

Journal: BIOMETRIKA

Article Id: ASR017

Article Title: Testing parametric assumptions of trends of nonstationary time series

First Author: T. Zhang

Corr. Author: T. Zhang

**AUTHOR QUERIES – TO BE ANSWERED BY THE CORRESPONDING AUTHOR**

The following queries have arisen during the typesetting of your manuscript. Please answer these queries by marking the required corrections at the appropriate point in the text. Failure to do so could result in delayed publication.

Q1	Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online. Please also check all author affiliations.	
Q2	Please review the suggested running head.	
Q3	Figures have been placed as close as possible to their first citation. Please check that they have no missing sections and that the correct figure legend is present.	
Q4	Please supply supplementary material.	
Q5	Please update reference Wu and Zhou (2010).	
Q6	Cheng, B & Tong, H (1998) not cited in text.	
Q7	Horowitz, J-L & Spokoing, V.G (2001) not cited in text	

# Testing parametric assumptions of trends of nonstationary time series

BY TING ZHANG AND WEI BIAO WU

Q1

Department of Statistics, The University of Chicago, Chicago, Illinois 60637, U.S.A.

tzhang@galton.uchicago.edu    wbwu@galton.uchicago.edu

5

## SUMMARY

The paper considers testing whether the mean trend of a nonstationary time series is of certain parametric forms. A central limit theorem for the integrated squared error is derived, and with that a hypothesis-testing procedure is proposed. The method is illustrated in a simulation study, and is applied to assess the mean pattern of lifetime-maximum wind speeds of global tropical cyclones from 1981 to 2006. We also revisit the trend pattern in the central England temperature series.

*Some key words:* Bias correction; Central limit theorem; Integrated squared error; Local linear estimation; Locally stationary process; Nonparametric hypothesis testing.

## 1. INTRODUCTION

The problem of testing whether the mean trend of a time series follows certain parametric forms has attracted considerable attention; see Dette (1999), Bissantz et al. (2005), Percival & Rothrock (2005), Wu & Zhao (2007) and Pawlak & Stadmüller (1996, 2007), among others. Parametric models have the advantage of ease of interpretation and prediction, but may suffer from misspecification, leading to erroneous conclusions. Hypothesis testing in nonparametric regression under independence has been discussed by Härdle & Mammen (1993), Hart (1997), Fan et al. (2001) and Van Keilegom et al. (2008). Other contributions can be found in Azzalini et al. (1989), Eubank & LaRiccia (1992), Aerts et al. (1999), Eubank (1999), Horowitz & Spokoiny (2001) and Fan & Jiang (2007). In this paper we adopt the following formulation: suppose we observe

20

25

$$Y_i = \mu(i/n) + e_i \quad (i = 1, \dots, n), \quad (1)$$

where  $\mu(t)$ ,  $t \in [0, 1]$ , is an unknown signal or trend function and  $(e_i)_{i=1}^n$  is a zero-mean noise sequence which can be nonstationary. We are interested in testing the hypothesis

$$H_0 : \mu(t) = f(\theta, t), \quad (2)$$

where the function  $f(\cdot, \cdot)$  has a known form and  $\theta \in \mathbb{R}^d$  is a parameter vector of  $f$  being identically zero, constant, and special cases  $f \equiv 0$ ,  $f \equiv a$  constant and  $f(\theta, t) = \theta_0 + \theta_1 t$  for some  $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$  correspond to testing whether a signal exists, is time-varying and nonlinear, respectively. A natural approach would be to compare a nonparametric estimator of  $\mu(\cdot)$  and the fitted parametric trend  $f(\hat{\theta}_n, t)$ , where  $\hat{\theta}_n$  might, for example, be the least squares estimator, which minimizes

30

$$\ell_n(\theta) = \sum_{i=1}^n \{Y_i - f(\theta, i/n)\}^2. \quad (3)$$

35 Nonparametric estimation of  $\mu$  can be performed in a number of ways including the Priestley–Chao (1972), the Nadaraya–Watson, local polynomial, spline and wavelet methods. Here we shall use the local linear estimator, which has a good boundary performance (Fan & Gijbels, 1996),

$$\hat{\mu}_n(t) = \sum_{i=1}^n Y_i w_i(t), \quad (4)$$

40 where  $w_i(t) = K\{(i/n - t)/b_n\}\{S_2(t) - (t - i/n)S_1(t)\}/\{S_2(t)S_0(t) - S_1^2(t)\}$  are the local linear weights,  $b_n$  is the bandwidth,  $K(\cdot)$  is a kernel function and  $S_j(t) = \sum_{i=1}^n (t - i/n)^j K\{(i/n - t)/b_n\}$ . To test  $H_0$ , we shall develop central limit theory for the integrated squared error (ISE)

$$\text{ISE} = \int_0^1 \{\hat{\mu}_n(t) - \mu(t)\}^2 dt.$$

Under  $H_0$ , ISE can be estimated by  $\Delta_2^2$ , where  $\Delta_2$  is the  $\mathcal{L}^2$  distance

$$\Delta_2 = \left[ \int_0^1 \{\hat{\mu}_n(t) - f(\hat{\theta}_n, t)\}^2 dt \right]^{1/2}.$$

We reject  $H_0$  if  $\Delta_2$  is too large. The asymptotic normality of ISE has been studied under different settings. See for example Bickel & Rosenblatt (1973), Hall (1984), Ioannides (1992) and Alcalá et al. (1999), among others. However, in those papers the errors are independent. González-Manteiga & Vilar Fernández (1995) and Biedermann & Dette (2000) considered the same problem for linear processes with independent and identically distributed innovations. The ISE is a quadratic form in the errors  $e_i$ . As commented in Pawlak & Stadtmüller (2007), existing results on quadratic forms for dependent processes are mostly confined to linear processes. It is unclear whether similar results hold for general nonlinear processes. Here we shall substantially generalize earlier results by allowing nonlinear and nonstationary error processes. Hence our central limit theory should be widely applicable.

## 2. PRELIMINARIES

55 In (1) we allow nonstationary noise processes, on which there is a huge literature. Priestley (1965, 1988) considered processes with time-varying spectral representations. Dahlhaus (1997) defined a class of locally stationary processes for which a rigorous asymptotic theory can be obtained. Mallat et al. (1998) modelled locally stationary processes with pseudo-differential operators that are time-varying convolutions. Cheng & Tong (1998) applied wavelet representations. Nason et al. (2000) proposed to use a set of discrete non-decimated wavelets rather than the Fourier complex exponentials as in Dahlhaus (1997). Giurcanu & Spokoiny (2004) treated nonstationarity by assuming that correlation functions could be well approximated by those of stationary processes. Ombao et al. (2005) generalized the framework of Dahlhaus (1997) by utilizing the smooth localized complex exponentials. Here we shall follow the framework in Draghicescu et al. (2009) and Zhou & Wu (2009) and assume that the error sequence  $\{e_i\}_{i=1}^n$  is generated from the model

$$e_i = G(i/n; \mathcal{F}_i), \quad (5)$$

where  $\mathcal{F}_i = (\dots, \epsilon_{i-1}, \epsilon_i)$  is a shift process of independent and identically distributed shocks  $\epsilon_k$ ,  $k \in \mathbb{Z}$ , and  $G : [0, 1] \times \mathbb{R}^\infty \rightarrow \mathbb{R}$  is a measurable function such that  $e_i$  is well defined. The

framework (5) covers a wide range of nonstationary processes and naturally extends many existing stationary time series models to their nonstationary counterparts. Following Wu (2005), we interpret  $\mathcal{F}_i$  and  $e_i$  as the input and output, and  $G$  as the transform that represents the underlying physical mechanism. Let  $\zeta_i(t) = G(t; \mathcal{F}_i)$ . Then  $G(t; \cdot)$  is the data-generating mechanism at  $t$ , possibly time-varying. If  $G(t; \cdot)$  does not depend on  $t$ , then (5) provides a very general framework for stationary processes; see Priestley (1988) and Tong (1990). Under a stochastic continuity condition, cf. Condition 2, equation (5) defines a locally stationary process.

