The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

INFERENCE OF TRENDS IN TIME SERIES

Wei Biao Wu and Zhibiao Zhao

TECHNICAL REPORT NO. 569

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

July 18, 2006

Inference of trends in time series

Wei Biao Wu and Zhibiao Zhao

Department of Statistics, The University of Chicago

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. Based on 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; Nonlinear time series; Nonparametric regression

1 Introduction

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

$$X_k = \mu(k/n) + e_k, \quad k = 1, \ldots, n, \tag{1}$$

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 it can be interpreted as a signal (μ) plus noise (e_k) model. The paper has two primary goals. The first one is to develop statistical procedures to test whether the trend μ in the model (1) has structural breaks or jumps. If the curve μ is smooth, our second goal is to construct simultaneous confidence bands (SCB) for μ. Throughout the paper we consider *a posteriori* or off-line inference, namely the data have already been collected before the analysis.

For the model (1), the classical change-point analysis concerns testing the null hypothesis $\mu_1 = \ldots = \mu_n$ versus the alternative of one or multiple change points

$$\mu_1 = \ldots = \mu_{k_1} \neq \mu_{k_1+1} = \ldots = \mu_{k_2} \neq \mu_{k_2+1} = \ldots \neq \mu_{k_J+1} = \ldots = \mu_n, \tag{2}$$

where $\mu_k = \mu(k/n)$ and $k_1, \ldots, k_J$ are called change points. 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.

Nonparametric 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 only mention 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 the presence of serial correlation can seriously affect the distributions of change-point statistics.

For our second goal of constructing SCB for μ, we assume that μ is smooth. SCB 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 SCB has been discussed by Johnston (1982), Härdle (1989), Knafl et al (1985), Hall and Titterington (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 Kolmós et al (1975)'s strong invariance principle and constructed SCB for μ with asymptotically correct nominal values. In the context of kernel density estimation, Bickel and Rosenblatt (1973) obtained SCB for density functions. The construction of SCB 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 (2005a), we shall provide a solution to the problem and construct SCB 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 $\mathcal{C}^m[0, 1]$, $m = 0, 1, \ldots$, denote the collection of functions having up to mth order derivatives. For a function g we say that g has bounded variation if $V(g) := \sup \sum_i |g(t_i) - g(t_{i-1})| < \infty$, where the sup is taken over all $\ldots < t_{i-1} < t_i < \ldots$. 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 $\lim_{n\to\infty} a_n/b_n = 1$ and $a_n \asymp b_n$ if $0 < \liminf_{n\to\infty} |a_n/b_n| \leq \limsup_{n\to\infty} |a_n/b_n| < \infty$.

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 of μ. Section 4 discusses the construction of SCB 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 estimate the long-run variance of (e_i). The latter 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 Section 8.

2 The error structure

We assume that the error process (e_i) in the 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

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

is a proper random variable with mean 0 and finite variance. Many processes fall within the framework of (3) [see Tong (1990) and Stine (1997) among others]. Prominent examples are linear process and many widely used nonlinear time series including threshold autoregressive models, bilinear autoregressive models and (generalized) autoregressive models with conditional heteroscedasticity; see Wu and Min (2005) for more examples.

As in Wiener (1958) and Wu (2005b), (3) can be interpreted as a physical system with $\eta_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$ being the input, G being a filter and e_n being the output. Let (ε'_j) be an iid copy of (ε_j) and $e_i^* = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i)$ a coupled process of e_i. Assume that $\mathbb{E}(|e_i|^p) < \infty$, $p > 2$, and

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

Wu (2005a) established the following strong approximation or strong invariance principle. Under (4), there exists a standard Brownian motion $I\!B$ such that on a richer probability space, S_i can be uniformly approximated by $I\!B(i)$:

$$\max_{i \le n} |S_i - \sigma I\!B(i)| = o_{\text{a.s.}}(n^{1/p'} \log n),\ p' = \min(4, p), \tag{5}$$

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 (5) holds with the optimal error bound $o_{\text{a.s.}}(n^{1/p})$ and $\sigma = \|e_i\|$. In our problem the results by Komlós et al are not applicable due to the dependence among e_i.

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. Condition (4) is easily verifiable since it is directly related to the data-generating mechanism of (e_i). Wu (2005b) 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 nonlinear time series models, $\|e_n - e_n^*\|_p = O(r^n)$ for some $r \in (0, 1)$ and hence (4) trivially holds. Consider the ARMA process $e_n - \sum_{i=1}^{l} \psi_i e_{n-i} = \sum_{j=0}^{q} \theta_j \varepsilon_{n-j}$, where $\psi_1, \ldots, \psi_l, \theta_0, \ldots, \theta_q$ 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 (4) also holds.

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

Nonparametric inference of μ in model (1) typically involves the quantity

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

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

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

and the Gaussian process

$$Y_n^\diamond(t) = \sum_{i=1}^{n} w_n(t,i)\sigma[I\!B(i) - I\!B(i-1)]. \tag{8}$$

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

$$|Y_n(t) - Y_n^\diamond(t)| \le \Omega_n(t) \max_{i\le n} |S_i - \sigma I\!B(i)| = o_{\text{a.s.}}[\Omega_n(t) n^{1/4} \log n] \tag{9}$$

and the uniform approximation

$$\max_{0\le t\le 1} |Y_n(t) - Y_n^\diamond(t)| = o_{\text{a.s.}}(\Omega_n n^{1/4} \log n). \tag{10}$$

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 (cf (18)), we have $w_n(t,i) = K[(t - i/n)/b_n]/(nb_n)$ and hence $\Omega_n = O((nb_n)^{-1})$ if K has bounded variation. For local linear estimates (cf 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, (1) can be reduced to the conventional model

$$X_k^\diamond = \mu(k/n) + \sigma Z_k, \quad k = 1, \ldots, n, \tag{11}$$

where Z_k are iid standard normals and σ and μ are unknown. This idea is implemented in Sections 3 and 4 below.

