Austral. & New Zealand J. Statist. **41**(4), 1999, 463–479

LOCAL LINEAR KERNEL REGRESSION WITH LONG-RANGE DEPENDENT ERRORS

VO ANH[1], RODNEY WOLFF[1], JITI GAO[*1] AND QUANG TIENG[1]

Queensland University of Technology

Summary

This paper considers the use of a local linear kernel regression method to test whether the mean function of a sequence of long-range dependent processes has discontinuities or change-points. It proposes a non-parametric estimation procedure and then establishes an asymptotic theory for the estimation procedure. Examples, simulated and real, illustrate the estimation procedure.

Key words: change-point; discontinuity; local linear kernel regression; long-range dependence; spectral density.

1. Introduction

Recent developments in non-parametric regression focus on the estimation of continuous non-parametric regression functions with independent and identically distributed (iid) errors (see e.g. Fan & Gijbels, 1996), the estimation of continuous time series regression functions (see e.g. Györfi *et al.*, 1989), the estimation of continuous regression functions with long-range dependent (LRD) errors (see e.g. Robinson, 1997), and the estimation of discontinuous regression functions with iid errors (see e.g. Müller, 1992). As indicated by recent studies (see Müller, 1992; Robinson, 1997), there is evidence of both discontinuity and non-stationarity in the mean function of some data. In a set of data for the Nile River, for example, Müller suggests that the mean function has a discontinuity (change-point) while the research of Robinson (1997) indicates evidence of non-stationarity in the mean function. Thus, the estimation of discontinuous regression functions with LRD errors is an important issue both in theory and in practice.

We apply the local linear (LL) kernel estimation method of Fan & Gijbels (1996) to test whether the mean function of a sequence of LRD processes has change-points, and we construct non-parametric estimates both for the locations of change-points and for the corresponding jump sizes. We establish asymptotic distributions of the constructed estimates. We compare our estimation procedure with the Gasser–Müller method and demonstrate how to implement our estimation procedure through simulated and real examples. This paper extends some results of Hall & Hart (1990), Müller (1991, 1992) and Gao, Pettitt & Wolff (1998).

Section 2 states the main results of the paper. Computational aspects are given in Section 3, and the paper concludes with a discussion in Section 4. Mathematical assumptions and proofs are given in the Appendix.

Received March 1998; revised November 1998; accepted April 1999.
* Author to whom correspondence should be addressed.
[1] School of Mathematical Sciences, Queensland University of Technology, Gardens Point, GPO Box 2434, Brisbane, QLD 4001, Australia. e-mail: j.gao@fsc.qut.edu.au
Acknowledgments. This paper was partly supported by the Australian Research Council Large Grant A89601825.

2. A Nonparametric Estimation Procedure

2.1. Estimation Based on a Regression Model

In this paper, we mainly consider the non-parametric regression of the form

$$Y_i = m(x_i) + e_i \qquad (1 \le i \le n), \tag{2.1}$$

where $x_i = i/n$, m has a change-point at τ, $0 < \tau < 1$, and e_i is a strictly stationary error process with $\mathrm{E}e_i = 0$, $\mathrm{E}e_i^2 = \sigma^2 < \infty$ and $\mathrm{E}(e_1 e_{1+j}) = r(j)$, in which r is the covariance function satisfying

$$r(j) \sim \frac{c_\alpha}{|j|^\alpha},$$

where $0 < \alpha < 1$ and $0 < c_\alpha < \infty$ are constants, and $\sim$ indicates that the ratio of the left-hand side and the right-hand side tends to 1 as $j \to \infty$.

To establish the main results of this paper, we introduce an assumption.

Assumption 2.1. *Assume that the long-range dependent errors e_t have the form*

$$e_t = \sum_{s=1}^{t-1} b_{t-s}\epsilon_s, \qquad b_s \sim \frac{d_\alpha}{|s|^{(1+\alpha)/2}}, \tag{2.2}$$

where $0 < d_\alpha < \infty$ is a constant, ϵ_t^2 is uniformly integrable, and for $t \ge 1$

$$\mathrm{E}(\epsilon_t \mid \mathcal{F}_{t-1}) = 0, \quad \mathrm{E}(\epsilon_t^2 \mid \mathcal{F}_{t-1}) = 1 \qquad a.s.,$$

in which $\mathcal{F}_t$ is the σ-field of events generated by $\{\epsilon_s : 1 \le s \le t\}$.

Remark 2.1. Assumption 2.1 is similar to Robinson (1997 Assumption 2). Equation (2.2) is analogous to Koul & Surgailis (1997 equation (2.14)), who establish asymptotic results for some partial sums. Lemma A.5 establishes the asymptotic normality of a weighted sum based on equation (2.2).

We consider only the case where m has a single jump (discontinuity) at τ. Let $\beta = \beta(\tau) = m_2(\tau) - m_1(\tau)$ be the jump size at τ, where $m_2(\tau) = \lim_{x \downarrow \tau} m(x)$ and $m_1(\tau) = \lim_{x \uparrow \tau} m(x)$. Without loss of generality, we assume that $\beta > 0$. The case $\beta = 0$ corresponds to the non-existence of a change-point at τ.

To construct estimates for τ and β we first construct estimates for $\{m_\ell(\cdot), \ell = 1, 2\}$.

The LL estimator $\hat{m}_\ell(x)$ of $m_\ell(x)$ is defined by

$$\hat{m}_\ell(x) = \frac{1}{nh} \sum_{j=1}^{n} K_{n\ell}(x, x_j) Y_j \qquad (x \in [0, 1]), \tag{2.3}$$

where

$$K_{n\ell}(x, x_j) = \frac{s_{2\ell}(x) - s_{1\ell}(x)(x_j - x)}{s_{2\ell}(x)s_{0\ell}(x) - s_{1\ell}(x)^2} K_\ell\Big(\frac{x_j - x}{h}\Big),$$

$$s_{r\ell}(x) = \frac{1}{nh} \sum_{j=1}^{n} K_\ell\Big(\frac{x_j - x}{h}\Big)(x_j - x)^r \qquad (r = 0, 1, 2, \quad \ell = 1, 2),$$

and $\{K_\ell, \ell = 1, 2\}$ are kernel functions and h is a bandwidth parameter.

In the case where $\mathrm{E}(e_j^2) = \sigma^2(x_j)$, $\hat{m}_\ell$ can be replaced by a weighted estimator $\bar{m}_\ell$,

$$\bar{m}_\ell(x) = \frac{1}{nh}\sum_{j=1}^{n} \frac{K_{n\ell}(x, x_j)}{\hat{\sigma}(x_j)} Y_j \,, \tag{2.4}$$

where

$$\hat{\sigma}_\ell(x)^2 = \frac{1}{nh}\sum_{i=1}^{n} K_{n\ell}(x, x_i)\big(Y_i - \hat{m}_\ell(x_i)\big)^2 \text{ for } \ell = 1, 2, \quad \text{and} \quad \hat{\sigma}^2(x) = \tfrac{1}{2}\big(\hat{\sigma}_2^2(x) + \hat{\sigma}_1^2(x)\big). \tag{2.5}$$

As in Müller & Stadtmüller (1987 Lemma 5.1), $\sup_x |\hat{\sigma}^2(x) - \sigma^2(x)| = O_p(n^{-1/3}\log(n))$. Thus, $\hat{\sigma}(x_j)$ in (2.4) and (2.5) can be replaced by $\sigma(x_j)$ in the large sample situation.

In this paper, we consider only the case where m_ℓ is estimated by the LL-based $\hat{m}_\ell$. The local linear smoothing method is one of the most efficient smoothing methods in nonparametric statistics; see Fan & Gijbels (1996) for more details.