We now introduce dependence measures that will be needed for our asymptotic theory. Let  $\epsilon'_i, \epsilon_j$  ( $i, j \in \mathbb{Z}$ ) be independent and identically distributed and define the coupled process  $\mathcal{F}_{k,\{0\}} = (\dots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \dots, \epsilon_k)$  and  $\zeta_{k,\{0\}}(t) = G(t, \mathcal{F}_{k,\{0\}})$ . For  $p > 0$ , assume that  $c_p = \sup_{t \in [0,1]} \|\zeta_0(t)\|_p < \infty$ , where  $\|X\|_p = E(|X|^p)^{1/p}$ , and define

$$\delta_{k,p} = \sup_{t \in [0,1]} \|\zeta_k(t) - \zeta_{k,\{0\}}(t)\|_p.$$

If  $G(t, \mathcal{F}_k)$  does not depend functionally on  $\epsilon_0$ , then  $\zeta_k(t) - \zeta_{k,\{0\}}(t) = 0$ . So  $\delta_{k,p}$  measures the dependence of  $G(t; \mathcal{F}_k)$  on the single input  $\epsilon_0$  over  $t \in [0, 1]$ . Let  $p' = \min(p, 2)$  and define

$$\Theta_{m,p} = \sum_{j=m}^{\infty} \delta_{j,p}, \quad \Psi_{m,p} = \left( \sum_{j=m}^{\infty} \delta_{j,p}^{p'} \right)^{1/p'}.$$

We can interpret  $\Theta_{m,p}$  as the cumulative dependence of  $\epsilon_0$  on  $\{\zeta_j(t)\}_{j=m}^{\infty}$ . Throughout the paper we assume that the short-range dependence condition holds for some  $p \geq 2$ :

$$\Theta_{0,p} = \sum_{j=0}^{\infty} \delta_{j,p} < \infty. \quad (6)$$

If (6) holds with  $p = 2$ , then the long-run variance function of the stationary process  $\{\zeta_k(t)\}_k$  is bounded:

$$g(t) = \sum_{k \in \mathbb{Z}} \text{cov}\{\zeta_0(t), \zeta_k(t)\} < \infty. \quad (7)$$

Write  $\delta_k = \delta_{k,2}$  and  $\|\cdot\| = \|\cdot\|_2$ . We shall impose the following regularity conditions:

*Condition 1.* Let  $\mu \in \mathcal{C}^3[0, 1]$ ;

*Condition 2.* There exists  $C > 0$  such that  $\|\zeta_i(t_1) - \zeta_i(t_2)\| \leq C|t_1 - t_2|$  holds for all  $t_1, t_2 \in [0, 1]$ ;

*Condition 3.* Let  $g \in \mathcal{C}^2[0, 1]$ .

The key, Condition (2) indicates that the underlying data-generating mechanism  $\zeta_i(t) = G(t; \mathcal{F}_i)$  changes smoothly in time, thus suggesting local stationarity. More specifically, a length  $k$  subsequence  $(e_j)_{j=i}^{i+k-1}$  can be approximated by the stationary process  $e_j^{\circ} = G(i/n, \mathcal{F}_j)$ ,  $j = i, i+1, \dots, i+k-1$ , in that  $\|e_j - G(i/n, \mathcal{F}_j)\| \leq Ck/n = o(1)$  if  $k = o(n)$ .

### 3. MAIN RESULTS

#### 3.1. Asymptotic normality

Throughout the paper we assume that in (4) the kernel function  $K(\cdot) \in \mathcal{K}$ , the collection of symmetric, bounded functions in  $\mathcal{C}^1[-1, 1]$  with  $\int_{-1}^1 K(v)dv = 1$ . For example,  $K$  can be the

100 Epanechnikov kernel  $K(v) = 3 \max(0, 1 - v^2)/4$ , the Bartlett kernel  $K(v) = \max(0, 1 - |v|)$ , or the rectangle kernel  $K(v) = I(|v| \leq 1)/2$ , where  $I(\cdot)$  is the indicator function. Define

$$K^*(x) = \int_{-1}^{1-2|x|} K(v)K(v+2|x|)dv.$$

Since  $K \in \mathcal{C}^1[-1, 1]$ , we have  $K^* \in \mathcal{K}$ , which is continuous on  $\mathbb{R}$ . We present below central limit theorems for

$$I_n = \text{ISE} - E(\text{ISE}) = \int_0^1 (\{\hat{\mu}_n(t) - \mu(t)\}^2 - E\{\{\hat{\mu}_n(t) - \mu(t)\}^2\})dt$$

when  $\mu \equiv 0$  and  $\mu \not\equiv 0$ , respectively. It turns out that the asymptotic distributions are different for these two cases.

105 **THEOREM 1.** *Let  $\mu \equiv 0$ . Assume Condition 2, Condition 3,  $\Theta_{0,4} < \infty$ ,  $K \in \mathcal{K}$ ,  $b_n \rightarrow 0$  and  $nb_n \rightarrow \infty$  as  $n \rightarrow \infty$ .*

(i) *Let  $g_1 = \int_0^1 g(t)dt$ ,  $\Gamma_k = \sum_{i=0}^{\infty} \delta_i \delta_{i+|k|}$  and  $l_n = \lfloor 2nb_n \rfloor$ . Then*

$$nb_n \int_0^1 E\{\hat{\mu}_n(t)^2\}dt - g_1 K^*(0) = \sum_{k=0}^{\infty} O\{\min(\Gamma_k, b_n) + \min(1, k/l_n)\Gamma_k\}. \quad (8)$$

(ii) *Let  $K_2^* = \int_{-1}^1 \{K^*(v)\}^2 dv$  and  $g_2 = \int_0^1 g^2(t)dt$ . If  $nb_n^{3/2} \rightarrow \infty$ , then as  $n \rightarrow \infty$ ,*

$$nb_n^{1/2} I_n \rightarrow N(0, 4g_2 K_2^*) \quad (9)$$

*in distribution.*

110 Theorem 1(i) deals with the mean integrated squared error  $E(\text{ISE}) = \int_0^1 E\{\hat{\mu}_n(t)\}^2 dt$ . Since  $\sum_{k=0}^{\infty} \Gamma_k \leq (\sum_{i=0}^{\infty} \delta_i)^2 < \infty$ ,  $b_n \rightarrow 0$  and  $l_n \rightarrow \infty$ , by the Lebesgue dominated convergence theorem, the right-hand side of (8) goes to zero as  $n \rightarrow \infty$ . The quantity  $g_1$  is the integrated long-run variance. For stationary processes we have  $g(t) \equiv g(0)$  for all  $t$ , and  $g_1 = g(0)$ ,  $g_2 = g_1^2$ . Based on Theorem 1, we present in § 3.2 a simulation-based testing procedure.

115 **THEOREM 2.** *Assume Conditions 1–3 and  $\Theta_{0,4} < \infty$ . Also assume  $K \in \mathcal{K}$ ,  $b_n \rightarrow 0$  and  $nb_n^{3/2} \rightarrow \infty$  as  $n \rightarrow \infty$ .*

(i) *If  $nb_n^5 \rightarrow 0$ , then (9) holds.*

(ii) *Let  $\kappa_2 = \int K(v)v^2 dv$ . If  $nb_n^5 \rightarrow \infty$ , then*

$$\frac{n^{1/2}}{b_n^2} I_n \rightarrow \kappa_2 N \left\{ 0, \int_0^1 g(t)\mu''(t)^2 dt \right\}$$

*in distribution.*

120 For both big and small bandwidths,  $I_n$  is asymptotically normal. However, for those two cases, both the normalization sequences and the asymptotic variances are different.

### 3.2. Implementation

We now present a testing procedure based on results in §3.1. Assume at the outset that the long-run variance function  $g(\cdot)$  is known and we want to test the hypothesis of nonexistence of a signal

$$H'_0 : \mu \equiv 0. \quad (10)$$

Section 3.4 concerns testing of the hypothesis (2) with general nonzero parametric forms, and §3.6 presents an estimator for the long-run variance function  $g(\cdot)$ . To test (10), by Theorem 1, it is natural to use  $\text{ISE} = \int_0^1 \hat{\mu}_n(t)^2 dt$  and we reject  $H'_0$  at level  $\alpha$ ,  $0 < \alpha < 1$ , if

$$\text{ISE} > g_1 K^*(0)(nb_n)^{-1} + z_{1-\alpha} n^{-1} b_n^{-1/2} (4g_2 K_2^*)^{1/2},$$

where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ th quantile of a standard normal distribution. However, our simulation study in § 4.1 shows that convergence in Theorem 1 can be quite slow. To overcome this, we propose a simulation-based procedure. Let  $Z_i$  ( $i \in \mathbb{Z}$ ) be independent standard normal random variables,  $Y_i^\diamond = e_i^\diamond = g(i/n)^{1/2} Z_i$ , and let  $\hat{\mu}_n^\diamond(\cdot)$  be the corresponding local linear estimator (4) with  $Y_i$  therein replaced by  $Y_i^\diamond$ . By Theorem 1, we have

$$nb_n \int_0^1 E\{\hat{\mu}_n^\diamond(t)^2\} dt - g_1 K^*(0) = O(b_n)$$

and, for  $\text{ISE}^\diamond = \int_0^1 \hat{\mu}_n^\diamond(t)^2 dt$ , the central limit theorem holds: as  $n \rightarrow \infty$ ,

$$nb_n^{1/2} \{\text{ISE}^\diamond - E(\text{ISE}^\diamond)\} \rightarrow N(0, 4g_2 K_2^*) \quad (11)$$

in distribution. Hence  $\text{ISE} = \int_0^1 \hat{\mu}_n(t)^2 dt$  and  $\text{ISE}^\diamond$  have the same asymptotic normal distribution, with mean  $g_1 K^*(0)n^{-1}b_n^{-1}$  and variance  $4g_2 K_2^* n^{-2} b_n^{-1}$  if the bound in the right-hand side of (8) is of order  $o(b_n^{1/2})$ . The latter observation suggests that, instead of using the central limit theorem, the cutoff value of  $\text{ISE}$  can be obtained by simulating  $\text{ISE}^\diamond$ .

