CrossMark

Econometric Theory, **34**, 2018, 754–789.
doi:10.1017/S0266466617000238

SPECIFICATION TESTING DRIVEN BY ORTHOGONAL SERIES FOR NONLINEAR COINTEGRATION WITH ENDOGENEITY

CHAOHUA DONG
Southwestern University of Finance and Economics

JITI GAO
Monash University

This paper proposes two simple and new specification tests based on the use of an orthogonal series for a considerable class of bivariate nonlinearly cointegrated time series models with endogeneity and nonstationarity. The first test is proposed for the case where the regression function is integrable, which fills a gap in the literature, and the second test, which nests the first one, deals with regression functions in a quite large function space that is sufficient for both theoretical and practical use. As a starting point of our asymptotic theory, the first test is studied initially and then the theory is extended to the second test. Endogeneity in two general forms is allowed in the models to be tested. The finite sample performance of the tests is examined through several simulated examples. Our experience generally shows that the proposed tests are easily implementable and also have stable sizes and good power properties even when the 'distance' between the null hypothesis and a sequence of local alternatives is asymptotically negligible.

1. INTRODUCTION

Econometric model estimation for nonlinear and nonstationary cointegrating models has become an increasingly active area in recent years. Contributions to the literature include Karlsen and Tjøstheim (2001) for a nonparametric kernel estimation for a class of null-recurrent time series; Karlsen, Mykelbust, and Tjøstheim (2007) for nonparametric kernel estimation in a null-recurrent

The authors acknowledge the constructive comments from Professor Peter Phillips, the Co–Editor and three referees on the earlier versions. The authors would also like to thank Professor Maxwell King for the proofreading and useful comments on a final version. Thanks also go to Dr Qiying Wang for making an earlier version of Wang and Phillips (2016) available to us while we were writing this paper as well as his constructive comments on an earlier version. The authors would like to acknowledge constructive comments and suggestions from Professor Oliver Linton and Professor Peter Robinson on earlier versions. Thanks from the authors also go to the seminar participants at various seminars for their comments and suggestions. The first author acknowledges financial support from the National Natural Science Foundation of China under Grant No. 71671143 and the second author thanks the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421. Address correspondence to Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia; e-mail: jiti.gao@monash.edu.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

cointegration model; Cai, Li, and Park (2009) for a class of functional coefficient time series models; Wang and Phillips (2009a) for estimation in a nonparametric cointegration model; Wang and Phillips (2009b) for estimation in a structural nonparametric cointegrating model under endogeneity; Gao and Phillips (2013a) for estimation in a class of non- and semi-parametric cointegration models; and Gao and Phillips (2013b) for semiparametric estimation in triangular system equations of nonstationary time series. To avoid possible model misspecification issues, such nonparametric or semiparametric estimation is normally used as the first step to suggest a possible parametric approximation when there is no prior information about which particular model should be used for a given set of data. Then, a rigorous specification procedure should be used to test whether the suggested parametric model may be accepted statistically. These contributions have led to the recent literature on nonparametric specification testing of nonlinear and nonstationary time series.

Related to the specification testing problem considered in this paper, there are some existing papers which study specification testing issues in regression models with nonstationary regressors (mainly integrated processes). Examples include Gao, King, Lu, and Tjøstheim (2009a, 2009b), Hong and Phillips (2010), and Wang and Phillips (2012). Gao et al. (2009a) propose a nonparametric kernel-based test for the parametric specification of nonstationary nonlinear models where the regressor, that is an integrated time series, is independent of the equation error term that is a martingale difference sequence. Gao et al. (2009b) consider a testing issue for a nonstationary nonlinear autoregression model where strict conditions are imposed on the density of the error term; Hong and Phillips (2010) test linearity of cointegrating relations with an application; and Wang and Phillips (2012) investigate parametric specification of nonstationary nonlinear regression models where the regression function is imposed to have a certain growth rate.

Indeed, there is a gap in the literature about specification testing for integrable regression functions (Wang and Phillips, 2012, p. 731). Also existing tests mentioned above are not applicable to the case where there is a type of endogeneity. From a practical point of view, there is need for a test that accommodates endogeneity of the type that it is inherited in the United States aggregate consumers' consumption expenditure and disposable income over the period of 1960–2009. This paper therefore proposes a test for the specification of integrable nonstationary regression models, in which two forms of endogeneity are allowed. Usually, an instrumental variable approach is needed to deal with such an endogeneity issue, although it is often difficult to find appropriate instruments in the nonlinear and nonstationary setting. As discussed in Section 3, without introducing instrumental variables, it is shown that each of the proposed tests has a simple and known limiting distribution under the null hypothesis, and is also consistent under a sequence of local alternatives.

Recently, Wang and Phillips (2016) extend a kernel based test originally proposed in Dong, Gao, Tjøstheim, and Yin (2014) to the case where there is

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

endogeneity, although their paper mainly focuses on nonparmetric kernel estimation. As discussed in Dong et al. (2014); however, the performance of such a kernel based test highly depends on the choice of an arbitrary weight function. By contrast, the tests proposed in the current paper are simple and easily implementable. Due to the integrability and orthogonality of the series employed, there is no need to involve an arbitrary weight function in the proposed tests. Moreover, our tests do not require smoothness of the regression function under consideration, and we only focus on the case where a class of local departure functions are integrable on the real line, which implies that any 'tiny' departure from the null is detectable by our tests.

Consider a nonparametric cointegration model of the form

$$\begin{aligned} y_t &= m(x_t) + e_t, \\ x_t &= D^{-\alpha_0}\{u_t I(t > 0)\}, \\ E[u_t] &= E[e_t] = 0, \end{aligned} \tag{1.1}$$

for $t = 1, \ldots, n$, where $m(\cdot)$ is an unknown function, x_t is a truncated fractional integrated nonstationary process, D is the difference operator, $\alpha_0 \in (1/2, 3/2)$, and (u_t, e_t) is endogenous as defined in Assumption A.

We are interested in testing the following hypotheses

$$H_0: \ m(x) = g(x;\theta_0) \quad \text{versus} \quad H_1: \ m(x) = g(x;\theta_1) + \triangle_n(x), \tag{1.2}$$

where $g(x;\cdot)$ is a known integrable function on $\mathbb{R}$, $\theta_0, \theta_1 \in \Theta \subset \mathbb{R}^d$ is a compact parameter space, and $\{\triangle_n(x)\}$ is a sequence of unknown functions satisfying $\lim_{n\to\infty} \triangle_n(x) = 0$ for every $x \in \mathbb{R}$. Here, we study a type of local alternative H_1, like Wang and Phillips (2012, p. 732), since a fixed alternative is of less interest (see, for example, Horowitz and Spokoiny 2001).

To broaden our research, the models in the form of (1.1) are allowed to accommodate endogeneity in two possible forms: Type (i), both $u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ and $e_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$ are linear processes driven by the same innovations $\{\epsilon_i : -\infty < i < \infty\}$; and Type (ii), $u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ is a linear process, while $e_t = \wp(\epsilon_t, \ldots, \epsilon_{t-m_0+1}; \eta_t, \ldots, \eta_{t-m_1+1})$ is a functional of some innovations drawn from (ϵ_i, η_j). There are considerable possibilities of e_t and u_t within these settings, depending on the choice of the coefficients in the linear processes and functional forms. Two extreme cases are (a) $e_t = u_t$ and (b) e_t and u_t are independent. Observe that by the definition of x_t in (1.1), the proposed tests of this paper break new ground in the nonparametric specification testing literature by being able to take both endogeneity and nonstationarity into account. These settings facilitate the application of our specification testing method and theory in practice.

Note that under the null of (1.2) we have $m(x) \in L^2(\mathbb{R})$. The proposed test statistic below basically measures the total difference of $y_t - g(x_t;\widehat{\theta})$, weighted by orthonormal basis functions in $L^2(\mathbb{R})$, at each observation point, where $\widehat{\theta}$ is a consistent estimator of θ_0 under H_0. Then, the approach is extended to the function space $L^2(\mathbb{R}, e^{-x^2})$. The main idea of the second test is to exploit a simple

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

transformation of the form $M(x) = m(x)\exp(-x^2/2)$ such that $M(x)$ is integrable on $\mathbb{R}$ even though $m(x)$ itself may not be integrable. As a result, a large class of functions, such as bounded functions, polynomials, and power functions, can be allowed for in $m(x)$. These tests clearly cover existing ones, such as those of Gao et al. (2009a) and Wang and Phillips (2012), and fill a gap in the literature.

Our experience shows that while each of the test statistics is of a simple quadratic form, it is not necessarily easy to establish and then prove its asymptotic distribution. This is mainly because existing central limit theorems available for standardised versions of quadratic forms (see, e.g., Theorem A.1 of Gao, 2007) are not applicable in our case where there is no martingale structure involved in model (1.1). For our test statistics, as shown in Section 3, asymptotic distributions are established without imposing any martingale structure and without the use of an instrumental variable approach. In equations (3.4)–(3.7) in Section 3, we shall discuss this key feature and the relationship between the stationary case and the cointegrated time series case considered in this paper.

The main contributions of the paper are summarised as follows.

- This paper proposes a new test driven by an orthogonal series for specification testing of integrable functions of fractional integrated nonstationary time series. This test is then extended to the nonintegrable function case.
- The proposed tests are applicable to models that embrace endogeneity. The endogeneity dealt with in the paper is of a quite general form.

The rest of the paper is organised as follows. Section 2 gives some preliminaries about Hermite orthogonal polynomials as well as assumptions for model (1.1). Our specification test for the case where $m(x)$ is integrable in $\mathbb{R}$ is proposed and then studied in Section 3.1. Section 3.2 discusses the nonintegrable case. Section 4 investigates the finite-sample performance of the proposed tests. Section 5 concludes the paper. Appendix A gives some useful technical lemmas. Appendix B outlines the proof of all theorems as well as Lemma 3.1 except that Theorem 3.2 is proved in detail. Detailed proofs are available from Appendices C and D in the supplemental material to this article, available at Cambridge Journals Online (journals.cambridge.org/ect), and additional simulations are shown in Appendix E to justify some implications of our theoretical results.

Throughout the paper the following notation is used, unless otherwise stated. $\|\cdot\|$ stands for Euclidean norm for vector and element-wise norm for matrix. For example, $\|A\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{1/2}$ if $A = (a_{ij})_{nm}$. $[a]$ is the maximum integer not exceeding a. $\mathbb{R}$ is the real line and $\mathbb{R}^+$ is the set of all positive real numbers. Integral $\int$ means integration on the entire real line.

2. ASSUMPTIONS

The following assumptions are made for the subsequent theoretical development. Let L be the usual lag operator and $D = 1 - L$ be the usual difference operator. Fractional difference operator $D^{-\zeta} = (1 - L)^{-\zeta}$ is given by

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

$D^{-\zeta} = \sum_{j=0}^{\infty} a_j(\zeta) L^j$, where $a_j(\zeta) = \frac{\Gamma(j+\zeta)}{\Gamma(\zeta)\Gamma(j+1)}$, with $\Gamma(\cdot)$ being the conventional Gamma function.

Let $\{\epsilon_j, j = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of independent and identically distributed (i.i.d.) continuous random variables satisfying $E[\epsilon_0] = 0$, $E[\epsilon_0^2] = 1$, and $E[\epsilon_0^4] < \infty$. Suppose further that the characteristic function of ϵ_0 satisfies $\int |u\, E e^{iu\epsilon_0}| du < \infty$.

Assumption A.

A.1. For $t \geq 1$, let $x_t = D^{-\alpha_0}\{u_t I(t > 0)\}$ with $x_0 = O_P(1)$ and $\alpha_0 \in (1/2, 3/2)$.

A.2. Suppose that $\{u_t\}$ is a linear process defined by $u_t = \Psi(L; \rho_0)\epsilon_t = \sum_{j=0}^{\infty} \psi_j(\rho_0)\epsilon_{t-j}$, where $\rho_0 \in \Xi$ is a p-dimensional vector of unknown parameters and Ξ is compact, $\Psi(s; \rho) = \sum_{j=0}^{\infty} \psi_j(\rho) s^j$, in which $\psi_0(\rho) \equiv 1$, $\Psi(s; \rho)$ is continuous in s, and $\psi := \Psi(1; \rho_0) = \sum_{j=0}^{\infty} \psi_j(\rho_0) \neq 0$. Also, $\Psi(s; \rho)$ satisfies

A.2(i) for all $\rho \neq \rho_0$ and $\rho \in \Xi \subset \mathbb{R}^p$, on a set $S \subset \{s : |s| = 1\}$ of positive Lebesgue measure, we have $|\Psi(s; \rho)| \neq |\Psi(s; \rho_0)|$.

A.2(ii) for all ρ, $\Psi(e^{iv}; \rho)$ is differentiable in $v \in \mathbb{R}$ with derivative in Lip(ς) for $\varsigma > 1/2$, that is, $|U(v_1; \rho) - U(v_2; \rho)| \leq c(\rho)|v_1 - v_2|^{\varsigma}$ for any $v_1, v_2 \in \mathbb{R}$ where $U(v; \rho) = \partial \Psi(e^{iv}; \rho)/\partial v$.

A.2(iii) for all $v \in \mathbb{R}$, $\Psi(e^{iv}; \rho)$ is continuous in ρ.

A.2(iv) for all $\rho \in \Xi$, $|\Psi(s; \rho)| \neq 0$, $|s| \leq 1$.

A.3. The error process $\{e_t\}$ is generated by

A.3(i) $e_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$, where $\phi_0 = 1$, $\lim_{j\to\infty} \phi_j j^{\gamma_0}$ exists with $\gamma_0 > 3/2$, and $\phi := \sum_{j=0}^{\infty} \phi_j \neq 0$, or

A.3(ii) $e_t = \wp(\epsilon_t, \ldots, \epsilon_{t-m_0+1}; \eta_t, \ldots, \eta_{t-m_1+1})$ where $\min(m_0, m_1) \geq 1$, $\{\eta_t\}$ is a sequence of iid(0,1) continuous variables independent of $\{\epsilon_t\}$, and the function $\wp(\cdots)$ is a measurable mapping from $\mathbb{R}^{m_0+m_1} \mapsto \mathbb{R}$ such that $E[e_t] = 0$ and $E[e_t^4] < \infty$ for all $t > \max(m_0, m_1)$; we define $e_t = 0$ for $t \leq \max(m_0, m_1)$.

Remark 2.1. This assumption exhibits the structure of regressor and equilibrium error processes in model (1.1) from building blocks involving ϵ_j and η_j.

Here, x_t is fractionally integrated by u_t, while both u_t and e_t are generated by the iid(0, σ^2) sequence $\{\epsilon_j\}$. This setup provides a suitable framework for accommodating both endogeneity and exogeneity in our model, depending on the choice of the coefficients in the linear processes and/or the form of $\wp$ function. For instance, if $\phi_j = \psi_j$ for each j, $e_t = u_t$ implying strong endogeneity; when the functional $\wp$ does not include any ϵ_j, e_t is independent of u_s and therefore of x_t, giving the exogenous case.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

In the definition of x_t, the infinite series is truncated, $x_t = \sum_{\ell=0}^{t-1} a_\ell(\alpha_0)u_{t-\ell}$, where $a_\ell(\alpha_0) = \Gamma(\ell+\alpha_0)/[\Gamma(\alpha_0)\Gamma(\ell+1)]$. Without truncation, the second moment of x_t would not exist. If $\alpha_0 = 1$, x_t is a unit root process.

The conditions in A.2 are used to recover the unknown parameter ρ_0 in the linear process u_t and hence the proposed test statistics can be used directly in practice. The procedure for estimating ρ_0 is introduced in the next section.

