mathbb{B}(i+k_n) - 2\mathbb{B}(i) + \mathbb{B}(i-k_n)| + \frac{o_{AS}(1)}{\sqrt{\log(m)}} \\ &= \frac{1}{\sqrt{k_n}} \sup_{s \in [k_n, n-k_n]} \left| \int_{\mathbb{R}} H\left(\frac{s-u}{k_n}\right) d\mathbb{B}(u) \right| + \frac{O(\Omega_n)}{\sqrt{k_n}} + \frac{o_{AS}(1)}{\sqrt{\log(m)}}, \end{aligned}$$

where $\Omega_n = \sup\{|\mathbb{B}(u) - \mathbb{B}(u')| : u, u' \in [0, n], |u - u'| \leq 1\} = O_{\mathbb{P}}\{\sqrt{\log(n)}\}$. By equation (14) and lemma 2, expression (16) follows.

A.2. Proof of theorem 2

Let $\chi_n = b_n^3 + n^{-1} b_n^{-1}$. By expression (21), $\chi_n \sqrt{(nb_n)} = o\{\log(n)^{-1/2}\}$. So expression (22) follows from lemmas 2 and 3, which concern the stochastic part $\mu_{b_n}(t) - \mathbb{E}\{\mu_{b_n}(t)\}$ and the bias $\mathbb{E}\{\mu_{b_n}(t)\} - \mu(t) = b_n^2 \beta \mu''(t) + O(\chi_n)$ respectively.

A.3. Proof of theorem 3

For the proof of part (a) of theorem 3, recall the proof of theorem 1 for the definition of Z_i . Let $Y_i = |Z_i - Z_{i-1}|$ and $M_n = \text{median}(Y_i, 1 \leq i \leq m-1)$. Then Y_i is stationary and m dependent with $m = 2$ and the median of Y_i is $\xi_0 = \sqrt{(2u_{1/4})} = 0.954 \dots$. Let $F(x)$ and $f(x)$ be the distribution and density functions of Y_i ,

$$F_m(x) = (m-1)^{-1} \sum_{i=1}^{m-1} \mathbf{1}_{Y_i \leq x}.$$

By considering odd and even indices i respectively, we have by the law of iterated logarithms that $F_m(\xi_0) - \frac{1}{2} = O_{AS}(m^{-1/2} [\log\{\log(m)\}]^{1/2})$. By Sen (1968), $(M_n - \xi_0)f(\xi_0) = \frac{1}{2} - F_m(\xi_0) + O_{AS}\{m^{-3/4} \log(m)\}$. By equation (32),

$$\text{median}_{1 \leq i \leq m-1} |A_i - A_{i-1}| = \frac{\sigma}{\sqrt{k_n}} M_n + O_{AS}\{k_n^{-1} n^{1/4} \log(n) + n^{-1} k_n\}. \quad (33)$$

So $\hat{\sigma}_2 = \sigma + O_{AS}\{n^{-1/16} \log(n)\}$. The other case $\hat{\sigma}_1$ can be similarly proved.

The proof of part (b) is more complicated. Let

$$W_l = \sum_{i=l-k+1}^l e_i - \sum_{i=l-2k+1}^{l-k} e_i$$

and

$$r_l = \sum_{i=l-k+1}^l \mu_i - \sum_{i=l-2k+1}^{l-k} \mu_i.$$

Since $\mu \in L[0, 1]$, we have $r_l = O(k^2/n)$ uniformly over $l = 2k, \dots, n$. Let $2 \leq i \leq m$. By lemma 4, $\|W_{ik}\| = O(\sqrt{k})$. So

$$\|(W_{ik} + r_{ik})^2 - W_{ik}^2\| = O(k^4/n^2) + O(k/n)O(\sqrt{k}) = O(n^{-1/2})$$

since $k_n \asymp n^{1/3}$. Hence

$$\left\| \sum_{i=2}^m \{(W_{ik} + r_{ik})^2 - W_{ik}^2\} \right\| = m O(n^{-1/2}) = O(n^{1/6}),$$

which by lemma 5 below entails part (b).