2.2. Asymptotic Distributions of Unknown Location Estimators

Define the following estimator

$$\hat{\beta}(x) = \hat{m}_2(x) - \hat{m}_1(x) \qquad \text{for } x \in (0, 1). \tag{2.6}$$

Let $C \subset (0, 1)$ be a closed interval such that $\tau \in C$. Define the estimators

$$\hat{\tau} = \inf\big\{c \in C\colon \hat{\beta}(c) = \sup_{x\in C} \hat{\beta}(x)\big\} \quad \text{and} \quad \hat{\beta}(\hat{\tau}) \tag{2.7}$$

for the location of the discontinuity point τ and for the jump size β. We apply a functional limit theorem for the stochastic process $\hat{\eta}$ defined below to establish the asymptotic distributions for the estimators $\hat{\tau}$ and $\hat{\beta}(\hat{\tau})$. Let

$$\hat{\beta}(\tau + hy) = \hat{m}_2(\tau + hy) - \hat{m}_1(\tau + hy) \qquad \text{for } -\infty < y < \infty,$$

and define for some $0 < U < \infty$, $-U \le u \le U$, the stochastic process

$$\hat{\eta}(u) = (nh)^{(1+\alpha)/2}\Big(\hat{\beta}\Big(\tau + \big(\frac{h}{n}\big)^{1/2}u\Big) - \hat{\beta}(\tau)\Big).$$

We now state the main results of this section; see the Appendix for proofs and for definitions of the quantities involved in Theorems 2.1 and 2.2.

Theorem 2.1. *Under Assumptions 2.1 and A.1–A.3 listed in the Appendix,*

$$\Big(\frac{n}{h}\Big)^{1/2}(\hat{\tau} - \tau) \xrightarrow{d} \mathrm{N}\Big(0, \frac{2c_\alpha V_{1\alpha}}{\beta^2 K_2'(0)^2 C_1(K_2)^2}\Big) \qquad \text{as } n \to \infty.$$

Theorem 2.2. *Under the conditions of Theorem 2.1,*

$$(nh)^{\alpha/2}\big(\hat{\beta}(\hat{\tau}) - \beta(\tau)\big) \xrightarrow{d} \mathrm{N}(0, 2c_\alpha V_{2\alpha}) \qquad \text{as } n \to \infty.$$

Remark 2.2. Theorem 2.2 not only establishes the asymptotic normality of $\hat{\beta}$ but also provides a test statistic for testing H_0: $\beta = 0$. Under H_0,

$$F_1(\hat{\tau}) = \frac{(nh)^\alpha \hat{\beta}(\hat{\tau})^2}{2c_\alpha V_{2\alpha}} \xrightarrow{d} \chi_1^2 \qquad \text{as } n \to \infty,$$

where χ_1^2 denotes the chi-squared distribution with one degree of freedom. Thus, the data-based $F_1(\hat{\tau})$ can be used to demonstrate asymptotically whether the non-parametric regression function has a discontinuity point at τ. In practice, we need to replace α by a consistent estimator if α is unknown.

2.3. Asymptotic Distributions of Known Location Estimators

Section 2.2 only gives the asymptotic distributions for the estimator $\hat{\tau}$ of the unknown location $\tau \in (0, 1)$ and the estimator $\hat{\beta}(\hat{\tau})$ of the jump size β.

As in Müller (1991), we can show that the estimators $\hat{m}_2(x)$ at $x = 0$ and $\hat{m}_1(x)$ at $x = 1$ are not asymptotically unbiased. Thus, we need to construct new estimators for the case where both $x = 0$ and $x = 1$ are viewed as change-points.

Based on the new kernel functions $K_\ell(\cdot, q)$ that satisfy Assumption A.5, we define $\hat{m}_\ell(x, q)$ for $\ell = 1, 2$ as follows:

$$\hat{m}_\ell(x, q) = \frac{1}{nb} \sum_{j=1}^{n} K_{n\ell}(x, x_j; q) Y_j \qquad (x \in [0, 1])$$

where

$$K_{n\ell}(x, x_j; q) = \frac{s_{2\ell}(x, q) - s_{1\ell}(x, q)(x_j - x)}{s_{2\ell}(x, q) s_{0\ell}(x) - s_{1\ell}(x)^2} K_\ell\Big(\frac{x_j - x}{b}, q\Big),$$

$$s_{r\ell}(x, q) = \frac{1}{nb} \sum_{j=1}^{n} K_\ell\Big(\frac{x_j - x}{b}, q\Big)(x_j - x)^r \qquad (r = 0, 1, 2, \quad \ell = 1, 2),$$

where b is a bandwidth parameter satisfying Assumption A.6 below.

For $q_1 \in [0, 1]$, we define $\hat{m}_2(q_1 b, q_1)$ as the estimator of $m(q_1 b)$; for $q_4 \in [0, 1]$, we define $\hat{m}_1(1 - q_4 b, q_4)$ as the estimator of $m(1 - q_4 b)$. We now state the main results of this section; proofs are given in the Appendix.

Theorem 2.3. *If Assumptions 2.1 and A.1–A.3, A.5–A.6 hold, then as $n \to \infty$,*

$$\frac{(nb)^{\alpha/2}}{\sqrt{c_\alpha V_{2\alpha}(K_2, q_1)}} \Big(\hat{m}_2(q_1 b, q_1) - m(q_1 b) - \tfrac{1}{2} m''(q_1 b) b^2\Big) \xrightarrow{d} \mathrm{N}(0, 1),$$

$$\frac{(nb)^{\alpha/2}}{\sqrt{c_\alpha V_{1\alpha}(K_1, q_4)}} \Big(\hat{m}_1(1 - q_4 b, q_4) - m(1 - q_4 b) - \tfrac{1}{2} m''(1 - q_4 b) b^2\Big) \xrightarrow{d} \mathrm{N}(0, 1).$$

Theorem 2.4. *Under the conditions of Theorem 2.3, as $n \to \infty$,*

$$\mathrm{E}[\hat{m}_2(q_1 b, q_1) - m(q_1 b)]^2 \sim \frac{c_\alpha V_{2\alpha}(K_2, q_1)}{(nb)^\alpha} + \frac{m''(q_1 b)^2 D_1(q_1)^2}{4} b^4,$$

$$\mathrm{E}[\hat{m}_1(1 - q_4 b, q_4) - m(1 - q_4 b)]^2 \sim \frac{c_\alpha V_{1\alpha}(K_1, q_4)}{(nb)^\alpha} + \frac{m''(1 - q_4 b)^2 D_2(q_4)^2}{4} b^4.$$

Remark 2.3. Theorems 2.3 and 2.4 provide some asymptotic properties of the proposed estimates. Theorem 2.3 corresponds to Theorem 2.2 for the unknown location case. Theorem 2.4 not only extends some related results of Müller (1991) to the case where the error process is LRD but also provides a theoretical selection for the bandwidth b.

Remark 2.4. In practice, an important problem is how to select the bandwidth parameters h and b. For the iid case, Gao *et al.* (1998) adopt the plug-in method used by Sheather & Jones (1991), Fan & Gijbels (1995), and Ruppert, Sheather & Wand (1995). For the case where the error process is LRD and the regression function is continuous, Hall, Lahiri & Polzehl (1995) consider the selection of a bandwidth parameter involved in a kernel regression and Gao & Anh (1999) suggest using a generalized cross-validation selection criterion to select a truncation parameter involved in a finite series approximation to the continuous regression function. See also Gao (1998). By combining the results of Gao *et al.* (1998) and Gao & Anh (1999), we can select the bandwidth parameters h and b for the LRD case. The details are technical and we report them elsewhere.