By Lemma A.1 below, as $n \to \infty$, $d_n^2 := E\left[x_n^2\right] \sim \psi^2\Gamma^{-2}(\alpha_0)n^{2\alpha_0-1}(1+o(1))$. In addition, it follows from Marinucci and Robinson (2000) that for $r \in [0,1]$,

$$x_{n[nr]} := \frac{1}{d_n}x_{[nr]} \to_D W(r;\alpha_0) := \int_0^r (r-s)^{\alpha_0-1}dB(s) \tag{2.1}$$

in the space $D[0,1]$, where $[\cdot]$ is the integer part of the argument, $W(r;\alpha_0)$ is a Type II fractional Brownian motion, defined via a standard Brownian motion $B(s)$. See Marinucci and Robinson (2000) for details about $W(r;\alpha_0)$. Moreover, the process $W(r;\alpha_0)$ possesses a continuous local time process $L_W(t,a)$ which measures the sojourning time of $W(r;\alpha_0)$ at the point a over the time interval $[0,t]$.

For unknown functions in $L^2(\mathbb{R}) = \{f(x) : \int f^2(x)dx < \infty\}$, we introduce an orthonormal basis in the space. Let $\{H_i(x), i = 0,1,2,\ldots\}$ be the Hermite polynomial system orthogonal with respect to $\exp(-x^2)$. As a classical polynomial orthogonal system, $\{H_i(x)\}$ is a complete orthogonal system in the Hilbert space $L^2(\mathbb{R},\exp(-x^2)) = \{f(x) : \int f^2(x)e^{-x^2}dx < \infty\}$. In addition, the orthogonality of the system is defined by the equation: $\int H_i(x)H_j(x)\exp(-x^2)dx = \sqrt{\pi}2^i i!\delta_{ij}$, where δ_{ij} is the Kronecker delta. Denote $\varphi(x) = \exp(-\frac{1}{2}x^2)$ and define

$$\mathscr{H}_i(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^i i!}}H_i(x)\varphi(x), \quad i \geq 0. \tag{2.2}$$

Note that $\mathscr{H}_i(x)$ are the so-called Hermite functions in the literature. Moreover, $\{\mathscr{H}_i(x)\}$ is uniformly bounded in both i and x. Consequently, $\sum_{i=0}^{k-1}\mathscr{H}_i^2(x) \leq O(1)k$ uniformly in x. In addition, the system is a complete orthonormal basis in $L^2(\mathbb{R})$ in which the inner product is defined as $(f,g) = \int f(x)g(x)dx$. Therefore, any function $g(x) \in L^2(\mathbb{R})$ admits an orthogonal expansion of the form $g(x) = \sum_{i=0}^{\infty}\beta_i\mathscr{H}_i(x)$ with $\beta_i = \int g(x)\mathscr{H}_i(x)dx$. Throughout, let k be a positive integer and define $g_k(x) = \sum_{i=0}^{k-1}\beta_i\mathscr{H}_i(x)$ to be the truncated series with truncation parameter k. It is known that $g_k(x) \to g(x)$ in L^2 sense as $k \to \infty$ even if $g(x)$ is not continuous. More detailed discussion on the Hermite system and orthogonal expansion can be found in Chen and Shen (1998), Dong, Gao, and Peng (2015), and Dong, Gao, and Tjøstheim (2016).

Assumption B.

B.1. Let $k = [c \cdot n^\kappa]$ for some $\kappa : \max(0, 2(1-\alpha_0)) < \kappa < \frac{3}{2} - \alpha_0$ and constant $c > 0$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

B.2. Under H_0, there exists a consistent estimator $\widehat{\theta}$ of θ_0 such that $P(\|\widehat{\theta}-\theta_0\| > M\zeta_n) < \varepsilon_0$ for any $\varepsilon_0 > 0$ and $M = M(\varepsilon_0)$, where $\zeta_n\sqrt{\frac{n}{kd_n}} = O(1)$. This, moreover, holds as well under H_1 with θ_0 replaced by θ_1.

B.3. Suppose that $g(x;\theta)$ is twice differentiable with respect to θ and $g(x;\theta) \in L^2(\mathbb{R})$ for every fixed $\theta \in \Theta$. Let $l_1(x,\theta) := \frac{\partial}{\partial\theta}g(x;\theta)$ and $l_2(x,\theta) := \frac{\partial^2}{\partial\theta\partial\theta'}g(x;\theta)$. Suppose further that $\|l_1(x,\cdot)\|, \|l_2(x,\cdot)\| \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and that there exists a positive function $l(x) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that $\|l_2(x,\cdot)\| \le l(x)$.

Remark 2.2. For the truncation parameter k we need to impose Condition B.1 on k such that $k \to \infty$ as $n \to \infty$ but with a slow rate κ. In addition, κ is also restricted by the fractional difference order α_0 such that $n/(kd_n) \to \infty$ as $n \to \infty$, where the restriction $\max(0, 2(1-\alpha_0)) < \kappa < \frac{3}{2} - \alpha_0$ is mainly due to the balance between the convergence rate d_n in (2.1) and the bias introduced by the truncation of infinite series. Particularly, in the unit root case of $\alpha_0 = 1$, we have $0 < \kappa < 1/2$ and $d_n \sim \sqrt{n}$.

Condition B.2 stipulates a certain rate of convergence for $\widehat{\theta} \to_P \theta_0$, where the rate of convergence of $\zeta_n \to 0$ can be as slow as $O(\sqrt{kd_n/n})$. Note that a general asymptotic theory for $\widehat{\theta}-\theta_0$ under the current setting is not available in the literature. It is worth pointing out that in several special cases the rate of convergence of $\widehat{\theta}-\theta_0$ to zero can be as fast as $O_P(\sqrt{d_n/n})$, which include: (i) e_t is independent of u_t; (ii) $(e_t, \mathcal{F}_t)$ is a martingale difference sequence for a given information flow $\mathcal{F}_t$ such that x_t is adapted with respect to $\mathcal{F}_{t-1}$, since there already exist such results in Park and Phillips (1999, 2001); and (iii) we provide one example to verify the fast rate of convergence of $O_P(\sqrt{d_n/n})$ achievable with the presence of endogeneity and the unit root for the regressor in Appendix C of the online supplemental material of this paper.

Meanwhile, Wang and Phillips (2016) show different rates of convergence of $\widehat{\theta}-\theta_0$ to zero under different conditions where x_t is nonstationary possessing either short memory or long memory (Theorem 4.2, p. 14). In view of this result, they impose the rate of convergence of $\widehat{\theta}-\theta_0$ to be $O_P(\sqrt{d_n/nh})$ in their testing using a kernel based test, where h is the bandwidth. This is comparable with our assumption in the case where $[1/h]$ is regarded as k. Taking into account the discussion, we impose Condition B.2 on the rate of convergence of $\widehat{\theta}-\theta_0$ to zero under H_0.

Notice that Chen, Gao, and Li (2011) study the estimation of θ_1 under $H_1 : y_t = g(V_t;\theta_1) + \Delta_n(V_t) + e_t$, where $V_t = H(t/n) + u_t$ and $\{u_t\}$ is an iid sequence. They propose the estimator of θ_1 to be the minimizer of $\sum_{t=1}^n (y_t - g(V_t;\theta))^2$ and show that the estimator is unbiased with square root n convergence rate, the same as the case without the presence of $\Delta_n(V_t)$. See Corollary 2.4 of Chen et al. (2011). Wang and Phillips (2016) impose $\theta_1 = \theta_0$ for which there exist estimators that possess the same convergence rate under both H_0 and H_1. It is clear that the class of our alternatives is larger than that in Wang and Phillips (2016).

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Condition B.3 is a set of standard requirements used in similar situations (see, for example, Assumption 2.4 of Gao, Wang, and Yin (2011) and Assumption 4 of Wang and Phillips (2012)). It is also easily seen that the following class of functions for $g(x;\theta)$ satisfies these conditions: (1) $g(x;\theta)=\theta g_1(x)$, where $g_1(x)$ is integrable and $\theta\in\mathbb{R}$ (so called linear-in-parameter regression function in Park and Phillips 2001); (2) $g(x;\theta)=\frac{1}{1+\theta x^2}$, where $\theta>c>0$; (3) $g(x;\alpha,\beta)=\alpha\exp(-\beta x^2)$ where $\theta=(\alpha,\beta)\in\Theta\subset\mathbb{R}\times\mathbb{R}^+$ and Θ is compact.

Assumption C. Let $\triangle_n(x)=\delta_n\triangle(x)$ satisfy the following conditions:

C.1. $\triangle(x)$ is a nonzero function and $\int\triangle^2(x)dx<\infty$.

C.2. Let $\delta_n\to 0$ and $\delta_n^2 n/(d_n k)\to\infty$ as $n\to\infty$, where k satisfies Assumption B.1.

Remark 2.3. Exclusion of a zero function $\triangle(x)$ is trivial and implies $\int\triangle^2(x)dx\neq 0$. This assumption on the one hand guarantees our test statistic consistent, and on the other hand keeps the local departures as small as possible. Indeed, the attenuation of $\delta_n\to 0$ and the integrability of $\triangle^2(x)$ force the departures $\triangle_n(x)$ to shrink as the sample size increases, whereas $\delta_n^2 n/(d_n k)\to\infty$ is the requirement that enables all the test statistics introduced later to possess power to detect whether there exists a departure from the null. To illustrate this idea, in the finite sample simulation below we choose δ_n to decay so fast that $\delta_n^2 n/(d_n k)=\frac{1}{4}\ln(n)$.

3. SPECIFICATION TESTING

In this section, we consider first the case where $m(x)$ is integrable and then the case where $m(x)$ may be nonintegrable.

3.1. Integrable Case

Suppose $m(\cdot)\in L^2(\mathbb{R})$. Then, $m(x)$ admits an orthogonal expansion $m(x)=\sum_{i=0}^{\infty}\beta_i\mathscr{H}_i(x)$. Let k be a positive integer and then define a truncated series $m_k(x)=\sum_{i=0}^{k-1}\beta_i\mathscr{H}_i(x)$ and residue $\gamma_k(x)=\sum_{i=k}^{\infty}\beta_i\mathscr{H}_i(x)$ for any real x. Denote $Z(x)=(\mathscr{H}_0(x),\ldots,\mathscr{H}_{k-1}(x))'$ and $\beta=(\beta_0,\ldots,\beta_{k-1})'$, giving $m_k(x)=Z(x)'\beta$. Then, model (1.1) under H_0 can be written as $y_t=Z(x_t)'\beta+\gamma_k(x_t)+e_t$, for $t=1,\ldots,n$. Denote $Y=(y_1,\ldots,y_n)'$, $Z=(Z(x_1),\ldots,Z(x_n))'$ an $n\times k$ matrix, $e=(e_1,\ldots,e_n)'$ and $\gamma=(\gamma_k(x_1),\ldots,\gamma_k(x_n))'$. The equations are formulated into the following matrix form:

$$Y=Z\beta+\gamma+e. \tag{3.1}$$

Accordingly, we have the ordinary least squares (OLS) estimator of β given by $\widehat{\beta}=(Z'Z)^{-1}Z'Y$. Thereby, $\widehat{m}(x)=Z(x)'\widehat{\beta}$ is an estimator for $m(x)$ at $\forall x\in\mathbb{R}$.

Since the inverse of $Z'Z$ is a large dimensional matrix of nonlinear and nonstationary time series, it is very difficult to deal with it both theoretically and computationally. In order to avoid involving it in $\widehat{\beta}$, we use $\widetilde{\beta}=Z'Z\widehat{\beta}$ in the construct

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

of the first test statistic below. Correspondingly, the estimator $\widehat{m}(x) = Z(x)'\widehat{\beta}$ can be replaced by $\widetilde{m}(x) = Z(x)'\widetilde{\beta}$ for testing purposes and, invoking the expression of $\widehat{\beta}$,

$$\widetilde{m}(x) = \sum_{t=1}^{n} [Z(x_t)'Z(x)]y_t. \tag{3.2}$$

As discussed in (3.8) below, a similar treatment has been performed in the test proposed in Dong et al. (2014) and studied in Wang and Phillips (2016). Our experience suggests that the removal of either the inverse of $Z'Z$ in the series case or the random denominator in the kernel case not only simplifies the theoretical discussion, but also improves the finite sample performance of the respective test in each case.

Analogously, we may also have a similar version, $\widetilde{g}(x;\theta)$, for $g(x;\theta)$ of the form:

$$\widetilde{g}(x;\theta) = \sum_{t=1}^{n} [Z(x_t)'Z(x)]g(x_t;\theta). \tag{3.3}$$

Let $\widehat{\theta}$ be a consistent estimator of θ_0 under $H_0: \ y_t = g(x_t;\theta_0) + e_t$. Instead of comparing $\widehat{m}(x)$ with $g(x,\widehat{\theta})$, we measure the distance between $\widetilde{m}(x)$ and $\widetilde{g}(x;\widehat{\theta})$ and then propose a test statistic of the form:

$$L_n = \int_{-\infty}^{\infty} \big(\widetilde{m}(x) - \widetilde{g}\big(x;\widehat{\theta}\big)\big)^2 dx = \int_{-\infty}^{\infty} \left(\sum_{t=1}^{n} [Z(x_t)'Z(x)]\big(y_t - g\big(x_t;\widehat{\theta}\big)\big) \right)^2 dx, \tag{3.4}$$

which, by virtue of the orthogonality of the basis, simplifies to

$$L_n = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\widehat{e}_t\widehat{e}_s, \tag{3.5}$$

where $\widehat{e}_t := y_t - g(x_t;\widehat{\theta})$.

Clearly, L_n can be interpreted as the distance in the space $L^2(\mathbb{R})$ between the proxies of $m(x)$ and $g(x,\widehat{\theta})$ based on the observation (y_t, x_t). Moreover, under the null $\widehat{e}_t = e_t + [g(x_t;\theta_0) - g(x_t;\widehat{\theta})]$ consists of two errors such that, as shown in Theorem 3.1, a suitably normalised version of L_n converges to a local-time random variable, while under a class of local alternatives, L_n with the same normalization diverges to ∞, because of the existence of the local departure.

Meanwhile, we have

$$L_n = \sum_{t=1}^{n} \|Z(x_t)\|^2\widehat{e}_t^2 + 2\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z(x_t)'Z(x_s)\widehat{e}_t\widehat{e}_s \equiv L_{an} + L_{bn}. \tag{3.6}$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

In the case where x_t is stationary, as shown in Gao, Tong, and Wolff (2002), a suitably standardised version of L_{an} converges in probability to a non-stochastic quantity. Thus, one has to use a standardised version of the form

$$\widehat{L}_{bn} = \frac{\sum_{t=2}^{n}\sum_{s=1}^{t-1}\widehat{e}_s Z(x_t)' Z(x_s)\widehat{e}_t}{\sqrt{\sum_{t=2}^{n}\sum_{s=1}^{t-1}\widehat{e}_s^2 (Z(x_t)' Z(x_s))^2 \widehat{e}_t^2}} \tag{3.7}$$

to derive a consistent test. As shown in the proof of Theorem 3.1 in Appendix B below for the nonstationary case, by comparison, L_{an} is the leading term that already converges in distribution to a local-time random variable, while L_{bn} becomes asymptotically negligible compared with L_{an}. Therefore, there is no need to standardise L_{bn} as in the stationary case. As a consequence, both the size and power properties of L_n become stable and robust. These, along with the large and small-sample properties discussed in Sections 3–4 below, are the key features of L_n.

Another test, proposed in Dong, Gao, Tjostheim and Yin (2014) and studied in Wang and Phillips (2016), is a nonparametric kernel test of the form

$$\widetilde{L}_n = \int_{-\infty}^{\infty} \left(\widehat{q}(x) - \widetilde{q}\left(x;\widehat{\theta}\right)\right)^2 \pi(x)dx, \tag{3.8}$$

where $\widehat{q}(x) = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) y_t$, $\widetilde{q}(x;\theta) = \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} K_1\left(\frac{x_t - x}{h}\right) g(x_t;\theta)$ and $\pi(\cdot)$ is a probability weight function. One disadvantage of $\widetilde{L}_n$ is that it involves an arbitrary weight function $\pi(\cdot)$. By contrast, a normalised version of L_n of the form $\frac{\widehat{d}_n}{nk\widehat{\sigma}_e^2} L_n$ (as defined in Theorem 3.3) involves neither a weight function nor a random denominator of a quadratic form of nonstationary time series. As a consequence, the simple form of L_n of (3.5) makes it possible to establish Theorems 3.1–3.3 under two types of endogeneity.