Lemma 4. Assume that condition (4) holds with $p=2$. Then $\|S_{2n} - 2S_n\|^2 = 2n\sigma^2 + O(1)$.

Proof. Recall that $e_i^* = G(\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_i)$ and $\gamma(k) = \mathbb{E}(e_0 e_k)$. Let $\delta_p(i) = \|e_i - e_i^*\|_p$ and $\mathcal{F}_k = (\dots, \varepsilon_{k-1}, \varepsilon_k)$. For $\xi \in \mathcal{L}^1$ define the projection operator $\mathcal{P}_k \xi = \mathbb{E}(\xi | \mathcal{F}_k) - \mathbb{E}(\xi | \mathcal{F}_{k-1})$. By theorem 1 in Wu (2005), $\|\mathcal{P}_0 e_l\| \leq \delta_2(l)$, $l \geq 0$. Since $e_k = \sum_{j \in \mathbb{Z}} \mathcal{P}_j e_k$ and $\gamma(l) = \mathbb{E}(e_0 e_l)$, by the orthogonality,

$$\gamma(l) = \sum_{j \in \mathbb{Z}} \mathbb{E}(\mathcal{P}_j e_0 \mathcal{P}_j e_l).$$

By Schwarz's inequality,

$$|\gamma(l)| \leq \sum_{j=-\infty}^0 \|\mathcal{P}_j e_0\| \|\mathcal{P}_j e_l\|.$$

So $\sum_{l=1}^{\infty} l |\gamma(l)| < \infty$ since condition (4) implies that $\sum_{l=1}^{\infty} l \delta_2(l) < \infty$. Since $\sigma^2 = \sum_{i \in \mathbb{Z}} \gamma(i)$,

$$\|S_n\|^2 = \sum_{k=1-n}^{n-1} (n - |k|) \gamma(k)$$

and

$$|\mathbb{E}\{(S_{2n} - S_n)S_n\}| \leq \sum_{i=n+1}^{2n} \sum_{j=1}^n |\gamma(i-j)| \leq \sum_{l=1}^{\infty} l |\gamma(l)| < \infty,$$

we have $\|S_n\|^2 = n\sigma^2 + O(1)$ and the lemma follows.

Lemma 5. Let

$$W_l = \sum_{i=l-k+1}^l e_i - \sum_{i=l-2k+1}^{l-k} e_i$$

and

$$T_m = \sum_{j=1}^m W_{jk}^2.$$

Assume that $\sum_{i=1}^{\infty} \|e_i - e_i^*\|_4 < \infty$. Then, as $m, k \rightarrow \infty$, $\|T_m - m \mathbb{E}(W_{2k}^2)\| = O(k\sqrt{m})$.

Proof. By theorem 1 in Wu (2005), under $\sum_{i=1}^{\infty} \|e_i - e_i^*\|_4 < \infty$, we have $\|S_n\|_4 = O(\sqrt{n})$. Hence $\|W_l\|_4 = O(\sqrt{k})$. Let $l \geq 2k$ and

$$W'_l = \sum_{i=l-k+1}^l e_i^* - \sum_{i=l-2k+1}^{l-k} e_i^*.$$

By the Jensen inequality,

$$\|\mathcal{P}_0 W_l^2\| = \|\mathbb{E}(W_l^2 - W_l'^2 | \mathcal{F}_0)\| \leq \|W_l^2 - W_l'^2\|.$$

By the Schwarz inequality,

$$\|W_l^2 - W_l'^2\| \leq \|W_l - W_l'\|_4 \|W_l + W_l'\|_4 = O(\sqrt{k}) \sum_{i=l-2k+1}^l \delta_4(i).$$

