UNIVERSITY *of York*

Discussion Papers in Economics

No. 14/19

Specification Testing in Nonstationary Time Series Models

Jia Chen, Jiti Gao, Degui Li and Zhengyan Lin

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Specification Testing in Nonstationary Time Series Models

Jia Chen[1], Jiti Gao[2], Degui Li[3] and Zhengyan Lin[4]

[1] *Department of Economics and Related Studies, University of York, York, UK*

[2] *Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia*

[3] *Department of Mathematics, University of York, York, UK*

[4] *Department of Mathematics, Zhejiang University, Hangzhou, China*

□ In this paper, we consider a specification testing problem in nonlinear time series models with nonstationary regressors and propose using a nonparametric kernel-based test statistic. The null asymptotics for the proposed nonparametric test statistic have been well developed in the existing literature such as Gao *et al* (2009b) and Wang and Phillips (2012). In this paper, we study the local asymptotics of the test statistic, i.e., the asymptotic properties of the test statistic under a sequence of general nonparametric local alternatives, and show that the asymptotic distribution depends on the asymptotic behaviour of the distance function which is the local deviation from the parametrically specified model in the null hypothesis. In order to implement the proposed test in practice, we introduce a bootstrap procedure to approximate the critical values of the test statistic and establish a novel result of Edgeworth expansion which is used to justify the use of such an approximation. Based on the approximate critical values, we develop a bandwidth selection method, which chooses the optimal bandwidth that maximises the local power of the test while its size is controlled at a given significance level. The local power is defined as the power of the proposed test for a given sequence of local alternatives. Such a bandwidth selection is made feasible by an approximate expression for the local power of the test as a function of the bandwidth. A Monte-Carlo simulation study is provided to illustrate the finite sample performance of the proposed test.

Keywords: Asymptotic distribution; Edgeworth expansion; local power function; nonlinear time series; quadratic form; size function; specification testing; unit root.

JEL Classifications: C12; C13; C14.

1. Introduction

In the past two decades or so, there exists a rich literature on specification testing of a parametric model versus a nonparametric/semiparametric alternative involving stationary time series. Many testing procedures have been proposed based on nonparametric kernel methods. The existing literature includes Fan and Li (1996), Li and Wang (1998), Li (1999), Fan and Linton (2003), Juhl and Xiao (2005a), and Chen and Gao (2007). It is shown that the leading term of each of these existing nonparametric kernel test statistics is of a quadratic form (c.f., Chapter 3 of Gao 2007). With the help of an Edgeworth expansion for quadratic forms, Gao and Gijbels (2008) developed an asymptotic theory to support a local power function-based bandwidth selection method for optimal testing purposes. Some general asymptotic distributions for nonparametric kernel test statistics have also been discussed in the books by Fan and Yao (2003), Gao (2007), and Li and Racine (2007).

As pointed out in the literature, stationarity might be a quite restrictive assumption on time series data in real-world problems. When tackling economic issues from a time perspective, we often deal with nonstationary components. For example, substantial empirical evidence suggests that many exchange rates, prices, aggregate consumption and other macroeconomic variables are nonstationary. Hence, practitioners might feel more comfortable avoiding restrictions like stationarity for time series data. In this respect, some papers already discussed parametric and nonparametric estimation in nonlinear time series models with possible nonstationarity. Such studies include Phillips and Park (1998), Park and Phillips (1999, 2001), Karlsen and Tjøstheim (2001), Karlsen, Myklebust and Tjøstheim (2007), Cai, Li and Park (2009), Lin, Li and Chen (2009), Wang and Phillips (2000a, 2009b), Chen, Li and Zhang (2010), Wang and Phillips (2011), Chen, Gao and Li (2012), and Gao and Phillips (2013).

Meanwhile, there are also some papers on model specification testing in nonstationary time series case. Juhl and Xiao (2005b) focused on testing for cointegration using a partially linear model. Marmer (2008) developed a functional form test in dealing with nonlinearity, nonstationarity and spurious forecasts. Kasparis (2008), and Hong and Phillips (2010) considered model specification testing in cointegration models. Recently, Gao *et al* (2009a) established an asymptotically consistent test for a nonparametric unit–root specification problem in a nonlinear time series autoregression. In a paper closely related to the current paper, Gao *et al* (2009b) proposed a nonparametric kernel test statistic for detecting whether the regression function is of a known parametric form indexed by an unknown parameter vector and then established an asymptotic null distribution for the proposed test statistic. More recently, Wang and Phillips (2012) studied this test statistic and established some asymptotic properties under a nonlinear null hypothesis and a sequence of local alternatives that are different from those in Gao *et al* (2009b). In this paper, we study the same test statistic as in Gao *et al* (2009b) and establish its asymptotic property under a sequence of local alternative hypotheses which will be specified later. Furthermore, since most of the existing literature did not thoroughly study bandwidth selection issue in the nonparametric kernel-based tests with nonstationarity, this paper aims to fill in this gap by developing a bandwidth selection method that chooses the optimal bandwidth as the one that maximises the local power of the test while the size is controlled at a given significance level.