THEOREM 3.1. *Suppose that Assumptions A and B hold. Under H_0, we have as $n \to \infty$*

$$\frac{d_n}{nk\,\sigma_e^2} L_n \to_D L_W(1,0), \tag{3.9}$$

where $\sigma_e^2 = E[e_1^2]$, $d_n = \sqrt{E(x_n^2)}$ and $L_W(1,0)$ is the local-time random variable associated with $W(r;\alpha_0)$ given in (2.1).

For the definition and properties of the local-time processes, Geman and Horowitz (1980) and Revuz and Yor (2005) are standard references.

THEOREM 3.2. *Suppose that Assumptions A–C hold. Under H_1, we have as $n \to \infty$*

$$\frac{d_n}{nk\,\sigma_e^2} L_n \to_P \infty. \tag{3.10}$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Remark 3.1. Theorem 3.1 gives the limit distribution of the proposed test statistic. Theorem 3.2 shows that the test statistic has nontrivial power against a sequence of local alternatives whose distance from the null approaches zero at a rate slower than $(kd_n/n)^{1/2}$, as required in Assumption C for δ_n. This is similar to the stationary case discussed in Gao (Chap. 3, 2007). As can be seen from the proof, under the alternative $\frac{d_n}{nk\sigma_e^2}L_n$ diverges with order $O_P(\delta_n^2 n/d_n k)$. The results in Theorems 3.1 and 3.2 ensure rejection of H_0 in favour of H_1 if $(d_n/(nk\sigma_e^2))L_n$ is large enough.

Under H_1, $\widehat{e}_t = e_t + [g(x_t;\theta_1) - g(x_t;\widehat{\theta})] + \triangle_n(x_t)$, where $\widehat{\theta}$ is the consistent estimator of θ_1. The divergence of the normalisation of L_n is due to the inclusion of the local departure $\triangle_n(x_t)$, although the function $\triangle_n(x)$ attenuates to zero uniformly in x when the sample size increases.

Note that there are nuisance parameters involved in the large sample theory of the proposed test, namely, ψ, α_0 in d_n and σ_e^2, which should be replaced by consistent estimates. The following assumption is necessary to obtain such estimators.

The set of conditions A.2(i)–A.2(iv) in Assumption A is sufficient for us to estimate the unknown parameters involved in the linear process u_t. See Hualde and Robinson (2011, p. 3155). More importantly, by A.2(ii) the derivative of $\Psi(e^{i\nu};\rho)$ has Fourier coefficient $j\psi_j(\rho) = O(j^{-\varsigma})$ as $j \to \infty$, for all ρ, so that, by the compactness of Ξ and continuity of $\psi_j(\rho)$ in ρ for all j,

$$\sup_{\rho\in\Xi}|\psi_j(\rho)| = O\big(j^{-(1+\varsigma)}\big), \qquad j \to \infty. \tag{3.11}$$

Write $w_t = \Psi^{-1}(L;\rho_0)u_t = \Psi^{-1}(L;\rho_0)\triangle^{\alpha_0}x_t$ for $t \geq 1$. For any admissible $\tau = (\rho', \alpha)'$, define

$$w_t(\tau) = \Psi^{-1}(L;\rho)\triangle^{\alpha}x_t. \tag{3.12}$$

Also, define

$$\widehat{\tau} = \underset{\tau\in\Xi}{\operatorname{argmin}}\frac{1}{n}\sum_{t=1}^{n} w_t^2(\tau) \tag{3.13}$$

as an estimator of $\tau_0 = (\rho_0', \alpha_0)'$.

LEMMA 3.1. *Define* $\widehat{\psi} = \psi(\widehat{\rho}) = \sum_{j=0}^{\infty}\psi_j(\widehat{\rho})$, $\widehat{d}_n = |\widehat{\psi}|\Gamma^{-1}(\widehat{\alpha})n^{\widehat{\alpha}-1/2}$, $\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}\widehat{e}_t^2$ *with* $\widehat{e}_t = y_t - g(x_t,\widehat{\theta})$, *and* $\widehat{L}_W(1,0) = \frac{\widehat{d}_n}{n}\sum_{t=1}^{n}\mathscr{H}_0^2(x_t)$. *Then, under Assumption A and* H_0, *as* $n \to \infty$

$$\widehat{\psi} \to_P \psi, \quad \frac{\widehat{d}_n}{d_n} \to_P 1, \quad \widehat{\sigma}_e^2 \to_P \sigma_e^2, \quad \textit{and} \quad \widehat{L}_W(1,0) - L_W(1,0) \to_P 0. \tag{3.14}$$

The proofs of Lemma 3.1 and Theorem 3.3 are given in Appendix B.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

THEOREM 3.3. *When ψ and σ_e^2 are both unknown, under H_0 and Assumptions A and B, we have*

$$\frac{\widehat{d}_n}{nk\widehat{\sigma}_e^2} L_n \to_D L_W(1,0), \tag{3.15}$$

where $\widehat{d}_n$, $\widehat{\sigma}_e^2$ are defined in Lemma 3.1. Under H_1 and Assumptions A–C, as $n \to \infty$,

$$\frac{\widehat{d}_n}{nk\widehat{\sigma}_e^2} L_n \to_P \infty. \tag{3.16}$$

Remark 3.2. Note that the quantity $\widehat{M}_n \equiv \frac{\widehat{d}_n}{nk\widehat{\sigma}_e^2} L_n$ not only has nice large sample properties as stated in Theorem 3.3, but also $\widehat{M}_n$ is now completely computable in practice. It follows from Theorem 3.3 that, at the significance level ν, we may reject H_0 if $\widehat{M}_n > l_{1-\nu}$ with $l_{1-\nu}$ satisfying $P(L_W(1,0) > l_{1-\nu}) = \nu$.

3.2. Beyond Integrability

We are about to relax the restriction on $m(x)$ from $L^2(\mathbb{R})$ to $L^2(\mathbb{R}, \exp(-x^2))$. Our motivation is as follows. If $m(x)$ satisfies $\int m^2(x)e^{-x^2}dx < \infty$, then $m(x)\exp(-x^2/2) \in L^2(\mathbb{R})$. As a result, after a simple transformation we may borrow L_n to test models that have nonintegrable regression functions.

Given observations $\{(x_t, y_t), t = 1, 2, \ldots, n\}$ from model (1.1), we are interested in testing

$$H_{10}: \ m(x) = g(x;\theta_0) \quad \text{versus} \quad H_{11}: \ m(x) = g(x;\theta_1) + \triangle_n(x), \tag{3.17}$$

where for any fixed $\theta \in \Theta \subset \mathbb{R}^d$, $g(x;\theta) \in L^2(\mathbb{R}, \exp(-x^2))$ is a known function, and $\triangle_n(x)$ is the same as in Assumption C.

In order to make use of L_n, multiplying both sides by $\varphi(x_t) := \exp(-\frac{1}{2}x_t^2)$ will give us an integrable model. Specifically, by denoting $Y_t = y_t\varphi(x_t)$, $M(x_t) = \varphi(x_t)m(x_t)$ and $\varepsilon_t = \varphi(x_t)e_t$, we have

$$Y_t = M(x_t) + \varepsilon_t, \qquad t = 1, \ldots, n. \tag{3.18}$$

Moreover, H_{10} is completely equivalent to $H'_{10}: \ M(x) = G(x;\theta_0)$ in which $G(x;\theta) = g(x;\theta)\varphi(x)$ and H_{11} is completely equivalent to $H'_{11}: \ M(x) = G(x;\theta_1) + \triangle_n(x)\varphi(x)$. Therefore, $M(x)$ is an unknown integrable function, while $G(x;\theta)$ is a known integrable function. Roughly speaking, we have transformed the dataset from 'non-integrable' (y_t, x_t) to 'integrable' (Y_t, x_t). Notice that the deviation function $\triangle_n(x)\varphi(x)$ in H'_{11} has the same rate in testing as $\triangle_n(x)$ in H_{11}, so is nonzero and integrability. After transformation, we are dealing with the same specification issue as in the last subsection.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Therefore, we propose the following test statistic for testing (3.17):

$$\Pi_n = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\widehat{\varepsilon}_t\widehat{\varepsilon}_s, \tag{3.19}$$

where $\widehat{\varepsilon}_t := \widehat{e}_t\varphi(x_t) = (y_t - g(x_t;\widehat{\theta}))\varphi(x_t) = Y_t - G(x_t;\widehat{\theta})$ and $\widehat{\theta}$ is a consistent estimator of θ_0.

It is clear that, though always $\Pi_n = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)\widehat{e}_t\widehat{e}_s$, under H_{10}, $\widehat{e}_t = y_t - g(x_t;\widehat{\theta}) = e_t + g(x_t;\theta_0) - g(x_t;\widehat{\theta})$, whereas under H_{11}, $\widehat{e}_t = y_t - g(x_t;\widehat{\theta}) = e_t + g(x_t;\theta_0) - g(x_t;\widehat{\theta}) + \triangle_n(x_t)$. Comparing with L_n, only factors $\varphi(x_t)\varphi(x_s)$ are adhered in each term. This is just a consequence of the transformation from $m(x_t)$ to $M(x_t)$.

Remark 3.3. As pointed out by one referee, the above rationale could be used in a more general weighted function space. Specifically, if $g(x;\theta) \in L^2(\mathbb{R}, w(x))$, where $w(x)$ is a nonnegative function, we could use $w(x_t)^{1/2}$ to replace $\varphi(x_t)$. As a result, a similar version of Π_n is obtained, for which the following results could still hold. Since the function space $L^2(\mathbb{R}, \exp(-x^2))$ is big enough for our econometric testing, we focus on it throughout the rest of this paper.

Before stating the limit theory for Π_n, we need to give the following assumption which is the counterpart of Assumption B for the case where m belongs to $L^2(\mathbb{R}, \exp(-x^2))$.

Assumption B*.

B*.1 Let k satisfy Assumption B.1.

B*.2 Under the null in (3.17), there exists a consistent estimator, $\widehat{\theta}$, of θ_0 such that for any $\varepsilon_0 > 0$ and some $0 < M = M(\varepsilon_0) < \infty$, $P\left(\|\widehat{\theta} - \theta_0\| > M\zeta_n\right) < \varepsilon_0$, where $\zeta_n\sqrt{\frac{n}{kd_n}} = O(1)$. This, moreover, holds under the alternative as well with θ_0 replaced by θ_1.

B*.3 Suppose that $g(x;\theta)$ is twicely differentiable with respect to θ. Let $l_1(x,\theta) := \frac{\partial}{\partial\theta}g(x;\theta)$ and $l_2(x,\theta) := \frac{\partial^2}{\partial\theta\partial\theta'}g(x;\theta)$. Suppose further that $\|l_1(x,\cdot)\|, \|l_2(x,\cdot)\| \in L^2(\mathbb{R}, e^{-x^2})$, while $\|l_2(x,\theta)\| \le l(x)$ uniformly over θ with positive function $l(x) \in L^2(\mathbb{R}, e^{-x^2})$.

Remark 3.4. There are two differences between Assumptions B* and B. First, the estimation of θ_0 is under H_{10} in (3.17). However, a fast rate of convergence is not required in B*.2 and B*.3, although in the existing literature, the convergence rate for the non-integrable regression function is much faster than $\sqrt{d_n/n}$. This is an advantage of the proposed method, because a non-integrable $m(x)$ has been transformed to an integrable function $M(x)$, so that Π_n works in the same way as L_n does, as can be seen, the only difference between Π_n and L_n is that $Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)$ is replaced by $Z(x_t)'Z(x_s)$. Thus, all the regression functions studied in the literature easily satisfy this assumption. For example, for functions like polynomials, power functions, $\theta\log|x|$, $\theta e^x/(1+e^x)$, and

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

$x(1+\theta x)^{-1}1(x \geq 0)$, the rates of convergence are faster than those for integrable functions. A detailed discussion can be found in Theorems 5.2 and 5.3 of Park and Phillips (2001, p. 135).

Another difference is the function space. As the regression function $m(x)$ is now in the class of $L^2(\mathbb{R}, e^{-x^2})$, we basically stipulate that its derivatives with respect to the parameter vector are also in $L^2(\mathbb{R}, e^{-x^2})$. Clearly, the examples given above satisfy this condition.

THEOREM 3.4. *Suppose Assumptions A and B* hold. Under H_{10} of (3.17), we have as $n \to \infty$*

$$\frac{d_n}{nk\sigma_e^2}\Pi_n \to_D \int \mathscr{T}(x)\varphi^2(x)dx \cdot L_W(1,0), \tag{3.20}$$

where $\sigma_e^2 = E[e_1^2]$ and $\mathscr{T}(x) = \frac{1}{2\pi}\sqrt{4-x^2}$ on $x \in [-2,2]$ and 0 elsewhere, $\varphi(x) = \exp\left(-\frac{1}{2}x^2\right)$, and d_n and $L_W(1,0)$ are the same as in the previous theorem. When d_n and σ_e^2 are replaced by $\widehat{d}_n$ and $\widehat{\sigma}_e^2$, respectively, defined similarly as that in Lemma 3.1, we have a feasible version of (3.20).

Let Assumptions A, B, and C hold. Under H_{11}, we have as $n \to \infty$*

$$\frac{d_n}{nk\sigma_e^2}\Pi_n \to_P \infty. \tag{3.21}$$

Similarly, when d_n and σ_e^2 are replaced by their consistent estimators $\widehat{d}_n$ and $\widehat{\sigma}_e^2$, respectively, we obtain a feasible version of (3.21).

The proof of Theorem 3.4 is given in Appendix B.

Remark 3.5. Notice that the result of (3.20) is comparable with that of (3.9), knowing that due to $\int \mathscr{T}(x)dx = 1$, the constant before $L_W(1,0)$ in (3.9) is one. Due to the involvement of $\varphi(\cdot)$, the proof of Theorem 3.4 is much similar to that of Theorems 3.1 and 3.2.

Notice also that Assumption C is used in this case too, in addition to being used in the integrable case. Because usually an integrable function is much smaller than a nonintegrable function, especially in the region far from the origin, if our test is workable for the integrable departure functions, it should also be workable when the departure function becomes larger. It is also fairly clear that the test Π_n is consistent as long as $\Delta^2(x)e^{-x^2}$ is integrable, implying that $\Delta(x)$ can be any polynomial, power function, or bounded function, let alone integrable function. For this reason, we keep Assumption C. However, this does not attenuate the power of Π_n at all and we actually still have $\frac{d_n}{nk\sigma_e^2}\Pi_n = O_P(\delta_n^2 n/d_n k)$ under the alternative, the same as Theorem 3.2 as can be seen from the proof.

Remark 3.6. From a function space point of view, $L^2(\mathbb{R}) \subset L^2(\mathbb{R}, \exp(-x^2))$. Accordingly, the test statistic Π_n is applicable even if we are testing $m(x) = g(x,\theta_0)$ where $g(x,\theta_0) \in L^2(\mathbb{R})$. To illustrate the point, Example E.1 is given in

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Appendix E. The results show that Π_n still has stable size and good power in testing H_0: $m(x) = g(x, \theta_0)$, where $g(x, \theta_0)$ is integrable.

However, the establishment of the asymptotic theory for L_n is fundamental, and the asymptotic theory for Π_n relies on it. We could not develop the theory for Π_n directly without using L_n, because we do not have a counterpart of Lemma A.5 in the nonintegrable case. Therefore, if there is no clear information about whether $g(x, \theta_0) \in L^2(\mathbb{R})$, we suggest Π_n for the practitioner.

The performance of both the proposed tests as well as the related theory will be evaluated using simulated data in Section 4.

4. FINITE SAMPLE EXPERIMENT

Monte Carlo simulations are conducted in this section to investigate the performance of the proposed statistics L_n and Π_n in the finite-sample situations. To be in concert with Assumption A, two circumstances of generating data are considered:

I) Suppose $\{\epsilon_j, j \in \mathbb{Z}\}$ is a sequence of iid $N(0, \sigma^2)$. Let $e_t = \beta_0 e_{t-1} + \epsilon_t$ with $|\beta_0| < 1$, $u_t = \alpha_0 u_{t-1} + \epsilon_t$ with $|\alpha_0| < 1$ and $x_t = x_{t-1} + u_t$ with $x_0 = O_P(1)$.