Remark 2.5. To construct a consistent estimator for α, we define a new estimate for m,

$$\begin{aligned}\tilde{m}(x,\hat{\tau}) &= \tilde{m}(x,\hat{\tau};b)\\ &= \hat{m}_2(x,q_1)I_{[0,b]}(x) + \hat{m}(x)I_{(b,\hat{\tau}-b]}(x) + \hat{m}_1\big(x,q_2(\hat{\tau})\big)I_{(\hat{\tau}-b,\hat{\tau}]}(x)\\ &\quad + \hat{m}_2\big(x,q_3(\hat{\tau})\big)I_{(\hat{\tau},\hat{\tau}+b]}(x) + \hat{m}(x)I_{(\hat{\tau}+b,1-b]}(x) + \hat{m}_1(x,q_4)I_{(1-b,1]}(x),\end{aligned}$$

where $\hat{m}$ is as defined in (2.3) with h replaced by b and K_j replaced by K satisfying Assumption A.4 below, $q_1 = x/b$, $q_2(\hat{\tau}) = (\hat{\tau}-x)/b$, $q_3(\hat{\tau}) = (x-\hat{\tau})/b$, and $q_4 = (1-x)/b$.

Define the following objective function

$$\Gamma_u(\alpha,\theta) = \frac{1}{\tau}\sum_{j=1}^{m}\left\{\log(\theta\lambda_j^{\alpha-1}) + \frac{\lambda_j^{1-\alpha}}{\theta}I_u(\lambda_j)\right\}, \tag{2.8}$$

where $\lambda_j = 2\pi j/n$, $1 \le j \le m$, $m/n \to 0$ as $n \to \infty$, $I_u(\lambda_j) = |\sum_{s=1}^{n} u_s e^{is\lambda_j}|^2/(2\pi n)$, and $u_s = Y_s - \tilde{m}(x_s,\hat{\tau})$ depends on α.

Define the estimator of (α,θ) by

$$(\hat{\alpha},\hat{\theta}) = \operatorname*{arg\,min}_{\theta\in\mathbb{R}^+,\ \alpha\in\Theta_{12}} \Gamma_u(\alpha,\theta),$$

provided that the estimator exists, where Θ_{12} is an interval. Under an additional condition, we can show that $\hat{\alpha}$ is a consistent estimator of α. Details are similar to those of Gao & Anh (1999 Assumption 2.7 and Theorem 2.5); see also Robinson (1997 Section 4). When calculating $\hat{\alpha}$ for practical purposes, it is preferable to exclude a neighbourhood around $\hat{\tau}$, although a single discontinuity as in (2.1) does not disturb the asymptotic consistency of $\hat{\alpha}$.

3. Examples and Implementations

This section provides a small sample study to support the asymptotic results presented in Theorems 2.1–2.2.

Example 3.1. Consider the model (2.1) where

$$m(x) = x^4 + I_{[0.5,1]}(x), \tag{3.1}$$

$$e_t = \sum_{s=1}^{t-1} b_{t-s}\epsilon_s, \qquad b_j = j^{-1/3}, \tag{3.2}$$

and ϵ_s is a sequence of iid $\mathrm{N}(0,1)$ random variables.

It follows from (3.1) that the location of the jump point is $\tau = 0.5$ and the corresponding size of the jump value is $\beta = 1$.

For calculating (2.6) and (2.7), we choose

$$K_2(x) = \begin{cases} 6(1-x)(1-2x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

$K_1(x) = K_2(-x)$ for $x \in [-1, 0]$, and $h = (n \log n)^{-1/3}$ as suggested by Müller (1992), where n is the number of observations. The conditions of Theorems 2.1 and 2.2 hold.

To demonstrate the superiority of the local linear kernel estimation method, we compare the estimator $\hat{m}_j$ in (2.3) with the Gasser–Müller (GM) estimator defined by

$$\tilde{m}_j(x) = \sum_{i=1}^{n} \tilde{K}_{nj}(x, x_i) Y_i \qquad (j = 1, 2),$$

$$\tilde{K}_{nj}(x, x_i) = \frac{1}{h} \int_{s_{i-1}}^{s_i} K_j\Big(\frac{x-z}{h}\Big)\, dz,$$

where $s_0 = 0$, $s_i = \frac{1}{2}(x_i + x_{i+1})$ for $1 \le i \le n-1$ and $s_n = 1$.

Define $\tilde{\beta}(x) = \tilde{m}_2(x) - \tilde{m}_1(x)$ for $x \in (0, 1)$. Choose a closed interval $C \subset (0, 1)$, and define the estimators

$$\tilde{\tau} = \inf\big\{c \in C, \tilde{\beta}(c) = \sup_{x \in C} \tilde{\beta}(x)\big\} \quad \text{and} \quad \tilde{\beta}(\tilde{\tau}). \tag{3.3}$$

which correspond to $\hat{\tau}$ and $\hat{\beta}(\hat{\tau})$ defined in (2.7).

Now we investigate the practical implications of Theorems 2.1 and 2.2; we choose the sample mean squared error (MSE) to see whether Theorems 2.1 and 2.2 work well numerically. By using the SPLUS functions including the *ms* function (see Chambers & Hastie, 1992), the sample MSEs of $\hat{\tau}$, $\tilde{\tau}$, $\hat{\beta}(\hat{\tau})$ and $\tilde{\beta}(\tilde{\tau})$ denoted by $\mathrm{MSE}(\hat{\tau})$, $\mathrm{MSE}(\tilde{\tau})$, $\mathrm{MSE}(\hat{\beta}(\hat{\tau}))$ and $\mathrm{MSE}(\tilde{\beta}(\tilde{\tau}))$, respectively, were calculated 1000 times and the means are given in Table 3.1. Four plots for $n = 100$ are given in Fig. 1: Fig. 1(a) gives a time plot of the data Y_i generated from (2.1) and (3.1)–(3.2); the LL estimates of m_2 (dashed) and m_1 (solid) are given in Fig. 1(b); the LL estimate of $m_2 - m_1$ is given in Fig. 1(c); Fig. 1(d) presents estimates of $m_2 - m_1$ based on the LL method (dotted) and the GM method (solid), respectively. Both Fig. 1(c) and Fig. 1(d) show that the estimate $\hat{\tau}$ for $n = 100$ is close to $\tau = 0.5$.

As the value of h (second column) decreases, the performance of both $\hat{\tau}$ and $\tilde{\tau}$ is improved (third and fourth columns). The results in the third and fourth columns also show that for each value of h, the value of $\mathrm{MSE}(\hat{\tau})$ is smaller than that of $\mathrm{MSE}(\tilde{\tau})$. This supports the