3 Inference of structural breaks

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

Important problems in the inference of structural breaks include (i) testing the hypothesis of no structural change $H_0 : \mu \in L[0,1]$ and (ii) estimating the locations and sizes of structural breaks. Various aspects of the second problem has 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 of the trend: for some constant $c > 0$, $|\mu_{k+1} - \mu_k| \le c/n$ holds for all $n \ge 2$ and $1 \le k < n$. In comparison, in the classical change-point inference, μ_k does not change.

The formulation $\mu \in PL[0,1]$ is more general than the one in the classical setting. On the other hand, 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 $k_n \to \infty$ and $k_n/n \to 0$. If the two averages are close, then t is unlikely a discontinuous point. A global measure of the discrepancy is

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

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

$$D_n = \max_{1 \le i \le m-1} |A_i - A_{i-1}|, \text{ where } A_i = A_{i,n} = \frac{1}{k_n} \sum_{j=1}^{k_n} X_{j+ik_n}. \tag{13}$$

Here $m = \lfloor n/k_n \rfloor$ is the largest integer not exceeding n/k_n. Let $\omega_i = \mathbb{E}(A_i)$. For $1 \le i \le m$ 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 one expects 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 certainly exist other ways to detect discontinuities; see the references 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 (5), we show that after proper centering and scaling both D_n and D_n^* have asymptotic extreme value distributions.

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

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

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

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

and

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

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. On the other hand, 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 $\beta \in (1/2, 2/3)$, then (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. This problem has a long history. In Section 5 several estimates of σ are proposed satisfying $\hat{\sigma} - \sigma = O_{\mathbb{P}}(n^{-\gamma})$ for some $\gamma > 0$ (cf. Theorem 3). 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 (15). By Theorem 1, we reject the null hypothesis $\mu \in L[0, 1]$ if

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

Now we consider the power of the test. Consider the local alternative in which there exists 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 (14) holds, since $\mu \in PL[0, 1]$, $\max_{1 \leq i \leq m-1} |\omega_i - \omega_{i-1}| \geq \frac{1}{2} |\delta_n| + O(k_n/n)$. 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 $f^{(\tau)} \in PL[0, 1]$ has discontinuous points $0 < t_1 < \ldots < t_J < 1$. We say that $t_1, \ldots, t_J$ 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 (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 $k_n \asymp n^\beta, 1/2 < \beta < (2\tau+2)/(2\tau+3)$. 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 (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 (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:

(i) generate n iid standard normal random variables $Z_1, \ldots, Z_n$

(ii) use formula (12) and obtain $D_{n,Z}^*$ (say).

By (9), we can approximate the distribution of D_n^*/σ by $D_{n,Z}^*$ if the block length k_n is large enough. (This idea is also implemented in the proof of Theorem 1; see Section 8). Clearly the distribution of $D_{n,Z}^*$ can be obtained by repeating steps (i) and (ii) for many times. The distribution of D_n/σ can be similarly dealt with.

4 Simultaneous confidence bands

There exists a huge literature on nonparametric estimation of the mean regression function $\mu(t)$, $t \in (0,1)$. Consider the popular Priestley-Chao estimate

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

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

Proposition 1. *Assume that K has bounded variation, $b_n \to 0$ and $(\log n)^2 = o(\sqrt{n} b_n)$. Then for fixed $0 < t_1 < \ldots < t_J < 1$, $\sqrt{nb_n}\{\mu_{b_n}(t_j) - \mathbb{E}[\mu_{b_n}(t_j)]\}$, $1 \le j \le J$, 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 a 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

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

If the trend is of certain parametric forms, then methods such as least squares can be applied to estimate μ and SCB can be constructed by Scheffé's (1959) procedure in conjunction with asymptotic normal theory. However, in many cases such parametric forms are unknown and one needs to resort to nonparametric techniques since they make very few structural assumptions. On the other hand, nonparametric estimates may suggest appropriate parametric models. With the SCB (19), it is possible to test the validity of parametric models. We now state some regularity conditions.

Definition 1. *Let $\mathcal{H}(\alpha)$, $1 \le \alpha \le 2$, be the set of bounded functions H with bounded support satisfying (i) $\int_{\mathbb{R}} \Psi_H(u;\delta)du = O(\delta)$ as $\delta \to 0$, where $\Psi_H(u;\delta) = \sup\{|H(y) - H(y')| : y, y' \in [u-\delta, u+\delta]\}$, and (ii) the limit $D_{H,\alpha} = \lim_{\delta\to 0} |\delta|^{-\alpha} \int_{\mathbb{R}} [H(x+\delta) - H(x)]^2 dx$ exists and $D_{H,\alpha} \neq 0$. Let $\kappa_H^2 = \int_{\mathbb{R}} H^2(s)ds$. For $m \ge 3$ define*

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

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 $\beta = \int K(u)u^2du/2$. Further assume that $\mu \in \mathcal{C}^3[0,1]$ and*

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

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 $n \to \infty$,

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

Condition (ii) in Definition 1 is useful in the extremal value theory of Gaussian processes (cf 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, 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 with the range $\omega b_n \le t \le 1 - \omega b_n$ in (22) replaced by a narrower one. For example, for the Gaussian kernel, if $b_n \log n \to 0$, (22) holds with $t \in [b_n \log n,\, 1 - b_n \log n]$; see Remark 3.