II) Suppose $\{(\epsilon_j, \eta_j), j \in \mathbb{Z}\}$ is a sequence of iid $N(0, \sigma^2 I_2)$. Let $e_t = a\epsilon_t + b\eta_t$, $u_t = \alpha_0 u_{t-1} + \epsilon_t$ with $|\alpha_0| < 1$ and $x_t = x_{t-1} + u_t$ with $x_0 = O_P(1)$.

All the constants in the above will be specified in the following tables. Note that in the first case if $\alpha_0 = \beta_0$, $e_t = u_t$, this setting reflects the highest possible dependence; in the second case when $a = 0$, $e_t = b\eta_t$ which is independent of any ϵ_j and hence of u_t, there is independence between e_t and x_t.

4.1. Bootstrap Simulation Procedure

A bootstrap schedule is proposed below to generate critical values l^*_α for tests with $\alpha = 1\%$ and 10%. The number of replications is $M = 5,000$ and, for the bootstrap procedure, $M_b = 250$. Since Case I) involves an AR(1) model, we propose using a block bootstrap method. A regression bootstrap method is used for Case II).

Step 1. Let $\widehat{e}_t = y_t - g(x_t, \widehat{\theta})$, where $\widehat{\theta}$ is a consistent estimator of θ_0 based on the original sample (x_t, y_t) under the null hypothesis. Generate e^*_t as follows.

- For Case I) above, generate e^*_t by the conventional block bootstrap method (see, for example, Hall, Horowitz, and Jing (1995)) given as follows: Estimate β_0 by $\widehat{\beta}_0$ based on $\widehat{e}_t = \beta_0 \widehat{e}_{t-1} + \epsilon_t$ and then estimate σ^2 by $\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \left(\widehat{e}_t - \widehat{\beta}_0 \widehat{e}_{t-1}\right)^2$.

  Let $l = \lfloor n^{\frac{1}{3}} \rfloor$ and choose integer λ such that $\lambda l = n$. Generate $e^*_{1l}(j) = [\widehat{e}_1(j), \ldots, \widehat{e}_l(j)]$, $\ldots$, $e^*_{Nl}(j) = \left[\widehat{e}_{(\lambda-1)l+1}(j), \ldots, \widehat{e}_{\lambda l}(j)\right]$ in step j for $N = n - l + 1$. Replicate the resample $J = 250$ times and obtain J bootstrap resamples $\{e^*_{sl}(j) : 1 \le s \le N; 1 \le j \le J\}$ and then take the average $e^*_{sl} = \frac{1}{J}\sum_{j=1}^{J} e^*_{sl}(j)$ to obtain a block bootstrap version of e_t of the form: $\left(e^*_1, \ldots, e^*_n\right) = \left(e^*_{1l}, \ldots, e^*_{Nl}\right)$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

- For Case II) above, generate e_t^* by a regression bootstrap method (see, for example, Li and Wang (1998)) of the form $e_t^* = \widehat{e}_t \tau_t$ with τ_t being iid sequence possessing a two-point distribution of the form $P(\tau_1 = \mp(\sqrt{5} \mp 1)/2) = (\sqrt{5} \pm 1)/2\sqrt{5}$.

Step 2. Generate y_t^* by $y_t^* = g(x_t, \widehat{\theta}) + e_t^*$. Use the dateset (x_t, y_t^*) to re-estimate θ_0 and denote by $\widehat{\theta}^*$ the resulting estimate. Compute L_n^* that is a corresponding version of L_n with $\widehat{\theta}$ and $\{(x_t, y_t), 1 \le t \le n\}$ being replaced by $\widehat{\theta}^*$ and $\{(x_t, y_t^*), 1 \le t \le n\}$ in the expression of L_n.

Step 3. Repeat the above steps M_b times and thereby produce M_b versions of L_n^*, signified by $L_{n\ell}^*$ for $\ell = 1, \ldots, M_b$. Use the M_b values of $L_{n\ell}^*$ to construct their empirical bootstrap distribution function. The bootstrap distribution of $\frac{d_n}{nk} L_n^*$ given $\mathcal{W}_n = \{(x_t, y_t), 1 \le t \le n\}$ is defined by $P^*(\frac{d_n}{nk} L_n^* < x) = P(\frac{d_n}{nk} L_n^* < x | \mathcal{W}_n)$. Let l_α^* be the quantile such that $P^*(\frac{d_n}{nk} L_n^* \ge l_\alpha^*) = \alpha$ and then l_α^* is used to approximate l_α.

Step 4. Define the size and power functions by

$$\alpha = P\left(\frac{d_n}{nk} L_n \ge l_\alpha^* | H_0\right) \text{ and } \beta = P\left(\frac{d_n}{nk} L_n \ge l_\alpha^* | H_1\right).$$

Sample sizes used in the experiments are $n = 200, 500$, and $1,200$, respectively. In addition, the truncation parameter is selected by $k = [c \cdot n^\kappa]$ with $\kappa = 1/5, 1/4.5, 1/4$, and $c = 2.2$, respectively. We choose $\triangle_n(x) = \delta_n \frac{1}{1+x^2}$, in which $\delta_n = \frac{1}{2}\sqrt{k \log(n)/\sqrt{n}}$. It is clear that we are about to keep both δ_n and $\triangle(x)$ very 'small' by choosing an integrable function $\triangle(x) = \frac{1}{1+x^2}$ and the fast decaying δ_n such that $\delta_n^2 \sqrt{n}/k = \frac{1}{4}\log(n)$ diverges slowly. Note that the divergence of $\delta_n^2 \sqrt{n}/k$ is a necessary condition for the proposed test statistics to be consistent, but we keep it very slow. We shall examine the finite-sample performance of L_n and Π_n with unknown ψ and σ_e^2. That is, $\widehat{\sigma}_e^2$ is used in each experiment in the place of σ_e^2, but because ψ is only involved in x_t and we use the proposed bootstrap procedure to simulate the critical value, $\widehat{\psi}$ does not show up in the experiments.

Example 4.1

This example examines L_n using two integrable regression functions. The model for simulation is $y_t = m(x_t) + e_t$.

In situation I), the null hypothesis is H_0: $P(m(x_t) = 10\exp(-\theta_0 x_t^2)) = 1$ with $\theta_0 = 1$; and the alternative is H_1: $P(m(x_t) = 10\exp(-\theta_1 x_t^2) + \triangle_n(x_t)) = 1$ with $\theta_1 = 1$.

In situation II), the null hypothesis is H_0: $P(m(x_t) = (1 + x_t^2)\exp(-\theta_0 x_t^2)) = 1$ with $\theta_0 = 1$; and the alternative is H_1: $P(m(x_t) = (1 + x_t^2)\exp(-\theta_1 x_t^2) + \triangle_n(x_t)) = 1$ with $\theta_1 = 1$.

The sizes and power values are reported in Tables 1–4. Tables 1 and 3 contain all the sizes. It is readily seen that almost all the sizes fluctuate reasonably around

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

TABLE 1. Size: $m(x) = 10\exp(-\theta_0 x^2)$, $\theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
				$\alpha_0 = -0.01$, $\beta_0 = -0.01$			
200		0.0160	0.0126	0.0110	0.1222	0.1162	0.1090
500		0.0112	0.0140	0.0120	0.1162	0.1156	0.1046
1,200		0.0110	0.0120	0.0106	0.1006	0.1002	0.0918
				$\alpha_0 = 0.01$, $\beta_0 = 0.05$			
200		0.0162	0.0130	0.0140	0.1236	0.1176	0.1172
500		0.0136	0.0130	0.0130	0.1122	0.1052	0.1080
1,200		0.0110	0.0098	0.0094	0.1030	0.1050	0.1010

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = \beta_0 e_{t-1} + \epsilon_t$ and $\sigma = 0.7$.

the given significance levels, although not uniformly so. In Table 3, slightly less than half of the experiments are undersized, while most in Table 1 are oversized. Note also that in Table 3, in spite of the independence between e_t and x_t due to $a = 0$, the sizes in the test perform similarly to that of the test with $a = 0.2$. Overall, the sizes become stable when the sample size increases.

Tables 2 and 4 cover some important cases. Although the local departure function $\triangle_n(x) = \delta_n \triangle(x)$ is asymptotically negligible as δ_n approaches zero, the power values are very strong. It is clear that in most cases the power increases as either the sample size or κ involved in the truncation parameter increases.

TABLE 2. Power: $m(x) = 10\exp(-\theta_1 x^2) + \triangle_n(x)$, $\theta_1 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
				$\alpha_0 = -0.01$, $\beta_0 = -0.01$			
200		0.7810	0.7906	0.8016	0.9040	0.9070	0.9126
500		0.8196	0.8300	0.8410	0.9156	0.9176	0.9210
1,200		0.8544	0.8580	0.8660	0.9328	0.9320	0.9344
				$\alpha_0 = 0.01$, $\beta_0 = 0.05$			
200		0.7840	0.7956	0.8070	0.9030	0.9047	0.9092
500		0.8112	0.8252	0.8326	0.9150	0.9170	0.9220
1,200		0.8528	0.8588	0.8694	0.9318	0.9330	0.9380

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = \beta_0 e_{t-1} + \epsilon_t$ and $\sigma = 0.7$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

TABLE 3. Size: $m(x) = (1+x^2)\exp(-\theta_0 x^2),\ \theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
		$\alpha_0 = 0.05,$	$a = 0.2,$	$b = 0.9$			
200		0.0146	0.0136	0.0130	0.1092	0.1052	0.1052
500		0.0124	0.0116	0.0118	0.0958	0.0956	0.0958
1,200		0.0110	0.0104	0.0104	0.1022	0.1004	0.1008
		$\alpha_0 = 0.2,$	$a = 0,$	$b = 1$			
200		0.0094	0.0090	0.0080	0.1040	0.1058	0.1024
500		0.0114	0.0102	0.0082	0.0950	0.1012	0.0980
1,200		0.0102	0.0098	0.0114	0.0988	0.0992	0.0984

$u_t = \alpha_0 u_{t-1} + \epsilon_t,\ e_t = a\epsilon_t + b\eta_t,\ (\epsilon_t, \eta_t) \sim iiN(0, \sigma^2 I_2)$ and $\sigma = 0.5$.

This is because the increase of the sample size or κ should bring more terms of the orthogonal basis into L_n and, as can be seen in the proof of Theorem 3.2, each extra term may make a contribution to the power performance. Meanwhile, comparing the independent and dependent cases in Table 4, the endogeity does not attenuate the powers and power is strong and robust throughout.

Example 4.2

This example examines the test Π_n using two nonintegrable regression functions. The model for this simulation is $y_t = m(x_t) + e_t$.

TABLE 4. Power: $m(x) = (1+x^2)\exp(-\theta_1 x^2) + \triangle_n(x),\ \theta_1 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
		$\alpha_0 = 0.1,$	$a = 0.02,$	$b = -0.3$			
200		0.9968	0.9974	0.9978	0.9980	0.9988	0.9990
500		0.9972	0.9972	0.9974	0.9984	0.9992	0.9994
1,200		0.9978	0.9976	0.9986	0.9990	0.9990	0.9996
		$\alpha_0 = 0.2,$	$a = 0,$	$b = 0.5$			
200		0.9846	0.9856	0.9858	0.9954	0.9964	0.9958
500		0.9860	0.9868	0.9870	0.9956	0.9958	0.9962
1,200		0.9864	0.9860	0.9872	0.9986	0.9968	0.9968

$u_t = \alpha_0 u_{t-1} + \epsilon_t,\ e_t = a\epsilon_t + b\eta_t,\ (\epsilon_t, \eta_t) \sim iiN(0, \sigma^2 I_2)$ and $\sigma = 0.5$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

In situation I), the null hypothesis is H_{10}: $P(m(x_t) = \theta_0 x_t^2(x_t + \sin(x_t))) = 1$ with $\theta_0 = 1$; and the alternative is H_{11}: $P(m(x_t) = \theta_1 x_t^2(x_t + \sin(x_t)) + \triangle_n(x_t)) = 1$ with $\theta_1 = 1$.

In situation II), the null hypothesis is H_{10}: $P(m(x_t) = \theta_0 x_t^2(2 + x_t + \sin(x_t))) = 1$ with $\theta_0 = 1$; and the alternative is H_{11}: $P(m(x_t) = \theta_1 x_t^2(2 + x_t + \sin(x_t)) + \triangle_n(x_t)) = 1$ with $\theta_1 = 1$.

Tables 5 and 7 record the sizes of the test according to the two scenarios of the data generation, respectively. As can be seen, undersized experiments account for a larger percentage, especially in Table 7. Note also in Table 7 that, at the 1%

TABLE 5. Size: $m(x) = \theta_0 x^2(x + \sin(x))$, $\theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
				$\alpha_0 = -0.05, \quad \beta_0 = -0.05$			
200		0.0068	0.0074	0.0070	0.0824	0.0868	0.0854
500		0.0104	0.0100	0.0094	0.0908	0.0994	0.0920
1,200		0.0100	0.0104	0.0102	0.0994	0.0990	0.0988
				$\alpha_0 = -0.01, \quad \beta_0 = 0.04$			
200		0.0040	0.0044	0.0044	0.0814	0.0818	0.0844
500		0.0084	0.0080	0.0094	0.0898	0.0918	0.0954
1,200		0.0098	0.0106	0.0112	0.1060	0.1046	0.0984

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = \beta_0 e_{t-1} + \epsilon_t$, $\sigma = 0.6$.

TABLE 6. Power: $m(x) = \theta_0 x^2(x + \sin(x)) + \triangle_n(x)$, $\theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
				$\alpha_0 = -0.05, \quad \beta_0 = -0.05$			
200		0.9174	0.9294	0.9386	0.9562	0.9632	0.9686
500		0.9270	0.9350	0.9496	0.9590	0.9660	0.9730
1,200		0.9474	0.9520	0.9638	0.9756	0.9776	0.9824
				$\alpha_0 = -0.01, \quad \beta_0 = 0.04$			
200		0.9050	0.9184	0.9324	0.9496	0.9564	0.9626
500		0.9280	0.9388	0.9536	0.9630	0.9694	0.9736
1,200		0.9526	0.9582	0.9650	0.9758	0.9788	0.9840

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = \beta_0 e_{t-1} + \epsilon_t$, $\sigma = 0.6$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

TABLE 7. Size: $m(x) = \theta_0 x^2(2 + x + \sin(x))$, $\theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
		$\alpha_0 = 0.01,\ a = -0.05,\ b = -0.1$					
200		0.0058	0.0054	0.0050	0.0968	0.0900	0.0920
500		0.0068	0.0070	0.0060	0.0944	0.0924	0.0972
1,200		0.0104	0.0082	0.0104	0.0980	0.0926	0.0986
		$\alpha_0 = 0.05,\ a = 0.2,\ b = 0.9$					
200		0.0060	0.0052	0.0054	0.0920	0.0872	0.0868
500		0.0070	0.0074	0.0076	0.1062	0.0950	0.0940
1,200		0.0086	0.0078	0.0096	0.1004	0.0970	0.0990

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = a\epsilon_t + b\eta_t$ and $\sigma = 0.8$.

level, when $\kappa = 1/4.5$ the sizes are not close to the significant level even when $n = 1,200$. However, almost all the sizes approach the significant level stably as the sample size increases. Overall, the size performance is satisfactory.

Tables 6 and 8 give all the power values of the experiments. Power values are still quite strong and robust in this example, in the sense that they are increasing with the increase of the sample size and κ. The reason is the same as in Example 4.1. Also, it is noteworthy that in Table 8 there is a clear difference in power between two experiments, possibly due to the choice of the parameters α_0, a and b, or some other reason.