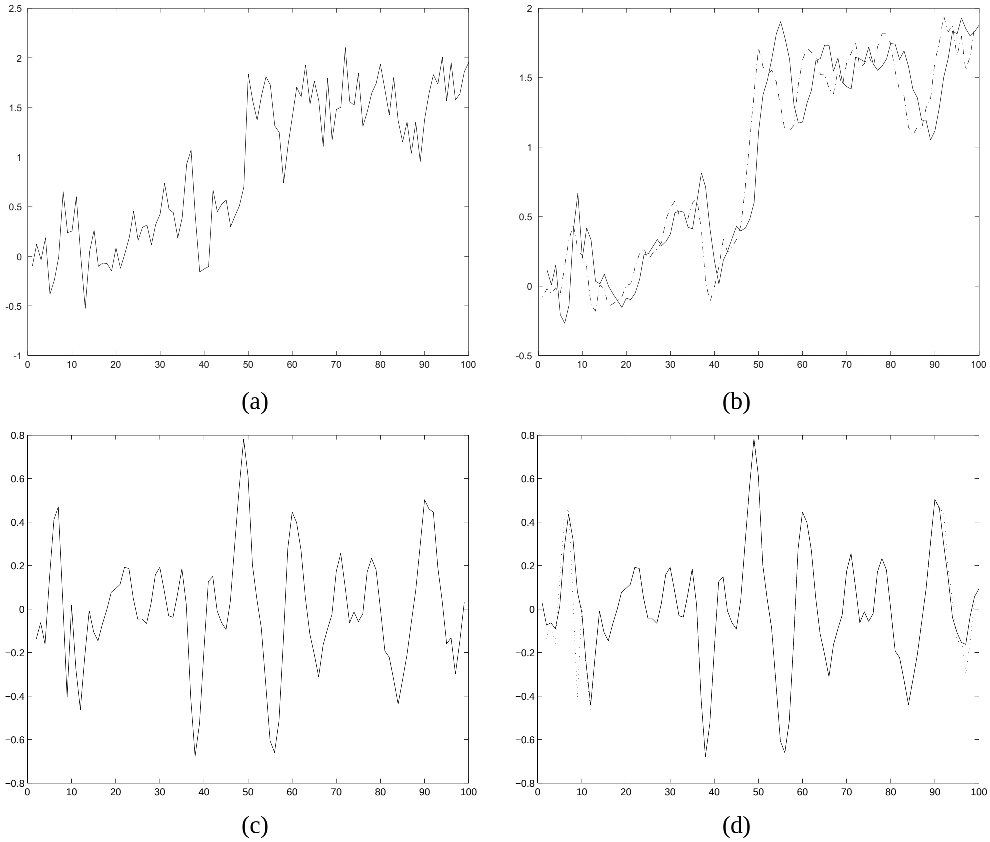


Fig. 1. (a) A time plot of the data Y_i; (b) the fitted values of m_2 (dashed) and m_1 (solid); (c) the fitted values of $m_2 - m_1$; (d) the LL-based fitted values of $m_2 - m_1$ (dotted) and the GM-based fitted values of $m_2 - m_1$ (solid), showing that the LL method and the GM method are identical on the interval $[10, 90]$ approximately; the x-axes show the time points $i = 1, \ldots, n$, where $n = 100$

TABLE 3.1

Numbers of observations (n), values of h, and sample mean squared errors of $\hat{\tau}$, $\tilde{\tau}$, $\hat{\beta}(\hat{\tau})$ and $\tilde{\beta}(\tilde{\tau})$

n	h	MSE($\hat{\tau}$)	MSE($\tilde{\tau}$)	MSE($\hat{\beta}(\hat{\tau})$)	MSE($\tilde{\beta}(\tilde{\tau})$)
100	0.1295	0.0684	0.0688	0.0867	0.0869
500	0.0685	0.0397	0.0403	0.0858	0.0867
1000	0.0525	0.0079	0.0083	0.0797	0.0805
1500	0.0450	0.0074	0.0078	0.0773	0.0778
2000	0.0404	0.00032	0.00032	0.0684	0.0693

asymptotic theory that the variance of the local polynomial estimator is smaller than that of the GM estimator (see Fan & Gijbels, 1996 Chap. 2). This is one of the properties which demonstrate that the local polynomial estimator is superior to the GM estimator. The performance of $\hat{\beta}(\hat{\tau})$ and $\tilde{\beta}(\tilde{\tau})$ can be discussed in a similar way. In this example, all the simulation results depend on the theoretical choice of h. However, as mentioned in Remark 2.4, a critical problem is how to select h in practice.

A complete discussion of the bandwidth parameters h and b is extremely technical, so we do not detail the small sample study of Theorems 2.3 and 2.4 in this paper.

Example 3.2. This example illustrates the estimation methods, using data on the annual volume of the Nile River from 1871 to 1970. A discussion of various approaches to non-parametric change-point modelling is given in Müller (1992). The question is whether and when there occurred an abrupt change in rainfall activity near the turn of the last century. Müller suggests that a change occurred in the year 1898.

To compute (2.6) and (2.7), we chose $h = (100 \log 100)^{-1/3} = 0.1295$,

$$K_2(x) = \begin{cases} 6(1-x)(1-2x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

and calculated $\hat{\tau}$ defined in (2.7) and $\tilde{\tau}$ defined in (3.3). The estimated jump sizes $\hat{\beta}(\hat{\tau})$ and $\tilde{\beta}(\tilde{\tau})$ were also calculated: $\hat{\tau} = 1898$, $\hat{\beta}(\hat{\tau}) = 264.92$, $\tilde{\tau} = 1898$, and $\tilde{\beta}(\tilde{\tau}) = 265.85$ were obtained. Both the LL method and the GM estimator indicate that an abrupt change in rainfall activity occurred in the year 1898, although the estimated jump sizes were slightly different.

As in Robinson (1997 equation (4.12)), we replace u_s in (2.8) by Y_s and determine whether the set of data is LRD, based on the estimator. Our research suggests that the spectral density of Y_s has the form

$$f(x) \sim \frac{\theta}{x^{1-\alpha}} \quad \text{as} \quad x \to 0, \qquad \text{where} \quad \theta = 0.073 \quad \text{and} \quad \alpha = 0.2066;$$

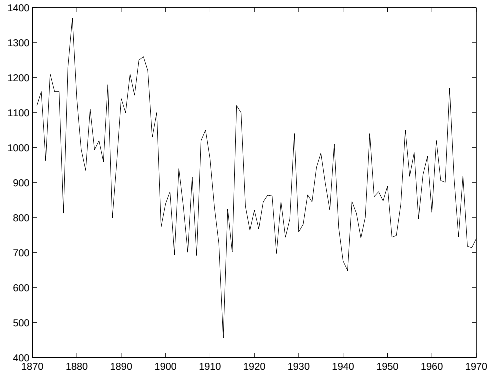
see Fig. 3(a). Our estimation procedure also shows that the set of data is actually long-range dependent. Figure 2(a) gives a time plot of the data Y_i. The LL-based fitted values of m_2 (dashed) and m_1 (solid) are given in Fig. 2(b). Figure 2(c) presents the LL-based fitted values of $m_1 - m_2$ (dotted) and the GM-based fitted values of $m_1 - m_2$ (solid). Figure 2(c) shows that the LL method and the GM method are identical on the interval 1885–1955 approximately; (see also Müller, 1992 Fig. 5).

Example 3.3. In this example, we illustrate our estimation procedure using the Nile River data listed in Beran (1994 Sect. 12.2.1). These data consist of readings of annual minimum levels at the Roda Gorge near Cairo, commencing in the year 622; often only the first 663 observations are used because missing observations occur after the year 1284.

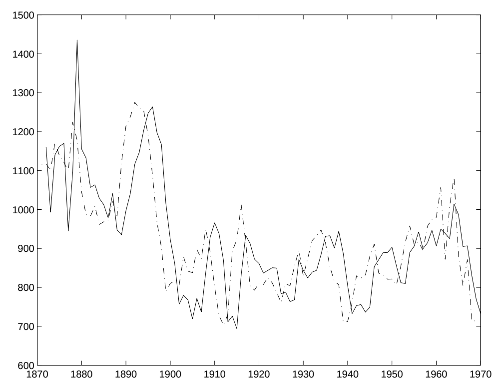
Robinson (1997) suggests that application of his methods to the Nile series between 622 and 1284 provides evidence of non-stationarity in the mean function. In this example, we apply equations (2.6) and (2.7) with $h = (663 \log 663)^{-1/3} = 0.0615$ to check whether there is evidence of discontinuity in the mean function. We find that $\hat{\tau} = 868$, $\hat{\beta}(\hat{\tau}) = 178.93$, $\tilde{\tau} = 868$ and $\tilde{\beta}(\tilde{\tau}) = 178.98$. Both the LL and GM estimators indicate that an abrupt change in rainfall activity occurred in the year 868, which means that the mean function has a discontinuity. We also apply (2.8) to find that the spectral density of Y_s has the form

