

Nonparametric Cointegrating Regression with Endogeneity and Long Memory*

Qiyang Wang

School of Mathematics and Statistics

The University of Sydney

Peter C. B. Phillips

Yale University, University of Auckland

University of Southampton & Singapore Management University

January 8, 2014

Abstract

This paper explores nonparametric estimation, inference, and specification testing in a nonlinear cointegrating regression model where the structural equation errors are serially dependent and where the regressor is endogenous and may be driven by long memory innovations. Generalizing earlier results of Wang and Phillips (2009a,b), the conventional non-parametric local level kernel estimator is shown to be consistent and asymptotically (mixed) normal in these cases, thereby opening up inference by conventional nonparametric methods to a wide class of potentially nonlinear cointegrated relations. New results on the consistency of parametric estimates in nonlinear cointegrating regressions are provided, extending earlier research on parametric nonlinear regression and providing primitive conditions for parametric model testing. A model specification test is studied and confirmed to provide a valid mechanism for testing parametric specifications that is robust to endogeneity. But under long memory innovations the test is not pivotal, its convergence rate is parameter dependent, and its limit theory involves the local time of fractional Brownian motion. Simulation results show good performance for the nonparametric kernel estimates in cases of strong endogeneity and long memory, whereas the specification test is shown to be sensitive to the presence of long memory innovations, as predicted by asymptotic theory.

Key words and phrases: Cointegration, endogeneity, long memory, nonparametric estimation, nonstationarity, specification tests.

AMS Subject Classifications: Primary: 62M10, 62G07; Secondary 60F05

JEL Classification: C14, C22.

*Wang acknowledges research support from the Australian Research Council. Phillips acknowledges support from the NSF under Grant Nos. SES 09-56687 and 12-58258.

1 Introduction

Most empirical econometric research with time series data still uses linear in variables models, particularly those involving vector autoregressions, error correction systems, or reduced rank regressions. These specifications are convenient for practical work and package software has many standard routines for dealing with such systems, encouraging extensive usage of the methods. While common in applications, there has been growing recognition of the limitations of linear systems and the need for nonlinearities in specification that accommodate such effects as thresholds, breaks, or nonlinear behavioral responses. Such extensions can be formulated in parametric and nonparametric ways. While parametric formulations have now been treated in some generality following Park and Phillips (2001), allowing for unknown nonlinearity and nonstationarity in potentially cointegrated systems has presented deeper technical challenges in the development of asymptotic theories of estimation, inference, and specification. Progress has therefore been slow in comparison with the rapid earlier development of inference in linear nonstationary systems. However, some recent critical advances have been made that are now opening up this field to the practitioner.

In an earlier paper Wang and Phillips (WP)(2009a) discovered that standard tools of kernel regression could be employed to estimate and conduct valid asymptotic inference in certain nonparametric cointegrating regression models. In particular, the standard normal limit theory for nonparametric regression estimates in stationary systems applies also to self-normalized kernel regression estimates even when the explanatory variable is integrated. In view of the complexities of nonstationary regression limit theory in linear models, this simple finding was unexpected. The results in WP (2009a) applied to a bivariate cointegrating regression without contemporaneous endogeneity. Somewhat surprisingly, the same result was found to apply in similar models with contemporaneous endogeneity (WP, 2009b), highlighting a major difference with the stationary case where kernel regression is inconsistent and ill posed inverse problems arise in the use of nonparametric instrumental variable approaches that are designed to address the endogeneity in stationary systems. In the predictive regression context, Kasparis et al (2013) have recently shown that the standard normal limit theory continues to apply when the regressor has a local to unity or nonstationary long memory generating mechanism.

The present paper shows that these advantages of the nonparametric approach in the nonstationary case extend to an even wider class of models than that considered in

WP (2009a,b). In particular, the regressor may be driven by long memory as well as short memory innovations and the structural equation errors may have rather general weak dependence characteristics and correlation with the regressor. These results give the nonparametric regression limit theory for nonstationary cointegrated models much the same level of generality as that of linear cointegrated regression, at least for bivariate systems.

A typical non-linear cointegrating regression model has the following form

$$y_t = f(x_t) + u_t, \quad t = 1, 2, \dots, n, \quad (1.1)$$

where u_t is a zero mean equilibrium error, x_t is a non-stationary regressor and $f(\cdot)$ is an unknown real function on \mathbb{R} . In the presence of more explicit prior information, the regression function $f(\cdot)$ may be specified in parametric form as

$$f(x) = g(x, \theta_0), \quad (1.2)$$

where $g(\cdot, \theta)$ represents a parametric family of functions with unknown true parametric value $\theta_0 \in \Theta$, a compact set in \mathbb{R}^m for some finite m . The past decade has witnessed progress in the development of an asymptotic theory of estimation and inference for both the non-parametric model (1.1) and the parametric model (1.2). Technical difficulties in the limit theory for nonlinear covariance functionals of nonstationary and stationary time series has confined much of the asymptotic theory to the case of strict exogeneity where the regressor x_t is uncorrelated with the regression errors u_t at all leads and lags. Exogeneity is a natural starting point for a pure cointegrated system and provides some useful insight into the properties of various estimates of nonlinear long run linkages between the system variables. But the assumption is restrictive, especially in a cointegrated framework where the driver variables may be expected to be temporally and contemporaneously correlated. Exogeneity therefore delimits potential applications as well as removing a central technical difficulty in the development of the asymptotics.

Further progress in the field is inhibited by these limitations. One contribution of the present paper is to address these technical difficulties. A second contribution is to expand the framework to include long memory process drivers in the regressors, thereby allowing for a wider class of regressors and temporal dependence properties within the system. A third contribution is to provide asymptotic properties of a specification test for evaluating parametric regression hypotheses of the form (1.2) under endogeneity and long memory. A further contribution is to develop new consistency results for parametric nonlinear

cointegrating regression. These developments widen the range of practical application for kernel regression methods and specification tests with nonstationary data.

The paper is organized as follow. Section 2 considers non-parametric estimation in a nonlinear cointegrating regression model such as (1.1). Unlike previous work in the nonlinear case, the current paper allows the regressor x_t to be driven by long memory innovations and permits serial dependence in the error u_t and the innovations driving x_t for all leads and lags. A limit theory is developed for local level and local linear nonparametric estimates and their self normalized versions. A model specification test is developed in Section 3 for testing parametric hypotheses such as (1.2). The limit distribution of the statistic depends on the local time of the Brownian motion or fractional Brownian motion limit process associated with the (standardized) nonstationary regressor. This test is convenient to apply in practice under endogeneity and short memory innovations and has power against local alternatives to the null. Under long memory driver innovations for the regressor, the test statistic has a non-pivotal limit theory and parameter dependence in its convergence rate, which complicate inference. Section 4 explores the limit theory of parametric estimates in nonlinear cointegrating regressions, extending some of the earlier results in Park and Phillips (2001) and providing support for a high level convergence condition used in the asymptotic theory of the specification test. Proofs of the main results in the paper are given in Section 6, which also presents several useful propositions. Proofs of these propositions are given in Section 7.

Throughout the paper, we denote constants by C, C_1, C_2, \dots , which may differ at each appearance. We use the notation $\|x\| = \max_j |x_j|$ for vector $x = (x_j)$ and $\|A\| = \max_{i,j} |a_{ij}|$ for matrix $A = [(a_{ij})]$. Other notation is standard.

2 Nonparametric Estimation

The local level kernel estimate of $f(x)$ in model (1.1) is given by

$$\hat{f}(x) = \frac{\sum_{t=1}^n y_t K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)},$$

where $K_h(s) = \frac{1}{h}K(s/h)$, $K(x)$ is a nonnegative real function, and the bandwidth parameter $h \equiv h_n \rightarrow 0$ as $n \rightarrow \infty$. The limit behavior of $\hat{f}(x)$ has been investigated in past work in some special situations, notably where the error process u_t is a martingale difference sequence and there is no contemporaneous correlation between x_t and u_t . See, Karlsen, et al. (2007), Cai et al.(2009), WP (2009a, 2011) and Wang (2013), for instance.

The treatment in WP (2009b) notably allowed for endogeneity in (1.1) so that the equation error u_t might be cross-correlated with x_s over some finite time horizon for which $|t - s| \leq m_0$ for some finite m_0 .

This section has a similar goal to WP (2009b) in terms of accommodating endogeneity, but provides more general results with advantages for empirical applications. First, our model allows for the regressor x_t to be driven by long memory innovations. Second, unlike WP (2009b) where only finite memory cross-correlation was considered, our assumptions permit dependence between the error process u_t and the innovations driving x_t for all leads and lags. These relaxations of the conditions in WP (2009b) are particularly important in nonlinear cointegrated systems because finite time horizon dependence between the regressor and the equation error will often be restrictive in practice and it is seldom realistic in analyzing co-movement to insist that system variables or regressors be exactly $I(1)$ time series.

Throughout the section we let $\eta_i \equiv (\epsilon_i, \nu_i)'$, $i \in \mathbb{Z}$, be a sequence of iid random vectors with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}(\eta_0\eta_0') = \Sigma$, and $\mathbb{E}\|\eta_0\|^\alpha < \infty$ for some $\alpha > 2$. Assume $\mathbb{E}\epsilon_0^2 = 1$, and let the characteristic function $\varphi(t)$ of ϵ_0 satisfy the integrability condition $\int_{-\infty}^{\infty} (1 + |t|) |\varphi(t)| dt < \infty$, which assures smoothness in the corresponding density. We make use of the following assumptions in the asymptotic development.

Assumption 2.1 $x_k = \sum_{j=1}^k \xi_j$, where $\{\xi_j, j \geq 1\}$ is a linear process defined by $\xi_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k}$, with coefficients ϕ_k , $k \geq 0$, satisfying $\phi_0 \neq 0$ and one of the following conditions:

- C1. $\phi_k \sim k^{-\mu} \rho(k)$, where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞ .
- C2. $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

Assumption 2.2 $u_k = \sum_{j=0}^{\infty} \psi_j \eta_{k-j}$, where the coefficient vector $\psi_k = (\psi_{k1}, \psi_{k2})$ satisfies $\sum_{k=0}^{\infty} k^{1/4} (|\psi_{1k}| + |\psi_{2k}|) < \infty$ and $\sum_{k=0}^{\infty} \psi_k \neq 0$.

Assumption 2.3 $K(x)$ is a nonnegative bounded continuous function satisfying $\int K(x) dx = 1$ and $\int |\hat{K}(x)| dx < \infty$, where $\hat{K}(x) = \int e^{ixt} K(t) dt$.

Assumption 2.4 For given x , there exists a real positive function $f_1(s, x)$ and $\gamma \in (0, 1]$ such that, when δ sufficiently small, $|f(\delta y + x) - f(x)| \leq \delta^\gamma f_1(y, x)$ for all $y \in R$ and $\int_{-\infty}^{\infty} K(s) [f_1(s, x) + f_1^2(s, x)] ds < \infty$.

Assumption 2.1 allows for short (under **C2**) and long (under **C1**) memory innovations ξ_j driving the regressor x_t . In the long memory case, the parameter $\mu = 1 - d$ with

$d \in (0, \frac{1}{2})$ measures the hyperbolic decay rate in the coefficients ϕ_k of the linear process defining ξ_j . In the special case where $(1-L)^d \xi_t = \epsilon_t$, d is the fractional differencing parameter. Set $d_n^2 = \mathbb{E}x_n^2$, $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^\infty x^{-\mu}(x+1)^{-\mu} dx$ and denote by $W_\beta(t)$ a fractional Brownian motion with Hurst parameter $0 < \beta < 1$. It is well-known that the asymptotic form of d_n as $n \rightarrow \infty$ is given by

$$d_n^2 \sim \begin{cases} c_\mu n^{3-2\mu} \rho^2(n), & \text{under } \mathbf{C1}, \\ \phi^2 n, & \text{under } \mathbf{C2}, \end{cases} \quad (2.1)$$

and on $D[0, 1]$ the following weak convergence applies (e.g., Wang, Lin and Gulati, 2003)

$$x_{[nt]}/d_n \Rightarrow \psi(t) := \begin{cases} W_{3/2-\mu}(t), & \text{under } \mathbf{C1} \\ W(t), & \text{under } \mathbf{C2}. \end{cases}, \quad (2.2)$$

where $[nt]$ is the floor function and $W = W_{1/2}$ is Brownian motion. Furthermore, the limit process $\psi(t)$ has continuous local time process $L_\psi(t, s)$ with dual (time and space) parameters (t, s) in $[0, \infty) \times \mathbb{R}$. The local time process $L_G(t, s)$ of a stochastic process $G(x)$ is defined by (e.g., Geman and Horowitz, 1980, Theorem 22.1)

$$L_G(t, s) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t I\{|G(r) - s| \leq \epsilon\} dr. \quad (2.3)$$

These notations are used throughout the rest of the paper without further explanation.

Assumption 2.2 allows the equation error u_t to be cross correlated with the regressor x_s for all $s \leq t$, thereby inducing endogeneity and giving the structural model more natural temporal dependence properties than those used in WP (2009b). By a simple calculation

$$\mathbb{E}u_0^2 = \sum_{k=0}^{\infty} \psi_k \Sigma \psi_k', \quad \text{where } \Sigma = \begin{pmatrix} 1 & E\epsilon_0\nu_0 \\ E\epsilon_0\nu_0 & E\nu_0^2 \end{pmatrix}.$$

We may have $cov(u_t, x_t) \neq 0$ under Assumptions 2.1 and 2.2, which differs from much previous work where x_t is often assumed to be adapted to \mathcal{F}_{t-1} and (u_t, \mathcal{F}_t) forms a martingale difference sequence. In that case, $cov(u_t, x_t) = \mathbb{E}[x_t \mathbb{E}(u_t | \mathcal{F}_{t-1})] = 0$.

Assumptions 2.3 and 2.4 are the same as in WP (2009b), are quite weak, and are easily verified for various kernels $K(x)$ and regression functions $f(x)$. Typical examples of $K(x)$ and $f(x)$ include the normal kernel, kernels with compact support for K , and functions $f(x) = |x|^\beta$ or $f(x) = 1/(1 + |x|^\beta)$ for some $\beta > 0$.

The following is the main result on local level kernel estimation of the unknown regression function in (1.1).

THEOREM 2.1. *Under Assumptions 2.1–2.4, for any h satisfying $nh/d_n \rightarrow \infty$ and $nh^{1+2\gamma}/d_n \rightarrow 0$, we have*

$$\left(\frac{nh}{d_n}\right)^{1/2} (\hat{f}(x) - f(x)) \rightarrow_D \tau L_\psi^{-1/2}(1, 0) N, \quad (2.4)$$

for any fixed x , where $\tau^2 = \mathbb{E}u_0^2 \int_{-\infty}^{\infty} K^2(s) ds$ and N is a standard normal variate independent of $L_\psi(1, 0)$. We also have the following self-normalized limit form

$$\left(h \sum_{t=1}^n K_h(x_t - x)\right)^{1/2} (\hat{f}(x) - f(x)) \rightarrow_D \tau N. \quad (2.5)$$

Under (C1) when $\rho(n)$ is constant, (2.1) is $d_n^2 \sim cn^{1+2d}$ for some constant c and the condition $nh/d_n \rightarrow \infty$ on the effective sample size required for consistency then reduces to $n^{\frac{1}{2}-d}h \rightarrow \infty$ or $\sqrt{nh} \rightarrow \infty$ when $d = 0$ as in WP (2009b). So the effective sample size falls as d increases. It follows that larger $d \in (0, \frac{1}{2})$ requires a larger bandwidth h to ensure that the effective sample size diverges. An intuitive explanation is that a fractionally integrated $I(1+d)$ series is smoother than an $I(1)$ series which correspondingly reduces the local signal inherent in the nonparametric regression signal $\sum_{t=1}^n K(\frac{x_t-x}{h})$. A larger bandwidth compensates for this reduction in the signal. Importantly, the self-normalized limit (2.5) is pivotal upon estimation of τ^2 and well-suited to inference and confidence interval construction.

As in WP (2011) (see also Wang (2013)), an explicit bias term may be incorporated into the limit theory (2.4) and (2.5) if we impose stronger smoothness conditions on f and K . Furthermore, the Nadaraya-Watson estimator $\hat{f}(x)$ has the same limit distribution (to the second order including bias) as the local linear nonparametric estimator (e.g., Fan and Gijbels, 1996), defined by

$$\hat{f}^L(x) = \sum_{i=1}^n w_i Y_i / \sum_{i=1}^n w_i, \quad w_i = K_h(x_i - x) \{V_{n,2} - (X_i - x)V_{n,1}\},$$

where $V_{n,j} = \sum_{i=1}^n K_h(x_i - x)(x_i - x)^j$. Explicitly, we have the following theorem.

THEOREM 2.2. *Suppose Assumptions 2.1–2.2 hold. Further assume that, for some $p \geq 2$,*

(i) $K(x)$ satisfies $\int K(y) dy = 1$,

$$\int y^p K(y) dy \neq 0, \quad \int y^i K(y) dy = 0, \quad i = 1, 2, \dots, p-1;$$

(ii) $K(x)$ has compact support and is twice continuously differentiable on \mathbb{R} ;

(iii) for given fixed x , $f(x)$ has a continuous $p + 1$ derivative in a small neighborhood of x .