TABLE 8. Power: $m(x) = \theta_0 x^2(2 + x + \sin(x)) + \triangle_n(x)$, $\theta_0 = 1$

		Nominal size 1%			Nominal size 10%		
n	$\kappa =$	1/5	1/4.5	1/4	1/5	1/4.5	1/4
		$\alpha_0 = 0.01,\ a = -0.05,\ b = -0.1$					
200		0.9978	0.9984	0.9988	0.9982	0.9988	0.9990
500		0.9988	0.9990	0.9992	0.9992	0.9990	0.9996
1,200		0.9990	0.9988	0.9990	0.9992	0.9992	0.9996
		$\alpha_0 = 0.05,\ a = 0.2,\ b = 0.9$					
200		0.7478	0.7934	0.8172	0.8234	0.8960	0.9014
500		0.8174	0.8260	0.8474	0.8778	0.9052	0.9182
1,200		0.8632	0.8844	0.8928	0.9114	0.9418	0.9416

$u_t = \alpha_0 u_{t-1} + \epsilon_t$, $e_t = a\epsilon_t + b\eta_t$ and $\sigma = 0.8$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

5. CONCLUSIONS

We have proposed two concise and computationally simple tests for parametric specification of structural time series models with endogeneity and nonstationarity in both cases of integrable and nonintegrable regression functions. Asymptotic theory for each of the proposed test statistics has been established. Several Monte Carlo simulation experiments have been used to evaluate the finite-sample performance of the proposed tests. Overall, the size and power values are found to be satisfactory.

There are some possible extensions. One of them is whether the proposed test is extendable to accommodate the case where x_t is a vector of multivariate regressors in a nonparametric multivariate time series case. Our experience suggests that it may be possible to construct a multivariate series expansion and then a multivariate version of the proposed test with corresponding limit theory. Another issue is how to develop a data-driven method for the choice of the truncated parameter k such that it is optimal for testing purposes. In the stationary time series case, Chapter 3 of Gao (2007) shows that one may choose a suitable truncated parameter such that the power function of the test under consideration is maximized while size is controlled by a significance level. Since such extensions require developing new techniques, we leave that for future research.

REFERENCES

Cai, Z., Q. Li, & J. Park (2009) Functional-coefficient cointegration models for nonstationary time series data. *Journal of Econometrics* 148, 101–113.

Chen, J., J. Gao, & D. Li (2011) Estimation in semiparametric time series regression. *Statistics and its Interface* 4, 243–251.

Chen, X. & X. Shen (1998) Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.

Dong, C., J. Gao, & B. Peng (2015) Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics* 188, 301–312.

Dong, C., J. Gao, & D. Tjøstheim (2016) Estimation for single-index and partially linear single-index nonstationary time series models. *Annals of Statistics* 44, 425–453.

Dong, C., J. Gao, D. Tjøstheim, & J. Yin (2014) Specification testing for nonlinear multivariate cointegrating regressions. Available at http://business.monash.edu/econometrics-and-business-statistics/research/publications/ebs/wp08-14.pdf.

Gao, J. (2007) *Nonlinear Time Series: Semi- and Non-Parametric Methods*. Chapman & Hall/CRC.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009a) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009b) Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics* 37(68), 3893–3928.

Gao, J. & P.C.B. Phillips (2013a) Functional Coefficient Nonstationary Regression with Non- and Semi-Parametric Cointegration. 13, 1–56. Working paper. Available at http://ideas.repec.org/p/msh/ebswps/2013-16.html.

Gao, J. & P.C.B. Phillips (2013b) Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics* 176, 59–79.

Gao, J., H. Tong, & R. Wolff (2002) Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis* 83, 324–359.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Gao, J., Q. Wang, & J. Yin (2011) Specification testing in nonlinear time series with long-range dependence. *Econometric Theory* 27, 260–284.
Geman, D. & J. Horowitz (1980) Occupation densities. *Annals of Probability* 8, 1–67.
Hall, P., J.L. Horowitz, & B. Jing (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82, 561–574.
Hong, S.H. & P.C.B. Phillips (2010) Testing linearity of cointegriting relations with an application to purchasing power parity. *Journal of Business and Economic Statistics* 28(1), 96–114.
Horowitz, J. & V. Spokoiny (2001) An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.
Hualde, J. & P.M. Robinson (2011) Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* 39, 3152–3181.
Karlsen, H.A., T. Mykelbust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35, 252–299.
Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics* 87, 145–165.
Marinucci, D. & P.M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* 86, 103–120.
Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.
Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regression with integreted time series. *Econometrica* 69(1), 117–161.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20(2), 971–1001.
Revuz, D. & M. Yor (2005) *Continuous Martingales and Brownian Motion*. A Series of Comprehensive Studies in Mathematics, vol. 293. Springer-Verlag.
Wang, Q. & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegreting regression. *Econometric Theory* 25, 710–738.
Wang, Q. & P.C.B. Phillips (2009b) Structure nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Wang, Q. & P.C.B. Phillips (2012) A specification test for nonlinear nonstationary models. *Annals of Statistics* 40(2), 727–758.
Wang, Q. & P.C.B. Phillips (2016) Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory* 32, 359–401.

APPENDIX A: Lemmas

Appendix A presents a few technical lemmas which are crucial for the proofs of the theorems. Their proofs can be found in Appendix C of the supplemental material.

We shall consider several versions of decomposition for x_t. Without loss of generality, in what follows let $x_0 = 0$ almost surely. Observe that, suppressing δ_0 from $a_i(\delta_0)$,

$$x_t = \sum_{i=0}^{t-1} a_i u_{t-i} = \sum_{i=1}^{t} a_{t-i} u_i = \sum_{i=1}^{t} a_{t-i} \sum_{j=-\infty}^{i} \psi_{i-j} \epsilon_j$$
$$= \sum_{j=-\infty}^{t} \left(\sum_{i=\max(1,j)}^{t} \psi_{i-j} a_{t-i} \right) \epsilon_j := \sum_{j=-\infty}^{t} b_{t,j} \epsilon_j,$$

where $b_{t,j} = \sum_{i=\max(1,j)}^{t} \psi_{i-j} a_{t-i}$ with $j \le t$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Let $j \le t$ be fixed. Thus we always have

$$x_t = b_{t,j}\epsilon_j + x_{t/j}, \qquad \text{with } x_{t/j} := \sum_{\ell=-\infty,\neq j}^{t} b_{t,\ell}\epsilon_\ell, \tag{A.1}$$

where obviously $x_{t/j}$ and ϵ_j are mutually independent.

Additionally, letting $1 \le s < j \le t$,

$$\begin{aligned} x_t - x_s &= \sum_{i=1}^{t} a_{t-i}u_i - \sum_{i=1}^{s} a_{s-i}u_i = \sum_{i=1}^{s}[a_{t-i} - a_{s-i}]u_i + \sum_{i=s+1}^{t} a_{t-i}u_i \\ &= \sum_{i=1}^{s}[a_{t-i} - a_{s-i}]u_i + \sum_{i=s+1}^{t} a_{t-i}\sum_{j=-\infty}^{i}\psi_{i-j}\epsilon_j \\ &= \sum_{i=1}^{s}[a_{t-i} - a_{s-i}]u_i + \sum_{j=-\infty}^{s}\epsilon_j\sum_{i=s+1}^{t} a_{t-i}\psi_{i-j} + \sum_{j=s+1}^{t}\epsilon_j\sum_{i=j}^{t} a_{t-i}\psi_{i-j} \\ &:= \bar{x}_s + x_{ts}, \end{aligned} \tag{A.2}$$

where $\bar{x}_s = \sum_{i=1}^{s}[a_{t-i} - a_{s-i}]u_i + \sum_{j=-\infty}^{s}\epsilon_j\sum_{i=s+1}^{t}\psi_{i-j}a_{t-i}$. For convenience, denote $x_s^* = x_s + \bar{x}_s$ which contains all information available up to s. Clearly, x_s^* and x_{ts} are mutually independent.

Meanwhile, for any $j : s < j \le t$, x_{ts} can be written as

$$x_{ts} = b_{t,j}\epsilon_j + x_{ts/j}, \qquad \text{with } x_{ts/j} = \sum_{\ell=s+1,\neq j}^{t}\epsilon_\ell\sum_{i=\ell}^{t}\psi_{i-\ell}. \tag{A.3}$$

Evidently, $x_{ts/j}$, x_s^* and ϵ_j are independent of each other whenever $s < j \le t$. Moreover, when $t - s$ is large, $d_{ts}^{-1}\bar{x}_s = o_P(1)$.

LEMMA A.1. *Let $d_t = (Ex_t^2)^{1/2}$. When t is large, $d_t^2 \sim \psi^2\frac{1}{\Gamma^2(\alpha_0)}t^{2\alpha_0-1}$ for $\alpha_0 \in (1/2, 3/2)$. Let $d_{ts} = [E(x_{ts}^2)]^{1/2}$. When $t-s$ is large, $d_{ts}^2 \sim \psi^2\frac{1}{\Gamma^2(\alpha_0)}(t-s)^{2\alpha_0-1}$.*

We say that a function $h(x)$ satisfies a uniform Lipschitz condition if $\sup_x |h(x+y) - h(x)| \le C|y|$ for any y and some constant $C > 0$. It is easily seen that if $h(x)$ has a uniformly bounded derivative, then it satisfies the uniform Lipschitz condition. In practical use, we are interested in small $|y|$, although the inequality still holds.

LEMMA A.2. *Suppose that Assumption A holds. Let t and $t-s$ be large.*

(*1*) *$d_t^{-1}x_t$, where d_t is given by Lemma A.1, has a uniformly bounded density $f_t(x)$ over all t and x satisfying a uniform Lipschitz condition with some constant $C > 0$ independent of t.*

(*2*) *Let $j \le t$. $d_t^{-1}x_{t/j}$, with $x_{t/j}$ given by (A.1), has a uniformly bounded density $f_{t/j}(x)$ over all t, j and x satisfying a uniform Lipschitz condition with some constant $C > 0$ independent of t and j.*

(*3*) *$d_{ts}^{-1}x_{ts}$, with x_{ts} and d_{ts} given by (A.2) and Lemma A.1, respectively, has a uniformly bounded density $f_{ts}(x)$ over all t, s and x satisfying a uniform Lipschitz condition with some constant $C > 0$ independent of t and s.*

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

(4) *Let* $1 \le s < j \le t$. $d_{ts}^{-1} x_{ts/j}$, *where* $x_{ts/j}$ *given by (A.3), has a uniformly bounded density* $f_{ts/j}(x)$ *over all* t, j *and* s, x *satisfying a uniform Lipschitz condition with some constant* $C > 0$ *independent of* t, s *and* j.

LEMMA A.3. *Suppose that Assumption A holds. Let* j *be a fixed positive integer such that* $j \le t$ *where* t *is large and let* C *be a generic constant. For any functions* U *and* $f : \mathbb{R} \mapsto \mathbb{R}$ *such that* $\int |U(w)| dw < \infty$ *and* $E|\epsilon_j f(\epsilon_j)| < \infty$, *we have*

(1) $E[U(x_t) f(\epsilon_j)] = E[U(x_{t/j})] E[f(\epsilon_j)] + c_{t,j} d_t^{-2}$, *where* $c_{t,j}$ *is such that* $|c_{t,j}| \le C|b_{t,j}| \int |U(w)| dw$. *In particular, if* $Ef(\epsilon_j) = 0$, *then* $E[U(x_t) f(\epsilon_j)] = c_{t,j} d_t^{-2}$; *also,* $E|U(x_t) f(\epsilon_j)| \le C d_t^{-1} E|f(\epsilon_j)| \int |U(w)| dw$, *which remains true for* $f(\cdot) \equiv 1$, *that is,* $E|U(x_t)| \le C d_t^{-1} \int |U(w)| dw$.

(2) *For any* ℓ: $j \ne \ell \le t$, $E[U(x_t) f(\epsilon_j) | \epsilon_\ell] = E[U(x_{t/j}) | \epsilon_\ell] E[f(\epsilon_j)] + d_t^{-2} \eta_{j\ell}$, *where* $\eta_{j\ell}$ *is a random variable depending on* ϵ_ℓ *such that* $|\eta_{j\ell}| \le C|b_{t,j}|$ *almost surely. If* $E[f(\epsilon_j)] = 0$, *then* $E[U(x_t) f(\epsilon_j) | \epsilon_\ell] = d_t^{-2} \eta_{j\ell}$. *Also,* $E[|U(x_t) f(\epsilon_j)| | \epsilon_\ell] \le C d_t^{-1} \int |U(w)| dw$ *almost surely.*

(3) *For* $1 \le s < j \le t$ *and* $t - s$ *large,* $E[U(x_t) f(\epsilon_j) | \mathcal{F}_s] = E[f(\epsilon_j)] E[U(x_{t/j}) | \mathcal{F}_s] + d_{ts}^{-2} \xi_{s,j}$, *where* $|\xi_{s,j}| \le C|b_{t,j}| \int |U(x)| dx$ *almost surely. In addition,* $E[|U(x_t) f(\epsilon_j)| | \mathcal{F}_s] \le C d_{ts}^{-1} \int |U(w)| dw$ *almost surely.*

(4) *Similarly, letting* $1 \le s < j \ne \ell \le t$, *where* $t - s$ *is large and let* $g(x)$ *be a function such that* $E|\epsilon_\ell g(\epsilon_\ell)| < \infty$, *we have* $E[U(x_t) f(\epsilon_j) g(\epsilon_\ell) | \mathcal{F}_s] = E[f(\epsilon_j) g(\epsilon_\ell)] E[U(x_{t/j\ell}) | \mathcal{F}_s] + d_{ts}^{-2} \xi_{s,j\ell}$, *where* $x_{t/j\ell}$ *is defined as before and* $\xi_{s,j\ell}$ *satisfies that* $|\xi_{s,j\ell}| \le C(|b_{t,j}| + |b_{t,\ell}|) \int |U(x)| dx$ *almost surely. Moreover, we have* $E|U(x_t) f(\epsilon_j) g(\epsilon_\ell)| \le C d_t^{-1} \int |U(w)| dw$.

(5) *For* t *large and* e_t *defined by a linear process in A.3(i) of Assumption A,* $E[U(x_t) e_t] = q_t d_t^{-2}$ *with* $|q_t| \le C \int |U(w)| dw$ *and* $|E[U(x_t) e_t^2]| \le C d_t^{-1} \int |U(w)| dw$.

Corresponding to e_t defined by A.3(ii) in Assumption A, the following decomposition develops a similar version of the preceding lemma in Lemma A.4 below. Observe that, if $t > m_0$, by (A.2) we have

$$x_t = x_{t-m_0}^* + x_{t-m_0,t}, \tag{A.4}$$

where $x_{t-m_0,t} = \sum_{j=t-m_0+1}^{t} b_{t,j} \epsilon_j$. Then, by Lemma A.1, $E[x_{t-m_0}^*]^2 = O(1) d_{t-m_0}^2$ for large t, and it follows similarly from Lemma A.2 that $d_{t-m_0}^{-1} x_{t-m_0}^*$ has a density $f_{t-m_0}(w)$ uniformly bounded and satisfying a uniform Lipschitz condition.