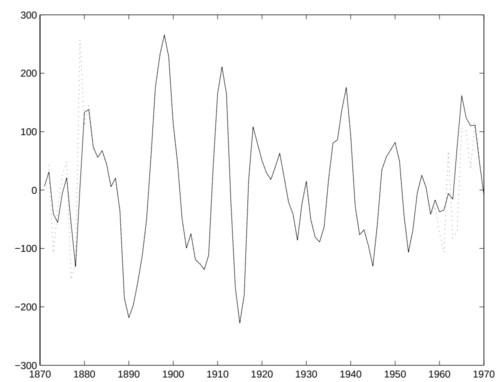
$$f(x) \sim \frac{\theta}{x^{1-\alpha}} \quad \text{as} \quad x \to 0, \qquad \text{where} \quad \theta = 0.077 \quad \text{and} \quad \alpha = 0.1994;$$

see Fig. 3(b). The results are similar to those of Robinson (1997). Figure 4(a) gives a time plot of the data Y_i. The LL-based fitted values of m_2 (dashed) and m_1 (solid) are given in



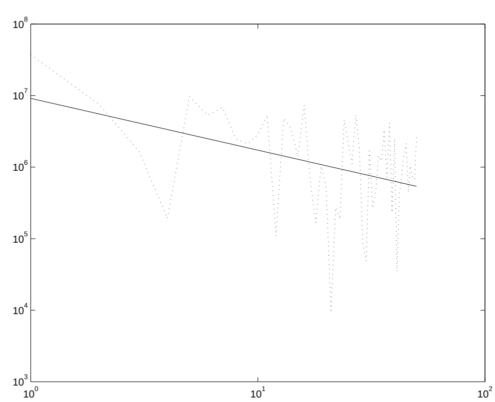
(a)

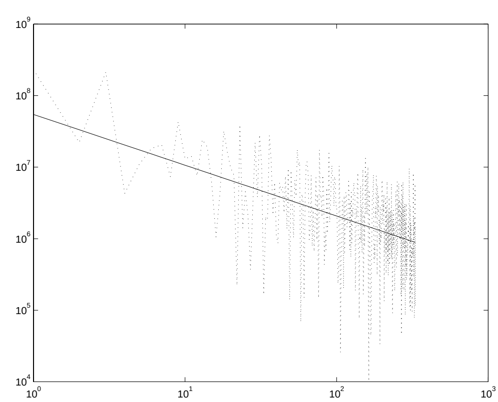




(b) (c)

Fig. 2. (a) A time plot of the Nile River data given in Example 3.2; (b) the fitted values of m_2 (dashed) and m_1 (solid); (c) the LL-based fitted values of $m_1 - m_2$ (dotted) and the GM-based fitted values of $m_1 - m_2$ (solid)





(a) (b)

Fig. 3. (a) The periodogram (dotted) in the log-logform of the data given in Example 3.2; (b) the periodogram (dotted) in the log-logform of the data given in Example 3.3; the solid lines are the corresponding estimates based on the model $f(x) \sim \theta x^{\alpha-1}$

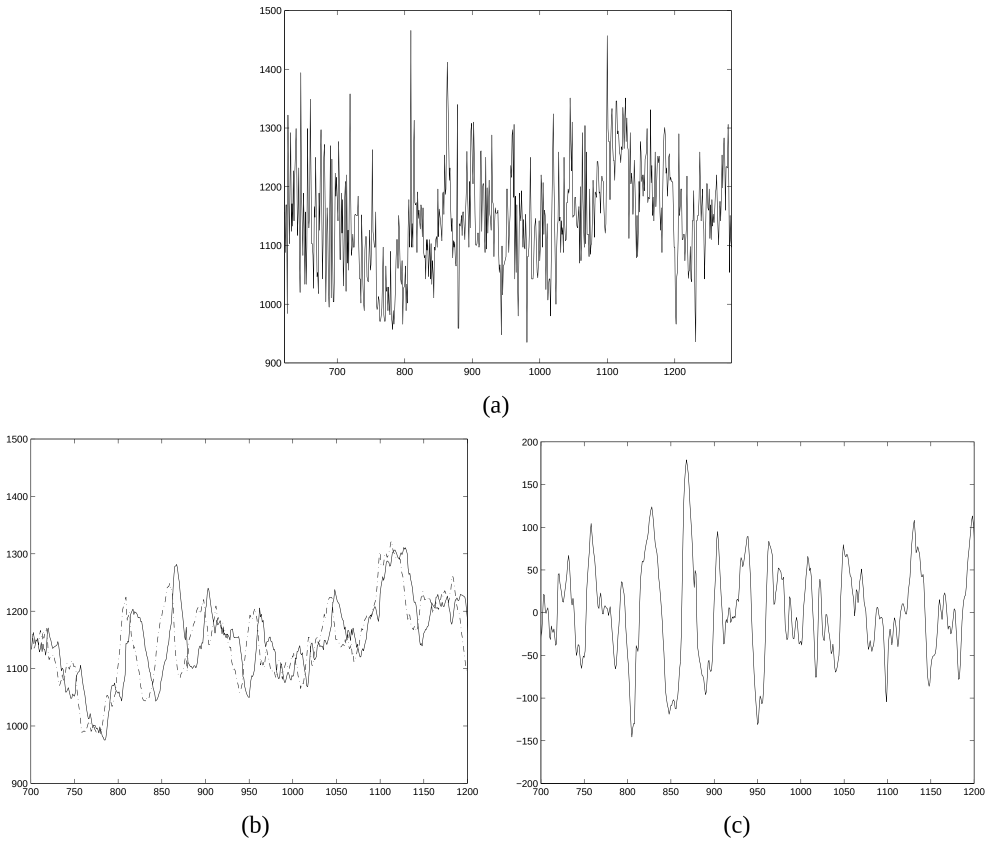


Fig. 4. (a) A time plot of the Nile River data given in Example 3.3; (b) the fitted values of m_2 (dashed) and m_1 (solid); (c) the LL-based fitted values of $m_1 - m_2$

Fig. 4(b). Figure 4(c) presents the LL-based fitted values of $m_1 - m_2$. Both Fig. 2(c) and Fig. 4(c) show that the estimator $\hat{\tau}$ of τ is found as the maximum of the function $\hat{m}_1 - \hat{m}_2$, because in this case the jump is from lower to higher levels.

4. Discussion

This paper considers only the case where the spectral density of the LRD error has the form

$$f(x) \sim \frac{d_\alpha}{x^{1-\alpha}} \quad \text{as } x \to 0, \qquad \text{where} \quad d_\alpha > 0 \text{ is an unknown parameter.}$$

A more general concept of LRD processes was suggested in Gray, Zhang & Woodward (1989), formalizing the initial work of Hosking (1981, 1984). Here, the spectral density is allowed to have the form

$$f(x) = \frac{\phi(x)}{|x-\tau|^{2\beta}} \qquad (0 < \beta < \tfrac{1}{2}),$$

where $\tau \in (-\pi, \pi)$ is regarded as an unknown parameter and ϕ is a slowly varying function in $-\pi < x \leq \pi$. In other words, the singularity of f may occur at a frequency $0 < \tau < \pi$ or $-\pi < \tau < 0$; and the model can be used to describe LRD periodicities of the data. More recently, Robinson (1997) discusses a general class of spectral densities and proposes consistent estimates for the parameters involved in the class. As a result of his research, both β and $\eta = \phi(0)$ can be estimated. However, the estimation of τ has not been considered in the literature to date. The main difficulty in estimating τ is that both $f(x)$ and

$$\log f(x) = \log \phi(x) - 2\beta \log |x - \tau|$$

are just continuous in τ and the first derivative of $\log(f(x))$ with respect to τ does not exist. For this case, a fundamental problem is how to construct a sequence of independent random errors and then derive a consistent estimator for τ based on either the Gaussian–Whittle contrast function (see Robinson, 1997 equation (4.9)) or a non-linear regression approach. Details for this case will be reported elsewhere.

Appendix

Assumption A.1.

(i) *The kernel function* K_1 *has the support interval* $[-1, 0]$ *and satisfies* $K_1(-1) = K_1(0) = 0$ *and* $|\int_{-1}^{0} u^i K_1(u)^j \, du| < \infty$ *for* $i = 0, 1, 2$ *and* $j = 1, 2$.

