Long-range dependent time series specification

JITI GAO[1], QIYING WANG[2] and JIYING YIN[3]

[1]*Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. E-mail: jiti.gao@monash.edu*
[2]*School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia. E-mail: qiying@maths.usyd.edu.au*
[3]*Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. E-mail: jiying.yin@monash.edu*

In this paper we propose using a nonparametric model specification test for parametric time series with long-range dependence (LRD). To establish asymptotic distributions of the proposed test statistic, we develop new central limit theorems for certain weighted quadratic forms of stationary time series with LRD. To implement our proposed test in practice, we develop a computer-intensive parametric bootstrap simulation procedure for finding simulated critical values. As a result, our finite-sample studies demonstrate that both the proposed theory and the simulation procedure work well, and that the proposed test has little size distortion and reasonable power.

Keywords: central limit theorem; Gaussian process; linear process; long-range dependence; parametric time series regression; specification testing

1. Introduction

Starting approximately two decades ago, model specification testing of short-range dependent stationary time series has become a very active research field in both econometrics and statistics. In the meantime, estimation of long-range dependent stationary time series models has also been quite active. To the best of our knowledge, however, model specification of stationary time series with long-range dependence (LRD) has not been discussed in the literature. This is probably related to the unavailability of certain central limit theorems for weighted quadratic forms of stationary time series with LRD.

Specification test statistics based on nonparametric and semi-parametric techniques for both independent and short-range dependent cases have been proposed and studied extensively over the last two decades. Recent studies have shown that some data sets may display LRD (see, e.g., Beran [5]; Cheng and Robinson [10]; Robinson [41]; Baillie and King [4]; Anh and Heyde [1]; Robinson [44]; Gao [16]). In addition, existing studies (e.g., Hidalgo [29]; Robinson [43]; Csörgő and Mielniczuk [11]; Mielniczuk and Wu [37]) have examined nonparametric regression analysis of data with LRD.

Recent studies (Cheng and Robinson [10]; Anh *et al.* [2]; Mikosch and Starica [38]; Gao and Hawthorne [18]) have shown that some real data in economics, environment, and finance applications with both LRD and nonlinearities may be modeled by $Y_t = m(X_t) + e_t$, $1 \le t \le n$, where

the regressors X_t are either fixed designs or stationary short-range dependent variables, $m(\cdot)$ is an unknown and probably nonlinear trend function, and the errors e_t are strictly stationary long-range dependent. In addition, the errors may be allowed for a more general dependent structure of the form $e_t = G(X_t, Z_t)$ as proposed by Cheng and Robinson [10] and discussed by others, including Mielniczuk and Wu [37], where $G(\cdot, \cdot)$ is a suitable function and $\{Z_t\}$ is a possibly long-range dependent process.

The key findings of these previous studies suggest that to avoid misrepresenting the mean function or the conditional mean function of long-range dependent data, we should let the data speak for themselves in terms of specifying the true form of the mean function or the conditional mean function. This is particularly important for data with LRD, because unnecessary nonlinearity or complexity in mean functions may cause erroneous LRD.

To address these issues, we propose to model data with possible LRD, nonlinearity, and non-stationary using a general nonparametric trend model. Thus, one of the objectives of the present work is to specify the trend by constructing a nonparametric kernel-based test. Consider a nonlinear time series model of the form

$$Y_t = m(X_t) + e_t, \qquad t = 1, 2, \ldots, n, \tag{1.1}$$

where n is the number of observations, $\{X_t\}$ is either a sequence of fixed designs of $X_t = \frac{t}{n}$ or random regressors, $m(\cdot)$ is an unknown function, and $\{e_t\}$ is a stationary long-range dependent linear process with $E[e_1] = 0$ and $0 < E[e_1^2] = \sigma^2 < \infty$. In addition, $\{X_s\}$ and $\{e_t\}$ are assumed to be independent for all $s, t \geq 1$ when $\{X_t\}$ is a sequence of random regressors. We develop a kernel-based test for the hypotheses

$$H_0\colon m(x) = m_{\theta_0}(x) \quad \text{versus} \quad H_1\colon m(x) = m_{\theta_1}(x) + c_n \Delta(x) \tag{1.2}$$

for all $x \in R = (-\infty, \infty)$, where θ_0 and θ_1 are vectors of unknown parameters, $m_\theta(x)$ is a known parametric function of x indexed by a vector of unknown parameters, θ, $\Delta(x)$ is some smooth function, and $\{c_n\}$ is a sequence of real numbers tending to 0 when $n \to \infty$. To ensure that the true model is identifiable, we also require that for all $\theta_1 \neq \theta_2$, there is a positive constant $\delta_1 > 0$ such that $E[m_{\theta_1}(X_1) - m_{\theta_2}(X_1)]^2 \geq \delta_1 > 0$ in the random design case and $\inf_{0 \leq x \leq 1} |m_{\theta_1}(x) - m_{\theta_2}(x)| \geq \delta_1 > 0$ in the fixed design case.

Note that under H_0, model (1.1) becomes a parametric model of the form

$$Y_t = m_{\theta_0}(X_t) + e_t, \tag{1.3}$$

which covers many important cases. For example, model (1.3) becomes a simple linear model with LRD as in (1.1) of Robinson and Hidalgo [45] when $m_{\theta_0}(X_t) = \alpha_0 + \beta_0 X_t$. For a given set of long-range dependent data, the acceptance of H_0 suggested by a test statistic may indicate that the mean function of the LRD data should be specified parametrically. In the case of the Nile river data as analyzed by Anh *et al.* [2], we would consider using a linear mean function of the form $m(\frac{t}{n}) = \alpha_0 + \beta_0 \cdot \frac{t}{n}$ if a suitable test suggested the acceptance of $H_0\colon m(x) = \alpha_0 + \beta_0 x$. Similarly, if a proper test suggested accepting a second-order polynomial function of the form $m(\frac{t}{n}) = \alpha_0 + \beta_0(\frac{t}{n}) + \gamma_0(\frac{t}{n})^2$ as the true trend of a financial data set $\{Y_t\}$, then we would need only to difference $\{Y_t\}$ twice to generate a stationary set of the data.

Although there is a long and rich literature in the field of model specification for time series models with stationarity (see, e.g., Gao [16]; Li and Racine [35]), little work has been reported on parametric specification testing of $m(\cdot)$ for the case where X_t, e_t, or both may be strictly stationary and long-range dependent time series. To the best of our knowledge, Gao, Wang and Yin [21] may be among the first to consider a parametric specification of $m(\cdot)$ for the case where $\{X_t\}$ is a sequence of stationary Gaussian time series regressors with possible LRD and $\{e_t\}$ is a sequence of i.i.d. random errors.

This paper is organized as follows. We present our proposed test for the hypothesis (1.2) in Section 2.1. To investigate the proposed specification test, we investigate the limiting theorems for the leading term $M_n(h)$ of our test in Section 2.2, where

$$M_n(h) = \sum_{s=1}^{n} \sum_{t=1,\neq s}^{n} e_s a_n(X_s, X_t) e_t, \tag{1.4}$$

with $a_n(X_s, X_t) = K(\frac{X_s - X_t}{h})$, in which $K(\cdot)$ is a probability kernel function and h is a bandwidth parameter. We mention that for the case where $\{X_t\}$ is a sequence of either fixed or random regressors but $\{e_t\}$ is a sequence of long-range dependent errors, the problem of establishing limiting distributions for $M_n(h)$ is difficult. Because there is an involvement of $h \to 0$ into the inside of $K(\cdot)$, existing central limit theorems for U-statistics of long-range dependent processes (see Hsing and Wu [33]) are not applicable. Thus, the limit theorems presented in Section 2.2 are interesting and useful in and of themselves. In Section 3 we discuss some important extensions and applications of the theory established in Section 2. We present a parametric bootstrap simulation procedure as well as some resulting properties for the fixed design situation in Section 4. In Section 4 we also provide an example to demonstrate how to implement the proposed test and the bootstrap simulation procedure in practice. In Section 5 we conclude the paper with some remarks on extensions. Mathematical details are relegated to Appendices A and B. Additional details are available in Appendix C of Gao and Wang [20]. Throughout the paper, we use let $a_n \sim b_n$ denote $\lim_{n\to\infty} a_n/b_n = 1$.

2. Asymptotic theory

We propose a test statistic for the hypothesis (1.2) in Section 2.1. To investigate the proposed test statistic, we develop some new limiting distributions of weighted quadratic forms of dependent processes with LRD in Section 2.2. Their proofs, along with other proofs, are relegated to Appendix A.

2.1. Model specification test

Let K be a one-dimensional bounded probability density function and h be a smoothing bandwidth. When $\{Y_t\}$ is a sequence of long-range dependent random variables, the conventional kernel estimator of $m(\cdot)$ is defined by

$$\widehat{m}(x) = \frac{(1/nh) \sum_{t=1}^{n} K((x - X_t)/h) Y_t}{\widehat{f}(x)}, \tag{2.1}$$

where $\widehat{f}(x) = \frac{1}{nh} \sum_{t=1}^{n} K(\frac{x-X_t}{h})$ is the density estimate of the marginal density function, $f(x)$, of $\{X_t\}$ when $\{X_t\}$ is a sequence of stationary random regressors. When $X_t = \frac{t}{n}$, $\widehat{f}(x) = \frac{1}{nh} \sum_{t=1}^{n} K(\frac{nx-t}{nh})$ is a sequence of functions of x. Various asymptotic properties for $\widehat{m}(x)$ have been studied in the literature (see, e.g., Cheng and Robinson [9]; Robinson [43]; Anh *et al.* [2]). The nonparametric test statistic for the hypothesis (1.2) might be expected to be related to $\widehat{m}(x)$. However, as demonstrated in the model specification literature for both the independent and short-range dependent stationary time series cases (see, e.g., Li and Wang [36]; Gao [16]; Li and Racine [35]), some classes of nonparametric test statistics have little involvement with nonparametric estimation. One of the advantages of such test statistics is that both their large-sample and finite-sample properties are much less sensitive to the use of individual nonparametric estimation as well as the resulting estimation biases. To test the hypothesis (1.2), we therefore propose a kernel-based test statistic of the form

$$\widehat{M}_n(h) = \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} \widehat{e}_s a_n(X_s, X_t) \widehat{e}_t \tag{2.2}$$

for the case where $\{X_t\}$ is a sequence of stationary random regressors, with $a_n(X_s, X_t) = K(\frac{X_s - X_t}{h})$ and $\widehat{e}_t = Y_t - m_{\widetilde{\theta}}(X_t)$, in which $\widetilde{\theta}$ is a consistent estimator of θ_0 under H_0. For the case of fixed-design mean with LRD errors, we also suggest the same form of (2.2) with $X_t = \frac{t}{n}$. As pointed out earlier, we chose (2.2) over those related to $\widehat{m}(x)$ mainly because our experience shows that such a form does not involve biases caused by nonparametric estimation and thus works well both theoretically and practically.

We need the following assumptions on the error process $\{e_t\}$, the kernel function $K(\cdot)$, the regressor $\{X_t\}$, and the regression function $m_\theta(x)$ for the main results of this paper.

Assumption 2.1. (i) $\{e_t\}$ *is a sequence of strictly stationary error processes defined by* $e_t = \sum_{j=-\infty}^{\infty} \psi_j \eta_{t-j}$, *where the innovations* $\{\eta_j\}$ *is a sequence of i.i.d. random variables with* $E[\eta_1] = 0$, $E[\eta_1^2] = 1$ *and* $E[\eta_1^6] < \infty$, *and the covariance* $\gamma(k) = E[e_t e_{k+t}] = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k}$ *satisfies that* $\gamma(0) = \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ *and* $\gamma(k) \sim \eta|k|^{-\alpha}$ *as* $k \to \infty$, *where* $\lambda = (\alpha, \eta)$ *with* $0 < \alpha < 1$ *and* $0 < \eta < \infty$ *is a vector of unknown parameters.*
(ii) *In addition, let* $\psi_j \geq 0$ *when* $\{X_t\}$ *is a sequence of fixed designs.*

Assumption 2.2. *When* $\{X_t\}$ *is a sequence of i.i.d. random regressors,* $\{X_t\}$ *and* $\{e_s\}$ *are assumed to be independent for all* $s \geq 1$ *and* $t \geq 1$, *and the density function* $f(x)$ *of* $\{X_t\}$ *is bounded and uniformly continuous.*

Assumption 2.3. *Let* $\lambda = (\alpha, \eta)$ *be defined as in Assumption* 2.1. *There exist some* $\widetilde{\lambda} = (\widetilde{\alpha}, \widetilde{\eta})$ *such that* $\|\widetilde{\lambda} - \lambda\| = O_P(w_n^{-1})$, *where* $\{w_n\}$ *is a sequence of positive numbers satisfying* $\lim_{n \to \infty} \frac{w_n}{\log n} = \infty$, *where* $\|\cdot\|$ *denotes the Euclidean norm.*

Assumption 2.4. (i) $K(\cdot)$ *is a bounded and symmetric probability kernel function over the real line* R. (ii) *Given* $\frac{1}{2} < \alpha < 1$, *there exists some* $0 < \beta < \alpha - \frac{1}{2}$ *such that* $K(x) = O(\frac{1}{1+|x|^{1-\beta}})$.

Assumption 2.5 (Random design). (i) *Under the null hypothesis* H_0, $\|\widetilde{\theta} - \theta_0\| = O_P(n^{-1/2})$.
(ii) *There exists some* $\varepsilon_0 > 0$ *such that* $\frac{\partial^2 m_\theta(x)}{\partial \theta^2}$ *is continuous in both* $x \in R$ *and* $\theta \in \Theta_0$, *where*
$\Theta_0 = \{\theta: \|\theta - \theta_0\| \leq \varepsilon_0\}$. *In addition,*

$$0 < E\left[\left\|\frac{\partial m_\theta(X_1)}{\partial \theta}\bigg|_{\theta=\theta_0}\right\|^2\right] < \infty \quad and \quad 0 < E\left[\left\|\frac{\partial^2 m_\theta(X_1)}{\partial \theta^2}\bigg|_{\theta=\theta_0}\right\|^2\right] < \infty.$$

Assumption 2.6 (Fixed design). (i) *Under the null hypothesis* H_0, $\|\widetilde{\theta} - \theta_0\| = O_P(n^{-1/2})$.
(ii) *There exists some* $\varepsilon_0 > 0$ *such that* $\frac{\partial m_\theta(x)}{\partial \theta}$ *and* $\frac{\partial^2 m_\theta(x)}{\partial \theta^2}$ *both are bounded and continuous*
in $0 \leq x \leq 1$ *and* $\theta \in \Theta_0$, *where* $\Theta_0 = \{\theta: \|\theta - \theta_0\| \leq \varepsilon_0\}$.

Assumption 2.1(i) is quite standard and covers some important cases. For instance, the case
where $\{e_t\}$ is a sequence of Gaussian errors is included. Assumption 2.1(ii) is required to establish Theorems 2.2 and 2.4 below. The positivity of ψ_j in Assumption 2.1(ii) may be replaced by
less restrictive conditions, such as that ψ_j are eventually positive. Also note that it is possible to
involve a slowly varying function into the form of $\gamma(k)$; however, because this is not essential to
either our theory or practice, we use the current Assumption 2.1 throughout this paper. For alternative conditions on $\{e_t\}$ in this kind of study, we refer to Cheng and Robinson [10], Robinson
[43], and Robinson and Hidalgo [45].