Then, for any h satisfying $nh/d_n \rightarrow \infty$ and $nh^{1+2(p+1)}/d_n \rightarrow 0$, we have

$$\left(\frac{nh}{d_n}\right)^{1/2} \left[\hat{f}(x) - f(x) - \frac{h^p f^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y) dy \right] \rightarrow_D \tau N L_{\psi}^{-1/2}(1, 0), \quad (2.6)$$

and

$$\left(h \sum_{t=1}^n K_h(x_t - x)\right)^{1/2} \left[\hat{f}(x) - f(x) - \frac{h^p f^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y) dy \right] \rightarrow_D \tau N. \quad (2.7)$$

where the notation follows Theorem 2.1. Furthermore, both results (2.6) and (2.7) (with $p = 2$) hold if we replace $\hat{f}(x)$ by $\hat{f}^L(x)$.

This finding provides further evidence that, under pointwise estimation, the bias reducing advantage of the local linear nonparametric estimator is lost when x_t is nonstationary, a phenomenon first discussed in WP (2011). In contrast to pointwise estimation, the local linear non-parametric estimator does have superior performance characteristics to the Nadaraya-Watson estimator in terms of uniform asymptotics over wide domains (Chan and Wang, 2013a; Duffy, 2013).

3 Model Specification Testing

The preceding theory deals with nonparametric estimation of a nonlinear cointegrating regression under general conditions of an endogenous regressor. Nonparametric function estimation is often the first step in analyzing data when there is no prior information on functional form. As is apparent from Theorems 1 and 2, nonparametric estimation has the merit of simplicity in terms of both practical implementation and asymptotics. In comparison to parametric counterparts (e.g., Park and Phillips, 2001; Chang, et al, 2001; Chan and Wang, 2013b), nonparametric estimators typically deliver slow convergence rates. Parametric estimation can therefore be attractive in practical work, whilst allowing for some potential functional misspecification. The latter possibility makes it desirable to perform a test of parametric specification. This section considers a parametric model specification test that is suited to nonlinear cointegrating regression with an endogeneous regressor.

In view of the maintained model (1.1) interest typically focuses on testing a specific parametric null hypothesis such as

$$H_0 : f(x) = g(x, \theta_0), \quad \theta_0 \in \Omega_0,$$

for $x \in \mathbb{R}$, where $g(x, \theta)$ is a given real function indexed by a vector θ of unknown parameters which lie in the parameter space Ω_0 . To test H_0 , Gao, et al. (2009) and WP (2012) considered kernel-smoothed U statistics of the form

$$S_n = \sum_{s,t=1, s \neq t}^n \hat{u}_t \hat{u}_s K[(x_t - x_s)/h], \quad (3.1)$$

where $\hat{u}_t = y_t - f(x_t, \hat{\theta})$, $K(x)$ is a non-negative real kernel function, h is a bandwidth satisfying $h \equiv h_n \rightarrow 0$ as the sample size $n \rightarrow \infty$ and $\hat{\theta}$ is a parametric estimator of θ under the null H_0 that is consistent whenever $\theta \in \Omega_0$. The behavior of the kernel weights $K((x_t - x_s)/h)$ depends on the self intersection properties of x_t . The U statistic asymptotics for S_n involve some new limit theory, developed by WP (2012), that depend on the self intersection local time of a Gaussian process (i.e., the local time for which the process intersects itself). The involvement of the kernel weights $K((x_t - x_s)/h)$ in the U statistic make the asymptotics for S_n complex and difficult to extend to the case of an endogenous regressor.

The present paper uses instead of (3.1) a normalized version of the following statistic

$$T_n = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] [y_k - g(x_k, \hat{\theta}_n)] \right\}^2 \pi(x) dx,$$

where $\pi(x)$ is a positive integrable function. The statistic T_n is a modification of the test statistic discussed by Härdle and Mammen (1993) for the random sample case. The test was used in Gao et al. (2012) for a nonlinear cointegrating model with a martingale error structure and no endogeneity. We proceed to show that the statistic T_n is asymptotically valid in a nonlinear cointegration model with endogeneity, as was indicated in their simulation results. Moreover, with changes in the convergence rate and the limit distribution, we demonstrate that the statistic remains valid under long memory input shocks to the regressor. But in that case the limit theory and convergence rate of the test both depend on the memory parameter of regressor, which complicates practical implementation. The alternative of proceeding under the assumption of short memory innovations when there are long memory input shocks to the regressor leads to a conservative test with zero asymptotic size and substantial reductions in power.

To proceed, we make the following additional assumptions on $K(x)$ and $g(x, \theta)$ to develop asymptotics for T_n .

Assumption 3.1. $K(x)$ has compact support and $|K(x) - K(y)| \leq C|x - y|$ whenever $|x - y|$ is sufficiently small.

Assumption 3.2. (i) There exist $g_1(x)$ and $g_2(x)$ such that, for each $\theta, \theta_0 \in \Omega_0$,

$$|g(x, \theta) - g(x, \theta_0)| \leq C \|\theta - \theta_0\| g_1(x),$$

and for some $0 < \beta \leq 1$,

$$|g_1(x + y) - g_1(x)| \leq C |y|^\beta g_2(x),$$

whenever y is sufficiently small. (ii) $\int_{-\infty}^{\infty} [1 + g_1^2(x) + g_2^2(x)] \pi(x) dx < \infty$.

Assumption 3.3. Under H_0 , $\|\hat{\theta}_n - \theta_0\| = o[(d_n/nh)^{1/2}]$.

Assumption 3.2 covers a wide class of functionals $g(x, \theta)$ and weight functions $\pi(x)$, including $g(x, \theta) = (x + \theta)^2$, $\theta e^x / (1 + e^x)$, $\theta \log |x|$, $\theta |x|^\alpha$ (α is fixed) and $\theta_0 + \theta_1 |x| + \dots + \theta_k |x|^k$ when $\pi(x) = e^{-x^2/2}$ or $\pi(x)$ has compact support. At this level of generality for $g(x, \theta)$, the condition on $\pi(x)$ is close to being necessary. Assumption 3.1 is slightly stronger than is necessary, and can be weakened to include the normal kernel function if more restrictions are imposed on the weight function $\pi(x)$. Under the current model with endogeneity, the stated consistency rate condition on $\hat{\theta}_n$ required in Assumption 3.3 is not presently available in the literature. For completeness therefore, we investigate convergence results of this kind and provide primitive conditions to validate the assumption in Section 4 of the paper. In particular, since $h \rightarrow 0$, Theorems 4.2 and 4.3 below show that Assumption 3.3 is achievable under stronger smoothness conditions on $g(x, \theta)$.

We have the following main result.

THEOREM 3.1. Suppose Assumptions 2.1–2.2 and 3.1–3.3 hold. Then, under H_0 , we have

$$T_{n,d} := \frac{d_n}{nh} T_n \rightarrow_D \tau_0 L_\psi(1, 0), \quad (3.2)$$

for any h satisfying $nh^2 \log n / d_n \rightarrow 0$ and $n^{1-\delta_0} h / d_n \rightarrow \infty$, where

$$\tau_0 = \mathbb{E} u_0^2 \int_{-\infty}^{\infty} K^2(s) ds \int_{-\infty}^{\infty} \pi(x) dx \quad (3.3)$$

and δ_0 can be as small as required.

Remark 3.1. As in the estimation theory, the condition on the bandwidth h that $n^{1-\delta_0}h/d_n \rightarrow \infty$ is close to being necessary. Similar to the discussions in WP (2012), the further condition $nh^2 \log^{1/2} n/d_n \rightarrow 0$ is used mainly to offset the impact of the dependence between u_t and x_t . See the proof of Proposition 7.3. It seems difficult to relax this condition under the current model.

Remark 3.2. The error variance $\mathbb{E}u_0^2$ in (3.3) can be estimated by

$$\hat{\sigma}_n^2 = \frac{\sum_{t=1}^n [y_k - g(x_k, \hat{\theta}_n)]^2 K_h(x_t - x)}{\sum_{t=1}^n K_h(x_t - x)}, \quad (3.4)$$

based on a localized version of the usual residual sum of squares. Routine calculations confirm that $\hat{\sigma}_n^2 \rightarrow_p \mathbb{E}u_0^2$, provided Assumptions 2.1–2.2, 3.1, 3.2(i) and 3.3 hold.

Importantly, the scaling in the statistic (3.2) relies on expansion rate parameter $d_n^2 \sim c_\mu n^{3-2\mu} \rho^2(n)$, which in turn relies on the unknown value of μ . Even in the simple case where $\rho(n)$ is constant and $d_n \sim cn^{\frac{1}{2}+d}$ for some constant c , the required scaling depends on the (typically unknown) value of the long memory parameter $d = 1 - \mu$. If d were estimated nonparametrically by \hat{d} using narrow band methods (e.g., by the exact local Whittle procedure in Shimotsu and Phillips, 2005) with convergence rate \sqrt{m} where $\frac{\log n}{\sqrt{m}} + \frac{m}{n} \rightarrow 0$, then a standard derivation shows that in this case $\sqrt{m} (\hat{d}_n - d_n) / (d_n \log n) = O_p(1)$. It follows that

$$\begin{aligned} \hat{T}_{n,d} & : = \frac{\hat{d}_n}{nh} T_n = \frac{d_n}{nh} T_n + \frac{\sqrt{m} (\hat{d}_n - d_n)}{d_n \log n} \left(\frac{d_n}{nh} T_n \right) \frac{\log n}{\sqrt{m}} \\ & = T_{n,d} + O_p \left(\frac{\log n}{\sqrt{m}} \right) \rightarrow_D \tau_0 L_\psi(1, 0), \end{aligned}$$

giving the same limit distribution as (3.2). However, the limit distribution still depends on d via the local time $L_\psi(1, 0)$ of the (unknown) fractional Brownian motion process ψ and is therefore non-pivotal. The local time $L_\psi(1, 0)$ may itself be estimated by kernel methods in view of the asymptotic approximation

$$\frac{d_n}{nh} \sum_{t=1}^n K \left(\frac{x_t - x}{h} \right) \sim L_\psi(1, 0), \quad (3.5)$$

which holds for all fixed x . However, this estimate also depends on the unknown value of d . Moreover, the self-normalized statistic

$$\frac{T_n}{\sum_{t=1}^n K \left(\frac{x_t - x}{h} \right)} \rightarrow_p \tau_0$$

is unsuited for inference, requires centering and a further limit theory for the recentred statistic, which again depends on the unknown value of d . The statistic $T_{n,d}$ is therefore not generally well-suited for practical implementation in specification testing.

In the special case where $d = 0$, we have $d_n = \phi\sqrt{n}$ and $T_{n,0}$ is proportional to $\frac{1}{\sqrt{nh}}T_n$, the statistic studied in Gao et al (2012) under exogeneity. In this case the limit distribution is

$$T_{n,0} = \frac{\phi}{\sqrt{nh}}T_n \rightarrow_D \tau_0 L_W(1, 0) \quad (3.6)$$

and the limit theory holds under endogeneity. But when $d \in (0, \frac{1}{2})$ the statistic $\frac{1}{\sqrt{nh}}T_n$ leads to a conservative test. In particular, for $d_n^2 \sim c_\mu n^{1+2d} \rho^2(n)$ we have

$$\frac{1}{\sqrt{nh}}T_n \sim \frac{d_n}{nh}T_n \times \frac{1}{c_\mu^{1/2} n^d \rho(n)} \rightarrow_p 0, \quad (3.7)$$

so that the size of a test based on $\frac{1}{\sqrt{nh}}T_n$ tends to zero as $n \rightarrow \infty$ whenever $d \in (0, \frac{1}{2})$.

To investigate asymptotic power we consider the following local alternative models

$$H_1 : f(x) = g(x, \theta_0) + \rho_n m(x), \quad (3.8)$$

where $\theta_0 \in \Omega_0$, ρ_n is a sequence of constants measuring local deviations from the null and $m(x)$ is a real function. Local alternatives of the form (3.8) are commonly used in the theory of non-parametric inference involving stationary data. See, for instance, Horowitz and Spokoiny (2001). We impose the following smoothness conditions on $m(x)$ to aid the asymptotic development here.

Assumption 3.4. (i) *There exist $m_1(x)$ and $\gamma \in (0, 1]$ such that, for any y sufficiently small,*

$$|m(x+y) - m(x)| \leq C |y|^\gamma m_1(x).$$

(ii) $\int_{-\infty}^{\infty} [1 + m^2(x) + m_1^2(x)] \pi(x) dx < \infty$.

THEOREM 3.2. *Suppose Assumptions 2.1–2.2 and 3.1–3.4 hold. Then, under H_1 , we have*

$$\lim_{n \rightarrow \infty} P\left(\frac{d_n}{nh} T_n \geq t_0\right) = 1,$$

for any $t_0 > 0$, any $h \rightarrow 0$ satisfying $n^{1-\delta_0} h / d_n \rightarrow \infty$ where δ_0 can be as small as required, and any ρ_n satisfying $nh\rho_n^2/d_n \rightarrow \infty$.

As in Assumption 3.2, the conditions on $m(x)$ imposed by Assumption 3.4 seem weak and are satisfied by a large class of real functions. The theorem shows that the T_n test

has nontrivial power against local alternatives of the form (3.8) whenever $\rho_n \rightarrow 0$ at a rate that is slower than $[d_n/(nh)]^{1/2}$, as $nh/d_n \rightarrow \infty$. This is different from the stationary situation where a test generally has nontrivial power only if $\rho_n \rightarrow 0$ at a rate slower than $n^{-1/2}$. Moreover, unlike the test used in WP (2012), the rate condition here is only related to the bandwidth h , not to the magnitude of $m(x)$. The explanation is that the weight function $\pi(x)$ in the test offsets any impact of the magnitude of $m(x)$ under the alternative.

Theorem 3.2 shows that specification tests based on T_n are valid and have non-trivial local power for alternatives of the form (3.8) under endogeneity and in the short memory case ($d = 0$). In that case the test statistic has the simple computable form $\frac{1}{\sqrt{nh}} T_n$, as in Gao et al (2012), and the limit theory is given by (3.6) where ϕ can be estimated by standard HAC methods and τ_0 is estimated using (3.4). Alternative HAR sieve methods (Phillips, 2005; Sun, 2011; Chen et al., 2014) or fixed-b kernel methods (Kiefer and Vogelsang, 2005; Sun, 2014) may be used after some changes to the limit theory to address the random limit theory involved in the estimation of ϕ but these methods are not explored here.

The situation is more complex in the long memory case where T_n does not produce a pivotal test and there are practical difficulties in implementing the test based on $\frac{d_n}{nh} T_n$. If the statistic $\frac{1}{\sqrt{nh}} T_n$ is used when $d \in (0, \frac{1}{2})$ the test has zero size asymptotically and its power function under alternatives of the form (3.8) depend on ρ_n and d_n . In particular, we note that

$$\frac{1}{\sqrt{nh}} T_n = \frac{d_n}{nh} T_n \times \frac{\sqrt{n}}{d_n} \sim \frac{d_n}{nh} T_n \times \frac{1}{c_\mu^{1/2} n^d \rho(n)},$$

so that limit behavior under the alternative depends on the divergence rate of $\frac{d_n}{nh} T_n$ in relation to d_n . The divergence rate of the statistic $\frac{d_n}{nh} T_n$ therefore needs to exceed $n^d \rho(n)$ for a test based on $\frac{1}{\sqrt{nh}} T_n$ to be consistent. Using arguments similar to those in the proof of Theorem 3.2 we find that test consistency is attained provided $\frac{n^{5/3} h}{d_n^{7/3}} \rho_n^2 \rightarrow \infty$. For small values of d , non-trivial power is then possible under local alternatives. For example, if $d_n \sim cn^{\frac{1}{2}+d}$ and $d = \frac{1}{8}$, then the test is consistent provided $n^{\frac{5}{12}} h \rho_n^2 \rightarrow \infty$. Larger values of d typically require more distant alternatives to ensure consistency. For example, if $d = \frac{1}{4}$, then the test is consistent if $n^{-\frac{1}{12}} h \rho_n^2 \rightarrow \infty$ which requires $\rho_n \rightarrow \infty$.

4 Parametric Consistency

This section considers extremum estimation of the unknown parameters θ in model (1.2) by nonlinear least squares (NLS). We provide primitive conditions for the verification of consistent parametric estimation of θ as given in Assumption 3.3.

Let $Q_n(\theta) = \sum_{t=1}^n (y_t - g(x_t, \theta))^2$. The NLS estimator $\hat{\theta}_n$ of θ is defined as the minimizer of $Q_n(\theta)$ over $\theta \in \Theta$, viz.,

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} Q_n(\theta).$$

Let \dot{Q}_n and \ddot{Q}_n be the first and second derivatives of $Q_n(\theta)$, so that $\dot{Q}_n = \partial Q_n / \partial \theta$ and $\ddot{Q}_n = \partial^2 Q_n / \partial \theta \partial \theta'$. Similar definitions are used for \dot{g} and \ddot{g} . We assume these quantities exist whenever they are introduced. To develop asymptotics for $\hat{\theta}_n$ the following framework is a generalization of Theorem 3.1 in Chan and Wang (2013b) and draws on Wooldridge (1994) and Andrews and Sun (2004).