(ii) *The first derivative,* $K_1^{(1)}$, *of* K_1 *is continuous on* $[-1, 0]$ *and* $K_1^{(1)}(0) < 0$.

(iii) *The kernel function* K_2 *defined by* $K_2(x) = K_1(-x)$ *satisfies the corresponding conditions.*

Assumption A.2.

(i) $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$.

(ii) $\lim_{n\to\infty} nh^3 = 0$.

Assumption A.3.

(i) m *has two continuous derivatives at each* $x \in (0, 1) - \{\tau\}$.

(ii) $m_1^{(i)}(\tau) = \lim_{x\uparrow\tau} m_1^{(i)}(x)$ *and* $m_2^{(i)}(\tau) = \lim_{x\downarrow\tau} m_2^{(i)}(x)$ *exist for* $i = 0, 1, 2$.

Assumption A.4.

(i) K *is a symmetric kernel function with support* $[-1, 1]$, *and* $K(-1) = K(1) = 0$.

(ii) *The first derivative* $K^{(1)}$ *of* K *is continuous on* $[-1, 1]$.

(iii) $\int_{-1}^{1} K(u)^2 \, du < \infty$ *and* $\int_0^1 u^\ell K(u) \, du < \infty$ *for* $\ell = 1, 2$.

Assumption A.5.

(i) *The kernel function* $K_1(x, q)$ *has the support interval* $[-1, q]$ *and satisfies* $K_1(-1, q) = K_1(q, q) = 0$ *and* $|\int_{-1}^{q} u^i K_1(u, q)^j \, du| < \infty$ *for* $0 \le q \le 1$, $i = 0, 1, 2$ *and* $j = 1, 2$.

(ii) *There exist absolute constants* C_1 *and* C_2 *such that* $\sup_x |K_1(x, q_1) - K_1(x, q_2)| \le C_1|q_1 - q_2|$ *and* $\sup_q |K_1(x_1, q) - K_1(x_2, q)| \le C_2|x_1 - x_2|$.

(iii) *The kernel function* $K_2(x, q)$ *defined by* $K_2(x, q) = K_1(-x, q)$ *satisfies the corresponding conditions.*

Assumption A.6.

(i) $b \to 0$ *and* $nb^2 \to \infty$ *as* $n \to \infty$.

(ii) $nhb^2 \to 0$ *as* $n \to \infty$.

(iii) $\limsup_{n\to\infty} (nb)^\alpha b^4 < \infty$.

Remark A.1. The kernel functions $K_1(x, q)$ and $K_2(x, q)$ that satisfy Assumption A.5 can be constructed based on a sequence of orthonormal polynomials. See Müller (1991) for more details.

Notation

For $j = 1, 2,\ r = 0, 1, 2,\ 0 \le q \le 1$,

$$\ell_{jr} = \int_{-1}^{1} u^r K_j(u)\, du, \qquad L_j = \ell_{j2}\ell_{j0} - \ell_{j1}^2,$$

$$C_1(K_2) = \frac{\ell_{22}}{L_2}, \qquad C_2(K_1) = L_2^{-2} \int_{-1}^{0} [\ell_{10} v - \ell_{11}]^2 K_1(v)^2\, dv,$$

$$V_{1\alpha} = L_2^{-2} \iint [\ell_{20} u - \ell_{21}][\ell_{20} v - \ell_{21}] \frac{K_2(u) K_2(v)}{|u-v|^\alpha}\, du\, dv,$$

$$V_{2\alpha} = L_2^{-2} \iint [\ell_{22} - u\ell_{21}][\ell_{22} - v\ell_{21}] \frac{K_2(u) K_2(v)}{|u-v|^\alpha}\, du\, dv,$$

$$A_{1r}(q) = \int_{-1}^{q} K_1(v, q) v^r\, dv, \qquad A_{2r}(q) = \int_{-q}^{1} K_2(v, q) v^r\, dv,$$

$$B_1(q) = A_{12}(q) A_{10}(q) - A_{11}(q)^2, \qquad B_2(q) = A_{22}(q) A_{20}(q) - A_{21}(q)^2,$$

$$D_1(q) = \frac{1}{B_2(q)} \int_{-q}^{1} [A_{22}(q) - v A_{21}(q)] v^2 K_2(v, q)\, dv,$$

$$D_2(q) = \frac{1}{B_1(q)} \int_{-1}^{q} [A_{12}(q) - v A_{11}(q)] v^2 K_1(v, q)\, dv,$$

$$C_3(K_2, q) = \frac{1}{B_2(q)^2} \int_{-q}^{1} [A_{22}(q) - v A_{21}(q)]^2 K_2(v, q)^2\, dv,$$

$$C_4(K_1, q) = \frac{1}{B_1(q)^2} \int_{-1}^{q} [A_{12}(q) - v A_{11}(q)]^2 K_1(v, q)^2\, dv,$$

$$V_{j\alpha}(K_j, q) = \iint \frac{[A_{j1}(q) u - A_{j2}(q)][A_{j1}(q) v - A_{j2}(q)]}{B_j(q)^2} \frac{K_j(u, q) K_j(v, q)}{|u-v|^\alpha}\, du\, dv,$$

where $K_1(v, q)$ and $K_2(v, q)$ are as defined in Assumption A.5.

Lemmas for the Proofs of Theorems 2.1–2.4

Lemma A.1. *Under Assumptions 2.1 and A.1–A.3, the process* $\hat{\eta}$ *converges weakly to* η *on* $C[-U, U]$, *where* η *is a continuous Gaussian process with moment structure*

$$\mathrm{E}(\eta(u)) = -\tfrac{1}{2}\beta u^2 K_2'(0) C_1(K_2), \qquad \operatorname{cov}(\eta(u_1), \eta(u_2)) = 2u_1 u_2 c_\alpha V_{1\alpha}. \tag{A.1}$$

The proof of Lemma A.1 follows from Billingsley (1968 Theorems 8.1 and 12.3), Hall & Heyde (1980 Theorems 4.1–4.4), and the following Lemmas A.2–A.7.

Lemma A.2.

(i) *Under Assumption 2.1, for* $\ell = 1, 2$,

$$\mathrm{E}\big(\hat{m}_\ell(x) - \mathrm{E}\,\hat{m}_\ell(x)\big)^2 \sim c_\alpha (nh)^{-\alpha} V_{2\alpha}. \tag{A.2}$$

(ii) *Under Assumption 2.3, for* $\ell = 1, 2$,

$$\mathrm{E}(\hat{m}_\ell(x, q) - \mathrm{E}\,\hat{m}_\ell(x, q))^2 \sim c_\alpha (nh)^{-\alpha} V_{\ell\alpha}(K_\ell, q). \tag{A.3}$$

Proof. Here we need only prove (A.2) for $\ell = 1$.

Let $a_{ni} = (1/nh)K_{n1}(x, x_i)$. The proof follows from Assumption 2.1, the definition of K_{n1} in (2.3) and

$$\begin{aligned}\mathrm{E}\big(\hat{m}_1(x) - \mathrm{E}[\hat{m}_1(x)]\big)^2 &= \mathrm{E}\Big(\sum_{i=1}^n a_{ni} e_i\Big)^2 = \sum_{i=1}^n \sum_{j=1}^n a_{ni} a_{nj} \mathrm{E}(e_i e_j) \\ &\sim c_\alpha (nh)^{-\alpha} L_2^{-2} \iint [\ell_{12} - \ell_{11} u][\ell_{12} - \ell_{11} v] K_1(u) K_1(v) |u - v|^{-\alpha}\, du\, dv \\ &\sim c_\alpha (nh)^{-\alpha} V_{2\alpha}.\end{aligned}$$

Lemma A.3. *Under the conditions of Lemma A.1, as $n \to \infty$,*

$$\mathrm{E}(\hat{\eta}(u)) = -\tfrac{1}{2}\beta u^2 K_2'(0) C_1(K_2) + o(1).$$

Proof. The proof is the same as that of Gao *et al.* (1998 Lemma A.2).