As in condition (14), the first part in condition (21) ensures the validity of the strong approximation and the second part controls the bias (cf Lemmas 2 and 3). 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 squares error (MSE) criterion.

Remark 2. In the literature strong invariance principles obtained for dependent random variables typically have rates of the form $o_{\text{a.s.}}(n^{1/2-\delta})$ for some arbitrarily small $\delta > 0$ [Philipp and Stout (1974), Eberlein (1986)]. As can be seen from (9) and the proof of Lemma 2, in our problem such error bounds are too crude to be useful. It requires $b_n \to 0$ and $n^{2\delta} b_n \to \infty$. 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 Kolmós et al's strong invariance principle, which has the optimal rate. In comparison with Kolmós et al's result, our bound in (5) is optimal up to a multiplicative logarithmic factor if $2 < p \le 4$ and it is sharp enough for asymptotic inference of nonparametric estimates.

4.1 Implementation

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

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

and $u_\alpha = -\log\log[(1-\alpha)^{-1/2}]$. In the independent error case, Eubank and Speckman (1993) proposed (23) with an estimated MSE optimal bandwidth $\hat{b}_n$. The estimation of μ'' is not easy especially when data points are not abundant.

To circumvent the difficulty, we adopt a jackknife-type bias correction scheme. Assume that the bias $\mathbb{E}[\mu_{b_n}(t)] - \mu(t) = b_n^2 \beta \mu''(t) + O(\chi_n)$ (cf. Lemma 3), where $\chi_n = b_n^3 + n^{-1} b_n^{-1}$. Consider the simple estimate of the form

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

Then $\mathbb{E}[\tilde{\mu}_{b_n}(t)] = \mu(t) + O(\chi_n)$. So one does not need to estimate the unpleasant term μ''. There certainly exists other forms, for example, a general form of (24) is $(\lambda^2 \mu_b - \mu_{b\lambda})/(\lambda^2 - 1)$, $\lambda > 0$. See Härdle (1986) for other forms. Here for simplicity we use

(24). As pointed out by a referee, using of (24) is equivalent to using the higher-order kernel $K^*(s) = 2K(s) - K(s/\sqrt{2})/\sqrt{2}$. 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, Sec 3.3) and Fan and Zhang (2003). For our K^*, the increase in variance is not severe. For example, for the Epanechnikov kernel we have $\kappa_{K^*}/\kappa_K \approx 1.53$.

Due to the presence of dependence, the optimal MSE bandwidth b_n is different from the one in the independence case (Herrmann et al, 1992). By the variance estimates in the latter paper 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 $w_n(t,i) = K[(t-i/n)/b_n][S_2(t) - (t-i/n)S_1(t)]/[S_2(t)S_0(t) - S_1^2(t)]$, where $S_j(t) = \sum_{i=1}^n (t-i/n)^j K[(t-i/n)/b_n]$, 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 us to choose 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. Based on $\tilde{b}_n$, in our applications we let $b_n' = \phi\rho^{1/5}b_n^*$, $1 \le \phi \le 4$.

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 $k_n \in (n^{1/3}, n^{1/2})$.

As mentioned in Section 3.1, the convergence in (22) is slow. We shall apply a similar simulation based approach which avoids using (23). For ease of application, we combine procedures in preceding paragraphs and list the details below. Simulation studies in Section 6.2 show that our approach has reasonably accurate coverage probabilities.

(i) Choose an appropriate $k_n \in (n^{1/3}, n^{1/2})$ and obtain an estimate $\hat{\sigma}$ of σ.

(ii) Let $b = 2\hat{\rho}^{1/5}b^*$, where b^* is calculated from Ruppert et al (1995), and $\hat{\rho} = \hat{\sigma}^2/\hat{\nu}$, $\hat{\nu} = n^{-1}\sum_{i=1}^n \hat{e}_i^2$, $\hat{e}_i = X_i - [2\mu_{b^*}(i/n) - \mu_{\sqrt{2}b^*}(i/n)]$.

(iii) Generate iid standard normals $Z_1, \ldots, Z_n$ and calculate $\sup_{0\le t\le 1} |\tilde{\mu}_b^\diamond(t)|$, where $\tilde{\mu}_b^\diamond(t) = 2\mu_b^\diamond(t) - \mu_{b\sqrt{2}}^\diamond(t)$ and $\mu_b^\diamond(t) = \sum_{i=1}^n w_n(t,i)Z_i$.

(iv) repeat (iii) and obtain the estimated quantile $\hat{q}_{0.95}$ of $\sup_{0\le t\le 1} |\tilde{\mu}_b^\diamond(t)|$.

(v) 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 1-2, one 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, smoothed periodogram estimates etc. See Bühlmann (2002) and Politis et al (1999) among others. The situation is slightly more complicated due 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) consider difference-based estimation of variance. See also Herrmann et al (1992). The latter paper assumes very strong moment conditions.

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

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

In $\hat{\sigma}_2$, $u_{1/4} = 0.674...$ is the 3rd quartile of the standard normal distribution. Carlstein (1986) considers strong mixing processes by using non-overlapping blocks. Our $\hat{\sigma}_3$ is closely related to Carlstein's subseries variance estimate.