### 3.3. Asymptotic power

We consider the power of our test for the local alternative model of the form

$$Y_i = a_n h(i/n) + e_i,$$

where  $h$  is a known nonzero  $\mathcal{C}^2$  function on  $[0, 1]$  and  $(a_n)_{n \geq 1}$  is a positive sequence with  $a_n \rightarrow 0$ . Here  $a_n$  indicates the magnitude of departure from the null hypothesis. By Proposition 1, the power goes to 1 if  $nb_n^{1/2} a_n^2 \rightarrow \infty$ .

**PROPOSITION 1.** *Assume Conditions 2 and 3 and  $\Theta_{0,4} < \infty$ . Also assume that  $K \in \mathcal{K}$ ,  $b_n \rightarrow 0$ ,  $nb_n^{3/2} \rightarrow \infty$  and  $nb_n^{1/2} a_n^2 \rightarrow c > 0$  as  $n \rightarrow \infty$ . Let  $\Phi$  be the standard normal distribution function. Then the power of our testing procedure converges to  $\Phi\{z_\alpha + (4g_2 K_2^*)^{-1/2} c \int_0^1 h(t)^2 dt\}$ .*

*Proof.* Let  $\hat{\mu}_n^0(t) = \sum_{i=1}^n w_i(t) e_i$  and  $h_n(t) = \sum_{i=1}^n w_i(t) h(i/n)$ . Then

$$\int_0^1 \{\hat{\mu}_n(t)\}^2 - \{\hat{\mu}_n^0(t)\}^2 dt = 2a_n \int_0^1 \hat{\mu}_n^0(t) h_n(t) dt + a_n^2 \int_0^1 h_n(t)^2 dt. \quad (12)$$

Since  $h \in \mathcal{C}^2[0, 1]$ ,  $|\int_0^1 w_i(t)h_n(t)dt| \leq C/n$  for some  $C > 0$ . By Lemma 1,

$$nb_n^{1/2}a_n \left\| \int_0^1 \hat{\mu}_n^0(t)h_n(t)dt \right\| = nb_n^{1/2}a_n \mathcal{O}(n^{-1/2}) \rightarrow 0.$$

So, Proposition 1 follows from Theorem 1 and (12) since  $\int_0^1 h_n(t)^2 dt \rightarrow \int_0^1 h(t)^2 dt$ .  $\square$

### 3.4. Hypothesis testing with general parametric forms

Under the null hypothesis (2), let  $\theta_0$  be the true value and  $\hat{\theta}_n$  be the minimizer of (3). Then a natural test statistic would be  $\int_0^1 \{\hat{\mu}_n(t) - f(\hat{\theta}_n, t)\}^2 dt$ . However, it is not convenient to use this directly since by Theorem 2,  $\int_0^1 \{\hat{\mu}_n(t) - f(\theta_0, t)\}^2 dt$  can have different asymptotic normal distributions for different bandwidths. It is nontrivial to determine which central limit theorem to use. To solve this problem, we use the modified version

$$\text{ISE}_M = \int_0^1 \{\hat{\mu}_n(t) - \mu_M(t)\}^2 dt,$$

where  $\mu_M(t) = \sum_{i=1}^n f(\theta_0, i/n)w_i(t)$  is the local linear smoothed version of  $f(\theta_0, t)$ . Since  $\hat{\mu}_n(t) - \mu_M(t) = \sum_{i=1}^n w_i(t)e_i$ ,  $\text{ISE}_M$  reduces to  $\text{ISE}$  with  $\mu \equiv 0$ . The bias then disappears on replacing  $\mu(t)$  by  $\mu_M(t)$ . Such a bias correction scheme was previously used in Härdle & Mammen (1993). Note that  $\text{ISE}_M$  is not directly usable since it depends on the unknown function  $\mu$ , which under  $H_0$  depends on the parameter  $\theta_0$ . It can be estimated by

$$\text{ISE}_{\hat{M}} = \int_0^1 \{\hat{\mu}_n(t) - \mu_{\hat{M}}(t)\}^2 dt = \int_0^1 \{\hat{\mu}_n^{(e)}(t)\}^2 dt, \quad (13)$$

where  $\mu_{\hat{M}}(t) = \sum_{i=1}^n f(\hat{\theta}_n, i/n)w_i(t)$  and  $\hat{\mu}_n^{(e)}(t) = \hat{\mu}_n(t) - \mu_{\hat{M}}(t)$  is the local linear smoother of the estimated residuals  $\hat{e}_i = Y_i - f(\hat{\theta}_n, i/n)$ .

PROPOSITION 2. Assume that  $\dot{f}(\theta, t) = \partial f(\theta, t)/\partial \theta$  exists at a neighbourhood of  $\theta_0$ , that

$$\sup_{t \in [0, 1]} \sup_{|\theta - \theta_0| \leq c} |\dot{f}(\theta, t)| < \infty \quad (14)$$

holds for some  $c > 0$ , and that  $f(\theta, t)$  admits the uniform Taylor expansion: as  $\theta \rightarrow \theta_0$ ,

$$\sup_{t \in [0, 1]} |f(\theta, t) - f(\theta_0, t) - (\theta - \theta_0)^T \dot{f}(\theta_0, t)| = \mathcal{O}(|\theta - \theta_0|^2). \quad (15)$$

Then under conditions of Theorem 1, if  $\hat{\theta}_n - \theta_0 = \mathcal{O}_p(n^{-1/2})$ , we have

$$n(\text{ISE}_{\hat{M}} - \text{ISE}_M) = \mathcal{O}_p(1). \quad (16)$$

Proposition 2 implies that the statistic  $\text{ISE}_{\hat{M}}$  can approximate  $\text{ISE}_M$  well, and it is also asymptotically normally distributed with same asymptotic mean and variance, given in Theorem 1.

### 3.5. Bandwidth selection

Choosing a bandwidth such that the test procedure based on  $\text{ISE}_{\hat{M}}$  has a good performance, is nontrivial, and in our case it is further complicated by the presence of dependence and nonstationarity. In the regression setting with independent errors, the problem was considered

by Härdle & Marron (1985), Härdle et al. (1988), Park & Marron (1990), Ruppert et al. (1995), Wand & Jones (1995) and Xia (1998), among others. Herrmann et al. (1992) and Wu & Zhao (2007) dealt with models with stationary errors. Hall & Hart (1990), Kulasekera & Wang (1997) and Gao & Gijbels (2008) considered bandwidth selection in the context of nonparametric hypothesis testing. 175

We propose using the asymptotic mean squared error optimal bandwidth  $b_n = cn^{-1/5}$ , where  $c > 0$  is a constant. Due to dependence and nonstationarity, it is difficult to estimate  $c$ . On the positive side, our simulation studies carried out in §4 suggests that the performance of our simulation-based test is relatively robust to the choice of  $c$ . Hence in our simulation and data analysis, we simply choose  $b_n = n^{-1/5}$  in computing  $\text{ISE}_{\hat{M}}$ , and interestingly, this simple choice performs quite well. 180

### 3.6. Estimation of variance functions

In the implementation of our testing procedure, a key issue is to estimate the pointwise long-run variance function,  $g(t)$ ,  $t \in [0, 1]$ . If the errors  $e_i$  were independent and identically distributed, then  $g(t) \equiv \sigma_e^2 = \|e_i\|^2$ , the variance of  $e_i$ . In this case we can apply difference-based variance estimators and there is a huge literature on the estimation of  $\sigma_e^2$ ; see Rice (1984), Hall et al. (1990) and Dette et al. (1998), among others. In our setting, however, due to the dependence, the difference-based approach is generally invalid. For example, assuming that  $\mu \in \mathcal{C}^2[0, 1]$  and  $e_i$  are stationary, as  $n \rightarrow \infty$ , Rice's (1984) estimator  $(2n - 2)^{-1} \sum_{i=2}^n (Y_i - Y_{i-1})^2 \rightarrow \gamma_0 - \gamma_1$ , by the ergodic theorem. Here  $\gamma_k = \text{cov}(e_0, e_k)$  is the auto-covariance function of  $(e_i)$ . Note that  $\sigma^2 = \sum_{k \in \mathbb{Z}} \gamma_k$ , which is generally different from  $\gamma_0 - \gamma_1$ . 185

To account for dependence and nonstationarity, we estimate  $g(t)$  by

$$\hat{g}(t) = \frac{\sum_{i=1}^n Q_i \mathbf{I}(|i/n - t| \leq b_n)}{\sum_{i=1}^n \mathbf{I}(|i/n - t| \leq b_n)}, \quad (17)$$

where  $Q_i = e_i \sum_{|j-i| \leq m_n} e_j$ . By Theorem 3, the above estimator is consistent.