If $t - s > m_0$,

$$x_t = x_s^* + x'_{s+1,t-m_0} + x_{t-m_0,t}, \tag{A.5}$$

where x_s^* is defined as before containing all information up to s, $x'_{s+1,t-m_0}$ is defined similarly containing all information on $[s+1, t-m_0]$, and $x_{t-m_0,t}$ is the same as before. By virtue of the structure of $x'_{s+1,t-m_0}$, $d_{t-s-m_0}^{-1} x'_{s+1,t-m_0}$ has a density $f_{s+1,t-m_0}(w)$ uniformly bounded and satisfying uniform Lipschitz condition, as $t - s$ is large.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

LEMMA A.4. *Let e_t be defined by A.3(ii) in Assumption A. Let $U(\cdot)$ be an integrable function on $\mathbb{R}$, i.e., $\int |U(w)|dw < \infty$, $f(\cdot)$ be a function such that $E|b(\epsilon_t, \ldots, \epsilon_{t-m_0+1})f(e_t)| < \infty$, where $b(\epsilon_t, \ldots, \epsilon_{t-m_0+1}) := x_{t-m_0,t}$. Denote by C a generic constant.*

(*1*) *For large t, $E[U(x_t)f(e_t)] = E[U(x^*_{t-m_0})]E[f(e_t)] + d^{-2}_{t-m_0}c_U$, where $x^*_{t-m_0}$ is given by (A.4) and c_U satisfies $|c_U| \le C\int |U(w)|dw$ almost surely. In particular, if $E[f(e_t)] = 0$, $E[U(x_t)f(e_t)] = d^{-2}_{t-m_0}c_U$; meanwhile, $|E[U(x_t)f(e_t)]| \le Cd^{-1}_{t-m_0}E|f(e_t)|\int |U(w)|dw$, where C is an absolute constant.*

(*2*) *For large $t-s$, $E[U(x_t)f(e_t)|\mathcal{F}_s] = E[U(x^*_s + x'_{s+1,t-m_0})|\mathcal{F}_s]E[f(e_t)] + d^{-2}_{t-s-m_0}\xi_s$ with*

$$|\xi_s| \le CE|b(\epsilon_t, \ldots, \epsilon_{t-m_0+1})f(e_t)| \int |U(w)|dw$$

almost surely, where x^*_s and $x'_{s+1,t-m_0}$ are defined by (A.5). If $E[f(e_t)] = 0$, $E[U(x_t)f(e_t)|\mathcal{F}_s] = d^{-2}_{t-s-m_0}\xi_s$; otherwise, we have $|E[U(x_t)f(e_t)|\mathcal{F}_s]| \le Cd^{-1}_{t-s-m_0}\int |U(w)|dwE|f(e_t)|$.

(*3*) *For $0 < t-s \le m_0$ where s is large, $|E[U(x_s)f(e_t)g(e_s)]| \le Cd^{-1}_{s-m_0}\int |U(x)|dxE|f(e_t)g(e_s)|$ for any function $g(\cdot)$ that satisfies the same condition as that for $f(\cdot)$.*

Let function sequence $\{\mathscr{T}_k(x)\}$ and function $\mathscr{T}(x)$ be defined as follows:

$$\mathscr{T}_k(x) := \frac{1}{k}\|Z(x)\|^2 = \frac{1}{k}\sum_{i=0}^{k-1}\mathscr{H}_i^2(x) \quad \text{and} \quad \mathscr{T}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{if } |x| \le 2, \\ 0, & \text{if } |x| > 2. \end{cases}$$

LEMMA A.5. (*1*) *The sequence $\mathscr{T}_k(x)$ converges to $\mathscr{T}(x)$ for any $x \in \mathbb{R}$ as $k \to \infty$. Moreover, as $k \to \infty$*

$$\int_{-\infty}^{\infty} |\mathscr{T}_k(x) - \mathscr{T}(x)|\,dx \to 0. \tag{A.6}$$

(*2*) *Furthermore, $Z'(x)Z(y) = \sum_{i=0}^{k-1}\mathscr{H}_i(x)\mathscr{H}_i(y) \to \delta(x-y)$ as $k \to \infty$, where $\delta(u)$ is the Deric delta function, i.e., $\delta(u) = 0$ if $u \ne 0$ and $\int \delta(u)du = 1$.*

APPENDIX B: Proofs of the Main Results

This appendix gives the full proof of Theorem 3.2, the outlines of the proofs of Lemma 3.1, Theorems 3.1 and 3.3–3.5. The full proofs of Lemma 3.1 and Theorems 3.1 and 3.3–3.5 are available from Appendix D of the supplemental material.

Outline of the Proof of Theorem 3.1. This proof is mainly for the case where e_t is a linear process stipulated by A.3(i) in Assumption A. Due to Lemma A.4, the proof for the case where e_t takes a functional form stipulated by A.3(ii) can be shown similarly. Under the hypothesis H_0, $y_t = g(x_t;\theta_0) + e_t$ for all $t = 1, \ldots, n$. Hence, it is easy

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

to rewrite $L_n = L_{1n} + L_{2n} + L_{3n}$, where, denoting $\widehat{g}(x) := g(x;\theta_0) - g(x;\widehat{\theta})$ for any real x, $L_{1n} = \sum_{t=1}^n \sum_{s=1}^n Z(x_t)'Z(x_s)e_t e_s$, $L_{2n} = 2\sum_{t=1}^n \sum_{s=1}^n Z(x_t)'Z(x_s)e_t \widehat{g}(x_s)$ and $L_{3n} = \sum_{t=1}^n \sum_{s=1}^n Z(x_t)'Z(x_s)\widehat{g}(x_t)\widehat{g}(x_s)$.

Note that $L_{1n} = \left\|\sum_{t=1}^n Z(x_t)e_t\right\|^2$ and $L_{3n} = \left\|\sum_{t=1}^n Z(x_t)\widehat{g}(x_t)\right\|^2$, while L_{2n} is the inner product of these vectors in L_{1n} and L_{3n}, multiplied by 2. Thus, both L_{1n} and L_{3n} are nonnegative, and Cauchy-Schwarz inequality entails $|L_{2n}| \le 2\sqrt{L_{1n}L_{3n}}$.

We first deal with L_{1n}. Observe that

$$\begin{aligned} L_{1n} &= \sum_{t=1}^n \sum_{s=1}^n Z(x_t)'Z(x_s)e_s e_t = \sum_{t=1}^n \|Z(x_t)\|^2 e_t^2 + 2\sum_{t=2}^n \sum_{s=1}^{t-1} Z(x_t)'Z(x_s)e_s e_t \\ &= \sigma_e^2 \sum_{t=1}^n \|Z(x_t)\|^2 + \sum_{t=1}^n \|Z(x_t)\|^2 (e_t^2 - \sigma_e^2) + 2\sum_{t=2}^n \sum_{s=1}^{t-1} Z(x_t)'Z(x_s)e_s e_t \\ &= \sigma_e^2 L'_{1n} + L''_{1n} + 2L'''_{1n}, \end{aligned}$$

where $\sigma_e^2 = Ee_t^2$. Using the notation in Lemma A.5,

$$\frac{d_n}{nk}L'_{1n} = \frac{d_n}{n}\sum_{t=1}^n \frac{1}{k}\|Z(x_t)\|^2 = \frac{d_n}{n}\sum_{t=1}^n \mathscr{T}_k(x_t) = \frac{d_n}{n}\sum_{t=1}^n \mathscr{T}(x_t) + \frac{d_n}{n}\sum_{t=1}^n \big[\mathscr{T}_k(x_t) - \mathscr{T}(x_t)\big].$$

It follows from Wang and Phillips (2009a) that, since $\int \mathscr{T}(x)dx = 1$,

$$\frac{d_n}{n}\sum_{t=1}^n \mathscr{T}(x_t) = \frac{d_n}{n}\sum_{t=1}^n \mathscr{T}(d_n x_{nt}) \to_D L_W(1,0).$$

Additionally, $\frac{d_n}{n}\sum_{t=1}^n [\mathscr{T}_k(x_t) - \mathscr{T}(x_t)] \to_P 0$. In fact, by Lemmas A.2 and A.5 we have

$$\begin{aligned} E\left|\frac{d_n}{n}\sum_{t=1}^n [\mathscr{T}_k(x_t) - \mathscr{T}(x_t)]\right| &\le \frac{d_n}{n}\sum_{t=1}^n E\,|\mathscr{T}_k(x_t) - \mathscr{T}(x_t)| \\ &= \frac{d_n}{n}\sum_{t=1}^n \int |\mathscr{T}_k(d_t x) - \mathscr{T}(d_t x)|\ f_t(x)dx \\ &= \frac{d_n}{n}\sum_{t=1}^n \frac{1}{d_t}\int |\mathscr{T}_k(x) - \mathscr{T}(x)|\ f_t(x/d_t)dx \\ &\le C\frac{d_n}{n}\sum_{t=1}^n \frac{1}{d_t}\int |\mathscr{T}_k(x) - \mathscr{T}(x)|dx \\ &= O(1)\int |\mathscr{T}_k(x) - \mathscr{T}(x)|dx \to 0. \end{aligned}$$

Hence, $\frac{d_n}{nk}L'_{1n} \to_D L_W(1,0)$. To complete the proof, it suffices to show that L''_{1n}, L'''_{1n}, L_{2n} and L_{3n} after normalisation are all $o_P(1)$. However, in order to save space, the proof of $\frac{d_n}{nk}L''_{1n} = o_P(1)$ and $\frac{d_n}{nk}L'''_{1n} = o_P(1)$ is given in Appendix D of the supplemental material of the paper, while that for $\frac{d_n}{nk}L_{2n} = o_P(1)$ and $\frac{d_n}{nk}L_{3n} = o_P(1)$ is given below.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

We shall prove $\frac{d_n}{nk}L_{3n} = o_P(1)$ first. As assumed in Assumption B, for any $\epsilon > 0$, there exists a fixed $M > 0$ such that $P(\|\widehat{\theta} - \theta_0\| > \zeta_n M) < \epsilon$. Then, for any $\delta > 0$,

$$P\left(\frac{d_n}{nk}L_{3n} > \delta\right) = P(\|\widehat{\theta} - \theta_0\| > \zeta_n M) + P\left(\frac{d_n}{nk}L_{3n} > \delta, \|\widehat{\theta} - \theta_0\| \le \zeta_n M\right)$$
$$\le \epsilon + \frac{d_n}{nk\delta}E[L_{3n}I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)],$$

by virtue of Markov's inequality and noting that L_{3n} is nonnegative, where $I(\cdot)$ stands for the conventional indicator function.

Using a Taylor expansion for $g(x;\theta)$ with respect to θ in a neighborhood of θ_0, we have

$$\widehat{g}(x_t) = g(x_t;\theta_0) - g(x_t;\widehat{\theta}) = l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta}) - \frac{1}{2}(\theta_0 - \widehat{\theta})'l_2(x_t;\bar{\theta})(\theta_0 - \widehat{\theta})$$

where $\bar{\theta}$ is between θ_0 and $\widehat{\theta}$. As a result,

$$\begin{aligned}
&\frac{d_n}{nk}E(|L_{3n}|I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M))\\
&= \frac{d_n}{nk}E\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)\widehat{g}(x_t)\widehat{g}(x_s)I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&= \frac{d_n}{nk}E\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta})l_1(x_s;\theta_0)'(\theta_0 - \widehat{\theta})I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&\quad - \frac{d_n}{nk}E\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta})(\theta_0 - \widehat{\theta})'l_2(x_s;\bar{\theta})(\theta_0 - \widehat{\theta})I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&\quad + \frac{d_n}{4nk}E\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)(\theta_0 - \widehat{\theta})'l_2(x_t;\bar{\theta})(\theta_0 - \widehat{\theta})(\theta_0 - \widehat{\theta})'l_2(x_s;\bar{\theta})(\theta_0 - \widehat{\theta})I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&= T_1 + T_2 + T_3, \qquad \text{say.}
\end{aligned}$$

Notice that

$$\begin{aligned}
T_1 &= \frac{d_n}{nk}E\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta})l_1(x_s;\theta_0)'(\theta_0 - \widehat{\theta})I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&= \frac{d_n}{nk}E\sum_{t=1}^{n}\|Z(x_t)\|^2[l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta})]^2I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&\quad + 2\frac{d_n}{nk}E\sum_{t=2}^{n}\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)l_1(x_t;\theta_0)'(\theta_0 - \widehat{\theta})l_1(x_s;\theta_0)'(\theta_0 - \widehat{\theta})I(\|\widehat{\theta} - \theta_0\| \le \zeta_n M)\\
&:= T_{11} + T_{12}, \quad \text{say.}
\end{aligned}$$

Using the Cauchy-Schwarz inequality, $\|Z(x)\|^2 \le O(1)k$ uniformly in x, and the boundedness of the densities $f_t(x)$ for $d_t^{-1}x_t$, we have

$$0 \le T_{11} \le M^2\zeta_n^2\frac{d_n}{n}E\sum_{t=1}^{n}\|l_1(x_t;\theta_0)\|^2 \le C\zeta_n^2\frac{d_n}{n}\sum_{t=1}^{n}d_t^{-1}\int\|l_1(x;\theta_0)\|^2dx \le C\zeta_n^2 = o(1).$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

For T_{12}, since $t > s$, the probability of $x_t = x_s$ is equal to zero, and hence in the calculation of the expectations in T_{12} we shall exclude these regions. Usually, such exclusion does not make any difference, but due to (2) of Lemma A.5 it really matters this time. Let $f_{ts}(x)$ and $f_s(y)$ stand for the densities of $d_{ts}^{-1}x_{ts}$ and $d_s^{-1}x_s$, respectively. Using the uniform boundedness of $f_{ts}(x)$ and $f_s(y)$ over x, y, t and s, and Lemma A.5,

$$
\begin{aligned}
|T_{12}| &\le C\frac{d_n}{nk}\zeta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}E[|Z(x_t)'Z(x_s)|\|l_1(x_t;\theta_0)\|\|l_1(x_s;\theta_0)\|] \\
&= C\frac{d_n}{nk}\zeta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}d_{ts}^{-1}d_s^{-1}\iint_{x\neq y}|Z(x)'Z(y)|\|l_1(x;\theta_0)\|\|l_1(y;\theta_0)\|dxdy \\
&\le C\frac{n}{d_nk}\zeta_n^2\iint_{x\neq y}|Z(x)'Z(y)|dxdy = O(1)\iint_{x\neq y}|Z(x)'Z(y)|dxdy \to 0,
\end{aligned}
$$

since $(n/d_nk)\zeta_n^2 \le C$ by Assumption B.

Similarly, again using Assumption B, we have as $n \to \infty$

$$
\begin{aligned}
0 \le T_3 &\le C\zeta_n^4\frac{d_n}{n}E\sum_{t=1}^{n}\|l_2(x_t;\bar{\theta})\|^2 + C\zeta_n^4\frac{d_n}{nk}E\sum_{t=2}^{n}\sum_{s=1}^{t-1}|Z(x_t)'Z(x_s)|\|l_2(x_t;\bar{\theta})\|\|l_2(x_s;\bar{\theta})\| \\
&\le O(1)\zeta_n^4\frac{d_n}{n}\sum_{t=1}^{n}d_t^{-1}\int l^2(x)dx + O(1)\zeta_n^4\frac{d_n}{nk}\sum_{t=2}^{n}\sum_{s=1}^{t-1}d_{ts}^{-1}d_s^{-1}\iint_{x\neq y}|Z(x)'Z(y)|l(x)l(y)dxdy \\
&\le O(1)\zeta_n^4 + o(1)\zeta_n^4(n/kd_n) \to 0.
\end{aligned}
$$

Notice that $T_1 = o_P(1)$ and $T_3 = o_P(1)$ imply $T_2 = o_P(1)$ by the Cauchy-Schwarz inequality. Hence, it follows that $\frac{d_n}{nk}L_{3n} \to_P 0$ and then by the Cauchy-Schwarz inequality it follows that $\frac{d_n}{nk}L_{2n} \to_P 0$ as well. This completes the outline of the proof of Theorem 3.1. ■

Proof of Theorem 3.2. We only show the theorem in the case where e_t is a linear process. In view of Lemma A.4, the case where e_t takes a functional form stipulated in A.3(ii) of Assumption A can be shown similarly.