Assumption 2.2 may be relaxed to include the case where $\{X_t\}$ is a sequence of strictly stationary and β-mixing random variables. The corresponding results and their proofs are given in
Section 5 and Appendix C of Gao and Wang [20].

Assumption 2.3 may be justified for certain w_n, like $w_n = n^{2/5}/\log n$; see the discussion for
the construction of $\widehat{\lambda}$ and Theorem 4.2 in Section 4. Assumption 2.4(i) is a standard condition on
the kernel function. To establish Theorems 2.2 and 2.4, we also need Assumption 2.4(ii). This
imposes some restrictions on α and $K(\cdot)$, but these restrictions are easily verifiable. For instance,
Assumption 2.4(ii) is satisfied when $K(\cdot)$ is the standard normal density function or belongs to a
class of probability kernel functions with compact support. Also note that under Assumption 2.4,
$A_\alpha < \infty$, where

$$\begin{aligned}
A_\alpha = \int_0^\infty \int_0^\infty \int_0^\infty x^{-\alpha} y^{-\alpha} [K(z)K(x+y-z) \\
+ K(z-x)K(z-y)]\,dx\,dy\,dz,
\end{aligned} \tag{2.3}$$

as shown in Lemma A.6.

The conventional $\sqrt{n}$ rate of convergence assumed in Assumptions 2.5(i) and 2.6(i) is achievable even in this kind of long-range dependent case (see, e.g., Robinson and Hidalgo [45]; Beran
and Ghosh [6]). It is obvious that Assumptions 2.5(i) and 2.6(i) imply that $\|\widetilde{\theta} - \theta_0\| \in o_P(n^{-\alpha/2})$,
which is crucial and is used in the proofs of Theorems 2.1 and 2.2. Assumption 2.5(ii) imposes
certain moment and smoothing conditions on both the form $m_{\theta_0}(\cdot)$ and the design $\{X_t\}$. Assumption 2.6 is used only for the fixed-design case in Theorem 2.2.

We now state the main results of this paper. Theorem 2.1 considers the case where $\{X_t\}$ is a
sequence of stationary random regressors. The case where $X_t = \frac{t}{n}$ is discussed in Theorem 2.2.
All notation is as before. The proofs of these results are given in Appendix A.

Theorem 2.1. *Suppose that Assumptions* 2.1(i), 2.2, 2.3, 2.4(i) *and* 2.5 *hold.*

(i) *If* $\lim_{n\to\infty} n^{2(1-\alpha)}h = 0$ *and* $\lim_{n\to\infty} nh = \infty$, *then, under* H_0,

$$\widehat{L}_{1n}(h) \equiv \frac{\widehat{M}_n(h)}{\widehat{\sigma}_{1n}(h)} \to_D N(0,1) \qquad as \ n \to \infty, \tag{2.4}$$

where $\widehat{\sigma}_{1n}^2(h) = 2n^2 h \int K^2(x) \, dx (\frac{1}{n}\sum_{t=1}^n \widehat{f}(X_t))(\frac{1}{n}\sum_{t=1}^n \widehat{e}_t^2)^2$.

(ii) *If* $\lim_{n\to\infty} h = 0$ *and* $\lim_{n\to\infty} n^{2(1-\alpha)}h = \infty$, *then, under* H_0,

$$\widehat{L}_{2n}(h) \equiv \frac{\widehat{M}_n(h)}{\widehat{\sigma}_{2n}(h)} \to_D \chi^2(1) \qquad as \ n \to \infty, \tag{2.5}$$

where $\chi^2(1)$ *is the chi-squared distribution with* 1 *degree of freedom, and*

$$\widehat{\sigma}_{2n}(h) = \frac{2n^{2-\widetilde{\alpha}} h \widetilde{\eta}}{(1-\widetilde{\alpha})(2-\widetilde{\alpha})} I_{(0<\widetilde{\alpha}<1)} \frac{1}{n} \sum_{t=1}^n \widehat{f}(X_t).$$

Theorem 2.2. *Suppose that Assumptions* 2.1, 2.3, 2.4 *and* 2.6 *hold. If* $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} w_n h^{1/2}/\log n = \infty$ *and* $nh \to \infty$ *hold, then, under* H_0,

$$\widehat{L}_{3n}(h) \equiv \frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n b_n(s,t)(\widehat{e}_s\widehat{e}_t - \widehat{\gamma}(s-t))}{\widehat{\sigma}_{3n}(h)} \to_D N(0,1) \tag{2.6}$$

as $n \to \infty$, *where* $\widehat{\sigma}_{3n}^2(h) = 8\widetilde{\eta}^2 n(nh)^{3-2\widetilde{\alpha}} A_{\widetilde{\alpha}}^*$ *with*

$$A_{\widetilde{\alpha}}^* = \int_{1/n}^n \int_{1/n}^n \int_{1/n}^n x^{-\widetilde{\alpha}} y^{-\widetilde{\alpha}} [K(z)K(x+y-z)$$

$$+ K(z-x)K(z-y)] \, dx \, dy \, dz,$$

$b_n(s,t) = K(\frac{s-t}{nh})$ *and*

$$\widehat{\gamma}(k) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n-|k|} \widehat{e}_i \widehat{e}_{i+|k|} & \text{for } |k| \le (nh)^{1/3}, \\[4mm] \widetilde{\eta}|k|^{-\widetilde{\alpha}} & \text{for } (nh)^{1/3} < |k| \le n-1. \end{cases}$$

Remark 2.1. (i) As expected, the limiting distributions of the test statistic $\widehat{M}_n(h)$ (under certain normalization) for the hypothesis (1.2) depend on the value of α and the choice of the bandwidth h. We require $1/2 < \alpha < 1$, together with $h^{-1} = o(n)$ and $h = o(n^{-2(1-\alpha)})$, in Theorem 2.1(i). In contrast, Theorem 2.1(ii) allows $0 < \alpha < 1$, but we have to restrict $h = o(1)$ and $h^{-1} = o(n^{2(1-\alpha)})$. These facts imply that to make the limiting distribution of $\widehat{M}_n(h)$ (under certain normalization) normal, the conditions $1/2 < \alpha < 1$ and $h = o(n^{-2(1-\alpha)})$ are essentially necessary for the case where $\{X_t\}$ is a sequence of stationary random regressors.

(ii) For the fixed-design case of $X_t = \frac{t}{n}$, we also require $1/2 < \alpha < 1$ for the asymptotical normality of $\widehat{M}_n(h)$ (under certain normalization) in Theorem 2.2. Furthermore, the range of the bandwidth h depends on the accuracy of $\|\widetilde{\lambda} - \lambda\| = O_P(w_n^{-1})$. Because it may be justified that $w_n = n^{2/5}/\log n$ (see Theorem 4.2), the optimal bandwidth $h \sim Cn^{-1/(4+\alpha)}$ in theory is included in Theorem 2.2. We also mention that the $\widehat{\gamma}(k)$ defined in Theorem 2.2 provides a consistent estimate of $\gamma(k)$ for each fixed k, but it is not possible to replace $\widehat{\gamma}(k) = \widetilde{\eta}|k|^{-\widetilde{\alpha}}$ by $\widehat{\gamma}(k) = \frac{1}{n}\sum_{i=1}^{n-|k|}\widehat{e}_i\widehat{e}_{i+|k|}$ when $(nh)^{1/3} < k \le n-1$, because (A.54) in the proof of Theorem 2.2 below is not true in the latter case. The limiting distribution of $\widehat{M}_n(h)$ (under certain normalization) for the case of $0 < \alpha \le 1/2$ in the fixed-design case is an open problem.

(iii) Our experience suggests that it should be possible to establish some corresponding results of Theorem 2.1 for the case where both $\{X_t\}$ and $\{e_t\}$ exhibit LRD. In the general case, existing studies (Hidalgo [29]; Csörgó and Mielniczuk [11]; Mielniczuk and Wu [37]; Guo and Koul [24, 25]; Zhao and Wu [47]; Kulik and Wichelhaus [34]) in nonparametric estimation have already shown that although similar techniques may be used to establish and prove such corresponding results, the corresponding conditions and proofs are much more technical than those involved in the present work. Our preliminary studies show that we may establish some corresponding results as in Theorem 2.1 for the case where $e_t = \sigma(X_t)\varepsilon_t$, in which $\{X_t\}$ is a sequence of strictly stationary regressors with LRD and $\{\varepsilon_t\}$ is a sequence of strictly stationary short-range dependent errors. To present the key ideas, we focus on the main cases in Theorems 2.1 and 2.2. Other cases are left for future research.

To motivate the need to establish some limit theorems for general quadratic forms of dependent processes with LRD, we observe that

$$\widehat{M}_n(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widehat{e}_s a_n(X_s, X_t)\widehat{e}_t = M_n(h) + 2R_{1n}(h) + R_{2n}(h), \tag{2.7}$$

where $M_n(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}e_s a_n(X_s, X_t)e_t$,

$$R_{1n}(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}a_n(X_s, X_t)e_s\big(m(X_t) - m_{\widetilde{\theta}}(X_t)\big),$$

$$R_{2n}(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}a_n(X_s, X_t)\big(m(X_s) - m_{\widetilde{\theta}}(X_s)\big)\big(m(X_t) - m_{\widetilde{\theta}}(X_t)\big).$$

We show in Appendix A that $2R_{1n}(h) + R_{2n}(h) = o_P(\sigma_{in}(h))$, where $\sigma_{1n}(h)$ and $\sigma_{2n}(h)$ are as defined in Theorem 2.3 below. Thus, to prove Theorem 2.1, we need to establish limit theorems for $M_n(h)$, which is a weighted quadratic form of $\{e_t\}$. Similar arguments also apply for Theorem 2.2. Because existing results are not directly applicable to proving Theorems 2.1 and 2.2, we establish our own results the the next section.

2.2. Limit theorems for quadratic forms

Because both the conditions and results for random designs differ from those for fixed designs, we establish the following two theorems separately. Their proofs are given in Appendix A.

Theorem 2.3. *Suppose that Assumptions* 2.1(i) *and* 2.2–2.4 *hold.*

(i) *If* $\lim_{n \to \infty} n^{2(1-\alpha)} h = 0$ *and* $\lim_{n \to \infty} nh = \infty$ *hold, then*

$$\frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n e_s a_n(X_s, X_t) e_t}{\sigma_{1n}(h)} \to_D N(0, 1) \qquad \text{as } n \to \infty, \tag{2.8}$$

where $\sigma_{1n}^2(h) = n^2 h A_{1\alpha}^2$, *in which* $A_{1\alpha}^2 = 2\gamma^2(0) \int K^2(x) \, dx \int f^2(y) \, dy$, *where* $f(\cdot)$ *denotes the marginal density of* $\{X_t\}$.

(ii) *If* $\lim_{n \to \infty} h = 0$ *and* $\lim_{n \to \infty} n^{2(1-\alpha)} h = \infty$ *hold, then*

$$\frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n e_s a_n(X_s, X_t) e_t}{\sigma_{2n}(h)} \to_D \chi^2(1) \qquad \text{as } n \to \infty, \tag{2.9}$$

where $\sigma_{2n}(h) = n^{2-\alpha} h A_{2\alpha}$, *in which* $A_{2\alpha} = \frac{2\eta}{(1-\alpha)(2-\alpha)}(\int f^2(y) \, dy)$.

Theorem 2.4. *Suppose that Assumptions* 2.1, 2.3 *and* 2.4 *hold. If in addition both* $\lim_{n \to \infty} h = 0$ *and* $\lim_{n \to \infty} nh = \infty$ *hold, then*

$$\frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n b_n(s,t)(e_s e_t - \gamma(s-t))}{\sigma_{3n}(h)} \to_D N(0, 1) \qquad \text{as } n \to \infty, \tag{2.10}$$

where $\sigma_{3n}^2(h) = 8\eta^2 n(nh)^{3-2\alpha} A_\alpha$, *with* A_α *defined as in* (2.3).

Remark 2.2. (i) Theorem 2.3 extends existing limit theorems for both i.i.d. and short-range dependent cases (see, e.g., Gao [16], Chapter 3) to the situation where $\{e_t\}$ is a long-range dependent linear process. Unlike the previously reported results, Theorem 2.3 shows that when $\{a_n(X_s, X_t)\}$ is a sequence of random functions, the limiting distribution of the random weighted quadratic form can be either a standard normal distribution or a chi-squared distribution.

(ii) In the situation where $\{a_n(X_s, X_t)\}$ is a sequence of non-random functions, related results on quadratic forms of long-range dependent time series have been given by Davydov [12], Fox and Taqqu [13,14], Avram [3], Giraitis and Surgailis [22], Giraitis and Taqqu [23], Ho and Hsing [30–32], Hsing and Wu [33], and others. Because the weighted coefficients in those results (see, e.g., Hsing and Wu [33]) are non-random and independent of n, they are not applicable for the establishment of Theorems 2.3 and 2.4. Thus, both Theorems 2.3 and 2.4 extend and complement various existing results.

3. Extensions and applications

In this section we show that each of the leading terms of many existing kernel-based test statistics may be represented by a quadratic form similar to (1.4). Theorems 2.1–2.4 show that it is possible

to establish corresponding results based on the long-range dependent errors for these existing test statistics. To avoid some repetitious arguments, here we only state some main steps; further details are available in Section 3 of Gao and Wang [20].

3.1. Existing kernel-based tests for conditional mean

A very simple idea for constructing a kernel test for H_0 is to compare the L_2 distance between a nonparametric kernel estimator of $m(\cdot)$ and a parametric counterpart. Denote the nonparametric estimator of $m(\cdot)$ by $\widehat{m}(\cdot)$ as in (2.1) and the parametric estimator of $m_{\theta_0}(\cdot)$ by $\widetilde{m}_{\widetilde{\theta}}(\cdot)$ given by

$$\widetilde{m}_{\widetilde{\theta}}(x) = \frac{\sum_{t=1}^{n} K_h(x - X_t) m_{\widetilde{\theta}}(X_i)}{\sum_{t=1}^{n} K_h(x - X_t)}, \tag{3.1}$$

where $\widetilde{\theta}$ is a consistent estimator of θ_0 as defined before and $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$. Härdle and Mammen [27] proposed a test statistic of the form

$$L_{n1}(h) = nh \int \{\widehat{m}(x) - \widetilde{m}_{\widetilde{\theta}}(x)\}^2 w(x) \, dx, \tag{3.2}$$

where $w(x)$ is a non-negative weight function. Recall the model (1.3). Under H_0, it is readily seen that

$$
\begin{aligned}
L_{n1}(h) &= nh \int \left(\frac{[\sum_{t=1}^{n} K_h(x - X_t)(e_t + m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t))]^2}{n^2 \widehat{f}^2(x)} \right) w(x) \, dx \\
&= nh \sum_{s=1}^{n} \sum_{t=1}^{n} \left(\int \frac{K_h(x - X_s) K_h(x - X_t)}{n^2 \widehat{f}^2(x)} w(x) \, dx \right) e_s e_t \\
&\quad + \text{remainder term,}
\end{aligned}
\tag{3.3}
$$

where the leading term is similar to (1.4). In a closely related study, Chen and Gao [8] constructed a test statistic based on empirical likelihood ideas. As they showed, the first-order approximation of their test is asymptotically equivalent to

$$L_{n2}(h) = nh \int [\widehat{m}(x) - \widetilde{m}_{\widetilde{\theta}}(x)]^2 w(x) \, dx. \tag{3.4}$$

It can be easily shown that the test statistic $L_{n2}(h)$ has a similar decomposition to (3.3), in which the leading term is similar to (1.4).