**THEOREM 3.** Assume Conditions 2 and 3,  $\Theta_{0,4} < \infty$ ,  $b_n \rightarrow 0$  and  $m_n^{-1} + m_n(nb_n)^{-1} \rightarrow 0$ . Then as  $n \rightarrow \infty$ , 195

$$(nb_n/m_n)^{1/2} [\hat{g}(t) - E\{\hat{g}(t)\}] \rightarrow N\{0, 2g^2(t)\} \quad (18)$$

in distribution for any  $t \in (0, 1)$ . Also, uniformly over  $t \in [b_n, 1 - b_n]$ , the bias

$$E\{\hat{g}(t)\} - g(t) = O(b_n^2) + \sum_{k=0}^{m_n} O\{\min(\Gamma_k, m_n/n)\} + \sum_{k>m_n} O(\Gamma_k). \quad (19)$$

If there exists  $\rho \in (0, 1)$  such that  $\Gamma_k = O(\rho^k)$ , letting  $m_n = \lfloor \log n / \log \rho^{-1} \rfloor$ , we have by (19) that  $E\{\hat{g}(t)\} - g(t) = O(b_n^2) + O(m_n^2/n)$ . Hence, by (18), the mean squared error of  $\hat{g}(t)$  is of order  $O(b_n^4 + m_n^4/n^2) + O\{m_n(nb_n)^{-1}\} = O\{(n^{-1} \log n)^{4/5}\}$  if  $b_n \asymp (n^{-1} \log n)^{1/5}$ . 200

To use (17), we suggest using the automatic bandwidth selector in Ruppert et al. (1995) to obtain a local linear fit  $\tilde{\mu}_n(\cdot)$  of the mean function and then the estimated residuals  $\tilde{e}_i = Y_i - \tilde{\mu}_n(i/n)$ . Then we replace  $e_i$  in (17) by  $\tilde{e}_i$ .

### 3.7. A simulation-based testing procedure

We summarize the testing procedure as follows. Its validity is justified by noting that  $\text{ISE}$  and  $\text{ISE}^\circ$  have the same asymptotic distribution, as argued in (11). Section 4 presents a simulation 205

study of its finite-sample performance.

- (i) Select a bandwidth  $b_n^*$  by the procedure in [Ruppert et al. \(1995\)](#); perform a local linear fit for  $\mu(\cdot)$  and obtain the estimated residuals  $\tilde{e}_i = Y_i - \tilde{\mu}_n(i/n)$ ; compute  $\hat{g}(t)$ ,  $t \in [0, 1]$ , via (17).
- 210 (ii) Obtain an estimator  $\hat{\theta}_n$  for the parameter  $\theta$ ; compute  $\tilde{e}_i = Y_i - f(\hat{\theta}_n, i/n)$  and  $\text{ISE}_{\hat{M}}$  via (13) with the local linear estimator (4) where  $b_n = n^{-1/5}$  is used.
- (iii) Generate independent  $Z_1, \dots, Z_n \sim N(0, 1)$  and let  $Y_i^\diamond = \hat{g}(i/n)^{1/2} Z_i$ . Then compute the corresponding  $\text{ISE}_{\hat{M}}^\diamond$  in the same manner as (ii).
- (iv) Let  $\alpha \in (0, 1)$  be the significance level. Repeat step (iii) and obtain the estimated quantile 215  $\hat{q}_{1-\alpha}$  of  $\text{ISE}_{\hat{M}}^\diamond$ , the bootstrap cutoff value.
- (v) Reject the null hypothesis at level  $\alpha$  if  $\text{ISE}_{\hat{M}} > \hat{q}_{1-\alpha}$ .

#### 4. A SIMULATION STUDY

##### 4.1. Approximations of distributions of test statistics

Consider model (1) with  $e_i = \zeta_i(i/n)$ , where for any  $t \in [0, 1]$ ,  $\{\zeta_i(t)\}_{i \in \mathbb{Z}}$  follows the recursion

$$\zeta_i(t) = \rho(t)\zeta_{i-1}(t) + \sigma\epsilon_i. \quad (20)$$

220 Here  $\epsilon_i, i \in \mathbb{Z}$ , are independent random variables with  $\text{pr}(\epsilon_i = -1) = \text{pr}(\epsilon_i = 1) = 1/2$ . Thus,  $(e_i)_{i \in \mathbb{Z}}$  is a first-order autoregressive process with time-varying coefficient. Calculations show that  $E\{\zeta_i(t)\} = 0$ ,  $g_0(t) = \text{var}\{\zeta_i(t)\} = \sigma^2/\{1 - \rho(t)^2\}$  and the long-run variance function  $g(t) = \sigma^2/\{1 - \rho(t)\}^2$ . We use the Epanechnikov kernel  $K(v) = 3 \max(0, 1 - v^2)/4$ . Then  $K^*(0) = 3/5$  and  $K_2^* = 167/770$ . We consider the problem of testing  $H_0': \mu \equiv 0$ . Choose  $\rho(t) =$  225  $0.1 + 0.4t$ ,  $\sigma = 1$  and  $n = 500$ , so the bandwidth  $b_n = n^{-1/5} = 0.289$ . We simulate 50 000 realizations of  $\text{ISE}_{\hat{M}}$ . Three different approximations of  $\text{ISE}_{\hat{M}}$  are considered: the normal approximation in Theorem 1;  $\text{ISE}^\sharp = \int_0^1 \hat{\mu}_n^\sharp(t)^2 dt$ , where  $\hat{\mu}_n^\sharp(t) = \sum_{i=1}^n Y_i^\sharp w_i(t)$ ,  $Y_i^\sharp = g_0(i/n)^{1/2} Z_i$  and  $Z_i$  are independent standard normal variables; and  $\text{ISE}^\diamond = \int_0^1 \hat{\mu}_n^\diamond(t)^2 dt$ , introduced in §3.2, which differs from the second in that the long-run variance function  $g$  is used instead of the 230 marginal variance function  $g_0$ . In the second scheme, the dependence is ignored. We use Q–Q plot to compare the distributions. The results are presented in Fig. 1, which shows that the normal approximation does not have a satisfactory finite-sample performance. A similar phenomenon was observed by [Härdle & Mammen \(1993\)](#). If we ignore the inherited dependence structure, then one may obtain an erroneous conclusion; see Fig. 1(b) of Fig. 1. As shown in Fig. 1(c), the 235 dependence-adjusted procedure provides a very good approximation of  $\text{ISE}_{\hat{M}}$ . The same conclusion applies to other parametric forms.

For a theoretical justification of the superiority of the simulation-based method, we use the Gaussian approximation principle in [Wu & Zhou \(2010\)](#). Consider the linear process  $X_i = \sum_{j=0}^\infty a_j(i/n)\eta_{i-j}$ , where  $\eta_i$  are independent and identically distributed with mean 0 and  $\eta_i \in$  240  $\mathcal{L}^p$  ( $p > 2$ ), and  $a_j(\cdot)$  are differentiable functions satisfying  $\sum_{j=0}^\infty \sup_{t \in [0,1]} |a_j'(t)| < \infty$  and  $\sum_{j=m}^\infty \sup_{t \in [0,1]} |a_j(t)| = O(m^{1/p-1/2})$ . Then on a richer probability space one can construct  $e_1^*, \dots, e_n^*$  and independent standard normal random variables  $Z_1^*, \dots, Z_n^*$  such that  $(e_i^*)_{i=1}^n$  and  $(e_i)_{i=1}^n$  have the same distribution and, for  $S_i^* = \sum_{j=1}^i e_j^*$  and  $T_i^* = \sum_{j=1}^i g(i/n)^{1/2} Z_j^*$ ,

$$\max_{i \leq n} |S_i^* - T_i^*| = O_p(n^{1/p} \log n). \quad (21)$$

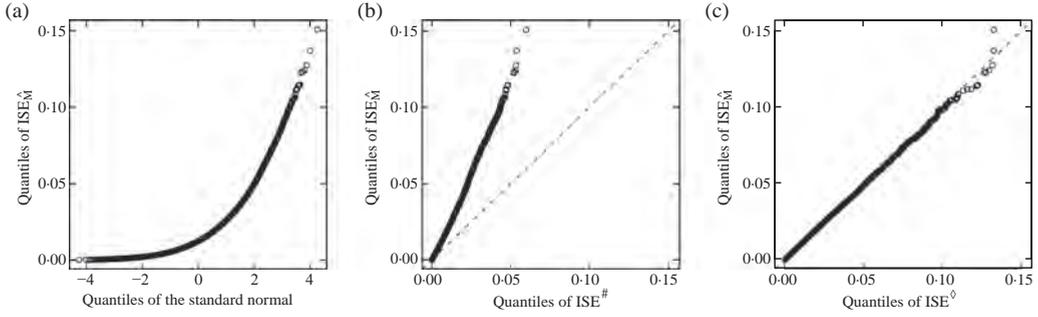


Fig. 1. Title Q-Q plots of the test statistic  $ISE_{\hat{M}}$  against normal approximation (a), approximation without dependence,  $ISE^{\#}$ , (b) and approximation with dependence,  $ISE^{\diamond}$ , (c). The dashed lines in (b) and (c) have unit slope and zero intercept.