Thus, for $j \geq 2$, by the orthogonality of projection operators,

$$\begin{aligned} \iota_j &:= \|\mathbb{E}(W_{jk}^2 | \mathcal{F}_0) - \mathbb{E}(W_{jk}^2 | \mathcal{F}_{-k})\|^2 \\ &= \sum_{t=1-k}^0 \|\mathcal{P}_t W_{jk}^2\|^2 \\ &= \sum_{t=1-k}^0 \|\mathcal{P}_0 W_{jk-t}^2\|^2 \end{aligned}$$

$$\begin{aligned}
&= O(k) \sum_{t=1-k}^0 \left\{ \sum_{i=jk-t-2k+1}^{jk-t} \delta_4(i) \right\}^2 \\
&= O(k^2) \left\{ \sum_{i=jk-2k+1}^{jk-1+k} \delta_4(i) \right\}^2.
\end{aligned} \tag{34}$$

By theorem 1 in Wu (2006),

$$\|T_m - m \mathbb{E}(W_{2k}^2)\| \leq \sqrt{m} \sum_{j=0}^{\infty} \sqrt{\iota_j}.$$

Elementary calculations show that equation (34) implies that $\|T_m - m \mathbb{E}(W_{2k}^2)\| = O(k\sqrt{m})$.

References

- Andel, J., Netuka, I. and Svara, K. (1984) On threshold autoregressive processes. *Kybernetika*, **20**, 89–106.
- Bickel, P. J. and Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1095.
- Bloomfield, P. and Nychka, D. (1992) Climate spectra and detecting climate change. *Clim. Change*, **21**, 275–287.
- Bühlmann, P. (1998) Sieve bootstrap for smoothing in nonstationary time series. *Ann. Statist.*, **26**, 48–83.
- Bühlmann, P. (2002) Bootstraps for time series. *Statist. Sci.*, **17**, 52–72.
- 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.
- Cobb, G. W. (1978) The problem of the Nile: conditional solution to a changepoint problem. *Biometrika*, **65**, 243–251.
- Cranston, M., Scheutzw, M. and Steinsaltz, D. (2000) Linear bounds for stochastic dispersion. *Ann. Probab.*, **28**, 1852–1869.
- Csörgő, M. and Horváth, L. (1997) *Limit Theorems in Change-point Analysis*. New York: Wiley.
- Cummins, D. J., Filloon, T. G. and Nychka, D. (2001) Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J. Am. Statist. Ass.*, **96**, 233–246.
- Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006) Structural break estimation for nonstationary time series models. *J. Am. Statist. Ass.*, **101**, 223–239.
- Dümbgen, L. (2003) Optimal confidence bands for shape-restricted curves. *Bernoulli*, **9**, 423–449.
- Eberlein, E. (1986) On strong invariance principles under dependence assumptions. *Ann. Probab.*, **14**, 260–270.
- Eubank, R. L. and Speckman, P. L. (1993) Confidence bands in nonparametric regression. *J. Am. Statist. Ass.*, **88**, 1287–1301.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
- Fan, J. and Zhang, C. (2003) A reexamination of diffusion estimators with applications to financial model validation. *J. Am. Statist. Ass.*, **98**, 118–134.
- Galambos, J. (1987) *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. Melbourne: Krieger.
- Gasser, T. and Müller, H.-G. (1979) Kernel estimation of regression functions. *Lect. Notes Math.*, **757**, 23–68.
- Gijbels, I. and Goderniaux, A.-C. (2004) Bootstrap test for change-points in nonparametric regression. *J. Nonparam. Statist.*, **16**, 591–611.
- Grégoire, G. and Hamrouni, Z. (2002) Two non-parametric tests for change-point problems. *J. Nonparam. Statist.*, **14**, 87–112.
- Hall, P., Kay, J. W. and Titterton, D. M. (1990) Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.
- Hall, P. and Titterton, D. M. (1988) On confidence bands in nonparametric density estimation and regression. *J. Multiv. Anal.*, **27**, 228–254.
- Härdle, W. (1986) A note on jackknifing kernel regression function estimators. *IEEE Trans. Inform. Theory*, **32**, 298–300.
- Härdle, W. (1989) Asymptotic maximal deviation of M -smoothers. *J. Multiv. Anal.*, **29**, 163–179.
- Härdle, W. and Marron, J. S. (1991) Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778–796.
- Herrmann, E., Gasser, T. and Kneip, A. (1992) Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, **79**, 783–795.
- Johnston, G. J. (1982) Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multiv. Anal.*, **12**, 402–414.
- Jones, P. D., Parker, D. E., Osborn, T. J. and Briffa, K. R. (2006) Global and hemispheric temperature anomalies—land and marine instrumental records. In *Trends: a Compendium of Data on Global Change*. Oak Ridge: Oak Ridge National Laboratory. (Available from <http://cdiac.ornl.gov/trends/temp/jonescr/jones.html>.)