THEOREM 4.1. *There exist a sequence of constants $\{k_n, n \geq 1\}$ and a sequence of $m \times m$ nonrandom nonsingular matrices D_n satisfying $k_n \rightarrow \infty$ and $k_n \|D_n^{-1}\| \rightarrow 0$, as $n \rightarrow \infty$, such that the following conditions hold:*

- (i) $\sup_{\theta: \|D_n(\theta - \theta_0)\| \leq k_n} \|(D_n^{-1})' \sum_{t=1}^n [\dot{g}(x_t, \theta) \dot{g}(x_t, \theta)' - \dot{g}(x_t, \theta_0) \dot{g}(x_t, \theta_0)'] D_n^{-1}\| = o_P(\delta_n^{-2})$,
- (ii) $\sup_{\theta: \|D_n(\theta - \theta_0)\| \leq k_n} \|(D_n^{-1})' \sum_{t=1}^n \ddot{g}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)] D_n^{-1}\| = o_P(\delta_n^{-2})$,
- (iii) $\sup_{\theta: \|D_n(\theta - \theta_0)\| \leq k_n} \|(D_n^{-1})' \sum_{t=1}^n \ddot{g}(x_t, \theta) u_t D_n^{-1}\| = o_P(\delta_n^{-2})$,
- (iv) $Y_n := (D_n^{-1})' \sum_{t=1}^n \dot{g}(x_t, \theta_0) \dot{g}(x_t, \theta_0)' D_n^{-1} \rightarrow_D M$, where $M > 0$, a.s., and

$$Z_n := (D_n^{-1})' \sum_{t=1}^n \dot{g}(x_t, \theta_0) u_t = O_P(\delta_n), \quad (4.1)$$

where $1 \leq \delta_n \leq k_n^{1-\epsilon_0}$ for some $\epsilon_0 > 0$. Then, there exists a sequence of estimators $\{\hat{\theta}_n, n \geq 1\}$ satisfying $\dot{Q}_n(\hat{\theta}_n) = 0$ with probability that goes to one and

$$D_n(\hat{\theta}_n - \theta_0) = Y_n^{-1} Z_n + o_P(1). \quad (4.2)$$

If we replace (iv) by the following condition (iv)', then $D_n(\hat{\theta}_n - \theta_0) \rightarrow_D M^{-1} Z$ and in (i)–(iii) we may take $\delta_n = 1$.

(iv)' for any $\alpha'_i = (\alpha_{i1}, \dots, \alpha_{im}) \in \mathbb{R}^m, i = 1, 2, 3$,

$$(\alpha'_1 Y_n \alpha_2, \alpha'_3 Z_n) \rightarrow_D (\alpha'_1 M \alpha_2, \alpha'_3 Z),$$

where $M > 0$, a.s. and $P(Z < \infty) = 1$.

Remark 4.1. Theorem 4.1 was established in Chan and Wang (2013b) with $\delta_n = 1$. The current result weakens the restriction on Z_n , which in turn allows us to establish consistency of the estimator $\hat{\theta}_n$ in model (1.1) - (1.2) under a more general framework with an endogenous regressor.

We next investigate an application of Theorem 4.1 when the regressor x_t and equation error u_t are defined as in Assumptions 2.1 and 2.2. To do so we require certain smoothness conditions on $g(x, \theta)$. We start with the case where $g(x, \theta)$ is a bounded and integrable function.

Assumption 4.1. Let $p(x, \theta)$ be any of g , \dot{g}_i or \ddot{g}_{ij} , $1 \leq i, j \leq m$.

- (i) $p(x, \theta_0)$ is a bounded and integrable real function;
- (ii) There exists a bounded and integrable function $T_p : R \rightarrow R$ such that $|p(x, \theta) - p(x, \theta_0)| \leq C \|\theta - \theta_0\| T_p(x)$, for each $\theta, \theta_0 \in \Theta$;
- (iii) $\Sigma = \int_{-\infty}^{\infty} \dot{g}(s, \theta_0) \dot{g}(s, \theta_0)' ds > 0$ for each $\theta_0 \in \Theta$, where $\dot{g}(\cdot) = (\dot{g}_1(\cdot), \dots, \dot{g}_m(\cdot))'$.

THEOREM 4.2. Under Assumptions 2.1–2.2 and 4.1, we have

$$\|\hat{\theta}_n - \theta_0\| = \left(\frac{d_n}{n}\right)^{1/2} \begin{cases} O_P(1), & \text{under } \mathbf{C1}, \\ O_P(\log^{1/2} n), & \text{under } \mathbf{C2}. \end{cases} \quad (4.3)$$

The conditions in Theorem 4.2 hold for a wide range of integrable regression functions, including $g(x, \theta_1, \theta_2) = \theta_1 |x|^{\theta_2} I(x \in [a, b])$, where a and b are finite constants, the Gaussian function $g(x, \theta_1, \theta_2) = \theta_1 e^{-\theta_2 x^2}$, and the Laplacian function $g(x, \theta_1, \theta_2) = \theta_1 e^{-\theta_2 |x|}$. The term $\log^{1/2} n$ under **C2** in (4.3) can be eliminated if stronger restrictions are imposed on the relationship between the error process u_t and regressor x_t . Park and Phillips (2001) and Chan and Wang (2013b) provide results in this case. However it is difficult to remove this term under the present model where there is general endogeneity. Consequently, we have not been able to establish a general limit distribution theory for $\hat{\theta}_n$ without further conditions. But Theorem 4.2 is sufficient for the purpose of the present paper and we leave this remaining challenge in nonlinear nonstationary asymptotics under general conditions for future work.

Assumption 4.2. Let $p(x, \theta)$ be any of g , \dot{g}_i or \ddot{g}_{ij} , $1 \leq i, j \leq m$. There exists a positive real function $v_p(\lambda)$ which is bounded away from zero as $\lambda \rightarrow \infty$ and a constant $\beta \geq 0$ such that, for each $\theta, \theta_0 \in \Theta$,

(i) $|p(x, \theta) - p(x, \theta_0)| \leq C \|\theta - \theta_0\|^\alpha T_{1p}(x)$ for some $0 < \alpha \leq 1$, where $T_{1p}(\lambda x) \leq C v_p(\lambda) (1 + |x|^\beta)$;

(ii) $p(\lambda x, \theta_0) \leq C v_p(\lambda) (1 + |x|^\beta)$ and, for $p(x, \theta_0) = \dot{g}_i(x, \theta_0), \ddot{g}_{ij}(x, \theta_0), 1 \leq i, j \leq m$,

$$|p(\lambda x, \theta_0) - p(\lambda y, \theta_0)| \leq C v_p(\lambda) [|x - y| + R_{1p}(\lambda x) + R_{2p}(\lambda x)],$$

whenever x and y are in a compact set, where $R_{1p}(z)$ and $R_{2p}(z)$ are bounded and integrable functions;

(iii) $\dot{g}_i(\lambda x, \theta_0) = v_{\dot{g}_i}(\lambda) h_i(x, \theta_0) + R_i(\lambda, x, \theta_0)$ for $1 \leq i \leq m$, where $R_i(\lambda, x, \theta_0) = o[v_{\dot{g}_i}(\lambda) h_i(x, \theta_0)]$ as $|\lambda| \rightarrow \infty$, and $h_i(x, \theta_0)$ is a locally bounded function (i.e., bounded on any compact set) satisfying $\sum_{\delta} = \int_{|s| \leq \delta} h(s, \theta_0) h(s, \theta_0)' ds > 0$ for all $\delta > 0$, where $h(\cdot) = (h_1(\cdot), \dots, h_m(\cdot))'$;

(iv) $\sup_{1 \leq i, j \leq m} \left| \frac{v(d_n) \ddot{v}_{ij}(d_n)}{\dot{v}_i(d_n) \dot{v}_j(d_n)} \right| < \infty$, where $v(\lambda) = v_g(\lambda)$, $\dot{v}_i(\lambda) = v_{\dot{g}_i}(\lambda)$ and $\ddot{v}_{ij}(\lambda) = v_{\ddot{g}_{ij}}(\lambda)$.

THEOREM 4.3. Under Assumptions 2.1–2.2, 4.2 and $\sum_{k=0}^{\infty} k^{1/2} (|\psi_{1k}| + |\psi_{2k}|) < \infty$, we have

$$\|D_n(\hat{\theta}_n - \theta_0)\| = O_P(1),$$

where $D_n = \text{diag}(\sqrt{n} \dot{v}_1(d_n), \dots, \sqrt{n} \dot{v}_m(d_n))$.

Assumption 4.2 allows for asymptotically homogeneous functions. Typical examples include $g(x, \theta) = (x + \theta)^2$, $\theta e^x / (1 + e^x)$, $\theta \log |x|$, $\theta |x|^\alpha$ (α is fixed) and $\theta_0 + \theta_1 |x| + \dots + \theta_k |x|^k$. The class of functions satisfying Assumption 4.2 is similar to but wider than the H_0 -regular functions on Θ imposed in Theorem 4.2 of Park and Phillips (2001). For instance, since $g'(x, \theta) = 2(x + \theta)$ and $g''(x, \theta) = 2$, Assumption 4.2 applies to the function $g(x, \theta) = (x + \theta)^2$ [with $v_g(\lambda) = \lambda$, $v_{\dot{g}}(\lambda) = \lambda$, $v_{\ddot{g}}(\lambda) = 1$ and $h(x, \theta_0) = 2x$]; but Theorem 4.2 of Park and Phillips (2001) does not directly apply for this function (e.g., see Example 4.1 (c) of that paper). While allowing for this extra generality here in establishing consistency of $\hat{\theta}_n$, it is nonetheless difficult to establish an asymptotic distribution theory for $\hat{\theta}_n$ under the current model, as remarked above.

5 Simulations

We report the results of a small Monte Carlo (MC) experiment to explore finite sample performance of estimation and inference under endogeneity and long memory. The

simulations are complementary to those in PW (2009b) for kernel regression in structural nonstationary models and focus on the impact of a long memory component in the regressor innovations. The generating mechanism follows (1.1) and has the explicit form

$$\begin{aligned} y_t &= f(x_t) + \sigma u_t, \quad \Delta x_t = \xi_t, \quad (1-L)^d \xi_t = \epsilon_t \\ u_t &= \theta u_{t-1} + \eta_t, \end{aligned}$$

where (ϵ_t, λ_t) are *iid* $N\left(0, \begin{bmatrix} 1 & \rho_{\epsilon\lambda} \\ \rho_{\epsilon\lambda} & 1 \end{bmatrix}\right)$. The following regression function from PW was used in the simulation:

$$f(x) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \sin(j\pi x)}{j^2},$$

where the function is truncated at $j = 4$ for numerical computation.

Kernel estimates of $f(x)$ together with the bias, standard deviation (Std), and root mean squared error (Rms) for these estimates were computed over the interval $[0, 1]$ on the equispaced grid $\{x = 0.01k; k = 0, 1, \dots, 100\}$ based on 50,000 replications. Simulations were performed for $\rho_{\epsilon\lambda} = E(\epsilon_t \lambda_t) \in \{0, 0.5, 1.0\}$, $\sigma = 0.2$, $\theta = 0.25$, long memory parameter $d \in \{0, 0.1, 0.2, 0.3, 0.4\}$, and for sample size $n = 500$. These specifications allow for endogeneity, serial dependence in u_t , and long memory in the innovation process ξ_t of the regressor. An Epanechnikov kernel was used with bandwidths $h = n^{-1/3}, n^{-1/4}, n^{-1/5}, n^{-1/6}$. These rates for h satisfy the condition $n^{\frac{1-d}{2}} h^{1+2\gamma} \rightarrow 0$ when $\gamma = 1$ and $d \in (0, \frac{1}{2})$, and the condition $n^{\frac{1}{2}-d} h \rightarrow \infty$ for subsets $d \in [0, \frac{1}{6})$, $d \in [0, \frac{1}{4})$, $d \in [0, \frac{3}{10})$, and $d \in [0, \frac{1}{3})$, respectively.

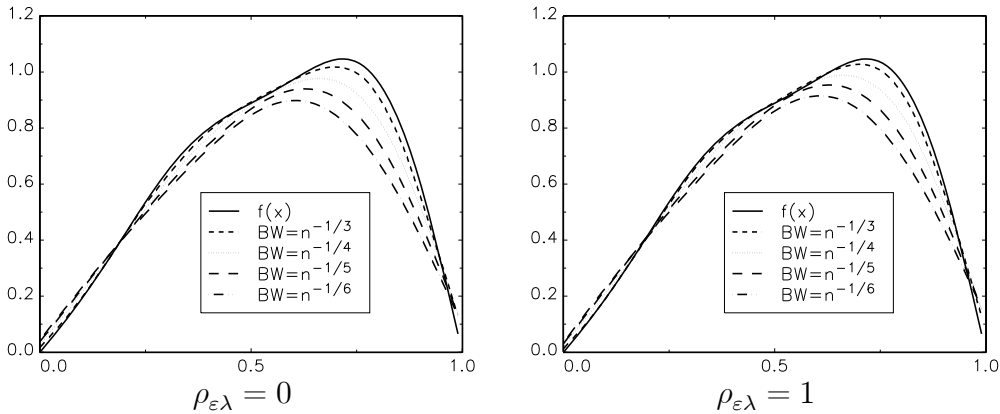


Fig. 1: MC estimates of $\mathbb{E}(\hat{f}(x))$ for $d = 0.4$, $n = 500$, and various bandwidths.

Figs. 1 (a) and (b) graph the mean simulated kernel estimates (broken lines) of f (solid line) under exogeneity ($\rho_{\varepsilon\lambda} = 0$) and strong endogeneity ($\rho_{\varepsilon\lambda} = 1$) for various bandwidth choices. Endogeneity evidently has a negligible effect on the performance of \hat{f} irrespective of bandwidth choice. As expected, smaller bandwidths lead to less bias in estimation but also higher variance, as is apparent from the summary statistics in Table 1 which reports the bias, standard deviation and root mean squared error of the estimates for various bandwidths and values of d . Figs. 2 (a) and (b) graph the mean simulated kernel estimates for various values of d and for $\rho_{\varepsilon\lambda} \in \{0, 1\}$, confirming that the performance of \hat{f} is robust to the presence of long memory drivers in the regressor innovations as well as to the degree of endogeneity in the regressor.

In the scale of Figs. 1 - 2, the difference in performance of the kernel estimates in terms of average location (and bias) is virtually indistinguishable in these two cases. The summary statistics on bias in Table 1 indicate that bias is slightly smaller on average under strong endogeneity than it is for exogenous regressors. But while endogeneity and long memory innovations in the regressor seem to have a negligible effect on bias in the kernel estimates, long memory innovations do affect variance. Table 1 shows that the standard deviation of the estimates increases by 25-30% as d increases from 0 to 0.4.

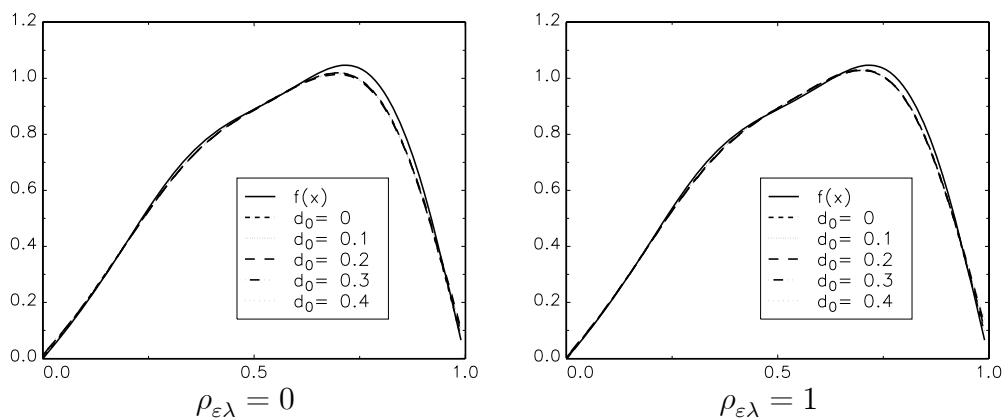


Fig. 2: MC estimates of $\mathbb{E}(\hat{f}(x))$ for $\rho = 1$, $n = 500$, and various d .