Q3

Let  $\mu_n^*(t) = \sum_{i=1}^n w_i(t)e_i^*$  and  $v_n^*(t) = \sum_{i=1}^n w_i(t)g(i/n)^{1/2}Z_i^*$ . For the local linear weights  $w_i(t)$  in (4), if  $K$  is Lipschitz continuous, by (21), we obtain

245

$$\sup_{0 \leq t \leq 1} |\mu_n^*(t) - v_n^*(t)| = O_p\{n^{1/p} \log n/(nb_n)\} \quad (22)$$

by a summation by parts technique. By Theorem 1,  $\int_0^1 \mu_n^*(t)^2 dt = O_p\{(nb_n)^{-1}\}$ . So, by (22),

$$\int_0^1 |\mu_n^*(t)^2 - v_n^*(t)^2| dt = O_p\{n^{1/p} \log n (nb_n)^{-3/2}\}. \quad (23)$$

When  $\mu \equiv 0$ , since  $(e_i^*)_{i=1}^n$  and  $(e_i)_{i=1}^n$  have the same distribution,  $\int_0^1 \mu_n^*(t)^2 dt$  is identically distributed as  $\int_0^1 \hat{\mu}_n(t)^2 dt$ . Note that  $\int_0^1 v_n^*(t)^2 dt$  corresponds to  $ISE_{\hat{M}}^{\diamond}$  in step (iii) in §3.7 if  $\hat{g}$  therein is replaced by the true  $g$ . For the mean squared error optimal bandwidth  $b_n \asymp n^{-1/5}$ , the error bound in (22) is  $O_p(n^{\tau-6/5})$ . Since  $\tau$  can be arbitrarily small if  $p$  is large,  $\int_0^1 \mu_n^*(t)^2 dt$  and  $\int_0^1 v_n^*(t)^2 dt$  can be very close; recalling Proposition 2 that the difference between  $ISE_M$  and the realized version  $ISE_{\hat{M}}$  has a larger order  $O_p(n^{-1})$ . So, we expect that the simulation-based method can have an excellent performance.

250

#### 4.2. Effect of bandwidths

We consider model (1) with error structure (20). With true mean function  $\mu(t) = 4t^2 - 4t + 3$ ,  $t \in [0, 1]$ , we consider testing whether the mean function has a quadratic form. To study how bandwidths affect the performance of our test, we choose  $\rho(t) = 0.3 - 0.5t$ ,  $\sigma = 1$  and 10 levels of  $b$ :  $b = 0.05j$  ( $j = 1, \dots, 10$ ). For each  $b$ , we calculate  $\hat{q}_{0.95}$  by repeating step (iii) in §3.7 50 000 times. Then we generate 50 000 realizations of the time-varying AR(1) process and compute the corresponding  $ISE_{\hat{M}}$  values for each realization. Simulated empirical rejection proportions with  $n = 100, 200$  and  $500$ , presented in Table 1, are reasonably close to the nominal level. If we choose  $b_n = n^{-1/5}$ , then  $b_{100} = 0.398$ ,  $b_{200} = 0.347$  and  $b_{500} = 0.289$ , and the corresponding empirical rejection probabilities are about 5.01, 4.97 and 5.03%, respectively. In addition, they become more robust to the change in bandwidths as  $n$  gets larger. So, in practice, we recommend using  $b_n = n^{-1/5}$ .

255

260

265

Table 1. Empirical rejection percentages with different bandwidths  $b$  and sample sizes  $n$ . The significance level is 5%.

$n$	$b$									
	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
100	2.8	4.0	4.5	4.7	4.7	5.0	5.2	5.0	5.1	5.1
200	3.8	4.6	4.6	5.1	4.6	5.1	5.0	5.2	5.1	5.2
500	4.6	4.7	5.0	4.9	5.1	5.0	5.0	5.1	5.2	5.1

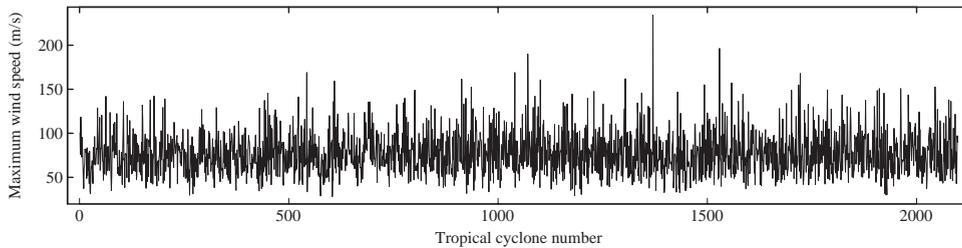


Fig. 2. Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006.

## 5. APPLICATIONS

### 5.1. Tropical cyclone data

According to Emanuel (1991), Holland (1997) and Bengtsson (2007), global warming is leading to increasing ocean temperatures and consequently to more tropical cyclones. Elsner et al. (2008) analysed patterns of tropical cyclone winds by fitting linear trends for quantiles of satellite-derived lifetime-maximum wind speeds of 2098 tropical cyclones over the globe during 1981–2006. The data set is available in the Supplementary material see Elsner et al. (2008) for a detailed description. Figure 2 shows these data.

We shall model the wind speed data by (1) and test the null hypothesis  $H_0 : \mu(\cdot)$  is linear. Zhou (2010) argued that the error process  $(e_i)$  is nonstationary. Hence  $g(t)$  is not a constant function. Using the procedure in Ruppert et al. (1995), we select  $b_n^* = 0.081$  and estimate the long-run variance function  $g(t)$  by using (17) with  $m_n = (nb_n^*)^{1/3}$ . For hypothesis testing, we choose  $b_n = n^{-1/5} = 0.217$ . The test statistic  $\text{ISE}_{\hat{M}} = 1.011$  with  $p$ -value 0.12 after 50 000 repetitions of step (iii) in §3.7. So, at the 5% significance level, we accept the linear trend hypothesis. The fitted trend, with standard error, is  $\mu(t) = 75.3(1.1) + 3.8(1.8)t$ ,  $t \in [0, 1]$ . For testing the hypothesis  $H_0^* : \mu$  is constant, using the same method we obtain  $\text{ISE}_{\hat{M}} = 2.091$  with  $p$ -value 0.04. Thus the mean constancy hypothesis is rejected at the 5% level. Zhou (2010) applied an  $\mathcal{L}^\infty$ -based method, and failed to reject  $H_0^*$ . Our  $\mathcal{L}^2$ -based testing procedure appears to be more powerful.

### 5.2. Central england temperature data

We consider the annual central England temperature series from year 1659 to 2009 by using model (1). The time series is plotted in Fig. 3 and the data are available in the Supplementary material. It was first constructed by Manley (1974) and is now routinely updated by the Hadley Center, U.K. See Jones & Hulme (1997) for a more detailed description. Jones & Hulme (1997) and Jones & Bradley (1992a) fitted linear trends, while Benner (1999) and Harvey & Mills (2003) fitted quadratic curves. Realizing that the quadratic trend assumption might not be appropriate, Harvey & Mills (2003) also tried local polynomial regression. Here we shall test the

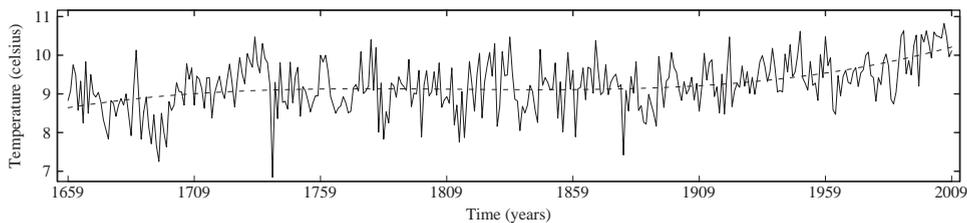


Fig. 3. Annual central England temperature series from 1659 to 2009. The dashed curve is the global cubic fit.

hypothesis  $H_0 : \mu(\cdot)$  is quadratic. As in the analysis of the tropical cyclone data, we follow the procedure in §3.7 and obtain the test statistic  $\text{ISE}_{\hat{M}} = 0.0059$  with  $p = 0.00002$ . Hence, the quadratic trend assumption is rejected at the 5% level. Interestingly, the cubic trend hypothesis is accepted: the test statistic  $\text{ISE}_{\hat{M}} = 0.000096$  and the corresponding  $p$ -value is 0.47. The fitted equation with standard error in parenthesis, is  $\mu(t) = 8.63(0.13) + 3.7(1.1)t - 8.9(2.5)t^2 + 6.8(1.7)t^3$ , where  $t \in [0, 1]$ . The cubic trend fit accords well with Benner's observation that the whole series has roughly three periods: the earliest part corresponds to the coldest weather which may represent the little ice age (Jones & Bradley, 1992b), the middle part fluctuates around the mean, while the last part exhibits a warming trend. This cannot be described by a quadratic fit.

#### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *Biometrika* online.

#### ACKNOWLEDGEMENT

We are grateful to the editor, an associate editor and three referees for their helpful comments and suggestions which render a much improved version. We thank Professor Wand for clarifications on bandwidth selection procedures. The research was supported by the US National Science Foundation.