Lemma A.4. *Under the conditions of Lemma A.1,*

$$\mathrm{cov}(\hat{\eta}(u_1), \hat{\eta}(u_2)) = 2u_1 u_2 c_\alpha V_{1\alpha} + O((nh)^{-\alpha/2}). \tag{A.4}$$

Proof. Define

$$\hat{\eta}(u) = (nh)^{(1+\alpha)/2} \Big\{ \Big[\hat{m}_2\Big(\tau + \Big(\frac{h}{n}\Big)^{1/2} u\Big) - \hat{m}_2(\tau)\Big] - \Big[\hat{m}_1\Big(\tau + \Big(\frac{h}{n}\Big)^{1/2} u\Big) - \hat{m}_1(\tau)\Big] \Big\}.$$

To approximate $\hat{\eta}(u)$, we first define the estimator $\hat{m}_j'(\tau)$ of $m_j'(\tau)$ by

$$\hat{m}_j'(\tau) = \frac{1}{nh^2} \sum_{i=1}^n K_j^*\Big(\frac{x_i - \tau}{h}\Big) Y_i,$$

where $K_j^*(u) = L_j^{-1}(\ell_{j0} u - \ell_{j1}) K_j(u)$ and $j = 1, 2$. Details about the estimators of derivatives can be found in Fan & Gijbels (1996 Section 3.2).

We now obtain

$$\begin{aligned}\hat{\eta}(u) &= (nh)^{(1+\alpha)/2} [\hat{m}_2'(\tau) - \hat{m}_1'(\tau)] \Big(\frac{h}{n}\Big)^{1/2} u + O_p\big(h^2 (nh)^{(\alpha-1)/2}\big) \\ &= \frac{u}{(nh)^{(1-\alpha)/2}} \sum_{i=1}^n \Big[K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big)\Big] Y_i + O_p\big(h^2 (nh)^{(\alpha-1)/2}\big) \\ &\equiv \tilde{\eta}^*(u) + O_p\big(h^2 (nh)^{(\alpha-1)/2}\big).\end{aligned}$$

Thus,

$$\tilde{\eta}^*(u) - \mathrm{E}[\tilde{\eta}^*(u)] = u (nh)^{\alpha/2 - 1} \sum_{i=1}^n \Big[K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big)\Big] e_i.$$

This implies that as $n \to \infty$,

$$\begin{aligned}&\mathrm{cov}\big(\tilde{\eta}^*(u_1), \tilde{\eta}^*(u_2)\big) \\ &= \frac{u_1 u_2}{(nh)^{2-\alpha}} \sum_{i=1}^n \sum_{j=1}^n \Big[K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big)\Big]\Big[K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big)\Big] \mathrm{E}(e_i e_j) \\ &= 2u_1 u_2 c_\alpha V_{1\alpha} + O(n^{-1}).\end{aligned}$$

Before establishing the asymptotic normality of $\hat{\eta}$, we need the following result.

Proposition. *Let $\{p_{ni} : i \geq 1\}$ be a sequence of real numbers. If Assumption 2.1 holds and if $q_{ni} = \sum_{j=i+1}^{n} b_{j-i} p_{nj}$ is such that*

$$\sum_{i=1}^{n} q_{ni}^2 \to \sigma_q^2 \qquad \text{and} \qquad \max_{i \geq 1} |q_{ni}| \to 0 \qquad \text{as } n \to \infty,$$

then

$$\sum_{i=1}^{n-1} p_{ni} e_i \xrightarrow{d} \mathrm{N}(0, \sigma_q^2) \qquad \text{as } n \to \infty.$$

Proof. The proof is similar to (3.18) and (3.19) of Gao & Liang (1995). See also Robinson (1997 Lemma 1).

Lemma A.5. *Under the conditions of Lemma A.1, for fixed $u \in [-U, U]$,*

$$\hat{\eta}(u) - \mathrm{E}\big(\hat{\eta}(u)\big) \xrightarrow{d} \mathrm{N}(0, 2u^2 c_\alpha V_{1\alpha}). \tag{A.5}$$

Proof. Let

$$b_{ni} = \frac{1}{(nh)^{(1-\alpha)/2}} \Big[K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big) \Big],$$

and $\quad \tilde{\eta}^*(u) - \mathrm{E}[\tilde{\eta}^*(u)] = u \sum_{i=1}^{n} b_{ni} e_i = \sum_{i=1}^{n-1} v_{ni} \epsilon_i, \quad$ where $v_{ni} = \sum_{j=i+1}^{n} b_{nj} b_{j-i}.$

To prove (A.5), noting Assumption 2.1 and applying the above Proposition, it suffices to show that as $n \to \infty$,

$$\sum_{i=1}^{n} v_{ni}^2 \to 2u^2 c_\alpha V_{1\alpha}, \tag{A.6}$$

$$\max_{i \geq 1} |v_{ni}| \to 0. \tag{A.7}$$

In the following, we only check whether (A.7) holds. The proof of (A.6) follows similarly from (A.2).

For $\delta_n = nh$, we have

$$\max_{i \geq 1} |v_{ni}| \leq \Big(\sum_{i \geq 1} b_{ni}^2 \sum_{j > \delta_n} b_j^2 \Big)^{1/2} + \max_{1 \leq i \leq n} |b_{ni}| \sum_{j \leq \delta_n} |b_j|.$$

As $n \to \infty$,

$$\begin{aligned} \sum_{i=1}^{n} b_{ni}^2 &= \frac{u^2}{(nh)^{2-\alpha}} \sum_{i=1}^{n} \Big(K_2^*\Big(\frac{x_i - \tau}{h}\Big) - K_1^*\Big(\frac{x_i - \tau}{h}\Big) \Big)^2 \\ &= 2u^2 \frac{1}{(nh)^{1-\alpha}} L_2^{-2} \int_{-1}^{1} [\ell_{20} v - \ell_{21}]^2 K_2(v)^2 \, dv + O(n^{-1}), \\ \sum_{j > \delta_n} b_j^2 &= o(1). \end{aligned} \tag{A.8}$$

Analogously,

$$\max_{i\geq 1}|b_{ni}| = O\big((nh)^{\alpha/2-1}\big) \tag{A.9}$$

$$\sum_{j\leq\delta_n}|b_j| = O\big(\delta_n^{(1-\alpha)/2}\big). \tag{A.10}$$

Equations (A.8)–(A.10) imply

$$\max_{i\geq 1}|v_{ni}| = O\big((nh)^{(\alpha-1)/2}\big) + O\big((nh)^{-1/2}\big) \to 0 \qquad \text{as } n\to\infty.$$

Lemma A.6. *Under the conditions of Lemma A.1, for fixed* $(u_1,\dots,u_\ell)$ *and* $u_i\in[-U,U]$,

$$\Big(\hat\eta(u_1) - \mathrm{E}\big(\hat\eta(u_1)\big),\dots,\hat\eta(u_\ell) - \mathrm{E}\big(\hat\eta(u_\ell)\big)\Big) \xrightarrow{d} \mathrm{N}(0,\Sigma),$$

where $\Sigma = (\sigma_{ij})_{1\leq i,j\leq\ell}$ *and* $\sigma_{ij} = 2u_iu_jc_\alpha V_{1\alpha}$.

Lemma A.7. *Under the conditions of Lemma A.1, the sequence* $\bar\eta(\cdot) = \tilde\eta(\cdot) - \mathrm{E}(\tilde\eta(\cdot))$ *is tight.*

Proof. To prove the tightness, it suffices to show that there exists a constant C such that, for n large enough,

$$\mathrm{E}\big(\bar\eta(u_1) - \bar\eta(u_2)\big)^2 \leq C|u_1-u_2|^2,$$

which follows from (A.4).