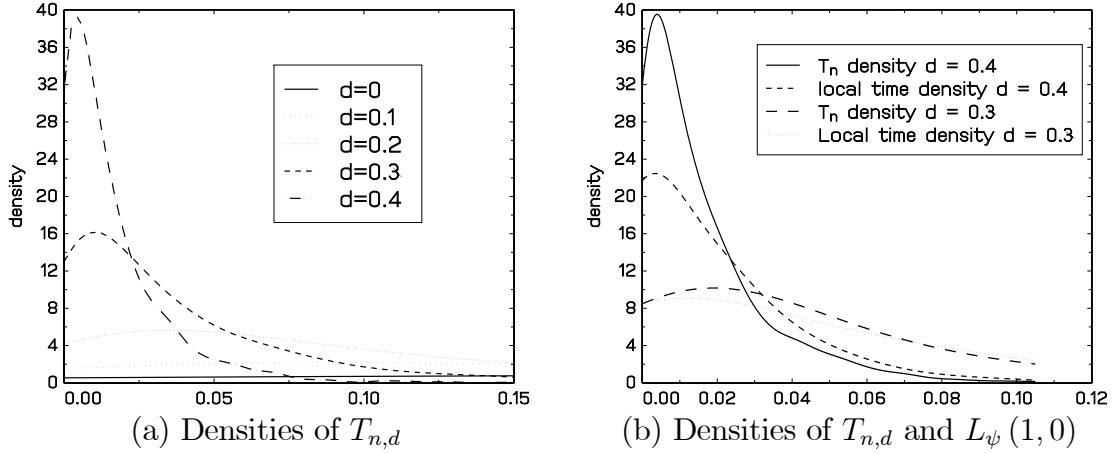


Fig. 3: Densities of $T_{n,d} = \frac{d_n}{nh} T_n$ and the kernel estimate (3.5) of $L_{\psi}(1,0)$ for various values of d , $\rho_{\varepsilon\lambda} = 0.5$, and $n = 500$

Table 1

		$\rho_{\varepsilon\lambda} = 0$			$\rho_{\varepsilon\lambda} = 1$		
d	h	Bias	Std	Rms	Bias	Std	Rms
0.4	$n^{-1/3}$	0.013	0.166	0.173	0.007	0.156	0.161
	$n^{-1/4}$	0.030	0.136	0.151	0.025	0.125	0.140
	$n^{-1/5}$	0.053	0.130	0.163	0.048	0.119	0.153
	$n^{-1/6}$	0.076	0.130	0.184	0.072	0.119	0.175
0.2	$n^{-1/3}$	0.012	0.151	0.155	0.008	0.144	0.149
	$n^{-1/4}$	0.032	0.125	0.141	0.028	0.117	0.132
	$n^{-1/5}$	0.058	0.119	0.156	0.054	0.111	0.147
	$n^{-1/6}$	0.083	0.119	0.179	0.080	0.111	0.171
0	$n^{-1/3}$	0.013	0.127	0.132	0.007	0.125	0.128
	$n^{-1/4}$	0.035	0.107	0.125	0.029	0.105	0.121
	$n^{-1/5}$	0.061	0.104	0.145	0.056	0.102	0.139
	$n^{-1/6}$	0.089	0.105	0.172	0.084	0.102	0.165

Figs.3 (a) and (b) graph simulation estimates of the densities of the standardized test statistic $T_{n,d} = \frac{d_n}{nh} T_n$ and the kernel estimate (3.5) of $L_{\psi}(1,0)$. Fig. 3(a) shows that, as the long memory parameter d increases, the density of $T_{n,d}$ is heavily concentrated close to the origin. By contrast, when $d = 0$ the density appears almost flat in the scale of the Fig.3(a). When $d = 0$, the limit distribution is proportional to that of the local time of standard Brownian motion, $L_W(1,0)$, whose distribution function is $\{2\Phi(x) - 1\} \mathbf{1}\{x \geq 0\}$, as given in Gao et al (2012). Fig. 3(b) shows the densities of $T_{n,d}$ against those of the kernel estimate (3.5) of the local time $L_{\psi}(1,0)$ of fractional Brownian motion for $d = 0.3, 0.4$.

The concentration in the distributions of $T_{n,d}$ and $L_\psi(1,0)$ close to the origin in relation to that of $L_W(1,0)$ is evident in these graphs. A consequence of the concentration of the distribution of $T_{n,d}$ in the long memory case is that tests based on $\frac{\hat{\phi}}{\sqrt{nh}}T_n$ will be highly conservative in the presence of long memory, corroborating the result shown earlier in (3.7) that the size of a test based on $\frac{1}{\sqrt{nh}}T_n$ tends to zero as $n \rightarrow \infty$ whenever $d \in (0, \frac{1}{2})$.

6 Conclusion

The results in the present paper provide a sequel to those in Wang and Phillips (2009b), bringing the limit theory for nonparametric nonstationary regression close to that of the linear cointegrated system in terms of functionality under general short memory innovations and a single endogeneous regressor. The nonparametric estimation and inference results are robust to long memory driver innovations in the regressors, which further widens the scope of potential applications. The specification test results of Gao et al (2012) are also shown to hold under short memory innovations and an endogenous regressor, confirming a conjecture based on simulations reported in their paper.

The presence of long memory driver innovations in the regressor does raise obstacles in specification testing for functional form. The specification test of Gao et al (2012) is no longer pivotal in this case and leads to tests with asymptotic size zero in the presence of long memory. Practical implementation of an appropriately re-scaled test statistic is inhibited by parameter dependence in the rate of convergence of the test and in the limit theory which depends on the local time of fractional Brownian motion with an unknown parameter. These findings in the long memory case suggest that further research on specification testing is warranted to develop procedures that are robust under these wider conditions.

7 Proofs of the Main Results

We start with several propositions. These provide certain key results which are used in the proofs of the main theorems and which are of interest in their own right. Their proofs are given in Section 8.

PROPOSITION 7.1. *Suppose Assumptions 2.1–2.2 hold and $g(x)$ is a bounded function satisfying $\int_{-\infty}^{\infty} |g(x)|dx < \infty$.*

(i) For any $c_n \rightarrow \infty$ and $c_n/n \rightarrow 0$, we have

$$\frac{c_n}{n} \sum_{k=1}^n g(c_n x_{n,k}) \rightarrow_D \int_{-\infty}^{\infty} g(x) dx L_\psi(1, 0), \quad (7.1)$$

where $x_{n,k} = x_k/d_n$ and $L_\psi(1, 0)$ is defined as in (2.3). Furthermore,

$$\frac{d_n}{n} \sum_{k=1}^n |g(x_k)| (1 + |u_k|) = O_P(1), \quad (7.2)$$

$$\left(\frac{d_n}{n}\right)^{1/2} \sum_{k=1}^n g(x_k) u_k = \begin{cases} O_P(1), & \text{under } \mathbf{C1}, \\ O_P(\log^{1/2} n), & \text{under } \mathbf{C2}. \end{cases} \quad (7.3)$$

(ii) If in addition Assumption 2.3, for any $h \rightarrow 0$ ($h \log n \rightarrow 0$ under $\mathbf{C2}$) satisfying $nh/d_n \rightarrow \infty$ and any fixed x , we have

$$\begin{aligned} & \left\{ \left(\frac{d_n}{nh}\right)^{1/2} \sum_{k=1}^n K[(x_k - x)/h] u_k, \frac{d_n}{nh} \sum_{k=1}^n g[(x_k - x)/h] \right\} \\ & \rightarrow_D \left\{ \tau N L_\psi^{1/2}(1, 0), \int_{-\infty}^{\infty} g(s) ds L_\psi(1, 0) \right\}, \end{aligned} \quad (7.4)$$

where $\tau^2 = \mathbb{E}u_0^2 \int_{-\infty}^{\infty} K^2(s) ds$ and N is a standard normal variate independent of $L_\psi(1, 0)$.

(iii) If in addition $|g(x) - g(y)| \leq C|x - y|$ whenever $|x - y|$ is sufficiently small, then for any $h \rightarrow 0$ satisfying $n^{1-\delta_0}h/d_n \rightarrow \infty$, where δ_0 can be as small as required,

$$\sup_{|x| \leq \log n} \left| \sum_{k=1}^n g[(x_k - x)/h] - \sum_{k=1}^n g(x_k/h) \right| = O_P(nh \log^{-1} n/d_n). \quad (7.5)$$

PROPOSITION 7.2. *Suppose Assumptions 2.1–2.2 hold.*

(i) For any locally bounded function $H(x)$ (i.e., bounded on any compact set), we have

$$\frac{1}{n} \sum_{k=1}^n H(x_k/d_n) \rightarrow_D \int_0^1 H[\psi(x)] dx, \quad (7.6)$$

$$\frac{1}{n} \sum_{k=1}^n |H(x_k/d_n)| (1 + |u_k|) = O_P(1). \quad (7.7)$$

(ii) Let $v(\lambda)$ be a positive real function which is bounded away from zero as $\lambda \rightarrow \infty$ and assume $\sum_{k=0}^{\infty} k^{1/2} (|\psi_{1k}| + |\psi_{2k}|) < \infty$. For any real function $g(x)$ satisfying $|g(\lambda x)| \leq C v(\lambda) (1 + |x|^\beta)$ for some $\beta > 0$ and

$$|g(\lambda x) - g(\lambda y)| \leq C v(\lambda) [|x - y| + R_1(\lambda x) + R_2(\lambda y)], \quad (7.8)$$

whenever x and y are in a compact set, where $R_1(z)$ and $R_2(z)$ are bounded and integrable functions, we have

$$\frac{1}{v(d_n)\sqrt{n}} \sum_{k=1}^n g(x_k)u_k = O_P(1). \quad (7.9)$$

PROPOSITION 7.3. *Suppose Assumptions 2.1–2.2 and 3.1 hold.*

(i) *If $\int_{-\infty}^{\infty} |\pi(x)|dx < \infty$ and $|\pi(x)|$ is bounded by a constant, for any h satisfying $h \rightarrow 0$ and $nh/d_n \rightarrow \infty$, we have*

$$\frac{d_n}{nh} \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h]u_k \right\}^2 \pi(x)dx = O_P(1). \quad (7.10)$$

If in addition $nh^2 \log n/d_n \rightarrow 0$, then

$$\frac{d_n}{nh} \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h]u_k \right\}^2 \pi(x)dx \rightarrow_D \tau_0 L_\psi(1, 0), \quad (7.11)$$

where $\tau_0 = \mathbb{E}u_0^2 \int_{-\infty}^{\infty} K^2(x)dx \int_{-\infty}^{\infty} \pi(x)dx$.

(ii) *If $m(x)$ and $\pi(x)$ satisfy Assumption 3.4, for any h satisfying $h \rightarrow 0$ and $n^{1-\delta_0}h/d_n \rightarrow \infty$, where δ_0 can be as small as required, we have*

$$\left(\frac{d_n}{nh}\right)^2 \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h]m(x_k) \right\}^2 \pi(x)dx \rightarrow_D \tau_1 L_\psi^2(1, 0), \quad (7.12)$$

where $\tau_1 = \left(\int_{-\infty}^{\infty} K(x)dx\right)^2 \int_{-\infty}^{\infty} m^2(x)\pi(x)dx$.

7.1. Proof of Theorems 2.1 and 2.2. In view of the joint weak convergence given in (7.4), the proofs of Theorems 2.1 and 2.2 follow in precisely the same way as WP (2009b, Theorem 3.1). See also Wang (2013). The details are therefore omitted. \square

7.2. Proof of Theorem 3.1. Under the null H_0 , we have $y_k = f(x_k, \theta_0) + u_k$. Simple calculation gives the decomposition

$$\begin{aligned} T_n &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] [f(x_k, \hat{\theta}_n) - f(x_k, \theta_0) + u_k] \right\}^2 \pi(x)dx \\ &= T_{1n} + T_{2n} + T_{3n}, \end{aligned} \quad (7.13)$$

where $T_{1n} = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h]u_k \right\}^2 \pi(x)dx$,

$$T_{2n} = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] [f(x_k, \hat{\theta}_n) - f(x_k, \theta_0)] \right\}^2 \pi(x)dx,$$

and $|T_{3n}|^2 \leq 4T_{1n} T_{2n}$ by Hölder's inequality. To prove (3.2), by using (7.11) of Proposition 7.3, it suffices to show that $T_{2n} = o_P(nh/d_n)$. Indeed, as $g_1(x)$ satisfies Assumption 3.4 with $\gamma = \beta$, it follows from Assumptions 3.2–3.3 and the result (7.12) in Proposition 7.3 that

$$\begin{aligned} T_{2n} &\leq C \|\hat{\theta}_n - \theta_0\|^2 \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] g_1(x_k) \right\}^2 \pi(x) dx \\ &= o_P(nh/d_n), \end{aligned} \quad (7.14)$$

as required. The proof of Theorem 3.1 is now complete. \square

7.3. Proof of Theorem 3.2. Under the alternative H_1 , the test statistic T_n can be written as

$$\begin{aligned} T_n &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] [u_k^* + \rho_n m(x_k)] \right\}^2 \pi(x) dx \\ &= T_{1n} + T_{2n} + T_{3n} + 2\rho_n T_{4n} + \rho_n^2 T_{5n}, \end{aligned}$$

where $u_k^* = u_k + f(x_k, \hat{\theta}_n) - f(x_k, \theta_0)$, $T_{jn}, j = 1, 2, 3$, are defined as in proof of Theorem 3.1,

$$T_{5n} = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] m(x_k) \right\}^2 \pi(x) dx,$$

and $|T_{4n}| \leq [(T_{1n} + T_{2n} + T_{3n})]^{1/2} (T_{5n})^{1/2}$ by Hölder's inequality. Let $\epsilon_n = nh\rho_n^2/d_n$. We have $\epsilon_n \rightarrow \infty$. Note that $T_{1n} + T_{2n} + T_{3n} = O_P(nh/d_n)$, due to (7.10), (7.14) and $|T_{3n}|^2 \leq 4T_{1n} T_{2n}$. This, together with $(\frac{d_n}{nh})^2 T_{5n} \rightarrow_D A_1 L_\psi^2(1, 0)$ and $P(0 < L_\psi^2(1, 0) < \infty) = 1$, yields

$$\begin{aligned} \Delta_n &:= \frac{d_n}{nh} |T_{1n} + T_{2n} + T_{3n} + 2\rho_n T_{4n}| \\ &= O_P(1) + O_P(\epsilon_n^{-1/2}), \end{aligned}$$

and for any $t_0 > 0$, as $n \rightarrow \infty$,

$$\begin{aligned} P\left(\frac{d_n}{nh} T_n \geq t_0\right) &\geq P\left(\frac{d_n}{nh} T_n \geq \epsilon_n^{3/4}\right) = P\left[\left(\frac{d_n}{nh}\right)^2 T_{5n} \geq \epsilon_n^{-1/4} - \epsilon_n^{-1} \Delta_n\right] \\ &\geq P\left[\left(\frac{d_n}{nh}\right)^2 T_{5n} \geq \epsilon_n^{-1/4}/2\right] \rightarrow 1, \end{aligned}$$

which proves Theorem 3.2. \square

7.4. Proof of Theorem 4.1. The proof follows the same argument as that of Lemma 1 in Andrews and Sun (2004). For convenience we outline the argument here. Let $\Theta_0 = \{\theta \in$

$\Theta : \| D_n(\theta - \theta_0) \| \leq k_n, \| \theta - \theta_0 \| \leq \delta$ for some $\delta > 0$ so that $\{\theta \in \Theta : \| \theta - \theta_0 \| \leq \delta\} \subset \Theta$ and $Q_n(\theta)$ is twice differentiable on $\theta \in \{\theta \in \Theta : \| \theta - \theta_0 \| \leq \delta\}$. Note that

$$\begin{aligned}\dot{Q}_n(\theta_0) &= - \sum_{t=1}^n \dot{g}(x_t, \theta_0)(y_t - g(x_t, \theta_0)) = - \sum_{t=1}^n \dot{g}(x_t, \theta_0)u_t, \\ \ddot{Q}_n(\theta) &= \sum_{t=1}^n \dot{g}(x_t, \theta)\dot{g}(x_t, \theta)' - \sum_{t=1}^n \ddot{g}(x_t, \theta)u_t - \sum_{t=1}^n \ddot{g}(x_t, \theta)[g(x_t, \theta) - g(x_t, \theta_0)],\end{aligned}$$

and recall the definitions of Y_n and Z_n . It follows by Taylor expansion that

$$\begin{aligned}Q_n(\theta) - Q_n(\theta_0) &= \dot{Q}_n(\theta_0)'(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'\ddot{Q}_n(\theta_0)(\theta - \theta_0) + R_n(\theta, \theta_0) \\ &= \frac{1}{2}(D_n(\theta - \theta_0) + Y_n^{-1}Z_n)'Y_n(D_n(\theta - \theta_0) + Y_n^{-1}Z_n) \\ &\quad + \frac{1}{2}Z_n'Y_n^{-1}Z_n + R_{1n}(\theta, \theta_0),\end{aligned}\tag{7.15}$$

for all $\theta \in \Theta_0$, where

$$\begin{aligned}|R_n(\theta, \theta_0)| &\leq \sup_{\theta_1 \in \Theta_0} |(\theta - \theta_0)'[\ddot{Q}_n(\theta_1) - \ddot{Q}_n(\theta_0)](\theta - \theta_0)| \quad \text{and} \\ R_{1n}(\theta, \theta_0) &= R_n(\theta, \theta_0) + \frac{1}{2}(\theta - \theta_0)' \left(\sum_{t=1}^n \ddot{g}(x_t, \theta_0)u_t \right) (\theta - \theta_0).\end{aligned}$$