#### APPENDIX

Recall that  $\zeta_k(t) = G(t, \mathcal{F}_k)$ . We assume that  $E\{G(t, \mathcal{F}_k)\} = 0$  for any  $t \in [0, 1]$ . Define

$$\tilde{\zeta}_k(t) = E\{\zeta_k(t) \mid \mathcal{F}_{k-m,k}\}, \quad m \in \mathbb{N}.$$

Then for all  $t \in [0, 1]$ ,  $\{\tilde{\zeta}_k(t)\}_k$  is  $m$ -dependent with mean zero. Define the projection operator

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

For a set  $T \subseteq \mathbb{Z}$ , let  $\epsilon_{i,T} = \epsilon_i'$  if  $i \in T$ , and  $\epsilon_{i,T} = \epsilon_i$  otherwise. Let  $\tilde{\mathcal{F}}_{i,j,T} = (\epsilon_{k,T}, k = i, \dots, j)$ . In the proof  $C$  denotes constants whose value may vary from place to place.

To prove Theorem 1, we shall apply the technique in Liu & Wu (2010). Lemmas 1, 2 and 3 provide bounds for  $m$ -dependent and martingale approximations for linear and quadratic forms. They can be proved by using the arguments in Lemma 1, Propositions 1 and 2 in Liu & Wu (2010), respectively. That paper deals with stationary processes, but there are no essential additional difficulties involved for generalization to nonstationary processes. Detailed proofs can be found in the online Supplementary Material.

LEMMA A1. Assume  $\Theta_{0,p} < \infty$ ,  $p \geq 2$ . Let  $\alpha_1, \alpha_2, \dots, \in \mathbb{R}$ ,  $A_n = (\sum_{i=1}^n \alpha_i^2)^{1/2}$ , and  $C_p = 18p^{3/2}(p-1)^{-1/2}$ . Then (i)  $\|\sum_{i=1}^n \alpha_i e_i\|_p \leq C_p A_n \Theta_{0,p}$ ; and (ii)  $\|\sum_{i=1}^n \alpha_i (e_i - \tilde{e}_i)\|_p \leq C_p A_n \Theta_{m+1,p}$ .

320 LEMMA A2. Assume  $\Theta_{0,4} < \infty$ . Let  $\alpha_j \in \mathbb{R}$ ,

$$L_n = \sum_{1 \leq i < j \leq n} \alpha_{j-i} e_i e_j, \quad \tilde{L}_n = \sum_{1 \leq i < j \leq n} \alpha_{j-i} \tilde{e}_i \tilde{e}_j. \quad (\text{A1})$$

Let  $d_{m,4} = \sum_{k=0}^{\infty} \min(\delta_{k,4}, \Psi_{m+1,4})$  and  $A_n = (\sum_{i=1}^{n-1} \alpha_i^2)^{1/2}$ . Then

$$\|\{L_n - E(L_n)\} - \{\tilde{L}_n - E(\tilde{L}_n)\}\| \leq C_4 \Theta_{0,4} d_{m,4} n^{1/2} A_n.$$

LEMMA A3. Let  $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$  and  $V_m(\alpha) = \max_{i < n} \alpha_i^2 + m \sum_{i=1}^{n-1} |\alpha_i - \alpha_{i-1}|^2$ ; let

$$\tilde{D}_{k,n} = \sum_{l=0}^{\infty} \mathcal{P}_k \tilde{e}_{k+l}, \quad \tilde{M}_n = \sum_{1 \leq i < j \leq n} \alpha_{j-i} \tilde{D}_{i,n} \tilde{D}_{j,n}.$$

Assume  $\Theta_{0,4} < \infty$ . Then  $\|\tilde{L}_n - E(\tilde{L}_n) - \tilde{M}_n\|^2 \leq C m^3 n V_m(\alpha)$ .

325 THEOREM A1. Assume Condition 2 and Condition 3, (A3),  $\Theta_{0,4} < \infty$ ,  $K \in \mathcal{K}$ ,  $b_n \rightarrow 0$  and  $nb_n^{3/2} \rightarrow \infty$ . Recall (A1) for  $L_n$  and let  $\alpha_j = n^{-1} b_n^{-1/2} K^* \{j/(2nb_n)\}$ . Then as  $n \rightarrow \infty$ ,  $L_n - E(L_n) \rightarrow N(0, g_2 K_2^*)$  in distribution.

*Proof.* Recall Lemma A2 for  $A_n$  and Lemma A3 for  $V_m(\alpha)$ . Since  $K^* \in \mathcal{K}$ ,  $V_m(\alpha) = O(mn^3 b_n^2)$  and  $A_n = O(n^{-1/2})$ . By Lemmas A2 and A3,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \|L_n - E(L_n) - \tilde{M}_n\| = 0. \quad (\text{A2})$$

330 In (7) we replace  $\zeta_i(t)$  by  $\tilde{\zeta}_i(t)$  and let  $\tilde{g}(t)$  be the long-run variance function for the latter  $m$ -dependent process. Set  $\tilde{g}_2 = \int_0^1 \tilde{g}^2(t) dt$ , by (A2) it suffices to verify that  $\tilde{M}_n \rightarrow N(0, \tilde{g}_2 K_2^*)$  in distribution. This can be proved by using the argument of Theorem 6 in Liu & Wu (2010), where the case of stationary processes is dealt with. Here we shall only detail the step, cf. (A3) below, that requires special attention of nonstationarity since all other steps similarly follow. A complete proof is available as supplementary material. We shall show that

$$\sum_{j=2m+1}^n E(J_{1,j,n}^2) E(\tilde{D}_{j,n}^2) \rightarrow \tilde{g}_2 K_2^*, \quad (\text{A3})$$

335 where  $J_{1,j,n} = \sum_{i=1}^{j-2m} \alpha_{j-i} \tilde{D}_{i,n}$ . Let  $\tilde{D}_{k,n}^* = \sum_{l=0}^{\infty} \mathcal{P}_k \tilde{\zeta}_{k+l}(k/n)$ . By Condition (2),

$$\|\|\tilde{D}_{k,n}\|^2 - \|\tilde{D}_{k,n}^*\|^2\| \leq (\|\tilde{D}_{k,n}\| + \|\tilde{D}_{k,n}^*\|) \|\tilde{D}_{k,n} - \tilde{D}_{k,n}^*\| \leq C m^3 n^{-1} = o(1).$$

Observe that  $\|\tilde{D}_{k,n}^*\|^2 = \tilde{g}(k/n)$ . Then (A3) follows from

$$\begin{aligned} \sum_{j=2m+1}^n \sum_{i=1}^{j-2m} \alpha_{j-i}^2 \|\tilde{D}_{i,n}^*\|^2 \|\tilde{D}_{j,n}^*\|^2 &= \sum_{1 \leq i < j \leq n} \alpha_{j-i}^2 \tilde{g}(i/n) \tilde{g}(j/n) + o(1) \\ &= \sum_{i=1}^{n-1} \tilde{g}(i/n) \sum_{j=1+i}^n \alpha_{j-i}^2 \tilde{g}(j/n) + o(1) \\ &= \sum_{i=1}^{n-1} \tilde{g}^2(i/n) \sum_{j=1+i}^n \alpha_{j-i}^2 + o(1) = \tilde{g}_2 K_2^* + o(1), \end{aligned}$$

340 since  $\sum_{j=1+i}^n \alpha_{j-i}^2 = n^{-1} K_2^* \{1 + o(1)\}$  and  $\sum_{j=1+i}^n \alpha_{j-i}^2 \{\tilde{g}(j/n) - \tilde{g}(i/n)\} = o(n^{-1})$ .  $\square$

*Proof of Theorem 1.* Let  $\check{\mu}_n(t) = \sum_{i=1}^n v_i(t)Y_i$  be the Priestley–Chao estimator, where  $v_i(t) = (nb_n)^{-1}K\{(i/n - t)/b_n\}$ . Note that  $w_i(t) = v_i(t) + \mathcal{O}\{(nb_n)^{-2}\}$  uniformly over  $(b_n, 1 - b_n)$ . Both  $\int_0^{b_n} E\{\hat{\mu}_n(t)^2\}dt$  and  $\int_{1-b_n}^1 E\{\hat{\mu}_n(t)^2\}dt$  are of order  $O(n^{-1})$ . Then

$$\left\| \int_0^1 \hat{\mu}_n(t)^2 dt - \int_0^1 \check{\mu}_n(t)^2 dt \right\|_1 \leq C\{n^{-1} + (nb_n)^{-2}\}.$$

Hence it suffices to prove the same results for the Priestley–Chao estimator  $\check{\mu}_n(t)$ . For (i), write  $l = l_n$ . If  $|i - j| \leq l$ , by Condition (2) and the Cauchy–Schwarz inequality,  $|E\{e_i\{e_j - \zeta_j(i/n)\}\}| \leq \|e_i\| \|\zeta_j(j/n) - \zeta_j(i/n)\| = O(l/n)$ . Let  $\gamma_k(t) = E\{\zeta_0(t)\zeta_k(t)\}$ . Then  $E\{e_i e_j\} - \gamma_{j-i}(i/n) = O(l/n)$ . Since  $\mathcal{P}_s, s \in \mathbb{Z}$ , are orthogonal, 345

$$\begin{aligned} |E\{e_i e_j\}| &= \left| E \left( \sum_{s \in \mathbb{Z}} \mathcal{P}_s e_i \sum_{s' \in \mathbb{Z}} \mathcal{P}_{s'} e_j \right) \right| \leq \sum_{s \in \mathbb{Z}} |E\{(\mathcal{P}_s e_i)(\mathcal{P}_s e_j)\}| \\ &\leq \sum_{s \in \mathbb{Z}} \|\mathcal{P}_s e_i\| \|\mathcal{P}_s e_j\| \leq \sum_{s \in \mathbb{Z}} \delta_{i-s} \delta_{j-s} = \Gamma_{i-j}. \end{aligned} \quad (\text{A4})$$