Under H_1 and Assumption C, we have $m(x_t) = g(x_t;\theta_1) + \triangle_n(x_t) = g(x_t;\theta_1) + \delta_n\triangle(x_t)$. Thus, $\widehat{e}_t = y_t - g(x_t;\widehat{\theta}) = e_t + \delta_n\triangle(x_t) + \widehat{g}(x_t)$, where $\widehat{g}(x_t) = g(x_t;\theta_1) - g(x_t;\widehat{\theta})$. Write $L_n = L_{1n} + L_{2n} + L_{3n}$, where

$$
L_{1n} = \sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)(e_t + \delta_n\triangle(x_t))(e_s + \delta_n\triangle(x_s)),
$$

$$
L_{2n} = 2\sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)(e_t + \delta_n\triangle(x_t))\widehat{g}(x_s), \quad L_{3n} = \sum_{t=1}^{n}\sum_{s=1}^{n}Z(x_t)'Z(x_s)\widehat{g}(x_t)\widehat{g}(x_s).
$$

Given B.3 in Assumption B, in exactly the same way as in the proof of Theorem 3.1, we can have $\frac{d_n}{nk}L_{3n} = o_P(1)$. Also the Cauchy-Schwarz inequality implies that $|L_{2n}| \le 2\sqrt{L_{1n}L_{3n}}$. Hence, to fulfill the proof, it suffices to show that $\frac{d_n}{nk}L_{1n} \to_P \infty$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

To begin with,

$$L_{1n} = \delta_n^2 \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\triangle(x_t)\triangle(x_s) + 2\delta_n \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\triangle(x_t)e_s + \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)e_t e_s := L'_{1n} + 2L''_{1n} + L'''_{1n}, \text{ say.}$$

From the proof of Theorem 3.1, we have $\frac{d_n}{nk}L'''_{1n} \to_D \sigma_e^2 L_W(1,0)$. Rewrite $L''_{1n} = L''_{1n1} + L''_{1n2} + L''_{1n3}$, where $L''_{1n1} = \delta_n \sum_{t=1}^{n} \|Z(x_t)\|^2 \triangle(x_t)e_t$, $L''_{1n2} = \delta_n \sum_{t=2}^{n}\sum_{s=1}^{t-1} Z(x_t)'Z(x_s)\triangle(x_t)e_s$ and $L''_{1n3} = \delta_n \sum_{t=1}^{n-1}\sum_{s=t+1}^{n} Z(x_t)'Z(x_s)\triangle(x_t)e_s$.

Note that $\frac{d_n}{nk}L''_{1n1} = o_P(1)$. In fact, by (4) and (5) of Lemma A.3 and the integrability of $\triangle^2(x)$,

$$\begin{aligned}
&E\left(\frac{d_n}{nk}L''_{1n1}\right)^2 \\
&= \frac{d_n^2}{n^2k^2}\delta_n^2 E\left(\sum_{t=1}^{n} Z(x_t)'Z(x_t)\triangle(x_t)e_t\right)^2 \\
&= \frac{d_n^2}{n^2k^2}\delta_n^2 E\sum_{t=1}^{n}\|Z(x_t)\|^4\triangle^2(x_t)e_t^2 + 2\frac{d_n^2}{n^2k^2}\delta_n^2 E\sum_{t=2}^{n}\sum_{s=1}^{t-1}\|Z(x_t)\|^2\triangle(x_t)e_t\|Z(x_s)\|^2\triangle(x_s)e_s \\
&\le O(1)\frac{d_n^2}{n^2}\delta_n^2\sum_{t=1}^{n} E\left[\triangle^2(x_t)e_t^2\right] \\
&\quad + 2\frac{d_n^2}{n^2k^2}\delta_n^2 E\sum_{t=2}^{n}\sum_{s=1}^{t-1}\|Z(x_t)\|^2\triangle(x_t)\|Z(x_s)\|^2\triangle(x_s)\sum_{j=-\infty}^{t}\phi_{t-j}\epsilon_j\sum_{\ell=-\infty}^{s}\phi_{s-\ell}\epsilon_\ell \\
&\le O(1)\delta_n^2 + 2\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{j=s+1}^{t}\phi_{t-j}\sum_{\ell=-\infty}^{s}\phi_{s-\ell}E\left[\|Z(x_t)\|^2\triangle(x_t)\|Z(x_s)\|^2\triangle(x_s)\epsilon_j\epsilon_\ell\right] \\
&\quad + 2\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\phi_{t-j}\sum_{\ell=-\infty}^{s}\phi_{s-\ell}E\left[\|Z(x_t)\|^2\triangle(x_t)\|Z(x_s)\|^2\triangle(x_s)\epsilon_j\epsilon_\ell\right] \\
&= O(1)\delta_n^2 + 2\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}d_{ts}^{-2}\sum_{j=s+1}^{t}\phi_{t-j}\sum_{\ell=-\infty}^{s}\phi_{s-\ell}E\left[\xi_s\|Z(x_s)\|^2\triangle(x_s)\epsilon_\ell\right] \\
&\quad + 2\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\phi_{t-j}\sum_{\ell=-\infty}^{s}\phi_{s-\ell}E\left[\|Z(x_t)\|^2\triangle(x_t)\|Z(x_s)\|^2\triangle(x_s)\epsilon_j\epsilon_\ell\right],
\end{aligned}$$

where for $j > s$, due to $E\epsilon_j = 0$, $E[\|Z(x_t)\|^2\triangle(x_t)\epsilon_j|\mathcal{F}_s] = d_{ts}^{-2}\xi_s$ and $|\xi_s| \le O(1)\int\|Z(x)\|^2|\triangle(x)|dx \le O(1)k$ a.s. and therefore $|E[\xi_s\|Z(x_s)\|^2\triangle(x_s)\epsilon_\ell]| \le O(1)d_s^{-1}k^2$ by (2) of Lemma A.3. As a result, the second term in absolute value is bounded by $O(1)\frac{d_n^2}{n^2}\delta_n^2\sum_{t=2}^{n}\sum_{s=1}^{t-1}d_{ts}^{-2}d_s^{-1} = O(1)n^{-\delta_0+1/2}\delta_n^2 = o(1)$. Meanwhile, we simply use (2) and (4) of Lemma A.3 to show that

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

$$\left|E\left[\|Z(x_t)\|^2\triangle(x_t)\|Z(x_s)\|^2\triangle(x_s)\epsilon_j\epsilon_\ell\right]\right|$$
$$\le O(1)d_{ts}^{-1}\int\|Z(x)\|^2|\triangle(x)|dx d_s^{-1}\int\|Z(x)\|^2|\triangle(x)|dx \le O(1)d_{ts}^{-1}d_s^{-1}k^2,$$

no matter where ϵ_j and ϵ_ℓ are located. This, together with $\sum_{j=-\infty}^{s}|\phi_{t-j}| \le O(1)(t-s)^{-\gamma_0+1}$ and $\gamma_0 > 3/2$ by A.3(i) of Assumption A, ensures the last term to be infinitesimal.

Note also that

$$\frac{d_n}{nk}L''_{1n2} = \frac{d_n}{nk}\delta_n\sum_{t=2}^{m_n}\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)\triangle(x_t)e_s + \frac{d_n}{nk}\delta_n\sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)\triangle(x_t)e_s,$$

where the first term is $o_P(1)$, given a certain condition on m_n. In effect,

$$\frac{d_n}{nk}\delta_n\sum_{t=2}^{m_n}\sum_{s=1}^{t-1}E|Z(x_t)'Z(x_s)\triangle(x_t)e_s| \le O(1)\frac{d_n}{n}\delta_n\sum_{t=2}^{m_n}\sum_{s=1}^{t-1}E|e_s| = O(1)\frac{d_n}{n}\delta_n m_n^2 = o(1)$$

as long as $\frac{d_n}{n}m_n^2 = O(1)$. Due to this reason, we only consider the case of $t > m_n$:

$$E\left(\frac{d_n}{nk}\delta_n\sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)\triangle(x_t)e_s\right)^2$$
$$= \frac{d_n^2}{n^2k^2}\delta_n^2E\left(\sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)\triangle(x_t)e_s\right)^2$$
$$= \frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t=m_n+1}^{n}\left(\sum_{s=1}^{t-1}Z(x_t)'Z(x_s)\triangle(x_t)e_s\right)^2$$
$$+2\frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+2}^{n}\sum_{s_1=1}^{t_1-1}Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})e_{s_1}\sum_{t_2=m_n+1}^{t_1-1}\sum_{s_2=1}^{t_2-1}Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})e_{s_2}$$
$$= \frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1}\left(Z(x_t)'Z(x_s)\triangle(x_t)e_s\right)^2$$
$$+2\frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)e_{s_1}e_{s_2}$$
$$+2\frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+2}^{n}\sum_{t_2=m_n+1}^{t_1-1}\sum_{s_2=1}^{t_2-1}Z(x_{t_1})'Z(x_{s_2})\triangle(x_{t_1})Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})e_{t_2}e_{s_2}$$
$$+2\frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s=1}^{t_2-1}Z(x_{t_1})'Z(x_s)\triangle(x_{t_1})Z(x_{t_2})'Z(x_s)\triangle(x_{t_2})e_s^2$$
$$+4\frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s_1=2}^{t_2-1}\sum_{s_2=1}^{s_1-1}Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})e_{s_1}Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})e_{s_2}$$
$$:= \sum_{i=1}^{5}T_i, \quad \text{say.}$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Observe that, given $\mathcal{F}_s$ and $t-s$ large, $d_{ts}^{-1}x_{ts}$ has a uniformly bounded density $f_{ts}(x)$, and hence

$$\begin{aligned}
T_1 =& \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1} E\{[E(Z(x_t)'Z(x_s)\triangle(x_t))^2|\mathcal{F}_s]e_s^2\} \\
\le& \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1} d_{ts}^{-1} E\left[\int (Z(x)'Z(x_s)\triangle(x))^2 dx e_s^2\right] \\
\le& O(1)\frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1} d_{ts}^{-1} E[\|Z(x_s)\|^2 e_s^2] \\
\le& O(1)\frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s=1}^{t-1} d_{ts}^{-1} d_s^{-1}\int \|Z(x)\|^2 dx = O(1)\frac{1}{k}\delta_n^2 = o(1),
\end{aligned}$$

by the orthogonality of the basis and (5) of Lemma A.3.

Ignoring the unimportant constants in each T_i in the following computation, we have

$$\begin{aligned}
T_2 =& \frac{d_n^2}{n^2k^2}\delta_n^2 E \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)e_{s_1}e_{s_2} \\
=& \frac{d_n^2}{n^2k^2}\delta_n^2 E \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t) \sum_{j=-\infty}^{s_1}\phi_{s_1-j}\epsilon_j \sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}\epsilon_\ell \\
=& \frac{d_n^2}{n^2k^2}\delta_n^2 E \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t) \sum_{j=s_2+1}^{s_1}\phi_{s_1-j}\epsilon_j \sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}\epsilon_\ell \\
&+ \frac{d_n^2}{n^2k^2}\delta_n^2 E \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1} Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t) \sum_{j=-\infty}^{s_2}\phi_{s_1-j}\epsilon_j \sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}\epsilon_\ell \\
=& \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\sum_{j=s_2+1}^{s_1}\phi_{s_1-j}\sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}E\left[Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)\epsilon_j\epsilon_\ell\right] \\
&+ \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\sum_{j=-\infty}^{s_2}\phi_{s_1-j}\sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}E\left[Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)\epsilon_j\epsilon_\ell\right] \\
:=& T_{21}+T_{22}.
\end{aligned}$$

Note that for $t > s_1 > s_2$ any two of the time series x_t, x_{s_1} and x_{s_2} cannot be completely identical with probability one. In the calculation of the expectations in T_{21} and T_{22}, we therefore exclude these regions. Given $\mathcal{F}_{s_1}$, $d_{ts_1}^{-1}x_{ts_1}$ has a uniformly bounded density $f_{ts_1}(x)$. Whence, in T_{21},

$$\begin{aligned}
&E[Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)|\mathcal{F}_{s_1}] \\
&\quad = d_{ts_1}^{-1}\int Z(x)'Z(x_{s_1})Z(x)'Z(x_{s_2})\triangle^2(x) f_{ts_1}\left(\frac{x-x_{s_1}^*}{d_{ts_1}}\right)dx.
\end{aligned}$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Then, $E[Z(x)'Z(x_{s_1})Z(x)'Z(x_{s_2})\triangle^2(x)f_{ts_1}\left(\frac{x-x^*_{s_1}}{d_{ts_1}}\right)\epsilon_j|\mathcal{F}_{s_2}] = d^{-2}_{s_1s_2}\xi_{s_2}(x)$ for any x by (4) of Lemma A.3, where $|\xi_{s_2}(x)| \le O(1)\int_{y\neq x}|Z(x)'Z(y)Z(x)'Z(x_{s_2})|\triangle^2(x)dy$. Thus, (2) of Lemma A.3 gives

$$
\begin{aligned}
|T_{21}| &= \left|\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\sum_{j=s_2+1}^{s_1}\phi_{s_1-j}\sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}E[Z(x_t)'Z(x_{s_1})Z(x_t)'Z(x_{s_2})\triangle^2(x_t)\epsilon_j\epsilon_\ell]\right| \\
&= \left|\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}d^{-1}_{ts_1}d^{-2}_{s_1s_2}\sum_{j=s_2+1}^{s_1}\phi_{s_1-j}\sum_{\ell=-\infty}^{s_2}\phi_{s_2-\ell}\int E[\xi_{s_2}(x)\epsilon_\ell]dx\right| \\
&\le O(1)\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t=m_n+1}^{n}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}d^{-1}_{ts_1}d^{-2}_{s_1s_2}d^{-1}_{s_2}\iiint_{x\neq y\neq z}|Z(x)'Z(y)Z(x)'Z(z)|\triangle^2(x)dxdydz \\
&= O(1)\frac{d_n^2}{n^2k^2}\delta_n^2n^{3-2\delta_0}\iiint_{x\neq y\neq z}|Z(x)'Z(y)Z(x)'Z(z)|\triangle^2(x)dxdydz \\
&= O(1)\frac{1}{k^2}\delta_n^2\iiint_{x\neq y\neq z}|Z(x)'Z(y)Z(x)'Z(z)|\triangle^2(x)dxdydz = o(1),
\end{aligned}
$$

where the triple integral is $o(1)$ as $n\to\infty$ by (2) of Lemma A.5.

Simply, an application of (2) of Lemma A.3 establishes that T_{22} is smaller than T_{21} in absolute value since $\sum_{j=-\infty}^{s_2}|\phi_{s_1-j}| \le O(1)(s_1-s_2)^{-\gamma_0+1}$ and $\gamma_0 > 3/2$. T_3 can be calculated similarly to T_2 because the only difference between them is that $\triangle(x_t)^2$ in T_2 is replaced by $\triangle(x_{t_1})\triangle(x_{t_2})$ in T_3. Obviously, the replacement does not affect the calculation of the expectation.