In view of conditions (i)–(iii), simple calculations show that, for all $\theta \in \Theta_0$,

$$\begin{aligned}|R_{1n}(\theta, \theta_0)| &\leq \| D_n(\theta - \theta_0) \|^2 \left\{ \sup_{\theta_1 \in \Theta_0} \| (D_n^{-1})' [\ddot{Q}_n(\theta_1) - \ddot{Q}_n(\theta_0)] D_n^{-1} \| \right. \\ &\quad \left. + \sup_{\theta_1 \in \Theta_0} \| (D_n^{-1})' \sum_{t=1}^n \ddot{g}(x_t, \theta_1) u_t D_n^{-1} \| \right\} \\ &= o_P(\delta_n^{-2}) \| D_n(\theta - \theta_0) \|^2.\end{aligned}\tag{7.16}$$

Let $\tilde{\theta}_n = \theta_0 - D_n^{-1}Y_n^{-1}Z_n$. It follows from (iv) and $k_n/\delta_n \rightarrow \infty$ that

$$P(\tilde{\theta}_n \notin \Theta_0) \leq P(\|Y_n^{-1}Z_n\| \geq k_n) + P(\|D_n^{-1}Y_n^{-1}Z_n\| \geq \delta) \rightarrow 0.$$

This, together with (7.15) and (7.16), yields

$$Q_n(\tilde{\theta}_n) - Q_n(\theta_0) = \frac{1}{2}Z_n'Y_n^{-1}Z_n + R_{1n}(\tilde{\theta}_n, \theta_0),$$

where $R_{1n}(\tilde{\theta}_n, \theta_0) = o_P(1)$. For any $\epsilon > 0$ and $n \geq 1$, let

$$\Theta_n(\epsilon) = \{\theta \in \Theta : \|D_n(\theta - \theta_0) + Y_n^{-1}Z_n\| \leq \epsilon\}.$$

Recall $\|Y_n^{-1}Z_n\| = O_P(\delta_n) = o_P(k_n)$ and $k_n\|D_n^{-1}\| = o(1)$. It follows that $P[\Theta_n(\epsilon) \subset \Theta_0] \rightarrow 1$, as $n \rightarrow \infty$ and $\sup_{\theta \in \Theta_n(\epsilon)} |R_{1n}(\theta, \theta_0)| = o_P(1)$. Consequently, for any $\theta \in \partial\Theta_n(\epsilon)$, where $\partial\Theta_n(\epsilon)$ denotes the boundary of $\Theta_n(\epsilon)$, we have

$$Q_n(\theta) - Q_n(\tilde{\theta}_n) = \frac{1}{2}\nu_n' Y_n \nu_n + o_P(1),$$

where ν_n is a vector with $\|\nu_n\| = \epsilon > 0$. Since $Y_n \rightarrow_D M > 0$, a.s., we have $P(\frac{1}{2}\nu_n' Y_n \nu_n > 0) \rightarrow 1$ as $n \rightarrow \infty$. Hence, for each $\epsilon > 0$, the event that the minimum of $Q_n(\theta)$ over $\Theta_n(\epsilon)$ is in the interior of $\Theta_n(\epsilon)$ has probability that goes to one as $n \rightarrow \infty$. In particular, for each $\epsilon > 0$, there exists a point $\hat{\theta}_n(\epsilon) \in \Theta_n(\epsilon)$ (not necessary unique) so that $P(\dot{Q}_n[\hat{\theta}_n(\epsilon)] = 0) \rightarrow 1$, as $n \rightarrow \infty$. In consequence, there exists a sequence of $\hat{\theta}_n = \hat{\theta}_n(1/J_n) \in \Theta_n(1/J_n)$ where $J_n \rightarrow \infty$ so that $P(\dot{Q}_n(\hat{\theta}_n) = 0) \rightarrow 1$, as $n \rightarrow \infty$, and (4.2) holds.

Finally, if we have (iv)', then (4.2) holds with $\delta_n = 1$. The asymptotic distribution follows immediately from (4.2) and (iv)'. \square

7.5. Proof of Theorem 4.2. It suffices to verify the conditions (i)–(iv) of Theorem 4.1, with $k_n = \log n$, $D_n = \sqrt{n/d_n} \mathbf{I}$ where \mathbf{I} is the identity matrix,

$$M = \Sigma L_\psi(1, 0) \quad \text{and} \quad \delta_n = \begin{cases} 1, & \text{under } \mathbf{C1}, \\ \log^{1/2} n, & \text{under } \mathbf{C2}. \end{cases}$$

Let $\Omega_n = \{\theta : \|\theta - \theta_0\| \leq d_n \log n/n\}$. (i)–(iii) of Theorem 4.1 will follow if we prove: for any $1 \leq i, j \leq m$,

$$\frac{d_n}{n} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\dot{g}_i(x_t, \theta) \dot{g}_j(x_t, \theta) - \dot{g}_i(x_t, \theta_0) \dot{g}_j(x_t, \theta_0)| = o_P(\log^{-2} n), \quad (7.17)$$

$$\frac{d_n}{n} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\ddot{g}_{ij}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)]| = o_P(\log^{-2} n), \quad (7.18)$$

$$\frac{d_n}{n} \sup_{\theta \in \Omega_n} \left| \sum_{t=1}^n \ddot{g}_{ij}(x_t, \theta) u_t \right| = o_P(\log^{-2} n). \quad (7.19)$$

We only prove (7.17) by using part (i) of Proposition 7.1. The other derivations are similar and the details are omitted. First note that $|\dot{g}_j(x_t, \theta)|$, $1 \leq j \leq m$, are uniformly bounded on Θ from Assumption 4.1 (i) and (ii). It follows from Assumption 4.1 (ii) that

$$\begin{aligned} & |\dot{g}_i(x_t, \theta) \dot{g}_j(x_t, \theta) - \dot{g}_i(x_t, \theta_0) \dot{g}_j(x_t, \theta_0)| \\ & \leq C |\dot{g}_j(x_t, \theta) - \dot{g}_j(x_t, \theta_0)| + C_1 |\dot{g}_i(x_t, \theta) - \dot{g}_i(x_t, \theta_0)| \\ & \leq C \|\theta - \theta_0\| (T_{\dot{g}_j}(x_t) + T_{\dot{g}_i}(x_t)), \end{aligned}$$

and (7.17) follows immediately from (7.2) of Proposition 7.1.

The required condition (iv) of Theorem 4.1 follows from (7.3) of Proposition 7.1 and the following fact: using (7.1) of Proposition 7.1 with $c_n = d_n$ and $g(x) = \alpha_1' \dot{g}(x, \theta_0) \dot{g}(x, \theta_0)' \alpha_2$, we have, for any $\alpha_i = (\alpha_{i1}, \dots, \alpha_{im}) \in \mathbb{R}^m, i = 1, 2$,

$$\frac{d_n}{n} \sum_{t=1}^n \alpha_1' \dot{g}(x_t, \theta_0) \dot{g}(x_t, \theta_0)' \alpha_2 \rightarrow_D \alpha_1' M \alpha_2.$$

The proof of Theorem 4.2 is now complete. \square

7.6. Proof of Theorem 4.3. Similar to the proof of Theorem 4.2, we verify conditions (i)–(iv) of Theorem 4.1 with $k_n = \log n$ and $\delta_n = 1$.

First for (iv). Let $M = \int_0^1 h[\psi(t), \theta_0] h[\psi(t), \theta_0]' dt$. It is readily seen that $M > 0$ a.s. due to $\sum_{|s| \leq \delta} h(s, \theta_0) h(s, \theta_0)' ds > 0$ for all $\delta > 0$. Furthermore, for any $\alpha = (\alpha_1, \dots, \alpha_m)^m$, it follows from Assumption 4.2 (iii) and (7.6) of Proposition 7.2 with $H(x) = \sum_{i,j=1}^m \alpha_i \alpha_j h_i(x, \theta_0) h_j(x, \theta_0)$ that

$$\begin{aligned} & (D_n^{-1} \alpha)' \sum_{k=1}^n \dot{g}(x_k, \theta_0) \dot{g}(x_k, \theta_0)' D_n^{-1} \alpha \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{i,j=1}^m \frac{\alpha_i \alpha_j}{v_i(d_n) v_j(d_n)} \dot{g}_i(x_k, \theta_0) \dot{g}_j(x_k, \theta_0) \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{i,j=1}^m \alpha_i \alpha_j h_i(x_k/d_n, \theta_0) h_j(x_k/d_n, \theta_0) [1 + o_P(1)] \\ &\rightarrow_D \alpha' M \alpha. \end{aligned}$$

Hence $(D_n^{-1})' \sum_{k=1}^n \dot{g}(x_k, \theta_0) \dot{g}(x_k, \theta_0)' D_n^{-1} \rightarrow_D M$. On the other hand, (7.9) of Proposition 7.2 implies (4.1) with $\delta_n = 1$. These facts yield the required condition (iv) of Theorem 4.1.

To verify (i)–(iii), we first show that, for all $1 \leq i, j \leq m$,

$$\frac{1}{n \dot{v}_i(d_n) \dot{v}_j(d_n)} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\dot{g}_i(x_t, \theta) \dot{g}_j(x_t, \theta) - \dot{g}_i(x_t, \theta_0) \dot{g}_j(x_t, \theta_0)| = o_P(1), \quad (7.20)$$

$$\frac{1}{n v(d_n) \ddot{v}_{ij}(d_n)} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\ddot{g}_{ij}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)]| = o_P(1), \quad (7.21)$$

$$\frac{1}{n \ddot{v}_{ij}(d_n)} \sup_{\theta \in \Omega_n} \left| \sum_{t=1}^n \ddot{g}_{ij}(x_t, \theta) u_t \right| = o_P(1). \quad (7.22)$$

where $\Omega_n = \{\theta : \|D_n(\theta - \theta_0)\| \leq \log n\}$. In fact, it follows from Assumption 4.2(i)–(ii)

that, for any $1 \leq i \leq m$ and $\theta \in \Theta$,

$$\begin{aligned} |\dot{g}_i(x_t, \theta)| &\leq C \dot{v}_i(d_n) (1 + \|\theta - \theta_0\|^\alpha) (1 + |x_t/d_n|^\beta), \\ |\dot{g}_i(x_t, \theta) - \dot{g}_i(x_t, \theta_0)| &\leq C \|\theta - \theta_0\|^\alpha \dot{v}_i(d_n) (1 + |x_t/d_n|^\beta). \end{aligned}$$

This, together with $\|\theta - \theta_0\| = o(1)$ whenever $\theta \in \Omega_n$, implies that

$$\begin{aligned} &\sup_{\theta \in \Omega_n} \sum_{t=1}^n |\dot{g}_i(x_t, \theta) \dot{g}_j(x_t, \theta) - \dot{g}_i(x_t, \theta_0) \dot{g}_j(x_t, \theta_0)| \\ &\leq o[\dot{v}_i(d_n) \dot{v}_j(d_n)] \sum_{t=1}^n (1 + |x_t/d_n|^{2\beta}). \end{aligned} \quad (7.23)$$

Now, by noting that $1 + |x|^{2\beta}$ is a locally bounded function, the result (7.20) follows from (7.7) of Proposition 7.2.

The proof of (7.21) is similar. As for (7.22), by noting

$$\begin{aligned} \left| \sum_{t=1}^n \ddot{g}_{ij}(x_t, \theta) u_t \right| &\leq \left| \sum_{t=1}^n \ddot{g}_{ij}(x_t, \theta_0) u_t \right| + \sum_{t=1}^n |\ddot{g}_{ij}(x_t, \theta) - \ddot{g}_{ij}(x_t, \theta_0)| |u_t| \\ &\leq \left| \sum_{t=1}^n \ddot{g}_{ij}(x_t, \theta_0) u_t \right| + C \|\theta - \theta_0\|^\alpha \ddot{v}_{ij}(d_n) \sum_{t=1}^n (1 + |x_t/d_n|^\beta) |u_t|, \end{aligned}$$

the result follows from (7.7) and (7.9) of Proposition 7.2.

Recalling $\sup_{1 \leq i, j \leq m} \left| \frac{v(d_n) \ddot{v}_{ij}(d_n)}{\dot{v}_i(d_n) \dot{v}_j(d_n)} \right| < \infty$, it follows from (7.21) that

$$\begin{aligned} &\frac{1}{n \dot{v}_i(d_n) \dot{v}_j(d_n)} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\ddot{g}_{ij}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)]| \\ &\leq \frac{C}{n \ddot{v}_{ij}(d_n) v(d_n)} \sup_{\theta \in \Omega_n} \sum_{t=1}^n |\ddot{g}_{ij}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)]| = o_P(1), \end{aligned}$$

for any $1 \leq i, j \leq m$. This fact implies that

$$\sup_{\theta: \|D_n(\theta - \theta_0)\| \leq k_n} \left\| (D_n^{-1})' \sum_{t=1}^n \ddot{g}(x_t, \theta) [g(x_t, \theta) - g(x_t, \theta_0)] D_n^{-1} \right\| = o_P(1),$$

which yields the condition (ii) of Theorem 4.1. By using (7.20) and (7.22), similar arguments provide conditions (i) and (iii) of Theorem 4.1. The proof of Theorem 4.3 is now complete. \square

8 Proofs of Propositions

8.1 Preliminaries

We may write, for any $s < m$,

$$\begin{aligned}
x_m &= \sum_{j=1}^m \sum_{i=-\infty}^j \epsilon_i \phi_{j-i} \\
&= x_s + \sum_{j=s+1}^m \sum_{i=-\infty}^s \epsilon_i \phi_{j-i} + \sum_{j=s+1}^m \sum_{i=s+1}^j \epsilon_i \phi_{j-i} \\
&=: x_{s,m}^* + x'_{s,m},
\end{aligned} \tag{8.1}$$

where $x_{s,m}^*$ depends only on $(\dots, \epsilon_{s-1}, \epsilon_s)$ and

$$x'_{s,m} = \sum_{j=1}^{m-s} \sum_{i=1}^j \epsilon_{i+s} \phi_{j-i} = \sum_{i=s+1}^m \epsilon_i \sum_{j=0}^{m-i} \phi_j.$$

Define $\sum_{j=k}^l = 0$ if $l < k$, and put $a_k = \sum_{j=0}^k \phi_j$. By the definition of ϕ_j and ϵ_j , elementary calculations show that

$$a_m \asymp \begin{cases} m^{1-\mu} \rho(m), & \text{under } \mathbf{C1}, \\ 1, & \text{under } \mathbf{C2}. \end{cases} \tag{8.2}$$

$$\mathbb{E}(x'_{s,m})^2 = \sum_{k=s+1}^m a_{m-k}^2 \asymp d_{m-s}^2 \asymp (m-s) a_{m-s}^2, \tag{8.3}$$

for $s < m$.¹ We further let $\tilde{x}_{s,m} = x'_{s,m} - \Lambda_m$ and $\tilde{x}_m = x_m - \Lambda_m$, where Λ_m is a functional of some ϵ_k , $k \leq m$, such that Λ_m is independent of $\tilde{x}_{s,m}$ and \tilde{x}_m , satisfying $\mathbb{E}\Lambda_m = 0$. In view of (8.2)–(8.3) and $\int_{-\infty}^{\infty} (1+|t|)|\mathbb{E}e^{it\epsilon_0}|dt < \infty$, the following fact holds for $m-s \geq 1$ and Λ_m satisfying $\sup_{m-s \geq 1} \mathbb{E}\Lambda_m^2 / \mathbb{E}x'_{s,m}{}^2 < 1$.

F: $\tilde{x}_{s,m}/d_{m-s}$ and \tilde{x}_m/d_m have density functions $\nu_{s,m}(x)$ and $g_m(x)$ respectively, and the functions $\nu_{s,m}(x)$ and $g_m(x)$ are uniformly bounded over x by a constant C , and

$$\sup_x |\nu_{s,m}(x+u) - \nu_{s,m}(x)| \leq C \min\{|u|, 1\}, \tag{8.4}$$

$$\sup_x |g_m(x+u) - g_m(x)| \leq C \min\{|u|, 1\}. \tag{8.5}$$

See Section 3.1 of WP (2009a) or Section 8.1 of WP (2012) with some routine modifications. In particular, the fact **F** holds true for all $m-s \geq 1$ if $\Lambda_m = 0$ and $a_0 = \phi_0 \neq 0$.

¹Here and below the notation $b_n \asymp c_n$ denotes $0 < \liminf_{n \rightarrow \infty} b_n/c_n < \limsup_{n \rightarrow \infty} b_n/c_n < \infty$.

Recall that $\eta'_j = (\epsilon_j, \nu_j)$ is a sequence of iid random variables with $\mathbb{E}\eta_0 = 0$ and $\mathbb{E}\|\eta_0\|^2 < \infty$. To introduce the following lemmas, let $\mathcal{F}_s = \sigma(\eta_s, \eta_{s-1}, \dots)$, $\Lambda(\dots)$ be a real function of its components, $p(x)$ be a bounded real function satisfying $\int |p(x)|dx < \infty$, and Assumptions 2.1–2.2 hold. We use C, C_1, \dots to denote constants which differ at each occurrence.