Similarly,  $|\gamma_{j-i}(i/n)| \leq \Gamma_{j-i}$ . Then 350

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=-l}^l \left| K^* \left( \frac{k}{2nb_n} \right) \{E\{e_i e_{i+k}\} - \gamma_k(i/n)\} \right| = \sum_{k=0}^l O\{\min(\Gamma_k, b_n)\}. \quad (\text{A5})$$

Let  $\bar{\gamma}_k = \int_0^1 \gamma_k(t) dt$ . Since  $\gamma_k(\cdot)$  is Lipschitz continuous and  $K^* \in \mathcal{C}[-1, 1]$ ,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \sum_{k=-l}^l K^* \left( \frac{k}{2nb_n} \right) \gamma_k(i/n) - K^*(0) \sum_{k=-l}^l \bar{\gamma}_k \\ &= \sum_{k=-l}^l K^* \left( \frac{k}{2nb_n} \right) \{\bar{\gamma}_k + O(n^{-1})\} - K^*(0) \sum_{k=-l}^l \bar{\gamma}_k = O(l/n) + \sum_{k=0}^l O(k/l)\Gamma_k. \end{aligned} \quad (\text{A6})$$

By Lemma A1(i),  $\sup_i \|\hat{\mu}_n(t)\|^2 = O\{(nb_n)^{-1}\}$ . Note that  $\int_{-b}^{1+b} K(x - t/b)K(y - t/b)dt = bK^*(y/2 - x/2)$  if  $0 \leq x \leq y \leq b^{-1}$ . Hence 355

$$\begin{aligned} nb_n \int_0^1 E\{\hat{\mu}_n(t)^2\}dt &= nb_n \int_{-b_n}^{1+b_n} E\{\hat{\mu}_n(t)^2\}dt + O(b_n) \\ &= \frac{1}{n} \sum_{1 \leq i, j \leq n} K^* \left( \frac{i-j}{2nb_n} \right) E\{e_i e_j\} + O(b_n). \end{aligned} \quad (\text{A7})$$

Since  $g_1 = \sum_{k \in \mathbb{Z}} \bar{\gamma}_k$ , by (A5), (A6) and (A7), (8) follows in view of

$$\frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=-l}^l - \sum_{\substack{\min(n-i, l) \\ \max(1-i, l)}} \right\} K^* \left( \frac{k}{2nb_n} \right) E\{e_i e_{i+k}\} = \frac{1}{n} O(l) = O(b_n).$$

For (ii), by Theorem A1, we have

$$nb_n^{1/2} \int_{-b_n}^{1+b_n} [\hat{\mu}_n(t)^2 - E\{\hat{\mu}_n(t)^2\}] dt \rightarrow N(0, 4g_2 K_2^*)$$

in distribution. So, (ii) follows in view of  $nb_n^{1/2} \int_{-b_n}^0 E\{\hat{\mu}_n(t)^2\}dt = nb_n^{1/2} O\{b_n(nb_n)^{-1}\} \rightarrow 0$  and similarly, for the right tail,  $nb_n^{1/2} \int_1^{1+b_n} E\{\hat{\mu}_n(t)^2\}dt \rightarrow 0$ . 360  $\square$

*Proof of Theorem 2.* Observe that  $I_n = II_n + III_n$ , where

$$II_n = \int_0^1 \{[\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}]^2 - E([\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}]^2)\} dt,$$

$$III_n = \int_0^1 2[\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}][E\{\hat{\mu}_n(t)\} - \mu(t)] dt.$$

365 By Theorem 1(ii),  $nb_n^{1/2}III_n \rightarrow N(0, 4g_2K_2^*)$  in distribution. Then it suffices to show that

$$\frac{n^{1/2}}{b_n^2}III_n = \sum_{i=1}^n q_{n,i}e_i \rightarrow \kappa_2 N(0, \sigma^2) \quad (\text{A8})$$

in distribution, where  $\sigma^2 = \int_0^1 g(t)\mu''(t)^2 dt$  and

$$q_{n,i} = \frac{2n^{1/2}}{b_n^2} \int_0^1 w_i(t)[E\{\hat{\mu}_n(t)\} - \mu(t)] dt.$$

Under Condition 1, the bias  $E\{\hat{\mu}_n(t)\} - \mu(t) = b_n^2\kappa_2\mu''(t)/2 + o(b_n^2)$ . Since  $K \in \mathcal{K}$ , by elementary calculations,  $r_{n,i} := q_{n,i} - n^{-1/2}\kappa_2\mu''(i/n)$  satisfies  $\sum_{i=1}^n |r_{n,i}|^2 = o(1)$ . By Lemma A1(i),  $\|\sum_{i=1}^n r_{n,i}e_i\| = o(1)$ , and (A8) follows if

$$\sum_{i=1}^n n^{-1/2}\mu''(i/n)e_i \rightarrow N(0, \sigma^2) \quad (\text{A9})$$

370 in distribution. To prove (A9) we apply the  $m$ -dependence approximation method. By Lemma A1(ii),

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \left\| \sum_{i=1}^n n^{-1/2}\mu''(i/n)(e_i - \tilde{e}_i) \right\| = 0.$$

Note that  $\tilde{e}_i = E(e_i | \mathcal{F}_{i-m,i})$  are  $m$ -dependent. Let  $\tilde{\sigma}^2 = \int_0^1 \tilde{g}(t)\mu''(t)^2 dt$ . Hence (A9) follows from Hoeffding & Robbins' (1948) central limit theorem for  $\sum_{i=1}^n n^{-1/2}\mu''(i/n)\tilde{e}_i$  for  $m$ -dependent random variables  $\tilde{e}_i$  in view of

$$E \left\{ \sum_{i=1}^n n^{-1/2}\mu''(i/n)\tilde{e}_i \right\}^2 = \frac{1}{n} \sum_{|i-j| \leq m} \mu''(i/n)\mu''(j/n)E(\tilde{e}_i\tilde{e}_j) \rightarrow \tilde{\sigma}^2.$$

To see the above relation, we note that  $E(\tilde{e}_i\tilde{e}_j) - \tilde{\gamma}_{i-j}(i/n) = o(1)$  if  $|i-j| \leq m$ , and  
375  $n^{-1} \sum_{i=1}^n \mu''(i/n)^2 \tilde{\gamma}_{i-j}(i/n) \rightarrow \int_0^1 \mu''(t)^2 \tilde{\gamma}_{i-j}(t) dt$ , by the continuity of  $\tilde{\gamma}_{i-j}(\cdot)$  and  $\mu''(\cdot)$ .  $\square$

*Proof of Proposition 2.* Write

$$\begin{aligned} \text{ISE}_{\hat{M}} - \text{ISE}_M &= 2 \int_0^1 \{\hat{\mu}_n(t) - \mu_M(t)\} \{\mu_M(t) - \mu_{\hat{M}}(t)\} dt \\ &\quad + \int_0^1 \{\mu_M(t) - \mu_{\hat{M}}(t)\}^2 dt = 2A_n + B_n. \end{aligned}$$

So, (16) follows if both  $A_n$  and  $B_n$  are of order  $n^{-1}$ . To this end, by (15),

$$\begin{aligned} \mu_M(t) - \mu_{\hat{M}}(t) &= \sum_{i=1}^n w_i(t) \{f(\theta_0, t) - f(\hat{\theta}_n, t)\} \\ &= (\theta_0 - \hat{\theta}_n)^\top \sum_{i=1}^n w_i(t) \dot{f}(\theta_0, t) + O_p(n^{-1}) \end{aligned} \quad (\text{A10})$$

holds uniformly over  $t \in [0, 1]$  since  $\hat{\theta}_n$  is  $n^{1/2}$ -consistent. Then by (14), we have  $B_n = O_p(n^{-1})$ . For  $A_n$ , note that  $\hat{\mu}_n(t) - \mu_M(t) = \sum_{j=1}^n w_j(t)e_j$ . By (A10), since  $\|\hat{\mu}_n(t) - \mu_M(t)\|n^{-1} = O(n^{-3/2}b_n^{-1/2})$  and, for some constant  $C > 0$ , by (14),

$$\left| \int_0^1 w_j(t) \sum_{i=1}^n w_i(t) f'(\theta_0, t) dt \right| \leq C \int_0^1 |w_j(t)| dt \leq \frac{C}{n},$$

then we also have  $A_n = O_p(n^{-1})$  in view of Lemma A1(i). □ 385

*Proof of Theorem 3.* Let  $\mathcal{I}_n(t) = \{i : |i/n - t| \leq b_n\}$ . Observe that

$$\sum_{i=1}^n Q_i \mathbf{I}(|i/n - t| \leq b_n) = \sum_{i, j \in \mathcal{I}_n(t)} e_i e_j \mathbf{I}(|i - j| \leq m_n) + R_n,$$

where by (A4),

$$R_n = \sum_{i \in \mathcal{I}_n(t), j \notin \mathcal{I}_n(t)} e_i e_j \mathbf{I}(|i - j| \leq m_n) = O_p(m_n).$$