Proofs of Theorems 2.1–2.4

Proof of Theorem 2.1. Since the Gaussian limit process η is determined by its first and second moments, according to (A.1), it can be written equivalently as

$$\eta(u) = -\tfrac{1}{2}\beta u^2K_2'(0)C_1(K_2) + Zu, \qquad \text{where } Z \stackrel{d}{=} \mathrm{N}(0, 2c_\alpha V_{1\alpha}).$$

Under Assumption A.1, η is seen to have a unique maximum at

$$U^* = \frac{Z}{\beta K_2'(0)C_1(K_2)}.$$

Let U_n be the location of the maximum of $\hat\eta$. By construction, we have

$$\hat\tau = \tau + U_n\Big(\frac{h}{n}\Big)^{1/2}.$$

By applying results from Whitt (1970) and Eddy (1980), we can show that U_n converges weakly to U^* as $n\to\infty$. This concludes the proof of Theorem 2.1.

Proof of Theorem 2.2. As in the proofs of (A.2) and (A.5), as $n \to \infty$,

$$(nh)^{\alpha} \mathrm{E}\big(\hat{m}_j(\tau) - \mathrm{E}[\hat{m}_j(\tau)]\big)^2 \to c_\alpha V_{2\alpha},$$
$$(nh)^{\alpha/2}\big(\hat{m}_j(\tau) - E[\hat{m}_j(\tau)]\big) \xrightarrow{d} \mathrm{N}(0, c_\alpha V_{2\alpha}),$$
$$(nh)^{\alpha/2}\big(\hat{\beta}(\tau) - \mathrm{E}[\hat{\beta}(\tau)]\big) \xrightarrow{d} \mathrm{N}(0, 2c_\alpha V_{2\alpha}). \tag{A.11}$$

As in the proof of Lemma A.3, by applying Taylor expansions and using Assumptions A.1–A.3, we have, as $n \to \infty$,

$$(nh)^{\alpha/2}(\mathrm{E}[\hat{\beta}(\tau)] - \beta)$$
$$= (nh)^{\alpha/2} h^2 [m_2''(\tau) - m_1''(\tau)]\big(1 + o(1)\big) L_2^{-1} \int_0^1 \int_0^1 \pi_2(u, v) v^2 \, du \, dv \to 0, \tag{A.12}$$
$$(nh)^{\alpha/2}[\hat{\beta}(\hat{\tau}) - \hat{\beta}(\tau)] = (nh)^{\alpha/2}(\hat{\tau} - \tau)\big(1 + o(1)\big)[m_2'(\tau) - m_1'(\tau)] \to_p 0. \tag{A.13}$$

Therefore, Assumption A.2 and equations (A.11)–(A.13) imply

$$(nh)^{\alpha/2}(\hat{\beta}(\hat{\tau}) - \beta) \xrightarrow{d} \mathrm{N}(0, 2c_\alpha V_{2\alpha}).$$

Proof of Theorem 2.3. We provide only an outline for the proof of Theorem 2.3. By Taylor expansions and using Assumptions A.3, A.5 and A.6, we obtain for $0 < x < b$,

$$\mathrm{E}[\tilde{m}_2(x, q_1)] - m(x)$$
$$\sim \frac{1}{B_2(q_1)} \int_{-q_1}^{1} [A_{22}(q_1) - vA_{21}(q_1)] K_2(v, q_1)[m(x + vb) - m(x)] \, dv$$
$$\sim \tfrac{1}{2} m''(x) D_2(q_1) b^2.$$

Let $D_n^2 = \sum_{i=1}^n c_{ni}^2$, $c_{ni} = (nb)^{-\alpha/2} K_{n2}(x, x_i; q_1)$, and $d_{ni} = c_{ni}/D_n$. Then we can write

$$(nb)^{\alpha/2}\big(\hat{m}_2(x, q_1) - \mathrm{E}[\hat{m}_2(x, q_1)]\big) = \sum_{i=1}^{n} d_{ni} e_i \equiv \sum_{i=1}^{n-1} w_{ni} \epsilon_i,$$

where $w_{ni} = \sum_{j=i+1}^n b_{nj} b_{j-i}$.

As in the proof of (A.5), it suffices to show that, as $n \to \infty$,

$$\sum_{i=1}^{n} w_{ni} \to c_\alpha V_{2\alpha}(K_2, q_1), \tag{A.14}$$
$$\max_{i \ge 1} |w_{ni}| \to 0. \tag{A.15}$$

The proofs of (A.14) and (A.15) are similar to those of (A.6) and (A.7), respectively.

Proof of Theorem 2.4. We prove the first equation. The second one follows similarly:

$$\mathrm{E}[\hat{m}_2(x, q_1) - m(x)]^2 = \mathrm{E}\big(\hat{m}_2(x, q_1) - \mathrm{E}[\hat{m}_2(x, q_1)]\big)^2 + \big(\mathrm{E}[\hat{m}_2(x, q_1)] - m(x)\big)^2$$
$$\sim \frac{c_\alpha}{(nb)^\alpha} V_{2\alpha}(K_2, q_1) + \tfrac{1}{4} m''(x)^2 D_1(q_1)^2 b^4.$$

This completes the proof.

References

BERAN, J. (1994). *Statistics for Long-Memory Processes*. London: Chapman and Hall.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

CHAMBERS, J.M. & HASTIE, T.J. (1992). *Statistical Models in S*. New York: Chapman & Hall.

EDDY, W.F. (1980). Optimum kernel estimates of the mode. *Ann. Statist.* **8**, 870–882.

FAN, J. & GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *J. Roy. Statist. Soc. Ser. B* **57**, 371–394.

— & — (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

GAO, J. (1998). Semiparametric regression modelling of nonlinear time series. *Scand. J. Statist.* **25**, 521–539.

— & ANH, V. (1999). Semiparametric regression with long-range dependent errors. *J. Statist. Plann. Inference* **80**, 37–57.

— & LIANG, H. (1995). Asymptotic normality of pseudo-LS estimator for partially linear autoregressive models. *Statist. Probab. Lett.* **23**, 27–34.

—, PETTITT, A. & WOLFF, R. (1998). Local linear kernel estimation and change-points problems. *Comm. Statist. A — Theory Methods* **27**, 2871–2894.

GRAY, H.L., ZHANG, N.F. & WOODWARD, W.A. (1989). On generalized fractional processes. *J. Time Ser. Anal.* **10**, 233–257.

GYÖRFI, L., HÄRDLE, W., SARDA, P. & VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series*. Lecture Notes in Statistics **60**. New York: Springer.

HALL, P. & HART, J. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36**, 339–351.

— & HEYDE, C.C. (1980). *Martingale Limit Theory and Its Applications*. New York: Academic Press.

HALL, P., LAHIRI, S.N. & POLZEHL, J. (1995). On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.* **23**, 1921–1936.

HOSKING, J.R. (1981). Fractional differencing. *Biometrika* **68**, 165–176.

— (1984). Modelling persistence in hydrological time series using fractional differencing. *Water Resources Res.* **20**, 1898–1908.

KOUL, H. & SURGAILIS, D. (1997). Asymptotic expansion of M-estimates with long memory errors. *Ann. Statist.* **25**, 818–850.

MÜLLER, H. (1991). Smooth optimum kernel estimates near endpoints. *Biometrika* **78**, 521–530.

— (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737–761.

— & STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610–625.

ROBINSON, P.M. (1997). Large-smaple inference for nonparametric regression with dependent errors. *Ann. Statist.* **25**, 2054–2083.

RUPPERT, D., SHEATHER, S.J. & WAND, M.P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257–1270.

SHEATHER, S.J. & JONES, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53**, 68–690.

WHITT, W. (1970). Weak convergence of probability measures on the space $C([0,\infty])$. *Ann. Math. Statist.* **41**, 939–949.