3.2. Testing conditional mean with conditional variance

Because the main objective of this paper is to specify the form of $m(\cdot)$ parametrically, we have assumed that the variance or conditional variance σ^2 is an unknown parameter. As can be seen

from (1.4), we can replace e_t by $e_t = \sigma(X_t)\varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of long-range dependent linear processes. In this case, the leading term of $\widehat{M}_n(h)$ in (2.2) becomes

$$L_{n3}(h) = \sum_{s=1}^{n} \sum_{t=1,\neq s}^{n} \varepsilon_s \sigma(X_s) K\left(\frac{X_s - X_t}{h}\right) \sigma(X_t)\varepsilon_t, \tag{3.5}$$

which is also a quadratic form of (X_t, ε_t).

3.3. Testing conditional mean in additive form

When $X_t = (X_{t1}, \ldots, X_{td})$ in (1.1) is a vector of d-dimensional designs, we may consider a hypothetical problem of the form

$$H_0': m(x) = \sum_{i=1}^{d} m_{i\theta_0}(x_i) \quad \text{versus} \quad H_1': m(x) = \sum_{i=1}^{d} m_{i\theta_1}(x_i) + c_n \sum_{i=1}^{d} \Delta_i(x_i), \tag{3.6}$$

where each $m_{i\theta_0}(\cdot)$ is a known function indexed by θ_0 and $\Delta_i(\cdot)$ is also a known function over $\mathbb{R}$. Various additive models have been discussed in the literature (see, e.g., Sperlich, Tjøstheim and Yang [46]; Gao, Lu and Tjøstheim [19]; Gao [16]). The construction of $\widehat{M}_n(h)$ suggests a test statistic of the form

$$L_{n4}(h) = \sum_{j=1}^{n} \sum_{i=1}^{n} \widetilde{Y}_i \prod_{k=1}^{d} K\left(\frac{X_{ik} - X_{jk}}{h}\right) \widetilde{Y}_j \tag{3.7}$$

for the hypothesis (3.6), where $\widetilde{Y}_i = (Y_i - \sum_{k=1}^{d} m_{k\widetilde{\theta}_0}(X_{ik}))$. Clearly, $L_{n4}(h)$ also has a leading term similar to (1.4).

As mentioned earlier, some corresponding results of Theorems 2.1–2.4 may be established accordingly for $L_{ni}(h)$, $1 \le i \le 4$.

4. Simulation procedure and an example of implementation

In this section we focus on an implementation procedure for $\widehat{L}_{3n}(h)$ in the fixed design case. Implementation of $\widehat{L}_{1n}(h)$ and $\widehat{L}_{2n}(h)$ in the random design case requires a different asymptotic theory, which involves developing a different bootstrap simulation procedure. We leave a discussion of this for future research. Before we implement $\widehat{L}_{3n}(h)$, we develop a simulation procedure for the choice of a simulated critical value and then propose an estimation procedure for λ involved in the proposed test.

Explicitly, Section 4.1 establishes a simulation procedure for the implementation of the test statistic $\widehat{L}_{3n}(h)$. An estimation procedure for λ is briefly mentioned in Section 4.2. Section 4.3 presents an example of implementation to check whether the theory works in practice.

4.1. Simulation scheme and asymptotic properties

When constructing the simulation, the covariance structure $\gamma_\lambda(k) = E[e_t e_{t+k}]$ must be replaced by an estimated version, $\gamma_{\widetilde{\lambda}}(k)$, with $\widetilde{\lambda} = (\widetilde{\alpha}, \widetilde{\eta})$ as consistent estimators of $\lambda = (\alpha, \eta)$. We assume the existence of $\widetilde{\lambda}$ at the moment, and describe its construction in Section 4.2.

Simulation Procedure 4.1. *Let l_r $(0 < r < 1)$ be the $1 - r$ quantile of the exact finite-sample distribution of $\widehat{L}_{3n}(h)$. Because l_r might not be evaluated in practice, we suggest replacing it with an approximate r-level critical value l_r^*, using the following bootstrap procedure:*

1. *Generate Y_t^* from $Y_t^* = m_{\widetilde{\theta}}(X_t) + e_t^*$ for $1 \leq t \leq n$, where $X_t = \frac{t}{n}$ and $\{e_t^*\}$ is a sequence of stationary LRD errors generated from $e_t^* = \sum_{j=-\infty}^{\infty} \psi_j^* \eta_{t-j}^*$, with $\{\psi_j^*\}$ chosen as $\psi_j^* = c(\widetilde{\alpha})|j|^{-(1+\widetilde{\alpha})/2}$ such that $\gamma_{\widetilde{\lambda}}(k) = E[e_t^* e_{k+t}^* | \mathcal{W}_n] = \sum_{j=-\infty}^{\infty} \psi_j^* \psi_{j+k}^* \sim \widetilde{\eta}|k|^{-\widetilde{\alpha}}$ as k large enough and $c(\widetilde{\alpha})$ also chosen such that $\sum_{j=-\infty}^{\infty} \psi^{*2} = \widetilde{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - m_{\widetilde{\theta}}(X_t))^2$, where $\{\eta_s^*\}$ is a sequence of independent and identically random variables with $E[\eta_1^*] = 0$, $E[\eta_1^{*2}] = 1$, $E[\eta_1^{*6}] < \infty$, and $\mathcal{W}_n = (Y_1, \ldots, Y_n)$.*
2. *Use the data set $\{Y_t^*: 1 \leq t \leq n\}$ to obtain an estimator $\widetilde{\theta}^*$ of $\widetilde{\theta}$, and construct a corresponding version, $\widehat{L}_{3n}^*(h)$, of $\widehat{L}_{3n}(h)$ under H_0; that is, $\widehat{L}_{3n}^*(h)$ is computed according to the same formula as $\widehat{L}_{3n}(h)$ but with $\{Y_t: 1 \leq t \leq n\}$ and $(\theta_0, \widetilde{\theta})$ replaced by $\{Y_t^*: 1 \leq t \leq n\}$ and $(\widetilde{\theta}, \widetilde{\theta}^*)$, respectively.*
3. *Repeat the foregoing step M times and produce M versions of $\widehat{L}_{3n}^*(h)$, denoted by $\widehat{L}_{3n,m}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{L}_{3n,m}^*(h)$ to construct their empirical distribution function. Define $P^*(\cdot)$ as the bootstrap distribution of $\widehat{L}_{3n}^*(h)$ given $\mathcal{W}_n$ by $P^*(\widehat{L}_{3n}^*(h) \leq x) = P(\widehat{L}_{3n}^*(h) \leq x | \mathcal{W}_n)$. Let l_r^* $(0 < r < 1)$ satisfy $P^*(\widehat{L}_{3n}^*(h) \geq l_r^*) = r$, and estimate l_r by l_r^*.*

Remark 4.1. Note that $\{\eta_i^*\}$ is a sequence of independent random variables generated from a prespecified distribution, and that its choice has little effect on the finite-sample performance of the bootstrap simulation method. In Example 4.1 we generate a sequence of $\{\eta_i^*\}$ from a sequence of the normalized χ_2^2 random variables of the form $\frac{\chi_2^2 - 2}{2}$ even when $\{\eta_i\}$ is actually a sequence of standard normal random variables. As discussed by Li and Wang [36], Bühlmann [7], Franke, Kreiss and Mammen [15], Chen and Gao [8], and others, in general we may use a wild bootstrap method to generate a sequence of resamples for $\{e_t^*\}$.

As suggested by the referee, in Section 4.3 we propose using a block bootstrap method (see, e.g., Hall, Horowitz and Jing [26]) to compare the finite-sample performance of the proposed test.

To investigate asymptotic properties of l_r^* and $\widehat{L}_{3n}^*(h)$, we need the following additional assumptions. Assumption 4.1 ensures that $\widehat{L}_{3n}(h)$ has some power under the alternative hypothesis. Assumption 4.2 is a bootstrap version of Assumption 2.6(i).

Assumption 4.1. *Let H_0 be false.* (i) *Assumption 2.6 holds, with θ_0 replaced by θ_1.* (ii) $\lim_{n \to \infty} n^\alpha h^{\alpha-1/2} c_n^2 = \infty$ *for $\frac{1}{2} < \alpha < 1$ and $0 < \int_0^1 \Delta^2(x)\,dx < \infty$, where c_n and $\Delta(x)$ are as defined in (1.2).*

Assumption 4.2. *Under H_0, $\|\widetilde{\theta}^* - \widetilde{\theta}\| = O_{P^*}(n^{-1/2})$.*

We now have the following theorem. Its proof is similar to that of Theorems 2.1 and 2.2 and is provided in Appendix B.

Theorem 4.1. (i) *If, in addition to the conditions of Theorem 2.2, Assumption 4.2 holds, then, under H_0,*

$$\sup_{x \in R} \left| P^*\left(\widehat{L}_{3n}^*(h) \leq x\right) - P\left(\widehat{L}_{3n}(h) \leq x\right) \right| = o_P(1), \tag{4.1}$$

and $\lim_{n \to \infty} P(\widehat{L}_{3n}(h) > l_r^*) = r$.

(ii) *If, in addition to the conditions of Theorem 2.2, Assumptions 4.1 and 4.2 hold, then, under H_1, $\lim_{n \to \infty} P(\widehat{L}_{3n}(h) > l_r^*) = 1$.*

Theorem 4.1 shows that the bootstrap approximation works well asymptotically. For both the independent and short-range dependent stationary cases, a discussion of existing results is available in Chapter 3 of Gao [16]. To the best of our knowledge, Theorem 4.1 is a new result in this kind of long-range dependent time series specification.

Note that l_r^* is a function of h. A natural problem raised in simulation is the choice of a suitable bandwidth h. To examine how to address this issue, define the size and power functions of $\widehat{L}_{3n}(h)$ as

$$\gamma_n(h) = P\left(\widehat{L}_{3n}^*(h) > l_r^* \big| H_0\right) \quad \text{and} \quad \beta_n(h) = P\left(\widehat{L}_{3n}^*(h) > l_r^* \big| H_1\right). \tag{4.2}$$

Clearly, a reasonable selection procedure for a suitable bandwidth is such that the size function $\gamma_n(h)$ is controlled by a significance level, but the power function $\beta_n(h)$ is maximized over such bandwidths that make $\gamma_n(h)$ is controllable. In this paper we propose choosing

$$\widehat{h}_{\text{test}} = \arg\max_{h \in \mathcal{H}_n} \beta_n(h) \quad \text{with } \mathcal{H}_n = \{h \colon r - \varepsilon < \gamma_n(h) < r + \varepsilon\}, \tag{4.3}$$

where $0 < \varepsilon < r$. Theoretically, we have not been able to study $\widehat{h}_{\text{test}}$ asymptotically. In Example 4.1 below, we instead combine our proposed Simulation Procedure 4.1 and Implementation Procedure 4.1 to numerically approximate $\widehat{h}_{\text{test}}$.

Implementation Procedure 4.1. *Use $\widehat{l}_r^* = l_r^*(\widehat{h}_{\text{test}})$ as the simulated critical value to compute the sizes and power values of $\widehat{L}_{3n}(\widehat{h}_{\text{test}})$.*

4.2. LRD parameter estimation

As mentioned at the beginning of Section 4, we need to estimate $\lambda = (\alpha, \eta)$ when λ is unknown. Let $u_t \equiv \widehat{e}_t = Y_t - m_{\widetilde{\theta}}(X_t)$ and $I_u(\omega) = \frac{1}{2\pi n} |\sum_{s=1}^n u_s e^{is\omega}|^2$.

Concerning the estimation of λ for the case where the errors are not necessarily Gaussian, several methods are available in the literature. Giraitis and Surgailis [22] established an asymptotic theory for a Whittle estimation method. Heyde and Gay [28] considered a multivariate case

and established several asymptotic results. Robinson [42] proposed an efficient Gaussian semi-parametric estimation method for the long-range dependence parameter, which is equivalent to α in the present paper. Ideally, we would adopt the estimation method proposed by Robinson [42] for our case. But because that method does not directly imply asymptotic normality for $\widetilde{\eta}$ to meet our Assumption 2.3, we modified the method of Heyde and Gay [28] for our case.

Define the following Gauss–Whittle objective function:

$$\Gamma_u(\lambda) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log(\psi_u(\omega; \lambda)) + \frac{I_u(\omega)}{\psi_u(\omega; \lambda)} \right) d\omega, \tag{4.4}$$

where $\psi_u(\omega; \lambda)$ is the spectral density function of $\{u_t\}$.

Let $\widetilde{\lambda}$ minimize $\Gamma_u(\lambda)$ over Λ_0, a compact subset of $\Lambda = \{\lambda: \lambda = (\alpha, \eta): \frac{1}{2} < \alpha < 1, 0 < \eta < \infty\}$. Define $L_i((-\pi, \pi]) = \{\psi: \int_{-\pi}^{\pi} |\psi(\omega)|^i d\omega < \infty\}$ for $i = 1, 2$.

We introduce the following assumption:

Assumption 4.3. (i) *Let $\{e_t\}$ satisfy Assumption 2.1 and let $\psi_e(\omega, \lambda)$ be its spectral density function.*

(ii) *Let $\lambda = (\lambda_1, \lambda_2) = (\alpha, \eta)$. The functions $\psi_e(\omega; \lambda)$ and $p_i(\omega, \lambda) = -\frac{\partial \psi_e^{-1}(\omega, \theta)}{\partial \lambda_i}$ for $i = 1, 2$ satisfy the following conditions:*

- *$\int_{-\pi}^{\pi} \log(\psi_e(\omega, \lambda)) d\omega$ is twice differentiable in λ under the integral sign. In addition, $\psi_e(\omega, \lambda)$ is continuous at all $\omega \neq 0$ and $\lambda \in \Lambda$, and $\psi_e^{-1}(\omega, \lambda)$ is continuous at all (ω, λ).*
- *$\psi_e^{-1}(\omega, \lambda)$, $\omega \in (-\pi, \pi]$ and $\lambda \in \Lambda$, is twice differentiable with respect to λ, and the functions $\frac{\partial \psi_e^{-1}(\omega, \lambda)}{\partial \lambda_i}$ and $\frac{\partial^2 \psi_e^{-1}(\omega, \lambda)}{\partial \lambda_i \partial \lambda_j}$ are continuous at all (ω, λ), $\omega \neq 0$ for all $1 \leq i, j \leq 2$.*
- *For all $1 \leq i \leq 2$, $p_i(\omega, \lambda)$ are symmetric about $\omega = 0$ and $p_i(\omega, \lambda) \in L_1((-\pi, \pi])$ for any $\lambda \in \Lambda$.*
- *$\psi_e(\omega, \lambda) p_i(\omega, \lambda)$ for $i = 1, 2$ are in $L_2((-\pi, \pi])$ for any $\lambda \in \Lambda$. In addition, there exists a constant $0 < q \leq 1$ such that $|\omega|^q \psi_e(\omega, \lambda)$ is bounded and $|\omega|^{-q} p_i(\omega, \lambda)$ for $i = 1, 2$ are in $L_2((-\pi, \pi])$ for any $\lambda \in \Lambda$.*
- *The matrix $(\frac{\partial \log(\psi_e(\omega, \lambda))}{\partial \lambda})(\frac{\partial \log(\psi_e(\omega, \lambda))}{\partial \lambda})^\tau$ is in $L_1((-\pi, \pi])$ for any $\lambda \in \Lambda$ and $\Sigma(\lambda) = \frac{1}{4\pi} \int_{-\pi}^{\pi} (\frac{\partial \log(\psi_e(\omega, \lambda))}{\partial \lambda})(\frac{\partial \log(\psi_e(\omega, \lambda))}{\partial \lambda})^\tau d\omega$ is positive definite for all $\lambda \in \Lambda$.*

Assumption 4.3 is equivalent to Conditions (A1)–(A3) of Heyde and Gay [28]. Assumption 4.3(ii) is satisfied in many cases. Using existing results (see, e.g., Beran [5]), we can verified that Assumption 4.3(ii) is satisfied when the spectral density function is of the form $\psi_e(\omega, \lambda) = d(\omega, \lambda)(\sin(\frac{\omega}{2}))^{-(1-\alpha)}$, where $d(\omega, \lambda) > 0$ satisfies certain conditions.