With elementary manipulations, (18) follows by applying the argument of Theorem 1 to  $(e_i)_{i \in \mathcal{I}_n(t)}$  with  $\alpha_{i-j} = \mathbf{I}(|i - j| \leq m_n)$ . For (19), write

$$E(Q_i) - g(i/n) = \sum_{j=i-m_n}^{i+m_n} E[\zeta_i(i/n)\{\zeta_j(j/n) - \zeta_j(i/n)\}] + \sum_{j:|j-i|>m_n} E\{\zeta_i(i/n)\zeta_j(i/n)\}.$$

By (A6) and (Condition 2), the first and second terms above are of order  $\sum_{k \leq m_n} O\{\min(\Gamma_k, m_n/n)\}$  and  $\sum_{k > m_n} O(\Gamma_k)$ , respectively. Then (19) follows from Condition 3. □ 390

## REFERENCES

- AERTS, M., CLAESKENS, G. & HART, J. (1999). Testing the fit of a parametric function. *J. Am. Statist. Assoc.* **94**, 869–79.
- ALCALÁ, J. T., CRISTÓBAL, J. A. & GONZÁLEZ-MANTEIGA, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Prob. Lett.* **42**, 39–46. 395
- AZZALINI, A., BOWMAN, A. W. & HÄRDLE, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76**, 1–11.
- BENGTSSON, L., HODGES, K. I., ESCH, M., KEENLYSIDE, N., KORNBUEH, L., LUO, J. J. & YAMAGATA, T. (2007). How may tropical cyclones change in a warmer climate? *Tellus A* **59**, 539–61. 400
- BENNER, T. C. (1999). Central England temperatures: Long-term variability and teleconnections. *Int. J. Climatol.* **19**, 391–403.
- BICKEL, P. J. & ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimate. *Ann. Statist.* **1**, 1071–95.
- BISSANTZ, N., HOLZMANN, H., & MUNK, A. (2005). Testing parametric assumptions on band- or time-limited signals under noise. *IEEE Trans. Info. Theory* **51**, 3796–805. 405
- BIEDERMANN, S. & DETTE, H. (2000). Testing linearity of regression models with dependent errors by kernel based methods. *Test* **9**, 417–38.
- CHENG, B. & TONG, H. (1998). k-stationarity and wavelets. *J. Statist. Plan. Infer.* **68**, 129–144. Q6
- DAHLHAUS, R. (1997). Fitting time series models to non-stationary processes. *Ann. Statist.* **25**, 1–37. 410
- DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012–40.
- DETTE, H., MUNK, A. & WAGNER, T. (1998). Estimating the variance in nonparametric regression by quadratic forms—what is a reasonable choice? *J. R. Statist. Soc. B.* **60**, 751–64.
- DRAGHICESCU, D., GUILLAS, S. & WU, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *J. Comp. Graph. Statist.* **18**, 1–20. 415
- ELSNER, J. B., KOSSIN, J. P. & JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–5.
- EMANUEL, K. A. (1991). The theory of hurricanes. *Annu. Rev. Fluid Mech.* **23**, 179–196. 420
- EUBANK, R. L. (1999). *Nonparametric Regression and Spline Smoothing*, 2nd ed. New York: Marcel Dekker.

- EUBANK, R. L. & LARICCIA, V. M. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20**, 2071–86.
- FAN, J. Q. & GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. London, U.K.: Chapman & Hall.
- FAN, J. Q. & JIANG, J. C. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test* **16**, 409–44.
- 425 FAN, J. Q., ZHANG, C. M. & ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153–93.
- GAO, J. & GIJBELS, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Am. Statist. Assoc.* **484**, 1584–94.
- GIURCANU, M. & SPOKOINY V. G. (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statist. Decis.* **22**, 283–300.
- 430 GONZÁLEZ-MANTEIGA, W. & VILAR FERNÁNDEZ, J. M. (1995). Testing linear regression models using non-parametric regression estimators when the errors are correlated. *Comput. Statist. Data Anal.* **20**, 521–41.
- HALL, P. (1984). Integrated square error properties of kernel estimators of regression functions. *Ann. Statist.* **12**, 241–60.
- HALL, P. & HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Am. Statist. Assoc.* **85**, 1039–49.
- 435 HALL, P., KAY, J. W. & TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521–28.
- HÄRDLE, W., HALL, P. & MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Am. Statist. Assoc.* **83**, 86–95.
- 440 HÄRDLE, W. & MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13**, 1465–81.
- HÄRDLE, W. & MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–47.
- HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. New York: Springer.
- 445 HARVEY, D. I. & MILLS, T. C. (2003). Modelling trends in central England temperatures. *J. Forecasting* **22**, 35–47.
- HERRMANN, E., GASSER, T. & KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* **79**, 783–95.
- HOEFFDING, W. & ROBBINS, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780.
- 450 HOLLAND, G. J. (1997). The maximum potential intensity of tropical cyclones. *J. Atmosph. Sci.* **54**, 2519–41.
- HOROWITZ J. L. & SPOKOINY, V. G. (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631
- IOANNIDES, D. A. (1992). Integrated square error of nonparametric estimators of regression function: The fixed design case. *Statist. Prob. Lett.* **15**, 85–94.
- 455 JONES, P. D. & BRADLEY, R. S. (1992a). Climatic variations in the longest instrumental records. In *Climate Since A.D. 1500*, Ed. R. S. Bradley and P. D. Jones, London: Routledge. pp. 246–68.
- JONES, P. D. & BRADLEY, R. S. (1992b). Climatic variations over the last 500 years. In *Climate Since A.D. 1500*, Ed. R. S. Bradley and P. D. Jones, London: Routledge. pp. 649–65.
- JONES, P. D. & HULME, M. (1997). The changing temperature of Central England. In *Climates of the British Isles, Present, Past and Future*, Ed. M. Hulme and E. Barrow, London: Routledge. pp. 173–95.
- 460 KULASEKERA, K. B. & WANG, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Am. Statist. Assoc.* **92**, 500–11.
- LIU, W. D. & WU, W. B. (2010). Asymptotics of spectral density estimates. *Economet. Theory* **26**, 1218–45.
- MALLAT, S., PAPANICOLAOU, G. & ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26**, 1–47.
- 465 MANLEY, G. (1974). Central England temperatures: monthly means 1659 to 1973. *Quart. J. R. Meteorol. Soc.* **100**, 389–405.
- NASON, G. P., VON SACHS, R. & KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. B.* **62**, 271–92.
- 470 OMBAO, H., VON SACHS, R. & GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Assoc.* **100**, 519–31.
- PARK, B. U. & MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Am. Statist. Assoc.* **85**, 66–72.
- PAWLAK, M. & STADTMÜLLER, U. (1996). Recovering bank-limited signals under noise. *IEEE Trans. Inf. Theory* **42**, 1425–38.
- 475 PAWLAK, M. & STADTMÜLLER, U. (2007). Signal sampling and recovery under dependent errors. *IEEE Trans. Info. Theory* **53**, 2526–41.
- PERCIVAL, D. B. & ROTHROCK, D. A. (2005). “Eyeballing” trends in climate time series: A cautionary note. *J. Climate* **18**, 886–91.
- PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B.* **27**, 204–237.
- 480 PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. London: Academic Press.
- PRIESTLEY, M. B. & CHAO, M. T. (1972). Non-parametric function fitting. *J. R. Statist. Soc. B.* **34**, 385–92.
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215–1230.

- RUPPERT, D., SHEATHER, S. J. & WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Am. Statist. Assoc.* **90**, 1257–70.
- TONG, H. (1990). *Non-linear Time Series: A Dynamic System Approach*. Oxford: Oxford University Press. 485
- VAN KEILEGOM, I., MANTEIGA, W. G. & SELLERO, C. S. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17**, 401–15.
- WAND, M. P. & JONES, M. C. (1995). *Kernel Smoothing*. London: Chapman and Hall.
- WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Nat. Acad. Sci.* **102**, 14150–54.
- WU, W. B. & ZHAO, Z. B. (2007). Inference of trends in time series. *J. R. Statist. Soc. B.* **69**, 391–410. 490
- WU, W. B. & ZHOU, Z. (2010). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, To Appear. Q5
- XIA, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Statist. Soc. B.* **60**, 797–811.
- ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38**, 2187–217.
- ZHOU, Z. & WU, W. B. (2009). Local linear quantile estimation for non-stationary time series. *Ann. Statist.* **37**, 2696–29. 495

[Received April 2010. Revised February 2011]