Theorem 3. *Assume $\mu \in L[0, 1]$. (i) Let $k_n \asymp n^{5/8}$. Then $\hat{\sigma}_1, \hat{\sigma}_2 = \sigma + O_{\mathbb{P}}(n^{-1/16} \log n)$. (ii) Let $k_n \asymp n^{1/3}$. Then $\mathbb{E}(|\hat{\sigma}_3^2 - \sigma^2|^2) = O(n^{-2/3})$.*

We conjecture that, as $\hat{\sigma}_3$, 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 (not reported here) shows that $\hat{\sigma}_2$ is more robust while $\hat{\sigma}_1$ and $\hat{\sigma}_3$ are vulnerable to large jumps in μ. For AR(1) models with iid normal innovations, Carlstein (1986) argues that his subseries variance estimate has the optimal MSE $O(n^{-2/3})$ if the block length $\asymp n^{1/3}$. A result of similar vein based on 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) while the other is to achieve the MSE-optimal variance estimate as in Theorem 3(ii).

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

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

If $\theta = 0$, then $e_i = \varepsilon_i$ are iid. Otherwise e_i forms a nonlinear autoregressive process. Since $|\theta| < 1$, by Theorem 2 in Wu and Shao (2004), (26) has a stationary distribution and $\|e_n - e_n^*\|_4 = O(|\theta|^n)$. Hence (4) holds with $p = 4$. Interestingly, $F(u) = \mathbb{P}(e_i \le u)$ has a skew-normal density of the form $f(u) = 2\phi(u)\Phi(\delta u)$, where ϕ (resp. Φ) is the standard normal density (resp. 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 $\mathbb{E}(e_i) = \int_{\mathbb{R}} uf(u)du = \theta\sqrt{2/\pi}$ and $\mathrm{Var}(e_i) = 1 - 2\theta^2/\pi$. Let $\sigma^2(\theta)$ be the long-run variance of e_i in (26). For each level of $\theta = 0.0, 0.1, \ldots, 0.9$, we apply $\hat{\sigma}_3$ of (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.

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

6.1 Power curves for the test of structural breaks

Insert Figure 1 about here.

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 $e'_k = (e_k - \theta\sqrt{2/\pi})/\sigma(\theta)$ and e_k is defined by (26). Let the sample size $n = 200$ and $k_n = \lfloor n^{0.6} \rfloor = 24$. We use the simulation procedure listed in Section 3.1 with 4×10^5 repetitions to obtain the .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. 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 p-value is 0.032. See Section 6.2 for more discussion on the relationship between bandwidths and dependence in the context of SCB.

6.2 Coverage probabilities of SCB

Table 2. Coverage probabilities of SCB.

			θ									
b	$q_{0.95}$	bias	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
.01	2.020	.001	.951	.953	.953	.955	.966	.969	.976	.985	.995	1.00
.02	1.611	.004	.954	.954	.954	.954	.962	.969	.970	.975	.986	.999
.03	1.366	.009	.954	.954	.954	.952	.957	.959	.963	.967	.978	.993
.04	1.210	.015	.955	.951	.951	.955	.955	.959	.960	.964	.970	.989
.05	1.101	.022	.957	.956	.954	.952	.954	.955	.958	.962	.966	.980
.06	1.003	.030	.948	.951	.950	.952	.952	.955	.950	.954	.959	.973
.07	0.940	.039	.950	.952	.952	.953	.953	.957	.954	.957	.958	.971
.08	0.878	.046	.954	.949	.953	.950	.950	.956	.949	.952	.952	.963
.09	0.838	.053	.948	.951	.952	.951	.952	.949	.955	.955	.948	.958
.10	0.790	.059	.946	.948	.947	.944	.948	.945	.947	.948	.942	.949
.11	0.769	.071	.947	.950	.948	.945	.950	.944	.943	.943	.941	.950
.12	0.721	.095	.936	.935	.933	.934	.935	.937	.933	.935	.932	.940
.13	0.692	.123	.925	.922	.923	.921	.917	.917	.915	.920	.913	.910
.14	0.680	.154	.918	.911	.908	.907	.907	.907	.911	.914	.906	.904
.15	0.655	.187	.896	.886	.886	.886	.882	.892	.893	.895	.893	.893

Consider the model $X_k = \mu(k/n) + e'_k$, where $e'_i := (e_i - \theta\sqrt{2/\pi})/\sigma(\theta)$ and the mean function $\mu(t) = \cos(2\pi t)$, $0 \le t \le 1$. Let $n = 200$. To study how bandwidths and dependence affect the coverage probabilities, we choose 10 levels of θ: $\theta = 0.0, 0.1, \ldots, 0.9$, and 15 levels of b: $b = 0.01, 0.02, \ldots, 0.15$. For each level of b, following Steps (iii) and (iv) 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 $\tilde{\mu}_b(t) = 2\mu_b(t) - \mu_{\sqrt{2}b}(t)$. The SCB is constructed as

$\tilde{\mu}_b \pm \hat{\sigma} q_{0.95}$. If μ lies within this band, namely $\sup_{0\le t\le 1} |\mu(t) - \tilde{\mu}_b(t)| \le \hat{\sigma} q_{0.95}$, then we say that the SCB covers μ. Columns 4-13 show the simulated coverage probabilities with the error process e'_i for $\theta = 0.0, 0.1, \ldots, 0.9$, respectively. The third column shows the bias $\max_{0\le t\le 1} |\mathbb{E}[\tilde{\mu}_b(t)] - \mu(t)|$.