- Knafl, G., Sacks, J. and Ylvisaker, D. (1985) Confidence bands for regression functions. *J. Am. Statist. Ass.*, **80**, 683–691.
- Komlós, J., Major, P. and Tusnády, G. (1975) An approximation of partial sums of independent RV's and the sample DF, I. *Z. Wahrsch. Ver. Geb.*, **32**, 111–131.
- Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1241.
- Lavielle, M. (1999) Detection of multiple changes in a sequence of dependent variables. *Stoch. Processes Appl.*, **83**, 79–102.
- Müller, H.-G. (1992) Change-points in nonparametric regression analysis. *Ann. Statist.*, **20**, 737–761.
- Müller, H.-G. and Stadtmüller, U. (1999) Discontinuous versus smooth regression. *Ann. Statist.*, **27**, 299–337.
- Pettitt, A. N. (1980) A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, **67**, 79–84.
- Philipp, W. and Stout, W. (1975) Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Am. Math. Soc.*, **161**, 1–140.
- Politis, D. N., Romano, J. P. and Wolf, M. (1999) *Subsampling*. New York: Springer.
- Priestley, M. B. and Chao, M. T. (1972) Non-parametric function fitting. *J. R. Statist. Soc. B*, **34**, 385–392.
- Qiu, P. (2003) A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *J. Nonparam. Statist.*, **15**, 437–453.
- Qiu, P. and Yandell, B. (1998) A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics*, **40**, 141–152.
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.
- Rust, B. W. (2003) Signal from noise in global warming. *Comput. Sci. Statist.*, **35**, 263–277.
- Scheffé, H. (1959) *The Analysis of Variance*. New York: Wiley.
- Sen, K. S. (1968) Asymptotic normality of sample quantiles for m -dependent processes. *Ann. Math. Statist.*, **39**, 1724–1730.
- Spokoiny, V. G. (1998) Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.*, **26**, 1356–1378.
- Sun, J. and Loader, C. R. (1994) Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.*, **22**, 1328–1345.
- Tang, S. M. and MacNeill, I. B. (1993) The effect of serial correlation on tests for parameter change at unknown time. *Ann. Statist.*, **21**, 552–575.
- Vogelsang, T. J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, **66**, 123–148.
- Woodward, W. A. and Gray, H. L. (1993) Global warming and the problem of testing for trend in time-series data. *J. Clim.*, **6**, 953–962.
- Wu, J. S. and Chu, C. K. (1993) Kernel-type estimators of jump points and values of a regression function. *Ann. Statist.*, **21**, 1545–1566.
- Wu, W. B. (2005) Nonlinear system theory: another look at dependence. *Proc. Natn. Acad. Sci. USA*, **102**, 14150–14154.
- Wu, W. B. (2006) Strong invariance principles for dependent random variables. *Ann. Probab.*, to be published.
- Wu, W. B. and Shao, X. (2004) Limit theorems for iterated random functions. *J. Appl. Probab.*, **41**, 425–436.
- Wu, W. B., Woodroffe, M. and Mentz, G. (2001) Isotonic regression: another look at the change point problem. *Biometrika*, **88**, 793–804.
- Xia, Y. (1998) Bias-corrected confidence bands in nonparametric regression. *J. R. Statist. Soc. B*, **60**, 797–811.