Theorem 4.2 establishes an asymptotic consistency result for $\widetilde{\lambda}$. Its proof is given in Appendix B.

Theorem 4.2. *Let the conditions of Theorem 2.2, except Assumption 2.1, hold. If in addition Assumption 4.3 holds, then, for n large enough,*

$$\|\widetilde{\lambda} - \lambda\| = o_P(\log n / n^{2/5}), \tag{4.5}$$

where $\widetilde{\lambda} = (\widetilde{\alpha}, \widetilde{\eta})$. Theorem 4.2 shows that Assumption 2.3 can be justified for $w_n = \frac{n^{2/5}}{\log(n)}$.

4.3. An example of implementation

In this section we implement our proposed simulation procedure to show how to assess the finite-sample properties of the proposed test $\widehat{L}_{3n}(h)$ by using a simulated example.

Before we examine the finite-sample performance of the size and power functions of our proposed test, we briefly state the following simulation procedure as an alternative to the proposed Simulation Procedure 4.1.

Simulation Procedure 4.2. *Let l_r $(0 < r < 1)$ be the $1 - r$ quantile of the exact finite-sample distribution of $\widehat{L}_{3n}(h)$. Because l_r might not be evaluated in practice, we suggest an approximate r-level critical value, l_r^*, to replace it, using the following bootstrap procedure*:

1. *Generate $\{\widetilde{e}_t\}$ by $\widetilde{e}_t = \sum_{j=1}^{\infty} \widetilde{\psi}_j \eta_{t-j}$, in which $\{\eta_k\}$ is a sequence of independent observations drawn from $N(0, 1)$, and $\widetilde{\psi}_j = c(\widetilde{\alpha}) j^{-(1+\widetilde{\alpha})/2}$, in which $c(\widetilde{\alpha}) = \sqrt{\widetilde{\alpha}}$. Let $l = [n^{1/3}]$, and choose b such that $bl = n$. Generate $e_{1l}^*(j) = (\widetilde{e}_1(j), \ldots, \widetilde{e}_l(j)), \ldots,$ and $e_{Nl}^*(j) = (\widetilde{e}_{(b-1)l+1}(j), \ldots, \widetilde{e}_{bl}(j))$ in step j for $N = n - l + 1$. Replicate the resampling $J = 250$ times and obtain J bootstrap resamples $\{e_{sl}^*(j): 1 \le s \le N; 1 \le j \le J\}$. Then take the average $e_{sl}^* = \frac{1}{J} \sum_{j=1}^{J} e_{sl}^*(j)$ to obtain a block bootstrap version of e_t of the form $(e_1^*, \ldots, e_n^*) = (e_{1l}^*, \ldots, e_{Nl}^*)$. Then generate Y_t^* from $Y_t^* = m_{\widetilde{\theta}}(X_t) + e_t^*$ for $1 \le t \le n$, where $X_t = \frac{t}{n}$ for $1 \le t \le n$.*

2. *Use the data set $\{Y_t^*: 1 \le t \le n\}$ to obtain an estimator $\widetilde{\theta}^*$ of $\widetilde{\theta}$, and construct a corresponding version $\widehat{L}_{1n}^*(h)$ of $\widehat{L}_{1n}(h)$ under H_0; that is, compute $\widehat{L}_{1n}^*(h)$ according to the same formula as for $\widehat{L}_{1n}(h)$, but with $\{Y_t: 1 \le t \le n\}$ and $(\theta_0, \widetilde{\theta})$ replaced by $\{Y_t^*: 1 \le t \le n\}$ and $(\widetilde{\theta}, \widetilde{\theta}^*)$, respectively.*

3. *Repeat the foregoing step M times and produce M versions of $\widehat{L}_{3n}^*(h)$, denoted by $\widehat{L}_{3n,m}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{L}_{1n,m}^*(h)$ to construct the empirical distribution function. Define $P^*(\cdot)$ as the bootstrap distribution of $\widehat{L}_{1n}^*(h)$ given $\mathcal{W}_n$ by $P^*(\widehat{L}_{1n}^*(h) \le x) = P(\widehat{L}_{1n}^*(h) \le x | \mathcal{W}_n)$. Let l_r^* $(0 < r < 1)$ satisfy $P^*(\widehat{L}_{1n}^*(h) \ge l_r^*) = r$, and estimate l_r by l_r^*.*

Example 4.1. Consider a linear model of the form

$$Y_t = \alpha_0 + \beta_0 X_t + e_t, \qquad t = 1, \ldots, n, \tag{4.6}$$

where (α_0, β_0) is a pair of unknown parameters, $X_t = \frac{t}{n}$ for $1 \le t \le n$, and $\{e_t\}$ is a sequence of dependent errors given by $e_t = \sum_{j=1}^{\infty} \psi_j \eta_{t-j}$, where $\{\eta_k\}$ is a sequence of independent observations drawn from $N(0, 1)$ and $\psi_j = c(\alpha) j^{-(1+\alpha)/2}$ for $\frac{1}{2} < \alpha < 1$, with $c(\alpha) = \sqrt{\alpha}$. We choose $\alpha = \frac{3}{4}$ in our simulation. Throughout this section, we use the standard normal kernel function $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

To compute the sizes and power values of $\widehat{L}_{3n}(h)$, we generate $\{Y_t\}$ from

$$H_0: Y_t = \alpha_0 + \beta_0 X_t + e_t \quad \text{or} \quad H_1: Y_t = \alpha_1 + \beta_1 X_t + \gamma_1 X_t^2 + e_t, \tag{4.7}$$

Table 1. Sizes and power values based on simulated critical values

Null hypothesis is true

Observation	T:	$r = 1\%$		$r = 5\%$		$r = 10\%$	
		f_{0cv}	f_{0test}	f_{0cv}	f_{0test}	f_{0cv}	f_{0test}
250		0.004	0.011	0.026	0.064	0.061	0.086
500		0.006	0.009	0.031	0.060	0.058	0.082
750		0.003	0.012	0.028	0.042	0.049	0.122

Null hypothesis is false

Observation	T:	$r = 1\%$		$r = 5\%$		$r = 10\%$	
		f_{1cv}	f_{1test}	f_{1cv}	f_{1test}	f_{1cv}	f_{1test}
250		0.084	0.160	0.121	0.242	0.181	0.366
500		0.076	0.144	0.153	0.296	0.179	0.304
750		0.104	0.216	0.142	0.320	0.234	0.426

where the parameters (α_0, β_0) are estimated by $(\widetilde{\alpha}_0, \widetilde{\beta}_0)$ under H_0 and the parameters $(\alpha_1, \beta_1, \gamma_1)$ are estimated by the ordinary least squares estimators $(\widetilde{\alpha}_1, \widetilde{\beta}_1, \widetilde{\gamma}_1)$ under H_1. When we generate $\{Y_t\}$, the initial values are $\alpha_i = \beta_i \equiv 1$ for $i = 0, 1$, and we use $\gamma_1 = n^{-\frac{1}{2}}\sqrt{\log\log(n)}$ to compute the power values in both Tables 1 and 2.

Table 2. Sizes and power values based on simulated critical values

Null hypothesis is true

Observation	T:	$r = 1\%$		$r = 5\%$		$r = 10\%$	
		f_{0cv}	f_{0test}	f_{0cv}	f_{0test}	f_{0cv}	f_{0test}
250		0.004	0.010	0.016	0.060	0.050	0.136
500		0.002	0.012	0.016	0.062	0.034	0.113
750		0.006	0.010	0.018	0.051	0.036	0.102

Null hypothesis is false

Observation	T:	$r = 1\%$		$r = 5\%$		$r = 10\%$	
		f_{1cv}	f_{1test}	f_{1cv}	f_{1test}	f_{1cv}	f_{1test}
250		0.080	0.168	0.130	0.324	0.190	0.422
500		0.070	0.332	0.164	0.526	0.224	0.628
750		0.114	0.334	0.180	0.544	0.246	0.654

We first apply Implementation Procedure 4.1 for the case where $\{\eta_i^*\}$ is a sequence of the normalized χ_2^2 random variable of the form $\frac{\chi_2^2 - 2}{2}$. When using Simulation Procedure 4.2, we start with the case where $\{\eta_i\}$ is a sequence of i.i.d. observations drawn from $N(0, 1)$. We then apply either Implementation Procedure 4.1 or Simulation Procedure 4.2 to find $\widehat{h}_{\text{test}}$. To assess whether the use of $\widehat{h}_{\text{test}}$ associated with the proposed bootstrap method could improve the finite-sample performance of $\widehat{L}_{3n}$, we also consider using an estimation-based optimal bandwidth of the form $\widehat{h}_{cv} = \arg\min_{h \in H_{cv}} \frac{1}{n} \sum_{t=1}^{n} (Y_t - \widehat{m}_{-t}(X_t, h))^2$, where $\widehat{m}_{-t}(X_t, h) = \sum_{s=1, \neq t}^{n} K(\frac{X_t - X_s}{h}) Y_s / \sum_{u=1, \neq t}^{n} K(\frac{X_t - X_u}{h})$ and $H_{cv} = [c_1 n^{-1}, c_2 n^{-(1-c_0)}]$ for some $0 < c_1 < c_2 < \infty$ and $0 < c_0 < 1$.

In the implementation of Simulation Procedure 4.1, we consider the case with $M = 500$ replications, each with $B = 250$ bootstrapping resamples. When implementing Simulation Procedure 4.2, we use the same number of replications $M = 500$, each with $J = 250$ block bootstrapping resamples ($n = bl$). All of these simulations were done for data sets of size $n = 250$, 500 and 750. Let $\widehat{h}_{i\text{test}}$ and $\widehat{h}_{icv}$ denote the corresponding versions of $\widehat{h}_{\text{test}}$ and $\widehat{h}_{cv}$ under H_i for $i = 0, 1$. Let $f_{i\text{test}}$ denote the frequency of $\widehat{L}_{3n}(\widehat{h}_{i\text{test}}) > l_r^*(\widehat{h}_{i\text{test}})$ (note that we repeat the test 10 times to compute $l_r^*(\widehat{h}_{i\text{test}})$), and let f_{icv} be the frequency of $\widehat{L}_{3n}(\widehat{h}_{icv}) > z_r$ under H_i for each of $i = 0, 1$, where $z_{0.01} = 2.33$ at the 1% level, $z_{0.05} = 1.645$ at the 5% level, and $z_{0.10} = 1.28$ at the 10% level.

Tables 1 and 2 present the simulated sizes and power values based on Simulation Procedures 4.1 and 4.2, respectively.

Tables 1 and 2 show that the simulated power values associated with $\widehat{h}_{\text{test}}$ are greater than those based on $\widehat{h}_{cv}$, and also that $\widehat{L}_{1n}(\widehat{h}_{cv})$ has some kind of size distortion when using z_r (equivalent to using the asymptotic normality as the sample distribution) in practice. This finding is not surprising, given that the the theory demonstrates that each $\widehat{h}_{\text{test}}$ is chosen such to maximize the resulting power function, with the corresponding size function computed using the simulated critical value $l_r^*(\widehat{h}_{\text{test}})$ in each case.

Comparing the regression bootstrap method in Table 1 and the block bootstrap method in Table 2 shows similar sizes for the two methods. Meanwhile, the power values in columns 3, 5 and 7 in Tables 1 and 2 show that $\widehat{L}_{3n}(\widehat{h}_{\text{test}})$ associated with the block-bootstrap method is more powerful than $\widehat{L}_{3n}(\widehat{h}_{\text{test}})$ based on the regression bootstrap method, whereas whether $\widehat{L}_{3n}(\widehat{h}_{cv})$ associated with the block-bootstrap method or $\widehat{L}_{3n}(\widehat{h}_{cv})$ based on the regression bootstrap method is uniformly more powerful is not so clear. This finding further supports our argument that the choice of a suitable bootstrap method for selecting an appropriate bandwidth to maximize the resulting power function is more relevant than the choice of the bootstrap method itself.

5. Conclusion

We have proposed a new nonparametric test for the parametric specification of the mean function of long-range dependent time series, and have established asymptotic distributions of the proposed test for both the fixed and random design cases. In addition, we have proposed Simulation Procedure 4.1 and Implementation Procedure 4.1 to implement the proposed test in practice. Our finite-sample results show that the proposed test, as well as the two procedures, are practically applicable and implementable. Further topics, including how to extend existing results

(Nishiyama and Robinson [39,40]; Gao [16]; Gao and Gijbels [17]) for Edgeworth expansions for the size and power functions of the proposed tests to the long-range dependence case, are left for future research.

Appendix A: Technical details

This appendix provides technical details for the asymptotic theory in Section 2. Appendix A.1 gives several preliminary lemmas. Appendix A.2 presents the proofs of Theorems 2.3 and 2.4, and A.3 presents the proofs of Theorems 2.1 and 2.2. Proofs of the lemmas are given in Appendix A of Gao and Wang [20]. Throughout this section, we denote constants by $C, C_1, \ldots,$ which may take different values at each appearance.