Regarding to T_4, by virtue of the conditional densities for $d^{-1}_{t_1t_2}(x_{t_1}-x_{t_2})$ given $\mathcal{F}_{t_2}$, $d^{-1}_{t_2s}(x_{t_2}-x_s)$ given $\mathcal{F}_s$ and making use of (6) of Lemma A.3 for the remaining expectation, we have

$$
\begin{aligned}
|T_4| &= \frac{d_n^2}{n^2k^2}\delta_n^2\left|E\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s=1}^{t_2-1}Z(x_{t_1})'Z(x_s)\triangle(x_{t_1})Z(x_{t_2})'Z(x_s)\triangle(x_{t_2})e_s^2\right| \\
&\le O(1)\frac{d_n^2}{n^2k^2}\delta_n^2\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s=1}^{t_2-1}d^{-1}_{t_1t_2}d^{-1}_{t_2s}d_s^{-1} \\
&\quad\times\iiint|Z(x)'Z(z)\triangle(x)Z(y)'Z(z)\triangle(y)|dxdydz \\
&= O(1)\frac{d_n^2}{n^2k^2}\delta_n^2n^{-3\delta_0+9/2}\int dz\left(\int|Z(x)'Z(z)\triangle(x)|dx\right)^2 \\
&\le O(1)\frac{1}{k^2}\delta_n^2n^{-\delta_0+3/2}\int dz\int|Z(x)'Z(z)|^2dx\int|\triangle(x)|^2dx = O(1)\frac{1}{k}\delta_n^2n^{-\delta_0+3/2},
\end{aligned}
$$

in view of the Cauchy-Schwartz inequality. Lastly,

$$
\begin{aligned}
T_5 &= \frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s_1=2}^{t_2-1}\sum_{s_2=1}^{s_1-1}Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})e_{s_1}Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})e_{s_2} \\
&= \frac{d_n^2}{n^2k^2}\delta_n^2E\sum_{t_1=m_n+3}^{n}\sum_{t_2=m_n+2}^{t_1-1}\sum_{s_1=2}^{t_2-1}\sum_{s_2=1}^{s_1-1}Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})
\end{aligned}
$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

$$\times \sum_{j=-\infty}^{s_1} \phi_{s_1-j}\epsilon_j \sum_{\ell=-\infty}^{s_2} \phi_{s_2-\ell}\epsilon_\ell$$

$$= \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t_1=m_n+3}^{n} \sum_{t_2=m_n+2}^{t_1-1} \sum_{s_1=2}^{t_2-1} \sum_{s_2=1}^{s_1-1} \sum_{j=s_2+1}^{s_1} \phi_{s_1-j} \sum_{\ell=-\infty}^{s_2} \phi_{s_2-\ell}$$

$$\times E[Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})\epsilon_j\epsilon_\ell]$$

$$+ \frac{d_n^2}{n^2k^2}\delta_n^2 \sum_{t_1=m_n+3}^{n} \sum_{t_2=m_n+2}^{t_1-1} \sum_{s_1=2}^{t_2-1} \sum_{s_2=1}^{s_1-1} \sum_{j=-\infty}^{s_2} \phi_{s_1-j} \sum_{\ell=-\infty}^{s_2} \phi_{s_2-\ell}$$

$$\times E[Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})\epsilon_j\epsilon_\ell] := T_{51} + T_{52}.$$

Let $f_{t_1t_2}(w)$ and $f_{t_2s_1}(x)$ stand for the conditional densities of $d_{t_1t_2}^{-1}(x_{t_1} - x_{t_2})$ given $\mathcal{F}_{t_2}$, $d_{t_2s_1}^{-1}(x_{t_2} - x_{s_1})$ given $\mathcal{F}_{s_1}$, respectively. To deal with T_{51}, we use (4) of Lemma A.3 to show

$$E[Z(x_{t_1})'Z(x_{s_1})\triangle(x_{t_1})Z(x_{t_2})'Z(x_{s_2})\triangle(x_{t_2})\epsilon_j\epsilon_\ell]$$

$$= d_{t_1t_2}^{-1}d_{t_2s_1}^{-1} \iint_{w\neq x} E\Bigg[Z(w)'Z(x_{s_1})\triangle(w)Z(x)'Z(x_{s_2})\triangle(x)$$

$$\times f_{t_1t_2}\left(\frac{w-x}{d_{t_1t_2}}\right) f_{t_2s_1}\left(\frac{x-x_{s_1}}{d_{t_2s_1}}\right)\epsilon_j\epsilon_\ell\Bigg] dwdx$$

$$= d_{t_1t_2}^{-1}d_{t_2s_1}^{-1}d_{s_1s_2}^{-2}E[\xi_{s_2}\epsilon_\ell],$$

due to $E\epsilon_j = 0$, where ξ_{s_2} satisfies $|\xi_{s_2}| \leq \iiint_{w\neq x\neq y} |Z(w)'Z(y)\triangle(w)Z(x)'Z(x_{s_2})\triangle(x)|dwdxdy$, a.s. by the proof of Lemma A.2 in the online supplementary material.

Since $\iint_{w\neq y} |Z(w)'Z(y)\triangle(w)|dwdy = o(1)$ as $n \to \infty$ by (2) of Lemma A.5, using (2) of Lemma A.3, we have

$$|E[\xi_{s_2}\epsilon_\ell]| \leq \iiint_{w\neq x\neq y} E|Z(w)'Z(y)\triangle(w)Z(x)'Z(x_{s_2})\triangle(x)\epsilon_\ell|dwdxdy$$

$$\leq O(1)d_{s_2}^{-1} \iiiint_{w\neq x\neq y\neq z} |Z(w)'Z(y)\triangle(w)Z(x)'Z(z)\triangle(x)|dwdxdydz$$

$$= O(1)d_{s_2}^{-1}\left(\iint_{w\neq y} |Z(w)'Z(y)\triangle(w)|dwdy\right)^2 = o(1)d_{s_2}^{-1},$$

Therefore, we obtain

$$|T_{51}| \leq o(1)\frac{1}{nk^2}\delta_n^2 \sum_{t_1=m_n+3}^{n} \sum_{t_2=m_n+2}^{t_1-1} \sum_{s_1=2}^{t_2-1} \sum_{s_2=1}^{s_1-1} d_{t_1t_2}^{-1}d_{t_2s_1}^{-1}d_{s_1s_2}^{-2}d_{s_2}^{-1}$$

$$= O(1)\frac{1}{k^2}n^{3/2-\delta_0}\delta_n^2 = o(1)\frac{1}{k}(n/d_n)\delta_n^2.$$

Because of $\sum_{j=-\infty}^{s_2} |\phi_{s_1-j}| = O(1)(s_1 - s_2)^{-\gamma_0+1}$, $\gamma_0 > 3/2$ and a routine derivation, it is easy to obtain $|T_{52}| = o(1)\frac{1}{k}(n/d_n)\delta_n^2$ in view of the proof for $|T_{51}|$. This gives $|T_5| = o(1)\frac{1}{k}(n/d_n)\delta_n^2$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

From the above derivations and noting $T_{41} = O(1)\frac{1}{k}(n/d_n)\delta_n^2$, it follows that $\frac{d_n}{nk}L''_{1n2} = O_P((n/d_n)^{1/2}k^{-1/2}\delta_n)$. Similar calculation yields the same result for $\frac{d_n}{nk}L''_{1n3}$. Therefore, $\frac{d_n}{nk}L''_{1n} = O_P((n/d_n)^{1/2}k^{-1/2}\delta_n)$.

We then have as $n \to \infty$

$$\begin{aligned}\frac{d_n}{nk}L'_{1n} &= \delta_n^2\frac{d_n}{nk}\sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\triangle(x_t)\triangle(x_s)\\ &= \frac{\delta_n^2 d_n}{nk}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=0}^{k-1}\mathscr{H}_i(x_t)\mathscr{H}_i(x_s)\triangle(x_t)\triangle(x_s) = \frac{\delta_n^2 d_n}{nk}\sum_{i=0}^{k-1}\left(\sum_{t=1}^{n}\mathscr{H}_i(x_t)\triangle(x_t)\right)^2\\ &= \frac{\delta_n^2 n}{d_n k}\sum_{i=0}^{k-1}\left(\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_i(x_t)\triangle(x_t)\right)^2 \geq \frac{\delta_n^2 n}{d_n k}\left(\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_{i_0}(x_t)\triangle(x_t)\right)^2 \to_P \infty\end{aligned}$$

because $\left(\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_{i_0}(x_t)\triangle(x_t)\right)^2 \to_D \left(\int \mathscr{H}_{i_0}(x)\triangle(x)dx L_W(1,0)\right)^2$ by the continuous mapping theorem and $\delta_n^2 n/(d_n k) \to \infty$ by Assumption C, where $i_0 \geq 0$ is the first integer such that $\int \mathscr{H}_{i_0}(x)\triangle(x)dx \neq 0$. Obviously, such i_0 does exist, otherwise for any i, $\int \mathscr{H}_i(x)\triangle(x)dx = 0$ so that $\triangle(x)$ is orthogonal with every $\mathscr{H}_i(x)$ in Hilbert space $L^2(\mathbb{R})$. This amounts to saying that $\triangle(x)$ is a zero function. This completes the proof. ■

Proof of Lemma 3.1. It follows from Hualde and Robinson (2011) that $\widehat{\rho}$ is consistent, $\widehat{\rho} \to_P \rho_0$. Meanwhile, the series $\sum_{j=0}^{\infty}\psi_j(\widehat{\rho})$ converges uniformly in $\widehat{\rho}$ because of A.2(i) in Assumption A and the compactness of Ξ. The uniform convergence gives the continuity of $\widehat{\psi}$ in $\widehat{\rho}$. Hence $\widehat{\psi} \to_P \psi$ and $\widehat{d}_n/d_n \to_P 1$ by the continuous mapping theorem.

Next, to prove $\widehat{\sigma}_e^2 \to_P \sigma_e^2$, note that under the null, $\widehat{e}_t = e_t + \widehat{g}(x_t)$ where $\widehat{g}(x_t) := g(x_t,\theta_0) - g(x_t,\widehat{\theta})$. Thus,

$$\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}(e_t+\widehat{g}(x_t))^2 = \frac{1}{n}\sum_{t=1}^{n}e_t^2 + \frac{1}{n}\sum_{t=1}^{n}\widehat{g}^2(x_t) + 2\frac{1}{n}\sum_{t=1}^{n}e_t\widehat{g}(x_t).$$

To begin with, we shall show that $\frac{1}{n}\sum_{t=1}^{n}e_t^2 \to_P \sigma_e^2$. First, suppose e_t is stipulated in A.3(i) of Assumption A. The assertion holds due to Theorem 3.7 of Phillips and Solo (1992).

Second, let $e_t = \wp(\epsilon_t,\ldots,\epsilon_{t-m_0+1};\eta_t,\ldots,\eta_{t-m_1+1})$ being stipulated in A.3(ii) of Assumption A. Let $\bar{m} = \max(m_0,m_1)$. Notice that whenever $|t-s| > \bar{m}$, e_t and e_s are independent. Hence,

$$\begin{aligned}E\left(\frac{1}{n}\sum_{t=1}^{n}e_t^2 - \sigma_e^2\right)^2 &= \frac{1}{n^2}\sum_{t=1}^{n}E(e_t^2-\sigma_e^2)^2 + 2\frac{1}{n^2}E\sum_{t=2}^{n}\sum_{s=1}^{t-1}(e_t^2-\sigma_e^2)(e_s^2-\sigma_e^2)\\ &= \frac{1}{n}Var(e_1^2) + 2\frac{1}{n^2}E\sum_{t=\bar{m}+1}^{n}\sum_{s=t-\bar{m}}^{t-1}(e_t^2-\sigma_e^2)(e_s^2-\sigma_e^2)\\ &\leq o(1) + 2\frac{1}{n^2}n\bar{m}Var(e_1^2) = o(1).\end{aligned}$$

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

To fulfill the proof, it suffices to show that $\frac{1}{n}\sum_{t=1}^{n}\widehat{g}^2(x_t) = o_P(1)$. However, this is similar to and easier than $\frac{d_n}{nk}L_{3n} = o_P(1)$ in the proof of Theorem 3.1 and we therefore omit it. Consequently, the last term $\frac{1}{n}\sum_{t=1}^{n} e_t\widehat{g}(x_t) = o_P(1)$. The assertion is proved.

Finally, since the assertion that $\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_0^2(x_t) - L_W(1,0) = o_P(1)$ is the consequence of equation (2.7) in Theorem 2.1 of Wang and Phillips (2009a, p. 716) and Lemma 2.3(b) Park and Phillips (1999, p. 271), we have $\widehat{L}_W(1,0) - L_W(1,0) \to_P 0$ as $n \to \infty$ due to $\widehat{d}_n/d_n \to_P 1$. ■

Proof of Theorem 3.3. This is obviously due to Lemma 3.1, Theorems 3.1 and 3.2. ■

Proof of Theorem 3.4. The proof for the first part is almost the same as that of Theorem 3.1 except that

$$\begin{aligned}\frac{d_n}{nk}\sum_{t=1}^{n}\|Z(x_t)\|^2\varphi^2(x_t) &= \frac{d_n}{n}\sum_{t=1}^{n}\mathscr{T}_k(x_t)\varphi^2(x_t)\\ &= \frac{d_n}{n}\sum_{t=1}^{n}\mathscr{T}(x_t)\varphi^2(x_t)\\ &\quad + \frac{d_n}{n}\sum_{t=1}^{n}[\mathscr{T}_k(x_t) - \mathscr{T}(x_t)]\varphi^2(x_t) \to_D \int \mathscr{T}(x)\varphi^2(x)dx L_W(1,0)\end{aligned}$$

by Lemma A.5 and Wang and Phillips (2009a). Thus, the limit is different from that of Theorem 3.1.

For the second part, the consistency follows exactly the same as for Theorem 3.2, as $\Delta(x)e^{-x^2/2}$ plays exactly the role of $\Delta(x)$ in the integrable case. Indeed, recall that

$$\Pi_n = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\widehat{\varepsilon}_t\widehat{\varepsilon}_s,$$

where $\widehat{\varepsilon}_t := \widehat{e}_t\varphi(x_t) = (y_t - g(x_t;\widehat{\theta}))\varphi(x_t)$.

Observe that under H_{11} in the paper, $\widehat{e}_t = y_t - g(x_t;\widehat{\theta}) = e_t + g(x_t,\theta_1) - g(x_t,\widehat{\theta}) + \triangle_n(x_t)$. We then write $\Pi_n = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)\widehat{e}_t\widehat{e}_s$ as $\Pi_n = \Pi_{1n} + \Pi_{2n} + \Pi_{3n}$, where

$$\Pi_{1n} = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)(e_t + \triangle_n(x_t))(e_s + \triangle_n(x_s)),$$

$$\Pi_{2n} = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)\widehat{g}(x_t)\widehat{g}(x_s),$$

$$\Pi_{3n} = 2\sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)(e_t + \triangle_n(x_t))\widehat{g}(x_s),$$

where we denote $\widehat{g}(x_t) := g(x_t,\theta_1) - g(x_t,\widehat{\theta})$.

Following exactly the same arguments as in the proof of Theorem 3.2 in the current paper, we may show

$$\frac{d_n}{nk}\Pi_{2n} = o_P(1) \quad \text{and} \quad \frac{d_n}{nk}\Pi_{3n} = o_P(1)$$

as $n \to \infty$.

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press

Further write $\Pi_{1n} = \Pi'_{1n} + \Pi''_{1n} + \Pi'''_{1n}$ where

$$\Pi'_{1n} = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)\triangle_n(x_t)\triangle_n(x_s),$$
$$\Pi''_{1n} = \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)e_t e_s,$$
$$\Pi'''_{1n} = 2\sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)e_t\triangle_n(x_s).$$

Notice that

$$\begin{aligned}\frac{d_n}{nk}\Pi'_{1n} &= \sum_{t=1}^{n}\sum_{s=1}^{n} Z(x_t)'Z(x_s)\varphi(x_t)\varphi(x_s)\triangle_n(x_t)\triangle_n(x_s)\\ &= \frac{\delta_n^2 d_n}{nk}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{i=0}^{k-1}\mathscr{H}_i(x_t)\mathscr{H}_i(x_s)\varphi(x_t)\varphi(x_s)\triangle(x_t)\triangle(x_s)\\ &= \frac{\delta_n^2 n}{d_n k}\sum_{i=0}^{k-1}\left(\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_i(x_t)\varphi(x_t)\triangle(x_t)\right)^2\\ &\geq \frac{\delta_n^2 n}{d_n k}\left(\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_{i_0}(x_t)\varphi(x_t)\triangle(x_t)\right)^2\end{aligned}$$

where $i_0 : 0 \leq i_0 \leq k-1$ such that $\int \mathscr{H}_{i_0}(x)\varphi(x)\triangle(x)dx \neq 0$. For the same reason as in Theorem 3.2, such i_0 does exist.

Moreover, $\frac{d_n}{n}\sum_{t=1}^{n}\mathscr{H}_{i_0}(x_t)\varphi(x_t)\triangle(x_t) \to_D L_B(1,0)\int \mathscr{H}_{i_0}(x)\varphi(x)\triangle(x)dx$ by Wang and Phillips (2009a), giving $\frac{d_n}{nk}\Pi'_{1n} = O_P(\frac{\delta_n^2 n}{d_n k})$.

Meanwhile, $\frac{d_n}{nk}\Pi''_{1n} = O_P(1)$ and $\frac{d_n}{nk}\Pi'''_{1n} = O_P(\delta_n(n/d_n k)^{1/2})$ using the same arguments as in Theorem 3.2 again. Hence, $\frac{d_n}{nk}\Pi_{1n} = O_P(\frac{\delta_n^2 n}{d_n k}) \to_P \infty$ by Assumption C. This completes the proof of the second part.

Furthermore, it is clear from Lemma 3.1 that when $\widehat{d}_n$ and $\widehat{\sigma}_e$ are used in the place of d_n and σ_e, respectively, the feasible versions of the limits hold. ■

https://doi.org/10.1017/S0266466617000238 Published online by Cambridge University Press