Table 2 shows that the coverage probabilities of our SCB are reasonably close to the set nominal level 95%, especially when the bandwidth $b \le 0.11$ and θ is not very large. For the model $X_k = \cos(2\pi k/n) + \varepsilon_k$, $1 \le k \le n$, where $n = 200$ and ε_k are iid standard normals, Ruppert et al (1995)'s automatic bandwidth selection procedure shows that the optimal bandwidth b for the local linear regression is around 0.07. In this case, for $\theta = 0.0, 0.1, \ldots, 0.8$, 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 relatively large bandwidth to ensure the validity of the strong approximation. The last column of Table 2 supports this claim. On the other hand, however, too large b increases the bias, hence the coverage probabilities decrease.

7 Applications

7.1 Nile river data

The data X_i, $1 \le i \le 100$, consist of measurements of the annual flow of the river Nile at Ashwan from 1871 to 1970. Since Cobb (1978), the Nile river data has 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 analysis of this data assumes that the observations are independent. Here we shall apply Theorem 1 and test the existence of jumps without assuming independence.

Since the convergence in (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 (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 get $D_n^* = 254.06$, which needs to be re-scaled 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 (25). Examining the plot of the 3 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 the sake of robustness. Therefore the value of the test statistic $D_n^*/\hat{\sigma}_2 = 254.06/162 = 1.57$. Since the latter value is larger than the 99% quantile 1.24, we conclude that the jump does exist at 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 the ARMA modelling. The mean levels before and after year 1899 are 1097.75 and 849.97, respectively. Then the estimated noises are $\hat{e}_i = X_i - 1097.75$ if $i \le 28$ and $\hat{e}_i = X_i - 849.97$ if $i > 28$. Using the Akaike information criterion, the estimate noises $\hat{e}_i$ 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 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 complied by Jones et al. See http://cdiac.esd.ornl.gov /ftp/trends/temp/jonescru/. It contains global monthly temperature anomalies from 1856 to 2000. Assuming that the trend is non-decreasing, Wu et al (2001) fitted an isotonic regression for the annual temperature sequence. However, the latter paper 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 (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 (12), as in Section 7.1, let $k = \lfloor n^{0.6} \rfloor$ such that (14) holds. Then $D_n^* = 0.218$ and $D_n^*/\hat{\sigma} = 0.495$. Based on the simulation method 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) analyzed the 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 in reality. There is no jump in the derivative since the p-value is 0.414.

The automatic bandwidth selection of Ruppert et al gives $b_n^* = 0.0124$. The variance correction factor $\hat{\rho} = \tilde{\sigma}^2/\hat{\nu} = 0.229/0.0208 \approx 11$. For the 95% SCB, we choose the bandwidth $b_n = 2 \times 0.0124 \times 11^{.2} \approx 0.04$. 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 of them are 0.55, 0.008 and 0.15, respectively. Therefore we reject H_{linear} at 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.

Insert Figure 2 about here.

8 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 (15) for V.

Lemma 1. *Let $Z_i, i \in \mathbb{Z}$, be iid standard normals, $Y_i = |Z_{i+1} - Z_i|$ and $\gamma_n = (4\log n - 2\log\log n)^{1/2}$. Then as $n \to \infty$, $\sqrt{\log n}\{\max_{1\le i\le n-1} Y_i - \gamma_n\} \Rightarrow V$.*

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

$$\mathbb{P}(Y_1 \ge t, Y_2 \ge t) \le \mathbb{P}(Y_2 \ge t, |Z_1| \ge \lambda t) + \mathbb{P}(Y_1 \ge t, |Z_1| < \lambda t) = o[\mathbb{P}(Y_1 \ge t)]$$

By Theorem 3.7.1 in Galambos (1987), the lemma follows. $\diamondsuit$

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 $b_n \to 0$ satisfy $\sqrt{n}b_n/(\log n)^3 \to \infty$. For $0 \le t \le 1$ define*

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

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

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

Proof. Let $I\!\!B$ be the Brownian motion in (5), $Y(s) = \int_{\mathbb{R}} H(s-u)dI\!\!B(u)$, $\tilde{Y}(s) = \int_0^m H(s - \lfloor 1 + k_n v \rfloor / k_n)\, dI\!\!B(v)$, where $k_n = nb_n$, and $Z_j = I\!\!B(j) - I\!\!B(j-1)$. Then 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 $u \in \mathbb{R}$,

$$\lim_{m\to\infty} \mathbb{P}\Big\{ \max_{s\in\mathcal{I}} |Y(s)| - B_H(m') \le (2\log m')^{-1/2} u \Big\} = \exp[-2\exp(-u)], \tag{28}$$

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, (28) also holds with m' therein replaced by m. Let $R_n = \max_{1\le j\le n} |S_j - \sigma I\!\!B(j)|/(\sigma\sqrt{nb_n})$ and

$$W_n(t) = \sum_{j=1}^n \frac{H(m(t - j/n))}{\sqrt{nb_n}} Z_j = \int_0^n \frac{H(mt - m\lfloor 1+u \rfloor/n)}{\sqrt{nb_n}} dI\!\!B(u).$$