A.1. Technical lemmas

Lemma A.1. *Let* $\{e_t\}$ *be a linear process defined by* $e_t = \sum_{j=-\infty}^{\infty} \psi_j \eta_{t-j}$, *where* $\{\eta_j\}$ *is a sequence of i.i.d. random variables with* $E[\eta_1] = 0$, $E[\eta_1^2] = 1$, $E[\eta_1^4] < \infty$ *and* $\gamma(0) < \infty$, *where* $\gamma(k) = E[e_t e_{k+t}] = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k}$. *Then, for all* j, k, s *and* t,

$$
E[e_j e_k e_s e_t] = (E[\eta_1^4] - 3) \sum_{m=-\infty}^{\infty} \psi_{j-m} \psi_{k-m} \psi_{s-m} \psi_{t-m}
$$
$$
+ \gamma(j-k)\gamma(s-t) + \gamma(j-s)\gamma(k-t) + \gamma(j-t)\gamma(k-s). \tag{A.1}
$$

In particular, we have that $E[e_1^4] < \infty$,

$$
|E[e_j^2 e_k^2] - \gamma^2(0)| \le C\gamma^2(j-k) \quad \text{and} \tag{A.2}
$$

$$
|E[e_j^2 e_k e_s]| \le C[\gamma(j-k) + \gamma(j-s) + \gamma(k-s)] \tag{A.3}
$$

for all $j \ne k \ne s$. *If, in addition,* $\psi_k \ge 0$, *then for all* j, k, s *and* t,

$$
\left| E\left[\left(e_j e_{j+s} - \gamma(s) \right) \left(e_k e_{k+t} - \gamma(t) \right) \right] \right|
$$
$$
\le C\gamma(j-k)\gamma(j-k+s-t) + \gamma(j-k+s)\gamma(j-k-t). \tag{A.4}
$$

Lemma A.2. *Let* $1/2 < \alpha < 1$ *and* $0 < \beta < \alpha - 1/2$. *Then, for all* $k \ge 3$ *and as* $n \to \infty$,

$$
I_n = \frac{1}{n^{k/2}} \int_1^n \int_1^n \cdots \int_1^n |x_1 - x_2|^{-\alpha} |x_2 - x_3|^{\beta-1} \cdots
$$
$$
\times |x_{2k-1} - x_{2k}|^{-\alpha} |x_{2k} - x_1|^{\beta-1} \, dx_1 \, dx_2 \cdots dx_{2k} \to 0. \tag{A.5}
$$

In Lemma A.3 below, let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. random variables with density function $f(x)$, and set $g_n(X_i, X_j) = K(\frac{X_i - X_j}{h}) - E[K(\frac{X_i - X_j}{h})]$,

$$g_{1n}(X_i) = E[g_n(X_i, X_j)|X_i] \quad \text{and}$$
$$g_{2n}(X_i, X_j) = g_n(X_i, X_j) - g_{1n}(X_i) - g_{1n}(X_j). \tag{A.6}$$

Lemma A.3. *Let $K(x)$ satisfy Assumption 2.4(i). If $f(x)$ is a bounded and uniformly continuous function on $\mathbb{R}$, then*

$$E\big[K[(X_1 - X_2)/h]\big] \sim c_1 h, \tag{A.7}$$

$$E[g_{2n}^2(X_1, X_2)] \sim E\big[K^2[(X_1 - X_2)/h]\big] \sim c_2 h, \tag{A.8}$$

$$E[g_{2n}^4(X_1, X_2)] \sim E\big[K^4[(X_1 - X_2)/h]\big] \sim c_4 h, \tag{A.9}$$

where $c_j = \int_{-\infty}^{\infty} K^j(s)\,\mathrm{d}s \int_{-\infty}^{\infty} f^2(y)\,\mathrm{d}y$ for $j = 1, 2, 4$. Furthermore,

$$E[g_{1n}^2(X_1)] \sim d_1 h^2, \tag{A.10}$$

$$E[g_{2n}(X_1, X_3) g_{2n}(X_1, X_4) g_{2n}(X_2, X_3) g_{2n}(X_2, X_4)] \sim d_2 h^3, \tag{A.11}$$

where

$$d_1 = \int_{-\infty}^{\infty} \left[f(x) - \int_{-\infty}^{\infty} f^2(y)\,\mathrm{d}y \right]^2 f(x)\,\mathrm{d}x$$

and

$$d_2 = \int \int \int K(s)K(t)K(x+s)K(x+t)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}x \int_{-\infty}^{\infty} f^4(y)\,\mathrm{d}y.$$

Our next lemma establishes a Berry–Esseen type bound for random weighted U statistics. This lemma is interesting and useful in itself.

Lemma A.4. *Let $\{\varepsilon_k, k \geq 1\}$ be a sequence of i.i.d. random variables, let $\{a_{nij}\}$ be a sequence of constants with $a_{nij} = a_{nji}$ for all $n \geq 1$, and let $\{\varphi_n(x, y)\}$ be a sequence of symmetric Borel-measurable functions such that for all $n \geq 1$,*

$$E[\varphi_n^2(\varepsilon_1, \varepsilon_2)] > 0, \qquad E[\varphi_n(\varepsilon_1, \varepsilon_2)|\varepsilon_1] = 0. \tag{A.12}$$

Then there exists an absolute constant $A > 0$ such that

$$\sup_x |P(B_n^{-1} S_n \leq x) - \Phi(x)| \leq A B_n^{-4/5} \big(A_{1n} E \varphi_n^4(\varepsilon_1, \varepsilon_2) + A_{2n} \mathcal{L}_n \big)^{1/5}, \tag{A.13}$$

where $S_n = \sum_{1 \leq i < j \leq n} a_{nij} \varphi_n(\varepsilon_i, \varepsilon_j)$, $B_n^2 = \sum_{1 \leq i < j \leq n} a_{nij}^2 E \varphi_n^2(\varepsilon_1, \varepsilon_2)$,

$$A_{1n} = \sum_{i=2}^{n} \left(\sum_{j=1}^{i-1} a_{nij}^2 \right)^2, \qquad A_{2n} = \sum_{i=2}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{i-1} a_{nik} a_{njk} \right)^2, \tag{A.14}$$

$$\mathcal{L}_n = E[\varphi_n(\varepsilon_1, \varepsilon_3)\varphi_n(\varepsilon_1, \varepsilon_4)\varphi_n(\varepsilon_2, \varepsilon_3)\varphi_n(\varepsilon_2, \varepsilon_4)]. \tag{A.15}$$

Lemma A.5. *Let $\{\varepsilon_k, k \geq 1\}$ be a sequence of i.i.d. random variables with $E[\varepsilon_1] = 0$ and $E[\varepsilon_1^6] < \infty$. Let $\{a_{nij}\}$ be a sequence of real numbers with $a_{nij} = a_{nji}$ and $\|A\|^2 \equiv \sum_{i,j=-\infty}^{\infty} a_{nij}^2 < \infty$ for all $n \geq 1$. If there exists an absolute constant $b_1^2 > 0$ such that $1 - \frac{V^2}{\|A\|^2} \geq b_1^2$ with $V^2 = \sum_{i=-\infty}^{\infty} a_{nii}^2$, then*

$$\sup_x |P(S_n/B_n \leq x) - \Phi(x)| \leq C \frac{\{\mathrm{Tr}(A^4)\}^{1/4}}{\|A\|}, \tag{A.16}$$

where $S_n = \sum_{i,j=-\infty}^{\infty} a_{nij}(\varepsilon_i\varepsilon_j - E[\varepsilon_i\varepsilon_j])$, $B_n^2 = 2(\|A\|^2 - V^2)\mu_2^2 + V^2(\mu_4 - \mu_2^2)$ with $\mu_j = E[|\varepsilon_1|^j]$, and A is the infinite matrix with a_{nij} as its (i, j)th element.

Lemma A.6. *Let $K(x)$ be a non-negative symmetric integrable function satisfying $K(x) = O[(1 + |x|^{1-\beta})^{-1}]$, where $0 < \beta \leq \alpha - 1/2$ and $1/2 < \alpha < 1$. Then*

$$\Delta_0 \equiv \int_0^\infty x^{-\alpha} K(x)\, \mathrm{d}x < \infty, \tag{A.17}$$

$$A_\alpha \equiv \int_0^\infty \int_0^\infty \int_0^\infty x^{-\alpha} y^{-\alpha} [I_1(x, y, w) + I_2(x, y, w)]\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}w < \infty, \tag{A.18}$$

and as $h \to 0$,

$$\Delta_1 \equiv \int_0^{1/h} \int_0^{1/h} \int_x^{1/h} x^{-\alpha} y^{-\alpha} \max\{w, y\}[I_1(x, y, w) + I_2(x, y, w)]\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}w$$
$$= o(1/h), \tag{A.19}$$

where $I_1(x, y, w) = K(w)K(x + y - w)$ and $I_2(x, y, w) = K(w - x)K(w - y)$.

A.2. Proofs of Theorems 2.3 and 2.4

Proof of Theorem 2.3. We may write

$$\sum_{1 \leq i \neq j \leq n} e_i e_j \left[K\left(\frac{(X_i - X_j)}{h} \right) \right] = \widetilde{Q}_{2n}^{(1)} + \widetilde{Q}_{2n}^{(2)} + \widetilde{Q}_{2n}^{(3)},$$

where

$$\widetilde{Q}_{2n}^{(1)} = \sum_{1 \leq i \neq j \leq n} e_i e_j E\left[K\left(\frac{(X_i - X_j)}{h} \right) \right], \tag{A.20}$$

$$\widetilde{Q}_{2n}^{(2)} = \sum_{1 \le i \ne j \le n} e_i e_j [g_{1n}(X_i) + g_{1n}(X_j)], \tag{A.21}$$

$$\widetilde{Q}_{2n}^{(3)} = \sum_{1 \le i \ne j \le n} e_i e_j g_{2n}(X_i, X_j), \tag{A.22}$$

where $g_{1n}(X_i)$ and $g_{2n}(X_i, X_j)$ are defined as in (A.6). Theorem 2.3 now follows readily if we prove the following: Whenever $h \to 0$,

$$(A_{2\alpha} n^{2-\alpha} h)^{-1} \widetilde{Q}_{2n}^{(1)} \to_D \chi^2(1), \tag{A.23}$$

$$\widetilde{Q}_{2n}^{(2)} = o_P\left(\max\{n^{2-\alpha}h, n\sqrt{h}\}\right), \tag{A.24}$$

and if in addition, $nh \to \infty$, then

$$\left(A_{1\alpha} n \sqrt{h}\right)^{-1} \widetilde{Q}_{2n}^{(3)} \to_D N(0, 1). \tag{A.25}$$

Actually, if $h \to 0$ and $\sqrt{h} n^{1-\alpha} \to \infty$, then $\widetilde{Q}_{2n}^{(2)} + \widetilde{Q}_{2n}^{(3)} = o_P(n^{2-\alpha}h)$ by virtue of (A.24) and (A.25). This and (A.23) yield Theorem 2.3(ii). Similarly, if $\sqrt{h} n^{1-\alpha} \to 0$ and $nh \to \infty$, then $\widetilde{Q}_{2n}^{(1)} + \widetilde{Q}_{2n}^{(2)} = o_P(n\sqrt{h})$ by virtue of (A.23) and (A.24). This and (A.25), yield Theorem 2.3(i).

We now prove (A.23)–(A.25), starting with (A.23). By (A.7),

$$(A_{2\alpha} n^{2-\alpha} h)^{-1} \widetilde{Q}_{2n}^{(1)} = \left(1 + o_P(1)\right) \left[\left(\frac{1}{d_n} \sum_{j=1}^{n} e_j\right)^2 - \frac{1}{d_n^2} \sum_{j=1}^{n} e_j^2 \right],$$

where $d_n^2 = \frac{2\eta}{(1-\alpha)(2-\alpha)} n^{2-\alpha}$. It can be readily seen that $\frac{1}{d_n^2} \sum_{j=1}^{n} e_j^2 \to 0$ a.s. by the stationary ergodic theorem. This, together with (A.2) and the continuous mapping theorem, yield that (A.23) will follow if we prove

$$\frac{1}{d_n} \sum_{j=1}^{n} e_j \to_D N(0, 1), \tag{A.26}$$

which follows from Lemma 1 of Robinson [43]. This proves (A.23).

We next prove (A.24). By (A.10), independence of e_i and X_i, and (A.3),

$$E\left(\widetilde{Q}_{2n}^{(2)}\right)^2 = E\left(\sum_{i=1}^{n} g_{1n}(X_i) e_i \sum_{1 \le j \ne i \le n} e_j \right)^2$$

$$= [d_1 h^2 + o(h^2)] \left(\sum_{1 \le i \ne j \le n} E[e_i^2 e_j^2] + \sum_{1 \le i \ne k \ne j \le n} E[e_i^2 e_j e_k] \right)$$

$$\le C h^2 \left[n^2 E e_1^4 + n \sum_{i,j=1}^{n} \gamma(i - j) \right] \le C h^2 (n^2 + n^{3-\alpha}).$$

Thus, equation (A.24) follows immediately from the Markov inequality.

Finally, we prove (A.25). Write

$$B_n^2 = \sum_{1 \le i < j \le n} e_i^2 e_j^2 E[g_{2n}^2(X_i, X_j)] = \frac{1}{2} E[g_{2n}^2(X_1, X_2)] \cdot \left[\left(\sum_{i=1}^n e_i^2 \right)^2 - \sum_{i=1}^n e_i^4 \right].$$

By (A.8) and the stationary ergodic theorem, which yields that $\frac{1}{n} \sum_{i=1}^n e_i^2 \to E[e_1^2] = \gamma(0)$ a.s. and $\frac{1}{n} \sum_{i=1}^n e_i^4 \to E[e_1^4] < \infty$ a.s., it can be readily seen that as $n \to \infty$

$$4A_{1\alpha}^{-2} n^{-2} h^{-1} B_n^2 \to 1 \qquad \text{a.s.}, \tag{A.27}$$

where $A_{1\alpha}$ is defined as in Theorem 2.3. Thus, to prove (A.25), it suffices to show that

$$\widetilde{Q}_{2n}^{(3)} / (2B_n) \to_D N(0, 1). \tag{A.28}$$

Lemma A.4 is used to establish (A.28). In fact, noting that

$$\widetilde{Q}_{2n}^{(3)} = 2 \sum_{1 \le i < j \le n} e_i e_j g_{2n}(X_i, X_j)$$

and $E[g_{2n}(X_i, X_j)|e_i] = 0$ for all $i \ne j$, it follows from the independence of e_i and X_i, Lemma A.4, (A.9) and (A.11) that

$$\sup_x \left| P\left(\widetilde{Q}_{2n}^{(3)} / 2B_n \le x | e_1, \dots, e_n \right) - \Phi(x) \right| \le A B_n^{-4/5} (c_4 h A_{1n} + d_2 h^3 A_{2n})^{1/5}, \tag{A.29}$$

where A is an absolute constant, c_4 and d_2 are defined as in (A.9) and (A.11), and

$$A_{1n} = \sum_{i=2}^n \left(\sum_{j=1}^{i-1} (e_i e_j)^2 \right)^2 \le \sum_{i=2}^n e_i^4 \left(\sum_{j=1}^n e_j^2 \right)^2,$$

$$A_{2n} = \sum_{i=2}^{n-1} \sum_{j=i+1}^n \left(\sum_{k=1}^{i-1} e_i e_j e_k^2 \right)^2 \le \left(\sum_{1 \le i < j \le n} e_i^2 e_j^2 \right)^2.$$

Again by the stationary ergodic theorem, for n large enough,

$$\frac{1}{n^3} A_{1n} \le 2E[e_1^4] \cdot (E[e_1^2])^2 \qquad \text{a.s.} \quad \text{and} \quad \frac{1}{n^4} A_{2n} \le 2(E[e_1^2])^4 \qquad \text{a.s.}$$

This, together with (A.27) and (A.29), implies that for n large enough,

$$\sup_x \left| P\left(\widetilde{Q}_{2n}^{(3)} / (2B_n) \le x | e_1, \dots, e_n \right) - \Phi(x) \right| \le C \left(\frac{1}{nh} + h \right)^{1/5} \qquad \text{a.s.}$$