LEMMA 8.1. *Let $\Lambda = \Lambda(\eta_{t_1}, \dots, \eta_{t_{m_0}})$, where t_1, t_2, \dots, t_{m_0} is a subset of $\{t, t-1, \dots\}$. There exists an $A_0 > 0$ such that*

(i) *for any $h > 0$ and $t \geq A_0$, we have*

$$\mathbb{E}\{|\Lambda| |p(x_t/h)|\} \leq \frac{C h \mathbb{E}|\Lambda|}{d_t} \int_{-\infty}^{\infty} |p(x)|dx. \quad (8.6)$$

(ii) *for any $h > 0$ and $t - s \geq A_0$, we have*

$$\mathbb{E}\{|\Lambda| |p(x_t/h)| | \mathcal{F}_s\} \leq \frac{C h \mathbb{E}|\Lambda|}{d_{t-s}} \int_{-\infty}^{\infty} |p(x)|dx, \quad (8.7)$$

provided $s + 1 \leq t_j \leq t, j = 1, \dots, m_0$. If in addition $\mathbb{E}\Lambda = 0$, then

$$\left| \mathbb{E}\{\Lambda p(x_t/h) | \mathcal{F}_s\} \right| \leq \frac{C \beta h}{d_{t-s}^2} \int_{-\infty}^{\infty} |p(x)|dx, \quad (8.8)$$

where $\beta = m_0 (\mathbb{E}\Lambda^2)^{1/2} \sum_{k=0}^{t-\min\{t_1, \dots, t_{m_0}\}} |\phi_k|$. Furthermore, if Λ is a constant, the result (8.7) remains to be true for any $h > 0$ and $t - s \geq 1$.

Proof. We only prove (8.8). The other derivations are similar and the details are omitted. Let $\Lambda_{m_0} = \sum_{j=1}^{m_0} \epsilon_{t_j} a_{t-t_j}$ and $\tilde{x}_{s,t} = x'_{s,t} - \Lambda_{m_0}$. Recall (8.2) and (8.3). There exists an $A_0 > 0$ such that, whenever $t - s \geq A_0$, $\mathbb{E}\Lambda_{m_0}^2 \leq \mathbb{E}x_{s,t}'^2/2$. It follows from fact **F**, the independence of ϵ_i and (8.1) that

$$\begin{aligned} \mathbb{E}\{\Lambda p(x_t/h) | \mathcal{F}_s\} &= \mathbb{E}\{\Lambda p[(x_{s,t}^* + \Lambda_{m_0} + \tilde{x}_{s,t})/h] | \mathcal{F}_s\} \\ &= \mathbb{E}\left\{\Lambda \int_{-\infty}^{\infty} p[(x_{s,t}^* + \Lambda_{m_0} + d_{t-s}y)/h] \nu_{s,t}(y) dy | \mathcal{F}_s\right\} \\ &= \frac{h}{d_{t-s}} \int_{-\infty}^{\infty} p(y) \nu_{s,t}^*(y) dy, \end{aligned} \quad (8.9)$$

where

$$\nu_{s,t}^*(y) = \mathbb{E}\left\{\Lambda \nu_{s,t}\left(\frac{-x_{s,t}^* - \Lambda_{m_0} + hy}{d_{t-s}}\right) | \mathcal{F}_s\right\}.$$

Note that $\mathbb{E}\left\{\Lambda \nu_{s,t}\left(\frac{-x_{s,t}^* + hy}{d_{t-s}}\right) | \mathcal{F}_s\right\} = 0$. By (8.4), we have

$$|\nu_{s,t}^*(y)| \leq C \mathbb{E}[|\Lambda| \min\{|\Lambda_{m_0}|/d_{t-s}, 1\}].$$

Taking this estimate into (8.9), simple calculations yield (8.8). \square

LEMMA 8.2. Suppose that $\mathbb{E}\Lambda(\eta_1) = 0$ and $g(x)$ is a bounded function satisfying $\int_{-\infty}^{\infty} |g(x)|dx < \infty$.

(i) For any integer $m \geq 1$, there exists a constant H_0 such that

$$\sup_x \mathbb{E} \left| \sum_{k=1}^n g[(x_k - x)/h] \right|^m \leq H_0^m (m+1)! (nh/d_n)^m. \quad (8.10)$$

(ii) For any $h > 0$, $nh/d_n \rightarrow \infty$ and $j \geq 0$, we have

$$\begin{aligned} & \sup_x \mathbb{E} \left| \sum_{k=1}^n \Lambda(\eta_{k-j}) g[(x_k - x)/h] \right|^2 \\ & \leq C \mathbb{E} \Lambda^2(\eta_1) (nh/d_n) \begin{cases} 1 + hj^{1/2}, & \text{under } \mathbf{C1}, \\ 1 + hj^{1/2} + h \log n, & \text{under } \mathbf{C2}. \end{cases} \end{aligned} \quad (8.11)$$

Consequently, for the u_k defined in Assumption 2.2, we have

$$\mathbb{E} \left| \sum_{k=1}^n u_k g(x_k) \right|^2 \leq C \mathbb{E} \|\eta_1\|^2 nh/d_n \begin{cases} 1, & \text{under } \mathbf{C1}, \\ \log n, & \text{under } \mathbf{C2}. \end{cases} \quad (8.12)$$

Furthermore, if $u_{k,m_0} = \sum_{j=m_0}^{\infty} \psi_j \lambda'_{k-j}$, where $\psi_j = (\psi_{1j}, \psi_{2j})$ and $\lambda_i = (\Lambda_1(\eta_i), \Lambda_2(\eta_i))$ with $\mathbb{E}\lambda_1 = 0$, then for all $h \rightarrow 0$ ($h \log n \rightarrow 0$ under $\mathbf{C2}$) and $nh/d_n \rightarrow \infty$,

$$\begin{aligned} & \sup_x \mathbb{E} \left| \sum_{k=1}^n u_{k,m_0} g[(x_k - x)/h] \right|^2 \\ & \leq C \mathbb{E} \|\lambda_1\|^2 (nh/d_n) \left[\sum_{j=m_0}^{\infty} j^{1/4} (|\psi_{1j}| + |\psi_{2j}|) \right]^2. \end{aligned} \quad (8.13)$$

Proof. For the proof of (8.10), see Lemma 5.1 of Chan and Wang (2013), which comes from an application of (8.7). We next prove (8.11). Let $\sum_{j=k}^l = 0$ for $k > l$ and write

$$\begin{aligned} \Delta_n & \equiv \left| \sum_{k=1}^n \Lambda(\eta_{k-j}) g[(x_k - x)/h] \right|^2 \\ & \leq 2 \left| \sum_{k=A_0}^n \Lambda(\eta_{k-j}) g[(x_k - x)/h] \right|^2 + C \left(\sum_{k=1}^{A_0} |\Lambda(\eta_{k-j})| \right)^2 \\ & = 2 \left(\sum_{k=A_0}^n \sum_{\substack{l=1 \\ |k-l| < A_0}}^n + 2 \sum_{k=A_0}^{n-1} \sum_{l=k+A_0}^n \right) \Lambda(\eta_{k-j}) \Lambda(\eta_{l-j}) g[(x_k - x)/h] g[(x_l - x)/h] \\ & \quad + C \left(\sum_{k=1}^{A_0} |\Lambda(\eta_{k-j})| \right)^2 \\ & = \Delta_{1n} + \Delta_{2n} + \Delta_{3n}, \quad \text{say,} \end{aligned} \quad (8.14)$$

where A_0 is chosen as in Lemma 8.1. Using $|g(x)| \leq C$ and Lemma 8.1 with $p(y) = g(y - x/h)$, we have that, for $k \geq A_0$ and $|k - l| < A_0$,

$$\begin{aligned} & |\mathbb{E}\Lambda(\eta_{k-j}) \Lambda(\eta_{l-j}) g[(x_k - x)/h] g[(x_l - x)/h]| \\ & \leq C \mathbb{E}|\Lambda(\eta_{k-j}) \Lambda(\eta_{l-j}) g[(x_k - x)/h]| \\ & \leq C \mathbb{E}\Lambda^2(\eta_1) h/d_k, \end{aligned}$$

and for $k \geq A_0$ and $l - k \geq A_0$,

$$\begin{aligned} & |\mathbb{E}\Lambda(\eta_{k-j}) \Lambda(\eta_{l-j}) g[(x_k - x)/h] g[(x_l - x)/h]| \\ & \leq \begin{cases} \mathbb{E}|\Lambda(\eta_{k-j}) \Lambda(\eta_{l-j}) g[(x_k - x)/h] \mathbb{E}\{g[(x_l - x)/h] | \mathcal{F}_k\}| & \text{if } l - j \leq k, \\ \mathbb{E}|\Lambda(\eta_{k-j}) g[(x_k - x)/h] \mathbb{E}\{\Lambda(\eta_{l-j}) g[(x_l - x)/h] | \mathcal{F}_k\}| & \text{if } l - j > k, \end{cases} \\ & \leq C \mathbb{E}\Lambda^2(\eta_1) h^2 d_k^{-1} \begin{cases} d_{l-k}^{-1} & \text{if } l - j \leq k, \\ \sum_{k=0}^j |\phi_k| d_{l-k}^{-2} & \text{if } l - j > k. \end{cases} \end{aligned}$$

It follows from these facts that

$$\begin{aligned} \sup_x \mathbb{E}|\Delta_{1n}| & \leq C \mathbb{E}\Lambda^2(\eta_1) h \sum_{k=1}^n \sum_{\substack{l=1 \\ |k-l| < A_0}}^n 1/d_k \leq C_1 \mathbb{E}\Lambda^2(\eta_1) nh/d_n, \\ \sup_x \mathbb{E}|\Delta_{2n}| & \leq C \mathbb{E}\Lambda^2(\eta_1) h^2 \sum_{k=A_0}^{n-1} d_k^{-1} \left(\sum_{l=k+A_0}^{n \wedge (k+j)} d_{l-k}^{-1} + \sum_{k=0}^j |\phi_k| \sum_{l=k+j}^n d_{l-k}^{-2} \right) \\ & \leq C \mathbb{E}\Lambda^2(\eta_1) (nh^2/d_n) \begin{cases} j/d_j + \sum_{k=0}^j |\phi_k|, & \text{under } \mathbf{C1}, \\ j/d_j + \sum_{k=0}^j |\phi_k| \log n, & \text{under } \mathbf{C2}, \end{cases} \\ & \leq C \mathbb{E}\Lambda^2(\eta_1) (nh^2/d_n) \begin{cases} j^{1/2}, & \text{under } \mathbf{C1}, \\ j^{1/2} + \log n, & \text{under } \mathbf{C2}. \end{cases} \end{aligned}$$

On the other hand, it is readily seen that $\sup_x \mathbb{E}|\Delta_{3n}| \leq C \mathbb{E}\Lambda^2(\eta_1)$. Taking these estimates into (8.14) and noting $nh/d_n \rightarrow \infty$, we obtain the required (8.11).

The result (8.13) follows from

$$\begin{aligned} & \mathbb{E} \left| \sum_{k=1}^n u_{k,m_0} g[(x_k - x)/h] \right|^2 = \mathbb{E} \left| \sum_{j=m_0}^{\infty} \sum_{k=1}^n \psi_j \lambda'_{k-j} K[(x_k - x)/h] \right|^2 \\ & \leq \sum_{j=m_0}^{\infty} j^{1/4} (|\psi_{1j}| + |\psi_{2j}|) \sum_{j=m_0}^{\infty} j^{-1/4} (|\psi_{1j}| + |\psi_{2j}|)^{-1} \mathbb{E} \left| \sum_{k=1}^n \psi_j \lambda'_{k-j} K[(x_k - x)/h] \right|^2 \end{aligned}$$

$$\begin{aligned}
&\leq 2 \sum_{j=m_0}^{\infty} j^{1/4} (|\psi_{1j}| + |\psi_{2j}|) \sum_{j=m_0}^{\infty} j^{-1/4} (|\psi_{1j}| + |\psi_{2j}|) \\
&\quad \left(\mathbb{E} \left| \sum_{k=1}^n \Lambda_1(\eta_{k-j}) K[(x_k - x)/h] \right|^2 + \mathbb{E} \left| \sum_{k=1}^n \Lambda_2(\eta_{k-j}) K[(x_k - x)/h] \right|^2 \right) \\
&\leq C \mathbb{E} \|\lambda_1\|^2 (nh/d_n) \left[\sum_{j=m_0}^{\infty} j^{1/4} (|\psi_{1j}| + |\psi_{2j}|) \right]^2,
\end{aligned}$$

where we employ Hölder's inequality, $h \rightarrow 0$ ($h \log n \rightarrow 0$ under **C2**) and (8.11) with $\Lambda(\cdot) = \Lambda_1(\cdot)$ and $\Lambda_2(\cdot)$, respectively. The proof of (8.12) is similar, and hence the details are omitted. The proof of Lemma 8.2 is now complete. \square

LEMMA 8.3. *Suppose that $\mathbb{E}\Lambda(\eta_1) = 0$ and $|\Lambda(x)| \leq C$. Let $z_k = \int_{-\infty}^{\infty} K^2[(x_k - x)/h] p(x) dx$ and $z_{k,j} = \int_{-\infty}^{\infty} K[(x_k - x)/h] K[(x_j - x)/h] p(x) dx$, where $K(x)$ has compact support. Then, for any h satisfying $nh^2 \log n/d_n \rightarrow 0$ and $nh/d_n \rightarrow \infty$, $0 \leq s, s_1 \leq m_0$, where m_0 is a fixed constant, we have*

$$\sum_{k=1}^n \Lambda(\eta_{k-s}) \Lambda(\eta_{k-s_1}) z_k = O_P[(nh/d_n)^{1/2}], \quad (8.15)$$

$$\sum_{1 \leq k < j \leq n} \Lambda(\eta_{k-s}) \Lambda(\eta_{j-s_1}) z_{k,j} = o_P(nh/d_n). \quad (8.16)$$

Consequently, if we let $\tilde{\eta}_i = \hat{\eta}_i - \mathbb{E}\hat{\eta}_i$, where $\hat{\eta}_i = \eta_i I(|\eta_i| \leq C)$, and $\tilde{u}_{1k} = \sum_{j=0}^{m_0} \psi_j \tilde{\eta}_{k-j}$, then

$$\sum_{k=1}^n (\tilde{u}_{1k}^2 - \mathbb{E}\tilde{u}_{1k}^2) z_k = O_P[(nh/d_n)^{1/2}], \quad (8.17)$$

$$\sum_{1 \leq k < j \leq n} \tilde{u}_{1k} \tilde{u}_{1j} z_{k,j} = o_P(nh/d_n). \quad (8.18)$$

Proof. We only prove (8.16). The proof of (8.15) is similar but simpler. The proofs of (8.17) and (8.18) follow easily from (8.15) and (8.16), respectively. We omit the details.

For notational convenience, write $\delta_{ij} = \Lambda(\eta_{i-j})$. By symmetry and $|\Lambda(x)| \leq C$, it follows that

$$\mathbb{E} \left| \sum_{1 \leq k < j \leq n} \delta_{ks} \delta_{js_1} z_{k,j} \right|^2 \leq c_1 I_1(n) + c_2 I_2(n) + c_3 I_3(n), \quad (8.19)$$

for some constants $c_1, c_2, c_3 > 0$, where

$$\begin{aligned}
I_1(n) &= \sum_{1 \leq k < j \leq n} \mathbb{E}(z_{k,j}^2), & I_2(n) &= \sum_{1 \leq k \neq j \neq l \leq n} \mathbb{E}\{z_{k,j} z_{k,l}\}, \\
I_3(n) &= \sum_{1 \leq k < j < l < m \leq n} \left| \mathbb{E}\{\delta_{ks} \delta_{js_1} \delta_{ls} \delta_{ms_1} z_{k,j} z_{l,m}\} \right|.
\end{aligned}$$

As $K(x)$ has compact support (say $K(x) = 0$ if $|x| \geq M$), it is readily seen that, for any $t \in \mathbb{R}$,

$$K(t - x/h)K(t - y/h) = 0,$$

whenever $|y - x| \geq Mh$. This implies that

$$\begin{aligned} \mathbb{E}(z_{k,j} z_{k,l}) &= \mathbb{E} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K[(x_k - x)/h] K[(x_k - y)/h] \right. \\ &\quad \left. K[(x_j - x)/h] K[(x_l - y)/h] p(x)p(y) dx dy \right\} \\ &\leq \int_{-\infty}^{\infty} \int_{|y-x| \leq Mh} \mathbb{E} \left\{ K[(x_k - x)/h] \right. \\ &\quad \left. K[(x_j - x)/h] K[(x_l - y)/h] \right\} p(x)p(y) dx dy. \end{aligned}$$

Now (8.10) of Lemma 8.2 yields

$$\begin{aligned} I_2(n) &\leq \int_{-\infty}^{\infty} \int_{|y-x| \leq Mh} \mathbb{E} \sum_{1 \leq k \neq j \neq l \leq n} \left\{ K[(x_k - x)/h] \right. \\ &\quad \left. K[(x_j - x)/h] K[(x_l - y)/h] \right\} p(x)p(y) dx dy \\ &\leq \sup_x \mathbb{E} \left(\sum_{k=1}^n K[(x_k - x)/h] \right)^3 \int_{-\infty}^{\infty} \int_{|y-x| \leq Mh} p(x)p(y) dx dy \\ &\leq C (n/d_n)^3 h^4. \end{aligned} \tag{8.20}$$