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

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

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 (27) follows from (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 $\Theta = O_{\mathbb{P}}(r_n)$. Note that $\|\Delta(s)\|^2 \le \int_{\mathbb{R}} \Psi_H^2(s-v; k_n^{-1})dv \le Ck_n^{-1}$ holds for some constant C. Similarly, if $|s - s'| \le 1/2$, we have $\|Y(s) - Y(s')\|^2 \le C|s-s'|$ and $\|\tilde{Y}(s) - \tilde{Y}(s')\|^2 \le C|s-s'|$. So $\|\Delta(s) - \Delta(s')\|^2 \le 4C|s-s'|$. Let $\epsilon = 2^{-1}$, $\epsilon_j = (j+3)^{-2}$, $\delta_j = (n^2 2^j)^{-1}$ and $\mathcal{X}_j = \{k\delta_j, k \in \mathbb{Z}\} \cap \mathcal{I}$, $j \ge 0$. Then the cardinality $|\mathcal{X}_j| \le m/\delta_j$ and $\epsilon + \sum_{j=0}^\infty \epsilon_j < 1$. Let $\lambda = 8\sqrt{C}$. By the Chaining Lemma (cf. Lemma 4.1 in Cranston et al (2000)), we have

$$\mathbb{P}(\Theta > \lambda r_n) \;\le\; \mathbb{P}[|\Delta(w)| > \lambda r_n \epsilon] + \sum_{j=0}^\infty |\mathcal{X}_j| \sup_{|s-s'|\le\delta_j} \mathbb{P}[|\Delta(s) - \Delta(s')| > \lambda r_n \epsilon_j]$$

$$\leq \quad 2[1-\Phi(\frac{\lambda r_n \epsilon}{\sqrt{C k_n^{-1}}})] + \sum_{j=0}^{\infty} \frac{m}{\delta_j} 2[1-\Phi(\frac{\lambda r_n \epsilon_j}{\sqrt{C\delta_j}})]. \tag{30}$$

Since $1-\Phi(t) = (1+o(1))\phi(t)/t$ as $t \to \infty$, elementary calculations show that $\mathbb{P}(\Theta > \lambda r_n) = O(n^{-2})$. Then $\Theta = O_{\text{a.s.}}(r_n)$ and the lemma follows. ♢

Lemma 3. *Let $K \in \mathcal{H}(\alpha)$ be a symmetric kernel with support $[-\omega, \omega]$ and $\mu \in \mathcal{C}^3[0,1]$. 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 (i) in Definition 1, elementary calculations show that, for $j = 0, 1, 2$, we have

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

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 $\delta \to 0$, the lemma follows. ♢

Remark 3. In Lemmas 2 and 3, the kernel K is assumed to have bounded support. Assume $b_n \log n \to 0$. 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. ♢

Proof of Theorem 1. We shall first prove (15). Let $I\!B$ be the Brownian motion in the strong invariance principle (5). Then $Z_{i,n} = k_n^{-1/2}(I\!B((i+1)k_n) - I\!B(ik_n))$, $i = 0, \ldots, m-1$, are iid standard normals. Let $R_{i,n} = S_{(i+1)k_n} - \sigma I\!B((i+1)k_n) - S_{ik_n} + \sigma I\!B(ik_n)$ and write

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

By (5), $\max_{i<m}|R_{i,n}| = o_{\text{a.s.}}(n^{1/4}\log n)$. 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 $m = \lfloor n/k_n \rfloor$. By (14),

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

So (15) follows from (32) and Lemma 1.

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

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

where $\Omega_n = \sup\{|I\!\!B(u) - I\!\!B(u')| : u, u' \in [0, n], |u-u'| \le 1\} = O_{\mathbb{P}}(\sqrt{\log n})$. By (14) and Lemma 2, (16) follows. ♢

Proof of Theorem 2. Let $\chi_n = b_n^3 + n^{-1} b_n^{-1}$. By (21), $\chi_n \sqrt{n b_n} = o((\log n)^{-1/2})$. So (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. ♢

Proof of Theorem 3. Recall the proof of Theorem 1 for the definition of Z_i. Let $Y_i = |Z_i - Z_{i-1}|$ and $M_n = \text{median}(Y_i, 1 \le i \le m-1)$. Then Y_i is stationary and m-dependent with $m = 2$ and the median of Y_i is $\xi_0 = \sqrt{2} u_{1/4} = 0.954...$. Let $F(x)$ and $f(x)$ be the distribution and density functions of Y_i, $F_m(x) = (m-1)^{-1} \sum_{i=1}^{m-1} \mathbf{1}_{Y_i \le x}$. By considering odd and even indices i respectively, we have by the law of iterated logarithm that $F_m(\xi_0) - 1/2 = O_{\text{a.s.}}(m^{-1/2} (\log\log m)^{1/2})$. By Sen (1968), $(M_n - \xi_0) f(\xi_0) = 1/2 - F_m(\xi_0) + O_{\text{a.s.}}(m^{-3/4} \log m)$. By (32),

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

So $\hat\sigma_2 = \sigma + O_{\text{a.s.}}(n^{-1/16} \log n)$. The other case $\hat\sigma_1$ can be similarly proved.

The proof of (iii) is much 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 $l = 2k, \dots, n$. Observe that

$$\begin{aligned}\left\| \sum_{i=2}^{m} [(W_{ik} + r_{ik})^2 - W_{ik}^2] \right\| &\le \sum_{i=2}^{m} r_{ik}^2 + 2 \sum_{i=2}^{m} |r_{ik}| \|W_{ik}\| \\ &= O(mk^4/n^2) + O(mk/n) O(\sqrt{k}) = O(n^{1/6}).\end{aligned}$$