Now if $h \to 0$ and $nh \to \infty$, then

$$\lim_{n\to\infty} \sup_x \left| P\left(\widetilde{Q}_{2n}^{(3)}/(2B_n) \leq x\right) - \Phi(x)\right|$$

$$\leq E\left[\lim_{n\to\infty} \sup_x \left| P\left(\widetilde{Q}_{2n}^{(3)}/(2B_n) \leq x | e_1, \ldots, e_n\right) - \Phi(x)\right|\right] = 0.$$

This proves (A.28), and also completes the proof of Theorem 2.3. □

Proof of Theorem 2.4. Let

$$\widetilde{Q}_{1n} = K(0) \sum_{i=1}^n e_i^2 \quad \text{and} \quad Q_{1n} = \sum_{i=1}^n \sum_{j=1, j\neq i}^n e_i e_j b_n(i, j).$$

We have that

$$\widetilde{Q}_n \equiv Q_{1n} - E[Q_{1n}] + \widetilde{Q}_{1n} - E[\widetilde{Q}_{1n}] = \sum_{i,j=1}^n \left(e_i e_j - \gamma(i-j)\right) b_n(i, j)$$

$$= \sum_{k,l=-\infty}^\infty a_{nkl}(\eta_k \eta_l - E[\eta_k \eta_l]),$$

(A.30)

where $a_{nkl} = \sum_{i,j=1}^n \psi_{i-k} \psi_{j-l} b_n(i, j)$, based on the fact that $E[\eta_k \eta_l] = 0$ for $k \neq l$, $E[\eta_k^2] = 1$, and

$$\sum_{k=-\infty}^\infty a_{nkk} = \sum_{i,j=1}^n K\left(\frac{i-j}{nh}\right) \sum_{k=-\infty}^\infty \psi_{i-k} \psi_{j-k} = \sum_{i,j=1}^n \gamma(i-j) b_n(i, j).$$

By Lemma A.5, to prove Theorem 2.4, it suffices to show that as $n \to \infty$,

$$\frac{(\widetilde{Q}_{1n} - E\widetilde{Q}_{1n})}{\tau_n} \to_P 0,$$

(A.31)

$$2\|A\|^2 \equiv 2 \sum_{k,l=-\infty}^\infty a_{nkl}^2 \sim A_0^2 \tau_n^2,$$

(A.32)

$$V^2 = \sum_{k=-\infty}^\infty a_{nkk}^2 = o(\tau_n^2),$$

(A.33)

$$\text{Tr}(A^4) = o(\tau_n^4),$$

(A.34)

where $\tau_n = n^{2-\alpha} h^{3/2-\alpha}$, and $A_0^2 = 8\eta^2 A_\alpha$ with A_α as defined in (2.3). Indeed, by virtue of (A.32)–(A.34), it follows from Lemma A.5 that

$$(\widetilde{Q}_n - E[\widetilde{Q}_n])/\tau_n \to_D A_0 N(0, 1).$$

This, together with (A.31), yields Theorem 2.4.

In what follows, we give the proofs of (A.31)–(A.34). We start with (A.31). Recall that $1/2 < \alpha < 1$ and $\gamma(k) \sim \eta|k|^{-\alpha}$. By virtue of (A.2), it can be readily seen that

$$E[(\widetilde{Q}_{1n} - E\widetilde{Q}_{1n})^2] = K^2(0)E\left[\sum_{k=1}^n (e_k^2 - Ee_k^2)\right]^2$$

$$\leq C \sum_{j,k=1}^n \gamma^2(j-k) \leq C \sum_{k=1}^n (n-k)\gamma^2(k) \leq Cn.$$

This, together with the Markov inequality and $nh \to \infty$, yields (A.31).

We next prove (A.32). We have

$$\|A\|^2 = \sum_{k,l=-\infty}^{\infty} \left(\sum_{i,j=1}^n \psi_{i-k}\psi_{j-l}K\left(\frac{i-j}{nh}\right)\right)^2$$

$$= \sum_{i,j,s,t=1}^n \sum_{k,l=-\infty}^{\infty} \psi_{i-k}\psi_{j-l}\psi_{s-k}\psi_{t-l}K\left(\frac{i-j}{nh}\right)K\left(\frac{s-t}{nh}\right)$$

$$= \sum_{i,j,s,t=1}^n K\left(\frac{i-j}{nh}\right)K\left(\frac{s-t}{nh}\right)\gamma(i-s)\gamma(t-j).$$

Write $f_n(x, y; z, w) = K(\frac{x-z}{nh})K(\frac{y-w}{nh}) + K(\frac{y-z}{nh})K(\frac{x-w}{nh})$. Clearly, $f_n(\cdot, \cdot; \cdot, \cdot)$ has the following symmetry in its indexes:

$$f_n(x, y; z, w) = f_n(y, x; z, w) = f_n(x, y; w, z) = f_n(y, x; w, z).$$

Also $f_n(x, y; z, w) = f_n(z, w; x, y)$. Noting that for any function $g(x, y)$ and symmetric function $b(x)$,

$$\sum_{i,j=1}^n b(i-j)g(i,j) = b(0)\sum_{i=1}^n g(i,i) + \sum_{i=1}^{n-1}\sum_{j=1}^{n-i} b(i)[g(j, j+i) + g(j+i, j)],$$

some algebra shows that (noting that $\gamma(k) = \gamma(-k)$)

$$2\|A\|^2 = \sum_{i_1,i_2,j_1,j_2=1}^n \gamma(i_1-i_2)\gamma(j_1-j_2)f_n(i_1, i_2; j_i, j_2)$$

$$= \gamma^2(0)\sum_{i,j=1}^n f_n(i,i;j,j) + 4\gamma(0)\sum_{i=1}^n\sum_{j=1}^{n-1}\sum_{k=1}^{n-j}\gamma(i)f_n(i,i;k,j+k) \quad (A.35)$$

$$+ 4 \sum_{i_1=1}^{n-1} \sum_{i_2=1}^{n-i_1} \sum_{j_1=1}^{n-1} \sum_{j_2=1}^{n-j_1} \gamma(i_1)\gamma(j_1) f_n(i_2, i_1 + i_2; j_2, j_1 + j_2)$$

$$\equiv \Delta_{1n} + 4\Delta_{2n} + 4\Delta_{3n}.$$

Recalling that $K(x)$ is a probability density function and $\gamma(x) \sim \eta x^{-\alpha}$ for $0 < \alpha < 1$ and $x > 0$, we have that

$$\Delta_{3n} \sim \int_1^{n-1} \int_1^{n-1} \int_1^{n-x} \int_1^{n-y} \gamma(x)\gamma(y) f_n(z, x+z; w, y+w) \, dx \, dy \, dz \, dw. \qquad (A.36)$$

Write $g_n(x, y) = \int_1^{n-x} \int_1^{n-y} f_n(z, x+z; w, y+w) \, dz \, dw$. By Fubini's theorem,

$$g_n(x, y) = \int_1^{n-x} \int_{1-z}^{n-y-z} f_n(z, x+z; w+z, y+w+z) \, dw \, dz$$

$$= \int_{-(n-x-1)}^{n-y-1} \int_{\max\{1, 1-w\}}^{\min\{n-x, n-y-w\}} f_n(0, x; w, y+w) \, dz \, dw$$

$$= \int_0^{n-y-1} (n - 1 - \max\{x, y+w\}) f_n(0, x; w, y+w) \, dw$$

$$+ \int_0^{n-x-1} (n - 1 - \max\{x+w, y\}) f_n(0, x; -w, y-w) \, dw$$

$$= \int_y^{n-1} (n - 1 - \max\{x, w\}) f_n(0, x; w-y, w) \, dw$$

$$+ \int_x^{n-1} (n - 1 - \max\{w, y\}) f_n(0, x; x-w, x+y-w) \, dw.$$

Substituting this into (A.36), simple calculations show that if $h \to 0$ and $nh \to \infty$, then

$$\Delta_{3n} \sim \int_1^{n-1} \int_1^{n-1} \gamma(x)\gamma(y) g_n(x, y) \, dx \, dy$$

$$\sim 2 \int_1^{n-1} \int_1^{n-1} \int_x^{n-1} \gamma(x)\gamma(y)(n - 1 - \max\{w, y\})$$

$$\times \left[K\left(\frac{w}{nh}\right) K\left(\frac{x+y-w}{nh}\right) + K\left(\frac{w-x}{nh}\right) K\left(\frac{w-y}{nh}\right) \right] dx \, dy \, dw$$

$$\sim 2n(nh)^{3-2\alpha} \eta^2 \int_0^{1/h} \int_0^{1/h} \int_x^{1/h} x^{-\alpha} y^{-\alpha} (1 - h \max\{w, y\})$$

$$\times [K(w)K(x+y-w)$$

$$+ K(w-x)K(w-y)] dx \, dy \, dw \qquad (A.37)$$

$$\sim 2n(nh)^{3-2\alpha}\eta^2 \int_0^\infty \int_0^\infty \int_x^\infty x^{-\alpha}y^{-\alpha}$$
$$\times [K(w)K(x+y-w)+K(w-x)K(w-y)]\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}w$$
$$= 2n(nh)^{3-2\alpha}\eta^2 A_\alpha = n(nh)^{3-2\alpha}A_0^2/4,$$

based on the fact that $K(x)$ is symmetric, $A_\alpha^2 < \infty$, and (A.19).

By a similar argument, if $h \to 0$ and $nh \to \infty$, then

$$\Delta_{1n} + 4\Delta_{2n} = \mathrm{O}(n^{3-\alpha}h^2) = \mathrm{o}(\Delta_{3n}). \tag{A.38}$$

By virtue of (A.35), (A.37) and (A.38), we obtain the proof of (A.32).

Third, we prove (A.33). Let

$$h(i,j,s,t) = \sum_{k=-\infty}^\infty \psi_{i-k}\psi_{j-k}\psi_{s-k}\psi_{t-k} = \sum_{k=-\infty}^\infty \psi_k\psi_{j-i+k}\psi_{s-i+k}\psi_{t-i+k}.$$

By $\psi_j \geq 0$ and $K(x) \geq 0$, it can be readily seen that for any $j \geq 0$, s, and t,

$$\sum_{i=1}^n h(i,j+i,s,t) \leq \sum_{k=-\infty}^\infty \psi_k\psi_{j+k}\sum_{i=1}^n \psi_{s-i+k}\psi_{t-i+k} \leq \gamma(j)\gamma(t-s).$$

Therefore, as in (A.35)–(A.37), it follows from (A.17) that

$$V^2 = \sum_{k=-\infty}^\infty \left(\sum_{i,j=1}^n \psi_{i-k}\psi_{j-k}K\left(\frac{i-j}{nh}\right)\right)^2$$
$$= \sum_{i,j,s,t=1}^n K\left(\frac{i-j}{nh}\right)K\left(\frac{s-t}{nh}\right)h(i,j,s,t)$$
$$\leq K(0)\sum_{i,s,t=1}^n K\left(\frac{s-t}{nh}\right)h(i,i,s,t)$$
$$+ 2\sum_{j=1}^n K\left(\frac{j}{nh}\right)\sum_{i,s,t=1}^n K\left(\frac{s-t}{nh}\right)h(i,j+i,s,t) \tag{A.39}$$
$$\leq \left[K(0)\gamma(0) + 2\sum_{j=1}^n K\left(\frac{j}{nh}\right)\gamma(j)\right]\sum_{s,t=1}^n K\left(\frac{s-t}{nh}\right)\gamma(s-t)$$
$$\leq Cn\left(K(0)\gamma(0) + 2\int_1^n x^{-\alpha}K\left(\frac{x}{nh}\right)\mathrm{d}x\right)\int_1^n x^{-\alpha}K\left(\frac{x}{nh}\right)\mathrm{d}x$$
$$\leq Cn^{3-2\alpha}h^{2-2\alpha} = \mathrm{o}(\tau_n^2)$$

because $nh \to \infty$. This proves (A.33).

Finally, we prove (A.34). Tedious calculations show that

$$
\mathrm{Tr}(A^4) = \sum_{i,j,l,m=-\infty}^{\infty} a_{nij} a_{njl} a_{nlm} a_{nmi}
$$

$$
= \sum_{i,j,l,m=-\infty}^{\infty} \sum_{j_1,j_2,\ldots,j_7,j_8=1}^{n} \psi_{j_1-i} \psi_{j_2-j} \psi_{j_3-j} \psi_{j_4-l} \psi_{j_5-l} \psi_{j_6-m} \psi_{j_7-m} \psi_{j_8-i}
$$

$$
\times K\left(\frac{j_1 - j_2}{nh}\right) K\left(\frac{j_3 - j_4}{nh}\right) K\left(\frac{j_5 - j_6}{nh}\right) K\left(\frac{j_7 - j_8}{nh}\right)
$$

$$
= \sum_{j_1,j_2,\ldots,j_7,j_8=1}^{n} K\left(\frac{j_1 - j_2}{nh}\right) \gamma(j_2 - j_3) \cdots K\left(\frac{j_7 - j_8}{nh}\right) \gamma(j_8 - j_1).
$$

Recall that $K(x) = O[(1+|x|^{1-\beta})^{-1}]$. Similar to the proof of (A.32), it follows from Lemma A.2 that

$$
\mathrm{Tr}(A^4) \sim \int_1^n \int_1^n \cdots \int_1^n K\left(\frac{x_1 - x_2}{nh}\right) \gamma(x_2 - x_3) \cdots
$$

$$
\times K\left(\frac{x_7 - x_8}{nh}\right) \gamma(x_8 - x_1) \, dx_1 \, dx_2 \cdots dx_7 \, dx_8
$$

$$
\sim \eta^4 (nh)^{4(2-\alpha)} \int_0^{1/h} \int_0^{1/h} \cdots \int_0^{1/h} K(x_1 - x_2)|x_2 - x_3|^{-\alpha}
$$

$$
\times K(x_7 - x_8)|x_8 - x_1|^{-\alpha} \, dx_1 \, dx_2 \cdots dx_7 \, dx_8
$$

$$
= O(1)(nh)^{4(2-\alpha)} \int_0^{1/h} \int_0^{1/h} \cdots \int_0^{1/h} |x_1 - x_2|^{\beta-1}|x_2 - x_3|^{-\alpha} \cdots
$$

$$
\times |x_7 - x_8|^{\beta-1}|x_8 - x_1|^{-\alpha} \, dx_1 \, dx_2 \cdots dx_7 \, dx_8
$$

$$
= o(1)(nh)^{4(2-\alpha)}(1/h)^2 = o(1)\tau_n^4.
$$

This yields (A.34) and thus completes the proof of Theorem 2.4. $\qquad \square$

A.3. Proofs of Theorems 2.1 and 2.2

Proof of Theorem 2.1. Note that for any α and $\widehat{\alpha}$,

$$
\left| n^{(\alpha-\widetilde{\alpha})} - 1 \right| \leq |\alpha - \widetilde{\alpha}| \log n \cdot \exp\{|\alpha - \widetilde{\alpha}| \log n\}. \tag{A.40}
$$

By (2.7), Theorem 2.3, and Assumption 2.2, the proof of Theorem 2.1 follows if we prove

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{f}(X_t) = \frac{1}{n^2 h}\sum_{i,j=1}^{n} K\left(\frac{X_i - X_j}{h}\right) \to_P \int_{-\infty}^{\infty} f^2(x)\,dx, \tag{A.41}$$

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_t^2 = \frac{1}{n}\sum_{t=1}^{n}\left(e_t + m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t)\right)^2 \to_P \gamma(0), \tag{A.42}$$

and under the relevant conditions of Theorem 2.1,

$$2R_{1n}(h) + R_{2n}(h) = o_P(\sigma_{in}(h)), \qquad i = 1, 2, \tag{A.43}$$

where $R_{1n}(h)$ and $R_{2n}(h)$ are defined as in (2.7).