Similarly, we have

$$\begin{aligned} I_1(n) &\leq \sup_x \mathbb{E} \left(\sum_{k=1}^n K[(x_k - x)/h] \right)^2 \int_{-\infty}^{\infty} \int_{|y-x| \leq Mh} p(x)p(y) dx dy \\ &\leq C (n/d_n)^2 h^3. \end{aligned} \tag{8.21}$$

We next consider I_{3n} . Note that $s_1 \leq m_0$ and, due to Lemma 8.1,

$$|\mathbb{E}(\delta_{ms_1} K[(x_m - y)/h] | \mathcal{F}_l)| \leq \begin{cases} Ch/d_{m-l}, & \text{if } m-l \leq m_0 + 1 \\ Ch/d_{m-l}^2, & \text{if } m-l > m_0 + 1. \end{cases}$$

We have that

$$\begin{aligned} &\sup_x \sum_{m=l+1}^n |\mathbb{E}(\delta_{ms_1} K[(x_m - y)/h] | \mathcal{F}_l)| \\ &\leq Ch \sum_{m=l+1}^{l+m_0} d_{m-l} + Ch \sum_{m=l+m_0}^m d_{m-l}^{-2} \leq Ch \log n. \end{aligned}$$

This result, together with (8.10), implies that

$$\begin{aligned}
I_{3n} &= \sum_{1 \leq k < j < l < m \leq n} |\mathbb{E}\{\delta_{ks} \delta_{js_1} \delta_{ls} \delta_{ms_1} z_{k,j} z_{l,m}\}| \\
&\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{1 \leq k < j < l < m \leq n} |\mathbb{E}\{\delta_{ks} \delta_{js_1} \delta_{ls} \delta_{ms_1} K[(x_k - x)/h] K[(x_j - x)/h] \\
&\quad K[(x_l - y)/h] K[(x_m - y)/h]\}| p(x)p(y) dx dy \\
&\leq C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{1 \leq k < j < l < m} \mathbb{E}\{K[(x_k - x)/h] K[(x_j - x)/h] K[(x_l - y)/h] \\
&\quad \sum_{m=l+1}^n |\mathbb{E}(\delta_{ms_1} K[(x_m - y)/h] | \mathcal{F}_l)|\} p(x)p(y) dx dy, \\
&\leq Ch \log n \sup_x \mathbb{E}\left(\sum_{k=1}^n K[(x_k - x)/h]\right)^3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x)p(y) dx dy \\
&\leq C (n/d_n)^3 h^4 \log n. \tag{8.22}
\end{aligned}$$

Taking the estimates (8.20)-(8.22) into (8.19), we obtain

$$\mathbb{E} \left| \sum_{1 \leq k < j \leq n} \delta_{ks} \delta_{js_1} z_{k,j} \right|^2 \leq C [(n/d_n)^3 h^4 \log n + (n/d_n)^2 h^3] = o[(nh/d_n)^2],$$

whenever $nh^2 \log n/d_n \rightarrow 0$. This proves (8.16), and also completes the proof of Lemma 8.3. \square

In the following Lemmas 8.4 and 8.5, let $r(x)$ be real function such that $\int_{-\infty}^{\infty} |r(x)| dx < \infty$ and $\int |\hat{r}(t)| dt < \infty$, where $\hat{r}(t) = \int e^{itx} r(x) dx$. Under this condition, we have

$$r(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \hat{r}(t) dt. \tag{8.23}$$

Recall the definitions (8.1)-(8.3). Except when mentioned explicitly, we still make use of the notation given there.

LEMMA 8.4. *Let $u_k(m_0) = \sum_{i=0}^{m_0} \psi_i \eta_{k-i}$ with m_0 being a fixed constant. There exist $\gamma > 0$ and $A_0 > 0$ such that for all $j - k \geq A_0, k \geq s$ and $u \in \mathbb{R}$,*

$$\begin{aligned}
\Lambda_k(y, j) &:= |\mathbb{E}\{r[(y + x'_{s,j})/h] u_j(m_0) \exp(i\mu \sum_{q=k+1}^n \epsilon_q/\sqrt{n}) | \mathcal{F}_k\}| \\
&\leq C(h/d_{j-k}^2 + e^{-\gamma(j-k)}); \tag{8.24}
\end{aligned}$$

for all $j - k \geq 1, k \geq s + 1$ and $u \in \mathbb{R}$,

$$\Lambda_k(y, j) \leq C h^{1/2} \{\mathbb{E}[u_j^2(m_0) | \mathcal{F}_k]\}^{1/2}. \tag{8.25}$$

In consequence, for any $u \in \mathbb{R}$, we have

$$\sup_{y, 0 \leq s < m \leq n} \left| \sum_{k=s+1}^m \mathbb{E} I_k(y, u) \right| = o[(nh/d_n)^{1/2}], \quad (8.26)$$

$$\sup_{y, 0 \leq s < m \leq n} \left| \sum_{s+1 \leq k < j \leq m} \mathbb{E} I_{k,j}(y, u) \right| = o(nh/d_n), \quad (8.27)$$

where

$$I_k(y, u) = r[(y + x'_{s,k})/h] u_k(m_0) \exp \left\{ i\mu \sum_{j=1}^k \epsilon_j / \sqrt{n} \right\},$$

$$I_{k,j}(y, u) = r[(y + x'_{s,k})/h] r[(y + x'_{s,j})/h] u_k(m_0) u_j(m_0) \exp \left\{ i\mu \sum_{j=1}^l \epsilon_j / \sqrt{n} \right\}.$$

Proof. We first prove (8.24). Let

$$z^{(1)} = \sum_{q=k+1}^{j-m_0} \epsilon_q (t a_{j-q} + u h / \sqrt{n}), \quad z^{(2)} = \sum_{q=j-m_0+1}^j \epsilon_q (t a_{j-q} + u h / \sqrt{n}).$$

It follows from (8.23) and the independence of ϵ_k that

$$\Lambda_k(y, j) \leq \frac{1}{2\pi} \int |\mathbb{E} e^{iz^{(1)}/h}| |\mathbb{E} \{ e^{iz^{(2)}/h} u_j(m_0) \}| |\hat{r}(t)| dt.$$

We may take n sufficiently large so that u/\sqrt{n} is as small as required. Without loss of generality we assume $u = 0$ in the following proof for convenience. Recall (8.2) and (8.3).

There exists an $A_0 > 0$ such that $j - m_0 \geq (j + k)/2$ and

$$\delta_1 |a_{j-k}| \leq |a_{j-q}| \leq \delta_2 |a_{j-k}|, \quad k \leq q \leq (j + k)/2,$$

for some $0 < \delta_1 < \delta_2$ and all $j - k \geq A_0$. On the other hand, there exist constants $\gamma_1 > 0$ and $\gamma_2 > 0$ such that

$$|\mathbb{E} e^{i\epsilon_1 t}| \leq \begin{cases} e^{-\gamma_1} & \text{if } |t| \geq 1, \\ e^{-\gamma_2 t^2} & \text{if } |t| \leq 1, \end{cases}$$

since $\mathbb{E}\epsilon_1 = 0$, $\mathbb{E}\epsilon_1^2 = 1$ and ϵ_1 has a density. By virtue of these facts, simple calculations show that, for any $\delta_1 \leq \delta \leq \delta_2$, there exists $\gamma > 0$ such that

$$\begin{aligned} \Lambda_k(y, j) &\leq \frac{1}{2\pi} \left(\int_{|t| \geq \delta h / |a_{j-k}|} + \int_{|t| \leq \delta h / |a_{j-k}|} \right) |\mathbb{E} e^{iz^{(1)}/h}| |\mathbb{E} \{ e^{iz^{(2)}/h} u_j(m_0) \}| |\hat{r}(t)| dt \\ &\leq C e^{-\gamma(j-k)} \int |\hat{r}(t)| dt + C \int_{|t| \leq \delta h / |a_{j-k}|} e^{-\gamma t^2 (d_{j-k}/h)^2} (|t| h^{-1}) dt \\ &\leq C(h/d_{j-k}^2 + e^{-\gamma(j-k)}), \end{aligned}$$

where we have used the following fact: due to $\mathbb{E}u_j(m_0) = 0$, we have

$$\begin{aligned} |\mathbb{E}\{e^{iz^{(2)}/h} u_j(m_0)\}| &= |\mathbb{E}\{(e^{iz^{(2)}/h} - 1) u_j(m_0)\}| \\ &\leq C |th^{-1}| \sum_{j=0}^{m_0} |\phi_j| \mathbb{E}(|\epsilon_1 u_j(m_0)|) \leq C_1 |th^{-1}|. \end{aligned}$$

This proves (8.24).

The proof of (8.25) is simple. Indeed, by noting that (8.7) remains to be true for any $t - s \geq 1$ if $\Lambda \equiv 1$, we have

$$\begin{aligned} \Lambda_k(y, j) &\leq (\mathbb{E}\{|r[(y + x'_{s,j})/h]|^2 \mid \mathcal{F}_k\})^{1/2} \{\mathbb{E}[u_j^2(m_0) \mid \mathcal{F}_k]\}^{1/2} \\ &\leq Ch^{1/2} \{\mathbb{E}[u_j^2(m_0) \mid \mathcal{F}_k]\}^{1/2}, \end{aligned}$$

as required.

We next prove (8.27). Due to $h \rightarrow 0$, it follows from (8.24)-(8.25) and Lemma 8.1 that, for $j - k \geq \log h^{-1}$,

$$\begin{aligned} |\mathbb{E}I_{k,j}(y, u)| &\leq \mathbb{E}\{|r[(y + x'_{s,k})/h]| |u_k(m_0)| |\Lambda_k(y, j)|\} \\ &\leq C(h/d_{j-k}^2 + e^{-\gamma(j-k)}) \mathbb{E}\{|r[(y + x'_{s,k})/h]| |u_k(m_0)|\} \\ &\leq C(h/d_{j-k}^2 + e^{-\gamma(j-k)}) \begin{cases} h^{1/2}, & \text{for } 1 \leq k - s \leq A_0, \\ h/d_k, & \text{for } k - s \geq A_0 + 1, \end{cases} \end{aligned}$$

and for $j - k \leq \log h^{-1}$,

$$\begin{aligned} |\mathbb{E}I_{k,j}(y, u)| &\leq Ch^{1/2} \mathbb{E}\{|r[(y + x'_{s,k})/h]| |u_k(m_0)| \{\mathbb{E}[u_j^2(m_0) \mid \mathcal{F}_k]\}^{1/2}\} \\ &\leq Ch^{1/2} \begin{cases} h^{1/2}, & \text{for } 1 \leq k - s \leq A_0, \\ h/d_k, & \text{for } k - s \geq A_0 + 1, \end{cases} \end{aligned}$$

where A_0 is a constant given as in Lemma 8.1. These facts imply that, for any $u \in \mathbb{R}$,

$$\begin{aligned} &\sup_{y, 0 \leq s < m \leq n} \left| \sum_{s+1 \leq k < j \leq m} \mathbb{E}I_{k,j}(y, u) \right| \\ &\leq Ch + \sum_{k=A_0+1}^{n-1} \left(\sum_{j=k+1}^{k+\log h^{-1}} + \sum_{j=k+\log h^{-1}+1}^n \right) \sup_y |\mathbb{E}I_{k,j}(y, u)| \\ &\leq Ch + Ch^{3/2} \sum_{k=A_0+1}^{n-1} d_k^{-1} + C \sum_{k=A_0+1}^{n-1} \sum_{j=k+\log h^{-1}+1}^n h d_k^{-1} (h/d_{j-k}^2 + e^{-\gamma(j-k)}) \\ &\leq Cnh^{3/2}/d_n + C(nh/d_n) \sum_{j=\log h^{-1}+1}^n (h/d_j^2 + e^{-\gamma j}) \\ &\leq C(nh/d_n) \begin{cases} h^{\min\{1/2, \gamma\}}, & \text{under } \mathbf{C1}, \\ h^{1/2} + h \log n, & \text{under } \mathbf{C2}, \end{cases} \\ &= o(nh/d_n), \end{aligned}$$

due to $h \rightarrow 0$ and $h \log n \rightarrow 0$ under **C2**, which yields (8.27). The proof of (8.26) is similar and hence the details are omitted. \square

LEMMA 8.5. *Let $u_k(m_0) = \sum_{i=0}^{m_0} \psi_i \eta_{k-i}$ with m_0 being a fixed constant. Write, for $0 \leq t \leq 1$,*

$$\begin{aligned}\psi_{1n}(t) &= \frac{d_n}{nh} \sum_{k=1}^{[nt]} r[(x_k - x)/h], & \psi_n(t) &= \frac{d_n}{nh} \sum_{k=1}^{[nt]} u_k^2(m_0) r[(x_k - x)/h], \\ \eta_n(t) &= \left(\frac{d_n}{nh}\right)^{1/2} \sum_{k=1}^{[nt]} u_k(m_0) r[(x_k - x)/h].\end{aligned}$$

(i) *For any $h \rightarrow 0$ and $nh/d_n \rightarrow \infty$, we have*

$$\psi_{1n}(t) \Rightarrow \int r(x) dx L_\psi(1, 0), \quad \psi_n(t) \Rightarrow \tau L_\psi(1, 0).$$

on $D[0, 1]$, where $\tau = \mathbb{E}u_1^2(m_0) \int r^2(x) dx$.

(ii) *For any fixed $0 \leq t \leq 1$, $\eta_n(t)$, $\eta_n^2(t)$, and $\psi_n(t)$, $n \geq 1$ are integrable.*

(iii) *$\eta_n(t)$ is tight on $D[0, 1]$.*

Proof. For the proof of part (i) see Proposition 3.2 of Wang and Phillips (2011). The proofs of parts (ii) and (iii) are similar to Propositions 7.3 and 7.4 of WP (2009b), respectively, only requiring the replacement of Lemmas 7.1 and 7.2 there by Lemma 8.4 of the present paper. We omit the details. \square

8.2 Proofs of Propositions

Proof of Proposition 7.1. Result (7.1) follows from WP (2009a), and (7.5) follows from a simple application of Theorem 2.3 in Chan and Wang (2013a). By virtue of (8.6) in Lemma 8.1 and (8.12) in Lemma 8.2, we have (7.2) and (7.3), respectively. To prove (7.4), let $u_k(m_0) = \sum_{j=0}^{m_0} \psi_j \eta_{k-j}$. The result (8.13) in Lemma 8.2 establishes that, as first $n \rightarrow \infty$ and then $m_0 \rightarrow \infty$,

$$\sum_{k=1}^n [u_k - u_k(m_0)] K[(x_k - x)/h] = o_P[(nh/d_n)^{1/2}].$$

This implies that (7.4) holds if we prove the following: for any $m_0 > 0$,

$$\begin{aligned}& \left\{ \left(\frac{d_n}{nh}\right)^{1/2} \sum_{k=1}^n u_k(m_0) K[(x_k - x)/h], \frac{d_n}{nh} \sum_{k=1}^n g[(x_k - x)/h] \right\} \\ \rightarrow_D & \left\{ \tau_0 N L_\psi^{1/2}(1, 0), \int_{-\infty}^{\infty} g(x) dx L_\psi(1, 0) \right\}\end{aligned}\tag{8.28}$$

where $\tau_0^2 = \mathbb{E}u_0^2(m_0) \int K^2(s)ds$, N is a standard normal variate independent of $L_\psi(1, 0)$. Recalling Lemmas 8.4 and 8.5, the outline for the proof of (8.28) is exactly the same as that of (3.8) in WP (2009b). We omit the details. \square

Proof of Proposition 7.2. The result (7.6) is well known. See, Berkes and Horváth (2006), for instance. Note that $\sum_{k=1}^n u_k^2 = O_P(n)$ and $H^2(x)$ is a locally bounded function. (7.7) follows from (7.6) and the Hölder inequality:

$$\sum_{k=1}^n |H(x_k/d_n)| (1 + |u_k|) \leq \left(\sum_{k=1}^n |H(x_k/d_n)|^2 \right)^{1/2} \left(\sum_{k=1}^n (1 + |u_k|)^2 \right)^{1/2}.$$

To prove (7.9), we write

$$\begin{aligned} \sum_{k=1}^n g(x_k)u_k &= \sum_{k=1}^n g(x_k) \left(\sum_{j=0}^k + \sum_{j=k+1}^{\infty} \right) \psi_j \eta_{k-j} \\ &= \sum_{j=0}^n \sum_{k=j+1}^n [g(x_k) - g(x_{k-j-1})] \psi_j \eta_{k-j} \\ &\quad + \sum_{j=0}^n \sum_{k=1}^{n-j} g(x_{k-1}) \psi_j \eta_k + \sum_{k=1}^n g(x_k) \sum_{j=1}^{\infty} \psi_{j+k} \eta_{-j} \\ &= S_{1n} + S_{2n} + S_{3n}, \quad \text{say.} \end{aligned} \tag{8.29}$$

Let $\Omega_M = \{x_j : \max_{1 \leq j \leq n} |x_j|/d_n \leq M\}$. It follows from (7.8) that

$$\begin{aligned} \mathbb{E}|S_{1n}I(\Omega_n)| &\leq C v(d_n) \sum_{j=0}^n \sum_{k=j+1}^n \mathbb{E}\{d_n^{-1} |x_k - x_{k-j-1}| |\psi_j \eta_{k-j}|\} \\ &\quad + C v(d_n) \sum_{j=0}^n \sum_{k=j+1}^n \mathbb{E}\{(|R_1(x_k)| + |R_2(x_{k-j})|) |\psi_j \eta_{k-j}|\} \\ &\leq \frac{Cv(d_n)}{d_n} \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|) \sum_{k=j+1}^n \mathbb{E}\{|x_k - x_{k-j-1}| (|\nu_{k-j}| + |\epsilon_{k-j}|)\} \\ &\quad + C v(d_n) \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|) \sum_{k=j+1}^n d_k^{-1} \\ &\leq \frac{Cnv(d_n)}{d_n} \sum_{j=0}^n d_j (|\psi_{1j}| + |\psi_{2j}|) \\ &\leq Cv(d_n) \sqrt{n} \sum_{j=0}^{\infty} j^{1/2} (|\psi_{1j}| + |\psi_{2j}|), \end{aligned}$$

where we have used (8.6) of Lemma 8.1 and the following fact from Hölder's inequality

$$\mathbb{E}\{|x_k - x_{k-j-1}| (|\nu_{k-j}| + |\epsilon_{k-j}|)\} \leq C (\mathbb{E}|x_k - x_{k-j-1}|^2)^{1/2} \leq Cd_j.$$

Similarly, it follows from $g(\lambda x) \leq C v(\lambda) (1 + |x|^\beta)$ that

$$\begin{aligned}
\mathbb{E}|S_{3n}I(\Omega_n)| &\leq C v(d_n) \sum_{k=1}^n \sum_{j=1}^{\infty} \mathbb{E}|\psi_{j+k}\eta_{-j}| \\
&\leq C v(d_n) \sum_{k=1}^n \sum_{j=k+1}^{\infty} (|\psi_{1j}| + |\psi_{2j}|) \\
&\leq C v(d_n) \sum_{k=1}^n k^{-1/2} \sum_{j=k+1}^{\infty} j^{1/2} (|\psi_{1j}| + |\psi_{2j}|) \\
&\leq C v(d_n) \sqrt{n} \sum_{j=0}^{\infty} j^{1/2} (|\psi_{1j}| + |\psi_{2j}|).
\end{aligned}$$

Hence $|S_{1n}| + |S_{3n}| = O_P[v(d_n)\sqrt{n}]$ due to $P(\max_{1 \leq j \leq n} |x_j|/d_n \geq M) \rightarrow 0$ as $M \rightarrow \infty$.