Simple calculations show that (iii) follows from Lemmas 4 and 5 below. ♢

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

Proof. Recall the coupled process $e_i^* = G(\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i)$ and $\gamma(k) = \mathbb{E}(e_0 e_k)$. Let $\delta_p(i) = \|e_i - e_i^*\|_p$ and $\mathcal{F}_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$. 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 (2005b), $\|\mathcal{P}_0 e_l\| \le \delta_2(l)$, $l \ge 0$. By the orthogonal expansion $e_k = \sum_{j\in\mathbb{Z}} \mathcal{P}_j e_k$, we have

$$\begin{aligned}|\gamma(l)| &= \Big|\mathbb{E}[(\sum_{j\in\mathbb{Z}} \mathcal{P}_j e_0)(\sum_{j'\in\mathbb{Z}} \mathcal{P}_{j'} e_l)]\Big| = \Big|\sum_{j\in\mathbb{Z}} \mathbb{E}[(\mathcal{P}_j e_0)(\mathcal{P}_j e_l)]\Big| \\ &\le \sum_{j=-\infty}^{0} \|\mathcal{P}_j e_0\|\|\mathcal{P}_j e_l\| \le \sum_{j=-\infty}^{0} \delta_2(-j)\delta_2(l-j).\end{aligned}$$

By exchanging the order of summations, we have $\sum_{l=1}^\infty l|\gamma(l)| < \infty$ since (4) implies $\sum_{l=1}^\infty l\delta_2(l) < \infty$. Since $\sigma^2 = \sum_{i\in\mathbb{Z}} \gamma(i)$, $\|S_n\|^2 = \sum_{k=1-n}^{n-1} (n - |k|)\gamma(k)$ and

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

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* $m, k \to \infty$, $\|T_m - m\mathbb{E}(W_{2k}^2)\| = O(k\sqrt{m})$.

Proof. By Theorem 1 in Wu (2005b), 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 $l \ge 2k$ and $W'_l = \sum_{i=l-k+1}^{l} e_i^* - \sum_{i=l-2k+1}^{l-k} e_i^*$. By the Jensen inequality, $\|\mathcal{P}_0 W_l^2\| = \|\mathbb{E}(W_l^2 - W_l'^2|\mathcal{F}_0)\| \le \|W_l^2 - W_l'^2\|$. By the Schwarz inequality, $\|W_l^2 - W_l'^2\| \le \|W_l - W'_l\|_4 \|W_l + W'_l\|_4 = O(\sqrt{k}) \sum_{i=l-2k+1}^{l} \delta_4(i)$. Thus, for $j \ge 2$, by the orthogonality of projection operators,

$$\begin{aligned}\|\mathbb{E}(W_{jk}^2|\mathcal{F}_0) - \mathbb{E}(W_{jk}^2|\mathcal{F}_{-k})\|^2 &= \sum_{t=1-k}^{0} \|\mathcal{P}_t W_{jk}^2\|^2 = \sum_{t=1-k}^{0} \|\mathcal{P}_0 W_{jk-t}^2\|^2 \\ &= O(k) \sum_{t=1-k}^{0} \left[\sum_{i=jk-t-2k+1}^{jk-t} \delta_4(i)\right]^2 = O(k^2) \left[\sum_{i=jk-2k+1}^{jk-1+k} \delta_4(i)\right]^2.\end{aligned}$$

By Theorem 1 in Wu (2005a),

$$\|T_m - m\mathbb{E}(W_{2k}^2)\| \le \sqrt{m} \sum_{j=0}^{\infty} \|\mathbb{E}(W_{jk}^2|\mathcal{F}_0) - \mathbb{E}(W_{jk}^2|\mathcal{F}_{-k})\|$$

$$= O(k\sqrt{m}) + O(k\sqrt{m}) \sum_{j=2}^{\infty} \sum_{i=jk-2k+1}^{jk-1+k} \delta_4(i) = O(k\sqrt{m}),$$

which completes the proof of the lemma. ♢

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

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. *Climatic 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., Scheutzow, 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. Amer. Statist. Assoc.*, **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. Amer. Statist. Assoc.*, **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. Amer. Statist. Assoc.*, **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. Amer. Statist. Assoc.*, **98**, 118–134.

Galambos, J. (1987) *The asymptotic theory of extreme order statistics*, Second edition. Melbourne, FL: Krieger.

Gijbels, I. and Goderniaux, A.-C. (2004) Bootstrap test for change-points in nonparametric regression. *J. Nonparametr. Stat.*, **16**, 591–611.

Grégoire, G and Hamrouni, Z. (2002) Two non-parametric tests for change-point problems. *J. Nonparametr. Stat.*, **14**, 87–112.

Hall, P. and Titterington, D. M. (1988) On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.*, **27**, 228–254.

Hall, P., Kay, J. W. and Titterington, D. M. (1990) Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, **77**, 521–528.

Härdle, W. (1986) A note on jackknifing kernel regression function estimators. *IEEE Trans. Inf. Theory*, **32**, 298–300.

Härdle, W. (1989) Asymptotic maximal deviation of M-smoothers. *J. Multivariate 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. Multivariate Anal.*, **12**, 402–414.

Knafl, G., Sacks, J. and Ylvisaker, D. (1985) Confidence bands for regression functions. *J. Amer. Statist. Assoc.*, **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. und Verw. Gebiete*, **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. *Stochastic Process. 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. Amer. Math. Soc.*, **2**, No. 161.

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

Qiu, P. (2003) A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *J. Nonparametr. Stat.*, **15**, 437–453.

Qiu, P. and Yandell, B. (1998) A local polynomial jump-detection algorithm in nonparametric regression. *Technometrics*, **40**, 141–152.

Rice, J. (1984) Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215–1230.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Assoc.*, **90**, 1257–1270.

Rust, B. W. (2003) Signal from noise in global warming, *Computing Science and Statistics*, **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.

Stine, R. A. (1997) Nonlinear time series. In *Encyclopedia of Statistical Sciences* (eds S. Kotz, C. B. Read and D. L. Banks), updated vol. 1, pp. 430–437.

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.