Recall that $E[K(\frac{X_1 - X_2}{h})] \sim h\int_{-\infty}^{\infty} f^2(x)\,dx$ by (A.7). The proof of (A.41) follows from standard methodologym and thus the details are omitted. By the stationary ergodic theorem, $\frac{1}{n}\sum_{t=1}^{n} e_t^2 \to \gamma(0)$, a.s. This implies that (A.42) will follow if we have

$$\frac{1}{n}\sum_{t=1}^{n}\left[2\varepsilon_t\left(m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t)\right) + \left(m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t)\right)^2\right] \to_P 0. \tag{A.44}$$

Because the proof of (A.44) is similar to that of (A.43), we only prove (A.43) what follows.

For $\forall \varepsilon > 0$, write $\Omega_n = \{\widetilde{\theta}: \|\widetilde{\theta} - \theta_0\| \leq \varepsilon n^{-\alpha/2}\}$, and let

$$J_1(s, t) = K\left(\frac{X_s - X_t}{h}\right)\frac{\partial m_\theta(X_s)}{\partial \theta}\bigg|_{\theta=\theta_0},$$

$$J_2(s, t) = K\left(\frac{X_s - X_t}{h}\right)\left\{m_{\theta_0}(X_s) - m_{\widetilde{\theta}}(X_s) + (\theta_0 - \widetilde{\theta})^\tau \frac{\partial m_\theta(X_s)}{\partial \theta}\bigg|_{\theta=\theta_0}\right\}$$

and

$$J_3(s, t) = K\left(\frac{X_s - X_t}{h}\right)\left(m_{\theta_0}(X_s) - m_{\widetilde{\theta}}(X_s)\right)\left(m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t)\right).$$

Using this notation, we have $R_{2n}(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} J_3(s, t)$ and

$$R_{1n}(h) = (\theta_0 - \widetilde{\theta})\sum_{t=1}^{n} e_t \sum_{s=1,\neq t}^{n} J_1(s, t) + \sum_{t=1}^{n} e_t \sum_{s=1,\neq t}^{n} J_2(s, t).$$

Recalling Assumptions 2.2, 2.4(i) and 2.5, it is readily apparent that

$$E|J_1(s, t)| \leq E\left[K\left(\frac{X_s - X_t}{h}\right)\left\|\frac{\partial m_\theta(X_s)}{\partial \theta}\bigg|_{\theta=\theta_0}\right\|\right]$$

$$\leq ChE\left\|\frac{\partial m_\theta(X_s)}{\partial \theta}\bigg|_{\theta=\theta_0}\right\| \leq C_1 h,$$

$$E[|J_1(s,t)J_1(s_1,t)|] \le E\left[K\left(\frac{X_s - X_t}{h} \right) K\left(\frac{X_{s_1} - X_t}{h} \right) \right.$$
$$\left. \times \left\| \frac{\partial m_\theta(X_s)}{\partial \theta}\Big|_{\theta=\theta_0} \right\| \left\| \frac{\partial m_\theta(X_{s_1})}{\partial \theta}\Big|_{\theta=\theta_0} \right\| \right]$$
$$\le Ch^2 E\left[\left\| \frac{\partial m_\theta(X_1)}{\partial \theta}\Big|_{\theta=\theta_0} \right\|^2 \right] \le C_1 h^2$$

for all different values of s, s_1, and t. These findings imply that for any $1 \le t \le n$,

$$E\left[\sum_{s=1,\ne t}^n J_1(s,t) \right]^2 \le C(nh + n^2h^2) \le 2C(nh)^2,$$

because $nh \to \infty$. Thus, by the independence of e_t and X_s,

$$E\left[\sum_{t=1}^n e_t \sum_{s=1,\ne t}^n J_1(s,t) \right]^2 \le C(nh)^2 \sum_{t_1,t_2=1}^n E[e_{t_1}e_{t_2}] \le C_1 n^{4-\alpha}h^2, \tag{A.45}$$

where we have used the fact that

$$\sum_{t_1,t_2=1}^n E[e_{t_1}e_{t_2}] = E\left(\sum_{t=1}^n e_t \right)^2 \le Cn^{2-\alpha},$$

as seen from (A.1). On the other hand, it follows from Taylor's expansion of $m_\theta(x)$ (with respect to θ) that under H_0, for all $s \ne s_1$, $s \ne t$ and $s_1 \ne t$ and for n large enough such that $\Omega_n \subseteq \Theta_0$,

$$E[|e_t||J_2(s,t)|I(\tilde{\theta} \in \Omega_n)]$$
$$\le C\varepsilon n^{-\alpha} E|e_t|E\left[K\left(\frac{X_s - X_t}{h} \right) \left\| \frac{\partial^2 m_\theta(X_s)}{\partial \theta^2}\Big|_{\theta=\theta_0} \right\| \right] \tag{A.46}$$
$$\le C_1 \varepsilon n^{-\alpha} h E\left[\left\| \frac{\partial^2 m_\theta(X_s)}{\partial \theta^2}\Big|_{\theta=\theta_0} \right\|^2 \right] \le C_2 \varepsilon n^{-\alpha} h.$$

It follows from (A.45) and (A.46) that

$$E[|R_{1n}(h)|I(\tilde{\theta} \in \Omega_n)] \le \varepsilon n^{-\alpha/2} E\left| \sum_{t=1}^n e_t \sum_{s=1,\ne t}^n J_1(s,t) \right|$$
$$+ \sum_{t=1}^n \sum_{s=1,\ne t}^n E[|e_t||J_2(s,t)I(\tilde{\theta} \in \Omega_n)] \tag{A.47}$$
$$\le C\varepsilon n^{2-\alpha} h.$$

This, together with Markov's inequality, yields the following results for $\forall \varepsilon > 0$ and n sufficiently large: (i) If $nh \to \infty$ and $n^{2(1-\alpha)}h \to 0$, then

$$
\begin{aligned}
P\left(|R_{1n}(h)| \geq \varepsilon^{1/2}\sigma_{1n}\right) &\leq P(\|\widetilde{\theta} - \theta_0\| > \varepsilon n^{-\alpha/2}) \\
&\quad + C\varepsilon^{-1/2}(n^2 h)^{-1/2}E[|R_{1n}(h)|I(\widetilde{\theta} \in \Omega_n)] \\
&\leq P(\|\widetilde{\theta} - \theta_0\| > \varepsilon n^{-\alpha/2}) + Cn^{1-\alpha}h^{1/2}\varepsilon^{1/2} \leq C_1 \varepsilon.
\end{aligned} \tag{A.48}
$$

(ii) If $h \to 0$ and $n^{2(1-\alpha)}h \to \infty$, then

$$
\begin{aligned}
P\left(|R_{1n}(h)| \geq \varepsilon^{1/2}\sigma_{2n}\right) &\leq P(\|\widetilde{\theta} - \theta_0\| > \varepsilon n^{-\alpha/2}) \\
&\quad + C\varepsilon^{-1/2}(n^{4-2\alpha}h^2)^{-1/2}E[|R_{1n}(h)|I(\widetilde{\theta} \in \Omega_n)] \\
&\leq P(\|\widetilde{\theta} - \theta_0\| > \varepsilon n^{-\alpha/2}) + C\varepsilon^{1/2} \leq C_1 \varepsilon^{1/2}.
\end{aligned} \tag{A.49}
$$

The results (A.48) and (A.49) yield that $R_{1n}(h) = o(\sigma_{jn})$, $1 \leq j \leq 2$, under the corresponding conditions of Theorems 2.1. Similarly, by noting

$$
\begin{aligned}
E[|J_3(s,t)|I(\widetilde{\theta} \in \Omega_n)] &\leq C\varepsilon n^{-\alpha}E\left[K\left(\frac{X_s - X_t}{h}\right)\left\|\frac{\partial m_\theta(X_s)}{\partial\theta}\bigg|_{\theta=\theta_0}\right\|\left\|\frac{\partial m_\theta(X_t)}{\partial\theta}\bigg|_{\theta=\theta_0}\right\|\right] \\
&\leq C_1 \varepsilon n^{-\alpha}h E\left[\left\|\frac{\partial m_\theta(X_1)}{\partial\theta}\bigg|_{\theta=\theta_0}\right\|^2\right] \leq C_2 \varepsilon n^{-\alpha}h,
\end{aligned}
$$

we obtain that for $1 \leq j \leq 2$ and n sufficiently large,

$$
\begin{aligned}
P\left(|R_{2n}(h)| \geq \varepsilon^{1/2}\sigma_{jn}\right) &\leq P(\|\widetilde{\theta} - \theta_0\| > \varepsilon n^{-\alpha/2}) \\
&\quad + C\varepsilon^{1/2}(\sigma_{jn})^{-1}n^2 E[|J_3(1,2)|I(\widetilde{\theta} \in \Omega_n)] \leq C\varepsilon^{1/2},
\end{aligned} \tag{A.50}
$$

implying that $R_{2n}(h) = o(\sigma_{jn})$ holds for $1 \leq j \leq 2$. Combining these findings, we obtain (A.43), and thus also complete the proof of Theorem 2.1. □

Proof of Theorem 2.2. As in (2.7), under H_0, we can write

$$
\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} b_n(s,t)\left(\widehat{e}_s\widehat{e}_t - \widehat{\gamma}(s-t)\right) = M_n^*(h) + 2R_{1n}^*(h) + R_{2n}^*(h) + R_{3n}(h), \tag{A.51}
$$

where $M_n^*(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} b_n(s,t)[e_s e_t - \gamma(s-t)]$,

$$
\begin{aligned}
R_{1n}^*(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} b_n(s,t)e_s[m_{\theta_0}(t/n) - m_{\widetilde{\theta}}(t/n)] \\
&= (\widetilde{\theta} - \theta_0)\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} b_n(s,t)e_s\frac{\partial m_\theta(t/n)}{\partial\theta}\bigg|_{\theta=\theta_0}
\end{aligned}
$$

$$+ \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t) e_s \left[m_{\theta_0}(t/n) - m_{\widetilde{\theta}}(t/n) + (\widetilde{\theta} - \theta_0)^\tau \frac{\partial m_\theta(t/n)}{\partial \theta} \bigg|_{\theta=\theta_0} \right],$$

$$R_{2n}^*(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t)[m_{\theta_0}(s/n) - m_{\widetilde{\theta}}(s/n)][m_{\theta_0}(t/n) - m_{\widetilde{\theta}}(t/n)],$$

and by the symmetries of $K(x)$, $\gamma(k)$, and $\hat{\gamma}(k)$,

$$R_{3n}(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t)[\widehat{\gamma}(s-t) - \gamma(s-t)] = 2 \sum_{s=1}^{n-1}(n-s)K\left(\frac{s}{nh}\right)[\widehat{\gamma}(s) - \gamma(s)]. \quad \text{(A.52)}$$

By (A.51) and Theorem 2.4, Theorem 2.2 will follow if we prove that

$$2R_{1n}^*(h) + R_{2n}^*(h) = \mathrm{o}_P[\sigma_{3n}(h)], \quad \text{(A.53)}$$

$$R_{3n}(h) = \mathrm{o}_P[\sigma_{3n}(h)], \quad \text{(A.54)}$$

$$\frac{\widehat{\sigma}_{3n}(h)}{\sigma_{3n}(h)} \to_P 1. \quad \text{(A.55)}$$

The proofs of (A.53)–(A.55) are quite technical and thus are omitted here, but can be derived from the rest of the proof of Theorem 2.2 in Gao and Wang [20]. $\qquad \square$

Appendix B: Proofs of Theorems 4.1 and 4.2

Proof of Theorem 4.1. We first prove (4.1). In view of Theorem 2.2, it suffices to show that

$$\sup_{x \in R} \left| P^*\left(\widehat{T}_n^*(h) \leq x\right) - \Phi(x) \right| = \mathrm{o}_P(1). \quad \text{(B.1)}$$

As in (A.51), we may rewrite $\widehat{T}_n^*(h)$ as

$$\widehat{T}_n^*(h) = \frac{1}{\widehat{\sigma}_{3n}(h)}[M_n^{**}(h) + 2R_{1n}^{**}(h) + R_{2n}^{**}(h) + R_{3n}^{**}(h)], \quad \text{(B.2)}$$

where $M_n^{**}(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t)[e_s^* e_t^* - \gamma_{\widetilde{\lambda}}(s-t)]$,

$$R_{1n}^{**}(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t) e_s^* [m_{\widetilde{\theta}}(t/n) - m_{\widetilde{\theta}^*}(t/n)],$$

$$= (\widetilde{\theta} - \widetilde{\theta}^*) \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t) e_s^* \frac{\partial m_\theta(t/n)}{\partial \theta} \bigg|_{\theta=\widetilde{\theta}}$$

$$+ \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} b_n(s,t) e_s^* \left[m_{\widetilde{\theta}}(t/n) - m_{\widetilde{\theta}*}(t/n) + (\widetilde{\theta}^* - \widetilde{\theta})^\tau \frac{\partial m_\theta(t/n)}{\partial \theta} \bigg|_{\theta = \widetilde{\theta}} \right],$$

$$R_{2n}^{**}(h) = \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} b_n(s,t) \big(m_{\widetilde{\theta}*}(s/n) - m_{\widetilde{\theta}}(s/n) \big) \big(m_{\widetilde{\theta}*}(t/n) - m_{\widetilde{\theta}}(t/n) \big),$$

$$R_{3n}^{**}(h) = \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} b_n(s,t) [\widehat{\gamma}^*(s-t) - \gamma_{\widetilde{\lambda}}(s-t)],$$

where $\widehat{e}_s^* = Y_s^* - m_{\widetilde{\theta}*}(X_s)$ and

$$\widehat{\gamma}^*(k) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{i=1}^{n-|k|} \widehat{e}_i^* \widehat{e}_{i+|k|}^* & \text{for } |k| \leq (nh)^{1/3}, \\ \widetilde{\eta} |k|^{-\widetilde{\alpha}} & \text{for } (nh)^{1/3} < |k| \leq n-1. \end{cases}$$

The result (B.1) will follow if we prove that

$$I_{(\widetilde{\alpha} \in \Lambda_n)} \sup_{x \in R} \left| P\left(\frac{M_n^{**}(h)}{\widehat{\sigma}_{3n}(h)} \leq x \bigg| \mathcal{W}_n \right) - \Phi(x) \right| = \mathrm{o}_P(1), \tag{B.3}$$

and, for any $\varepsilon > 0$,

$$I_{\{\widetilde{\alpha} \in \Lambda_n, \widetilde{\theta} \in \Omega_{1n}\}} P\big(|2R_{1n}^{**}(h) + R_{2n}^{**}(h) + R_{3n}^{**}(h)| \geq \varepsilon \widehat{\sigma}_{3n}(h) | \mathcal{W}_n \big) = \mathrm{o}_P(1), \tag{B.4}$$

where $\mathcal{W}_n = (Y_1, \ldots, Y_n)$, $\Lambda_n = \{\widetilde{\alpha} : |\widetilde{\alpha} - \alpha| \leq C w_n^{-1/2}\}$ and $\Omega_{1n} = \{\widetilde{\theta} : \|\widetilde{\theta} - \theta_0\| \leq C_0 n^{-\alpha/2}\}$, with C_0 chosen so that $P(\|\widetilde{\theta} - \theta_0\| \geq C_0 n^{-\alpha/2}) \leq \varepsilon$. Indeed, recalling Assumptions 2.3 and 2.6(ii), we have $I(\widetilde{\alpha} \notin \Lambda_n \text{ or } \widetilde{\theta} \notin \Omega_{1n}) = \mathrm{o}_P(1)$. This, together with (B.3) and (B.4), proves the statement.