To estimate S_{2n} , let $\tilde{x}_k = x_k I(\max_{1 \leq j \leq k} |x_j|/d_n \leq M)$ and

$$\tilde{S}_{2n} = \sum_{j=0}^n \sum_{k=1}^{n-j} g(\tilde{x}_{k-1}) \psi_j \eta_k.$$

Due to the independence of η_k and $g(\lambda x) \leq C v(\lambda) (1 + |x|^\beta)$, we have

$$\begin{aligned}
\mathbb{E}|\tilde{S}_{2n}| &\leq \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|) \left\{ \mathbb{E} \left| \sum_{k=1}^{n-j} g(\tilde{x}_{k-1}) \nu_k \right| + \mathbb{E} \left| \sum_{k=1}^{n-j} g(\tilde{x}_{k-1}) \epsilon_k \right| \right\} \\
&\leq C \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|) \left\{ \sum_{k=1}^{n-j} \mathbb{E} g^2(\tilde{x}_{k-1}) \right\}^{1/2} \\
&\leq C v(d_n) \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|) \left\{ \sum_{k=1}^{n-j} \mathbb{E} [1 + |\tilde{x}_{k-1}/d_n|^\beta] \right\}^{1/2} \\
&\leq C v(d_n) \sqrt{n} \sum_{j=0}^n (|\psi_{1j}| + |\psi_{2j}|),
\end{aligned}$$

so that $\tilde{S}_{2n} = O_P[v(d_n)\sqrt{n}]$. This yields $|S_{2n}| = O_P[v(d_n)\sqrt{n}]$ as $S_{2n} = \tilde{S}_{2n}$ on Ω_M and $P(\max_{1 \leq j \leq n} |x_j|/d_n \geq M) \rightarrow 0$ as $M \rightarrow \infty$.

Combining all the above estimates, we obtain that

$$\left| \sum_{k=1}^n g(x_k/d_n) u_k \right| \leq |S_{1n}| + |S_{2n}| + |S_{3n}| = O_P[v(d_n)\sqrt{n}],$$

as required. \square

Proof of Proposition 7.3. Result (7.10) follows by direct application of (8.13) in Lemma 8.2. To prove (7.11), for a fixed $A > 0$, let $\hat{\eta}_i = \eta_i I(|\eta_i| \leq A)$,

$$\tilde{\eta}_i = \hat{\eta}_i - \mathbb{E}\hat{\eta}_i, \quad \check{\eta}_i = \eta_i - \tilde{\eta}_i, \quad \tilde{u}_k = \sum_{j=0}^{\infty} \psi_j \tilde{\eta}_{k-j}, \quad \check{u}_k = \sum_{j=0}^{\infty} \psi_j \check{\eta}_{k-j}$$

Note that $\mathbb{E}\check{\eta}_1 = 0$. It follows from (8.13) of Lemma 8.2 that

$$\begin{aligned} \mathbb{E} \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \check{u}_k \right\}^2 \pi(x) dx &\leq \mathbb{E} \|\check{\eta}_1\|^2 nh/d_n \\ &\leq C \mathbb{E} \|\eta_1\|^2 I(\|\eta_1\| > A) nh/d_n. \end{aligned}$$

As $\mathbb{E} \|\eta_1\|^2 I(\|\eta_1\| > A) \rightarrow 0$ as $A \rightarrow \infty$ and $u_k = \tilde{u}_k + \check{u}_k$, simple calculations show that (7.11) will follow if we prove

$$\begin{aligned} \frac{d_n}{nh} T_n &:= \frac{d_n}{nh} \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \tilde{u}_k \right\}^2 \pi(x) dx \\ &\rightarrow_D \mathbb{E} \tilde{u}_0^2 \int_{-\infty}^{\infty} K^2(x) dx \int_{-\infty}^{\infty} \pi(s) ds L_\psi(1, 0), \end{aligned} \quad (8.30)$$

for each $A > 0$.

To prove (8.30), we let $\tilde{u}_{1k} = \sum_{j=0}^{m_0} \psi_j \tilde{\eta}_{k-j}$ and $\tilde{u}_{2k} = \tilde{u}_k - \tilde{u}_{1k}$. It is readily seen that

$$T_n = T_{1n} + T_{2n} + T_{3n}, \quad (8.31)$$

where $T_{3n} \leq 2T_{1n}^{1/2} T_{2n}^{1/2}$, $T_{2n} = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \tilde{u}_{2k} \right\}^2 \pi(x) dx$ and

$$\begin{aligned} T_{1n} &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \tilde{u}_{1k} \right\}^2 \pi(x) dx \\ &= \mathbb{E} \tilde{u}_{10}^2 \sum_{k=1}^n \int_{-\infty}^{\infty} K^2[(x_k - x)/h] \pi(x) dx + \sum_{k=1}^n (\tilde{u}_{1k}^2 - \mathbb{E} \tilde{u}_{1k}^2) \int_{-\infty}^{\infty} K^2[(x_k - x)/h] \pi(x) dx \\ &\quad + 2 \sum_{1 \leq k < j \leq n} \tilde{u}_{1k} \tilde{u}_{1j} \int_{-\infty}^{\infty} K[(x_k - x)/h] K[(x_j - x)/h] \pi(x) dx \\ &= R_{1n} + R_{2n} + R_{3n}, \quad \text{say.} \end{aligned}$$

Due to (8.13) of Lemma 8.2, we have

$$\begin{aligned} \mathbb{E} T_{2n} &\leq C \sup_x \mathbb{E} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \tilde{u}_{2k} \right\}^2 \\ &\leq C \mathbb{E} \|\eta_1\|^2 (nh/d_n) \sum_{k=m_0}^{\infty} k^{1/4} (|\psi_{1k}| + |\psi_{2k}|), \end{aligned}$$

i.e., $T_{2n} = o_P(nh/d_n)$, as $n \rightarrow \infty$ first and then $m_0 \rightarrow 0$. Furthermore it follows from (8.17) and (8.18) of Lemma 8.3 that $|R_{2n}| + |R_{3n}| = o_P(nh/d_n)$ for each $A > 0$. Now, by virtue of (8.31) and $\mathbb{E} \tilde{u}_{10}^2 \rightarrow \mathbb{E} \tilde{u}_0^2$ as $m_0 \rightarrow \infty$, (8.30) will follow if we prove

$$\frac{d_n}{nh} R_{1n} \rightarrow_D \mathbb{E} \tilde{u}_{10}^2 \int_{-\infty}^{\infty} K^2(x) dx \int_{-\infty}^{\infty} \pi(s) ds L_\psi(1, 0). \quad (8.32)$$

Let $g_h(y) = h^{-1} \int_{-\infty}^{\infty} K^2(y-x/h)\pi(x)dx$. It is readily seen that $\int g_h(y)dy = \int_{-\infty}^{\infty} K^2(z)dz \int_{-\infty}^{\infty} \pi(x)dx$ and $|g_h(y)| \leq C \int_{-\infty}^{\infty} |\pi(x)|dx < \infty$. Then (8.32) follows from theorem 2.1 of WP (2009a) with minor modification. This completes the proof of (7.11).

We next prove (7.12). We may write

$$\int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] m(x_k) \right\}^2 \pi(x)dx = V_{1n} + V_{2n} + 2V_{3n}, \quad (8.33)$$

where $V_{3n} \leq V_{1n}^{1/2}V_{2n}^{1/2}$, $V_{2n} = \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] [m(x_k) - m(x)] \right\}^2 \pi(x)dx$ and

$$\begin{aligned} V_{1n} &= \int_{-\infty}^{\infty} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 m^2(x)\pi(x)dx, \\ &= \left\{ \sum_{k=1}^n K(x_k/h) \right\}^2 \int_{|x| \leq \log n} m^2(x)\pi(x)dx \\ &\quad + \int_{|x| \leq \log n} \left[\left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 - \left\{ \sum_{k=1}^n K(x_k/h) \right\}^2 \right] m^2(x)\pi(x)dx \\ &\quad + \int_{|x| \geq \log n} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 m^2(x)\pi(x)dx \\ &= V_{1n}^{(1)} + V_{2n}^{(2)} + V_{3n}^{(3)}, \quad \text{say.} \end{aligned}$$

Since $K(x)$ has compact support, it follows from Assumption 3.4 and (8.10) that

$$\begin{aligned} \mathbb{E}V_{2n} &\leq Ch^\gamma \int_{-\infty}^{\infty} \mathbb{E} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 m_1^2(x)\pi(x)dx \\ &\leq Ch^\gamma \sup_x \mathbb{E} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 \int_{-\infty}^{\infty} m_1^2(x)\pi(x)dx \\ &\leq Ch^\gamma (nh/d_n)^2. \end{aligned}$$

This yields $V_{2n} = o_P[(nh/d_n)^2]$. Similarly, by virtue of part (iii) of Proposition 7.1 and (8.10), we have

$$\begin{aligned} |V_{2n}^{(2)}| &\leq \sup_{|x| \leq \log n} \left| \sum_{k=1}^n \{K[(x_k - x)/h] - K(x_k/h)\} \right| \times \\ &\quad \int_{|x| \leq \log n} \sum_{k=1}^n \{K[(x_k - x)/h] + K(x_k/h)\} m^2(x)\pi(x)dx \\ &= o_P[(nh/d_n)^2], \\ |V_{2n}^{(3)}| &\leq \sup_x \mathbb{E} \left\{ \sum_{k=1}^n K[(x_k - x)/h] \right\}^2 \int_{|x| \geq \log n} m^2(x)\pi(x)dx \\ &= o_P[(nh/d_n)^2]. \end{aligned}$$

Taking these estimates into (8.33), (7.12) will follow if

$$\left(\frac{d_n}{nh}\right)^2 V_{1n}^{(1)} \rightarrow_D \tau_1 L_\psi^2(1, 0).$$

But this follows from $\frac{d_n}{nh} \sum_{k=1}^n K(x_k/h) \rightarrow_D \int_{-\infty}^{\infty} K(x) dx L_\psi(1, 0)$ and the continuous mapping theorem. The proof of Proposition 7.3 is now complete. \square

REFERENCES

- Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, **72**, 569-614.
- Cai, Z., Li, Q. and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101–113.
- Chan, N. and Wang, Q. (2013a). Uniform convergence for nonparametric estimators with non-stationary data. *Econometric Theory*, forthcoming.
- Chan, N. and Wang, Q. (2013b). Nonlinear regression with nonstationary time series, *Journal of Econometrics*, revised.
- Chang, Y. and Park, J. Y. (2001). Index models with integrated time series. *Journal of Econometrics*, **114**, 73–106.
- Chang, Y. and Park, J. Y. (2011). Endogeneity in nonlinear regressions with integrated time series, *Econometric Reviews*, **30**, 51–87.
- Chang, Y., Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, **4**, 1–36.
- Chen, X., Liao, Z., and Sun, Y. (2014). Sieve inference on possibly misspecified semi-nonparametric time series models, *Journal of Econometrics*, **178**, 639-658.
- Choi, I., Saikkonen, P., (2004). Testing linearity in cointegrating smooth transition regressions. *Econometrics Journal*, **7**, 341–365.
- Choi, I. and Saikkonen, P., (2010). Tests for nonlinear cointegration. *Econometric Theory*, **26**, 682–709.
- Duffy, J. (2013). Uniform convergence rates on a maximal domain for structural non-parametric cointegrating regression. Unpublished paper, Yale University.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. London: Chapman and Hall.
- Gao, J., K., Maxwell, Lu, Z., Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, **25**, 1869–1892.
- Gao, J., K., Maxwell, Lu, Z., Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics*, **37**, 3893–3928.
- Gao, J., K., Tjøstheim, D. and Jiying Yin (2012). Model specification between parametric and nonparametric cointegration. Working paper.

- Geman, D. and Horowitz, J. (1980). Occupation densities. *Annals of Probability*, **8**, 1–67.
- Granger, C. W. J. and Teräsvirta, T (1993), *Modelling nonlinear economic relationships*, Oxford University Press, New York.
- Guerre, E. (2004). Design-Adaptive pointwise nonparametric regression estimation for recurrent Markov time series. *Unpublished manuscript*.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*. **21**, 1926-1947.
- Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599-631.
- Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics*. **29**, no. 2, 372–416.
- Karlsen, H. A., Myklebust, T. and Tjostheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model, *Annals of Statistics*, **35**, 252-299.
- Kasparis, I. and Phillips, P.C.B. (2009). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*, **168**, 270-284.
- Kasparis, I., Andreou, and Phillips, P.C.B. (2013). Nonparametric Predictive Regression. Unpublished Working Paper, Yale University.
- Marmer, V. (2008). Nonlinearity, nonstationarity, and spurious forecasts. *Journal of Econometrics* **142**, 1–27.
- Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for nonlinear transformation of integrated time series, *Econometric Theory*, **15**, 269-298.
- Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series, *Econometrica*, **69**, 117-161.
- Phillips, P. C. B. (2005). "HAC Estimation by Automated Regression." *Econometric Theory*, **21**, 116-142.
- Phillips, P. C. B. and J. Y. Park (1998). Nonstationary Density Estimation and Kernel Autoregression. *Cowles Foundation Discussion Paper No. 1181*.
- Schienze, M. (2008). Nonparametric Nonstationary Regression, Unpublished Ph.D. Thesis, University of Mannheim.
- Shimotsu, K. and P. C. B. Phillips (2005). "Exact Local Whittle Estimation of Fractional Integration," *Annals of Statistics* **33**(4) 1890-1933.

- Sun, Y. (2011). “Robust trend inference with series variance estimator and testing-optimal smoothing parameter.” *Journal of Econometrics*, 164, 345–366.
- Sun, Y. (2014). “Let’s fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference”, *Journal of Econometrics*, 178, 659–677.
- Sun, Yiguo, Cai, Z. and Li, Q. (2013). Consistent nonparametric test on parametric smooth coefficients with nonstationary data, *Manuscript*.
- Wang, Q. (2013). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, forthcoming.
- Wang, Q. and Chan, N. (2013). Uniform convergence rates for a class of martingales with application in non-linear co-integrating regression. *Bernoulli*, forthcoming.
- Wang, Q. and P. C. B. Phillips (2009a). Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression, *Econometric Theory*, **25**, 710-738.
- Wang, Q. and P. C. B. Phillips (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **77**, 1901-1948.
- Wang, Q. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, **27**, 235-259.
- Wang, Q. and Phillips, P. C. B. (2012). A Specification Test for Nonlinear Nonstationary Models. *Annals of Statistics*, **40**, 727-758.
- Wang, Q., Lin, Y. X., Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, **19**, 143–164.
- Wang, Q. and Wang, R. (2013). Non-parametric cointegrating regression with NNH errors. *Econometric Theory*, **29**, 1–27.