Tong, H. (1990) *Nonlinear time series. A dynamical system approach.* New York: Oxford University Press.

Vogelsang, T. J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, **66**, 123–148.

Wiener, N. (1958) *Nonlinear Problems in Random Theory.* Cambridge: MIT Press.

Woodward, W. A. and Gray, H. L. (1993) Global warming and the problem of testing for trend in time-series data. *J. Climate*, **6**, 953–62.

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. (2005a) On strong convergence for sums of dependent random variables. Preprint.

Wu, W. B. (2005b) Nonlinear system theory: another look at dependence. *Proc. Natl Acad. Sci. USA*, **102**, 14150–14154.

Wu, W. B. and Min, W. (2005) On linear processes with dependent innovations. *Stochastic Process. Appl.*, **115**, 39–958.

Wu, W. B. and Shao, X. (2004) Limit theorems for iterated random functions. *J. Appl. Probab.*, **41**, 425–436.

Wu, W. B., Woodroofe, 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. Stat. Soc. Ser. B*, **60**, 797–811.

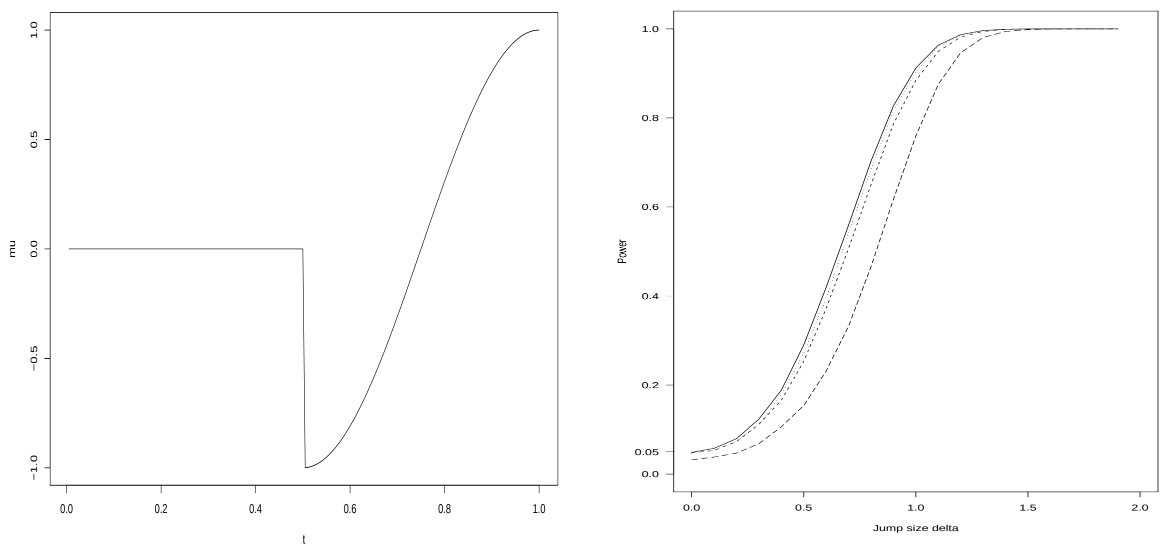

Figure 1: Left panel: mean curve $\mu(t) = \cos(2\pi t)\mathbf{1}_{t>0.5}$. Right panel: power curves for the test statistic D_n^* proposed in Section 3 under $\mu_\delta(t) = \delta \times \mu(t)$, $0 \le \delta \le 2$. The solid, dotted, dashed and long-dash lines correspond to $\theta = 0, 0.3, 0.6, 0.9$, respectively.

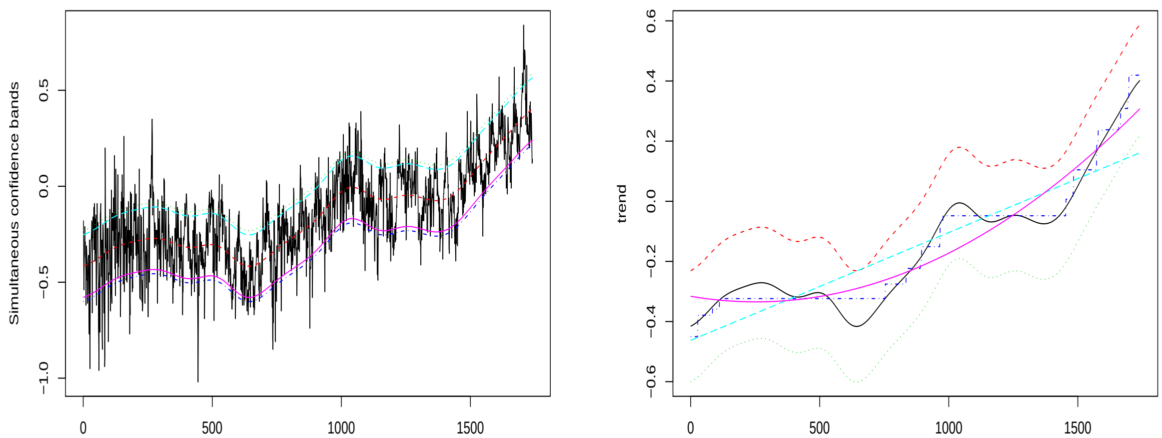

Figure 2: Simultaneous confidence bands for the global temperature data. Left panel: global monthly temperature anomalies from 1856 to 2000, local linear estimate of the trend curve and its .95 and .99 SCBs. Right panel: fitted linear, quadratic and isotonic trends, local linear estimate of the trend and its .95 SCB.