We next prove (B.3) and (B.4). We start with (B.3). As constructed in Section 4.1, $\{e_t^*\}$ is written as

$$e_t^* = \sum_{j=-\infty}^{\infty} \psi_j^* \eta_{t-j}^* \tag{B.5}$$

such that $E[e_t^* e_{t+k}^* | \mathcal{W}_n] = \gamma_{\widetilde{\lambda}}(k)$. Also note that $\hat{\alpha} \in \Lambda_n$ and $1/2 < \alpha < 1$ imply that there exists a $\delta_0 > 0$ such that $1/2 + \delta_0 < \widetilde{\alpha} < 1 - \delta_0$ whenever n is sufficiently large. Based on these facts, given $\mathcal{W}_n$, $M_n^{**}(h)$ has the same structure as that of $Q_{1n} - E[Q_{1n}]$ defined in the proof of Theorem 2.4. It follows from the same routine as in the proof of Theorem 2.4 that

$$I_{(\widetilde{\alpha} \in \Lambda_n)} \sup_{x \in R} \left| P\big(M_n^{**}(h) \leq \widehat{\tau}_n \widehat{A}_0 x | \mathcal{W}_n \big) - \Phi(x) \right| = \mathrm{o}_P(1), \tag{B.6}$$

where $\widehat{\tau}_n = n^{2-\tilde{\alpha}}h^{3/2-\tilde{\alpha}}$ and $\hat{A}_0^2 = 8\tilde{\eta}^2 A_{\tilde{\alpha}}$ with

$$A_{\tilde{\alpha}} = \int_0^\infty \int_0^\infty \int_0^\infty x^{-\tilde{\alpha}} y^{-\tilde{\alpha}} [K(z)K(x+y-z) + K(z-x)K(z-y)] \, dx \, dy \, dz.$$

Recall that Assumption 2.4(ii) implies that for any $u \in \mathbb{R}$,

$$\int_0^\infty K(w)K(w+u) \, dw \le C/(1+|u|^{1-\beta}). \tag{B.7}$$

It now can be readily seen that whenever $1/2 + \delta_0 < \tilde{\alpha} < 1 - \delta_0$,

$$A_{\tilde{\alpha}} = \int_{1/n}^n \int_{1/n}^n \int_{1/n}^n x^{-\tilde{\alpha}} y^{-\tilde{\alpha}} [K(z)K(x+y-z) + K(z-x)K(z-y)] \, dx \, dy \, dz + o_P(1)$$

$$= A_{\tilde{\alpha}}^*[1 + o_P(1)],$$

that is, $\widehat{\tau}_n \widehat{A}_0 = \widehat{\sigma}_{3n}(h)[1 + o_P(1)]$, where we have used the fact that $A_{\tilde{\alpha}}^* - A_\alpha = o_P(1)$ and $0 < A_\alpha < \infty$. Substituting this into (B.6), we get (B.3).

The proof of (B.4) follows from the same arguments as in the proofs of (A.53) and (A.54). Details are given in the rest of the proof of the first part of Theorem 4.1(i) of Gao and Wang [20].

We next prove the second part of Theorem 4.1(i). In view of Theorem 2.2, it suffices to show that

$$l_r^* - l_r = o_P(1). \tag{B.8}$$

In fact, recalling the definitions of l_r^* and l_r, it can be readily seen from (B.1) and Theorem 2.2 that $\Phi(l_r^*) - \Phi(l_r) = o_P(1)$, which implies (B.8), because $\Phi(x)$ is a bounded continuous function.

Finally, we prove Theorem 4.1(ii). In view of (B.8), it suffices to show that under H_1,

$$P\big(\widehat{L}_{3n}(h) \ge l_r\big) = 1, \tag{B.9}$$

with l_r satisfying $\Phi(l_r) = 1 - r + o(1)$ with $0 < r < 1$. To prove (B.9), as in (A.51), under H_1, we may rewrite $\widehat{L}_{3n}(h)$ as

$$\widehat{L}_{3n}(h) = \frac{1}{\widehat{\sigma}_{3n}(h)}[S_n(h) + 2Q_{1n}(h) + Q_{2n}(h) + R_{3n}(h)], \tag{B.10}$$

where

$$S_n(h) = \sum_{t=1}^n \sum_{s=1, \neq t}^n b_n(s,t)[\zeta_s \zeta_t - \gamma(s-t)],$$

$$Q_{1n}(h) = c_n \sum_{t=1}^n \sum_{s=1, \neq t}^n b_n(s,t) \zeta_s \Delta\left(\frac{t}{n}\right),$$

$$Q_{2n}(h) = c_n^2 \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} b_n(s,t) \Delta\left(\frac{s}{n}\right) \Delta\left(\frac{t}{n}\right),$$

in which $\zeta_t = e_t + [m_{\theta_1}(\frac{t}{n}) - m_{\widetilde{\theta}}(\frac{t}{n})]$ and $R_{3n}(h)$ is defined as in (A.51). Simple calculations show that

$$\begin{aligned}
Q_{2n}(h) &\sim c_n^2 \int_1^n \int_1^n K\left(\frac{x-y}{nh}\right) \Delta\left(\frac{x}{n}\right) \Delta\left(\frac{y}{n}\right) dx\, dy \\
&\sim A_0[1 + o(1)] c_n^2 (nh)^2 = A_0 d_n \sigma_{3n}(h),
\end{aligned} \tag{B.11}$$

where $d_n = c_n^2 n^\alpha h^{\alpha-1/2} \to \infty$ and $A_0 > 0$. Recalling Assumption 4.1(i), the same arguments as in the proof of Theorem 2.2 yield that

$$S_n(h)/\sigma_{3n}(h) \to_D N(0,1), \tag{B.12}$$

and $Q_{1n}(h) = O_P(c_n n^{-\alpha/2}(nh)^2) = o_P(Q_{2n}(h))$. These findings, together with (A.54) and (A.55) (i.e., $R_{3n} = o_P(\sigma_{3n}(h))$ and $\hat{\sigma}_{3n}(h)/\sigma_{3n}(h) \to_P 1$), imply that

$$\widehat{L}_{3n}(h) - A_0 d_n^2 \to_D N(0,1),$$

with $d_n^2 \to \infty$ and $A_0 > 0$. We now have (B.9), because l_r is finite for $0 < r < 1$. The proof of Theorem 4.1 is completed. $\qquad\square$

Proof of Theorem 4.2. Observe that under H_0,

$$\begin{aligned}
u_t \equiv \widehat{e}_t &= Y_t - m_{\widetilde{\theta}}(X_t) = e_t + m_{\theta_0}(X_t) - m_{\widetilde{\theta}}(X_t) \\
&= e_t + (\theta_0 - \widetilde{\theta})^\tau \left.\frac{\partial m_\theta(X_t)}{\partial \theta}\right|_{\theta=\theta_0} + o_P(\|\theta_0 - \widetilde{\theta}\|),
\end{aligned} \tag{B.13}$$

using Assumption 2.6.

By straightforward calculations, we then have, for n large enough,

$$\begin{aligned}
I_u(\omega) &= \frac{1}{2\pi n} \left|\sum_{s=1}^{n} u_s e^{is\omega}\right|^2 \\
&= \frac{1}{2\pi n} \left(\sum_{s=1}^{n} u_s^2 + 2\sum_{s=1}^{n-1}\sum_{t=1}^{n-s} \cos(\omega s) u_{s+t} u_s\right) \\
&= \frac{1}{2\pi n} \left(\sum_{s=1}^{n} e_s^2 + 2\sum_{s=1}^{n-1}\sum_{t=1}^{n-s} \cos(\omega s) e_{s+t} e_s\right) + o_P(1) \\
&= I_e(\omega) + o_P(1)
\end{aligned} \tag{B.14}$$

using Assumptions 2.1 and 2.6.

Thus, for n large enough,

$$
\begin{aligned}
\Gamma_u(\lambda) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log(\psi(\omega; \lambda)) + \frac{I_u(\omega)}{\psi(\omega; \lambda)} \right) d\omega \\
&= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log(\psi(\omega; \lambda)) + \frac{I_e(\omega)}{\psi(\omega; \lambda)} \right) d\omega + o_P(1) \\
&= \Gamma_e(\lambda) + o_P(1).
\end{aligned}
\tag{B.15}
$$

Therefore, by Assumption 4.3 and Theorem 1(ii) of Heyde and Gay [28], we have, as $n \to \infty$,

$$
\sqrt{n}(\widetilde{\lambda} - \lambda) \to_D N(0, \Sigma^{-1}(\lambda)),
\tag{B.16}
$$

which implies for n large enough,

$$
\frac{n^{2/5}}{\log(n)}(\widetilde{\lambda} - \lambda) \sim \frac{1}{n^{1/10}\log(n)} \sqrt{n}(\widetilde{\lambda} - \lambda) \to_P 0.
\tag{B.17}
$$

This completes the proof of Theorem 4.2. $\qquad\square$

Acknowledgements

References

[1] Anh, V. and Heyde, C., eds. (1999). Special Issue on Long-Range Dependence. *Journal of Statistical Planning & Inference* **80** 1.

[2] Anh, V., Wolff, R., Gao, J. and Tieng, Q. (1999). Local kernel regression with long-range dependent errors. *Aust. N. Z. J. Stat.* **41** 463–480.

[3] Avram, F. (1988). On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields* **79** 37–45. MR0952991

[4] Baillie, R. and King, M.L., eds. (1996). Special Issue of Journal of Econometrics. *Annals of Econometrics* **73** 1.

[5] Beran, J. (1994). *Statistics for Long-memory Processes. Monographs on Statistics and Applied Probability* **61**. New York: Chapman & Hall. MR1304490

[6] Beran, J. and Ghosh, S. (1998). Root-n-consistent estimation in partial linear models with long-memory errors. *Scand. J. Statist.* **25** 345–357. MR1649038

[7] Bühlmann, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074

[8] Chen, S.X. and Gao, J. (2007). An adaptive empirical likelihood test for parametric time series regression models. *J. Econometrics* **141** 950–972. MR2394208

[9] Cheng, B. and Robinson, P.M. (1991). Density estimation in strongly dependent nonlinear time series. *Statist. Sinica* **1** 335–359. MR1130123

[10] Cheng, B. and Robinson, P.M. (1994). Semiparametric estimation from time series with long-range dependence. *J. Econometrics* **64** 335–353. MR1310526

[11] Csörgő, S. and Mielniczuk, J. (1999). Random-design regression under long-range dependent errors. *Bernoulli* **5** 209–224. MR1681695

[12] Davydov, J.A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.* **15** 487–498.

[13] Fox, R. and Taqqu, M.S. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.* **13** 428–446. MR0781415

[14] Fox, R. and Taqqu, M.S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74** 213–240. MR0871252

[15] Franke, J., Kreiss, J.P. and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8** 1–37. MR1884156

[16] Gao, J. (2007). *Nonlinear Time Series*: *Semiparametric and Nonparametric Methods*. *Monographs on Statistics and Applied Probability* **108**. Boca Raton, FL: Chapman & Hall/CRC. MR2297190

[17] Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **103** 1584–1594. MR2504206

[18] Gao, J. and Hawthorne, K. (2006). Semiparametric estimation and testing of the trend of temperature series. *Econom. J.* **9** 332–355. MR2324973

[19] Gao, J., Lu, Z. and Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *Ann. Statist.* **34** 1395–1435. MR2278362

[20] Gao, J. and Wang, Q. (2010). Long-range dependent time series specification. Available at http://www.adelaide.edu.au/directory/jiti.gao.

[21] Gao, J., Wang, Q. and Yin, J. (2011). Specification testing in nonlinear time series with long-range dependence. *Econometric Theory* **27** 260–284. MR2782039

[22] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104. MR1061950

[23] Giraitis, L. and Taqqu, M.S. (1997). Limit theorems for bivariate Appell polynomials. I. Central limit theorems. *Probab. Theory Related Fields* **107** 359–381. MR1440137

[24] Guo, H. and Koul, H.L. (2007). Nonparametric regression with heteroscedastic long memory errors. *J. Statist. Plann. Inference* **137** 379–404. MR2298945

[25] Guo, H. and Koul, H.L. (2008). Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *Ann. Statist.* **36** 458–487.

[26] Hall, P., Horowitz, J.L. and Jing, B.Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282

[27] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

[28] Heyde, C.C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Process. Appl.* **45** 169–182. MR1204868

[29] Hidalgo, F.J. (1997). Nonparametric estimation with strongly dependent multivariate time series. *J. Time Series Anal.* **8** 95–122.

[30] Ho, H.C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

[31] Ho, H.C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636–1669. MR1487431

[32] Ho, H.C. and Hsing, T. (2003). A decomposition for generalized U-statistics of long-memory linear processes. In *Theory and Applications of Long-range Dependence* 143–155. Boston, MA: Birkhäuser. MR1956048

[33] Hsing, T. and Wu, W.B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.* **32** 1600–1631. MR2060311

[34] Kulik, R. and Wichelhaus, C. (2011). Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors. *Electron. J. Stat.* **5** 856–898. MR2831519

[35] Li, Q. and Racine, J.S. (2007). *Nonparametric Econometrics*: *Theory and Practice*. Princeton, NJ: Princeton Univ. Press. MR2283034

[36] Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87** 145–165. MR1648892

[37] Mielniczuk, J. and Wu, W.B. (2004). On random-design model with dependent errors. *Statist. Sinica* **14** 1105–1126. MR2126343

[38] Mikosch, T. and Starica, C. (2004). Nonstationarities in financial time series, the long–range dependence and the IGARCH effects. *Rev. Econom. Statist.* **86** 378–390.

[39] Nishiyama, Y. and Robinson, P. (2000). Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* **68** 931–980.

[40] Nishiyama, Y. and Robinson, P.M. (2005). The bootstrap and the Edgeworth correction for semiparametric averaged derivatives. *Econometrica* **73** 903–948. MR2135145

[41] Robinson, P.M. (1994). Time series with strong dependence. In *Advances in Econometrics*, *Sixth World Congress*, *Vol. I* (*Barcelona*, 1990) (C.A. Sims, ed.). *Econom. Soc. Monogr.* **23** 47–95. Cambridge: Cambridge Univ. Press. MR1278267

[42] Robinson, P.M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301

[43] Robinson, P.M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25** 2054–2083. MR1474083

[44] Robinson, P.M., ed. (2003). *Time Series with Long Memory*. *Advanced Texts in Econometrics*. Oxford: Oxford Univ. Press. MR2083220

[45] Robinson, P.M. and Hidalgo, F.J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77–104. MR1429918

[46] Sperlich, S., Tjøstheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* **18** 197–251. MR1891823

[47] Zhao, Z. and Wu, W.B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36** 1854–1878. MR2435458