This paper is concerned with a nonlinear time series model of the form

$$Y_t = g(V_t) + e_t, \quad t = 1, \cdots, n, \tag{1.1}$$

where $g(\cdot)$ is a smooth regression function, $\{V_t\}$ is an integrated process of order one

$$V_t = V_{t-1} + v_t, \quad t \geq 1, \quad V_0 = 0, \tag{1.2}$$

in which $v_t = \sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i}$ is a stationary linear process with $\{\varepsilon_t\}$ being a white noise sequence, and $\{e_t\}$ is a sequence of stationary martingale differences. Equation (1.1) provides a very general regression framework for studying the relationship between Y_t and V_t. While nonparametric smoothing techniques can be used to estimate the regression function $g(\cdot)$ when no information about its functional form is known, in a lot of economic studies researchers often specify, based on related economic theories or prior information, a particular nonlinear parametric form for $g(\cdot)$, i.e. $g(v) = g(v, \theta)$, where $\theta \in \Theta$ is a vector of unknown parameters and Θ is a suitable parameter space. The exact form of $g(\cdot)$ will depend on the question under consideration. Although these model specifications have the support of economic theories, they may not fit the real world very well and hence, small local deviations from them are possible. In this paper, we are interested in testing

$$\mathrm{H}_0 : g(v) = g(v, \theta) \tag{1.3}$$

against a sequence of local alternative hypotheses

$$\mathrm{H}_1^L : g(v) = g(v, \theta) + \Delta_n(v), \tag{1.4}$$

where $g(v, \theta)$ is a given real function indexed by the unknown parameter vector θ belonging to the parameter space Θ, and $\Delta_n(\cdot)$ is an unknown function satisfying $\Delta(v) \to 0$ for any v as $n \to \infty$ and represents the small deviation of the regression function from the specified functional form in the null H_0. Since $\Delta_n(v)$ diminishes as n increases, it is hence described as a local deviation. In what follows we will call $\Delta_n(v)$ the *distance function* as it measures the magnitude of departure of the model from the null specification. We assume that there is a unique parameter vector $\theta_0 \in \Theta$ such that $\mathsf{E}[Y_t|\mathcal{F}_{t-1}] = g(V_t, \theta_0)$ almost surely (a.s.) under the null H_0.

The reason for choosing a nonparametric distance function $\Delta_n(v)$ is that when there exists evidence that suggests we reject the null hypothesis, the level of the departure from the null is often unknown. Instead of establishing an asymptotic distribution of the proposed test statistic under the null hypothesis H_0, as has been done in Gao *et al* (2009b), the current paper focuses on studying the asymptotic properties of the test statistic under a sequence of general local alternatives of the form (1.4). As shown in Theorems 2.1 and 2.2 in Section 2 below, we find that the asymptotic distribution of the proposed test statistic under a given sequence of local alternatives depends on the asymptotic behaviour of the unknown distance function $\Delta_n(\cdot)$. Furthermore, our results show that for different distance functions, the nonparametric kernel test in the nonstationary time series case can detect alternatives with either smaller or larger magnitude of deviations from the null than its counterpart in stationary time series case. This is mainly because the rate of convergence of an estimator in the nonstationary case depends heavily on the functional form of the function being estimated. Similar observations have been made in Park and Phillips (2001) where the authors dealt with parameter estimation in nonlinear parametric regression models with integrated regressors.

For the choice of bandwidth in the nonparametric test statistic, we propose a computer-intensive bootstrap simulation procedure. The main idea of this bandwidth selection method is to maximise the local power of the proposed test while the size is controlled at a given significance level, where the local power is defined as the power of the test under a given sequence of local alternative hypotheses. We also establish the asymptotic behaviour of the bootstrap scheme under some mild conditions and obtain an Edgeworth expansion for a nonstationary kernel-weighted quadratic form. Furthermore, we obtain the asymptotic approximation for the local power function of the proposed test in Section 3, which enables us to construct a feasible procedure for selecting an appropriate bandwidth for the kernel-based test statistic.

The rest of the paper is organized as follows. Section 2 proposes using a nonparametric kernel test statistic for the testing problem (1.3)–(1.4), introduces some basic definitions for regular functions, gives some technical assumptions and then establishes the asymptotic properties of the test statistic under the local alternatives. Section 3 proposes using a simulated bootstrap procedure for bandwidth selection and establishes an Edgeworth expansion for a nonstationary kernel weighted quadratic form. Section 4 illustrates the finite sample performance of the proposed test through a simulated example. Section 5 concludes this paper with some comments. Appendix A gives some useful lemmas and then a sketch of the proofs of the main results. The full proofs of the main results and the lemmas are given in Appendix B and Appendix C, respectively, which are relegated to the supplementary appendix Chen *et al* (2014). Throughout the paper, we use $\to_D$ to denote convergence in distribution, $\to_P$ to denote convergence in probability, $a_n \sim b_n$ to denote $\frac{a_n}{b_n} \to 1$ as $n \to \infty$, and $a_n \overset{P}{\sim} b_n$ to denote $\frac{a_n}{b_n} \to_P 1$ as $n \to \infty$.

2. The test statistic and its asymptotic properties under local alternatives

In this section, we present a nonparametric kernel test statistic, introduce some basic definitions of regular functions, give some regularity conditions and then establish asymptotic properties of the test statistic under a sequence of local alternative hypotheses H_1^L.

2.1. The test statistic

It has been shown in both stationary and nonstationary time series cases (c.f., Gao, 2007; Li and Racine, 2007; Gao *et al*, 2009b; and Wang and Phillips, 2012) that a nonparametric kernel test statistic of the quadratic form

$$Q_n(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_t K\left(\frac{V_t - V_s}{h}\right) \widehat{e}_s \tag{2.1}$$

works well in both large and finite samples, where $K(\cdot)$ is a kernel function, $h = h_n$ is a bandwidth that tends to 0 as n tends to ∞, and $\widehat{e}_t = Y_t - g(V_t, \widehat{\theta})$, in which $\widehat{\theta}$ is the nonlinear least squares (NLS) estimator of θ_0 that is defined by

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \sum_{s=1}^{n} \left[Y_s - g(V_s, \theta)\right]^2. \tag{2.2}$$

Note that when H_0 holds (i.e., there is no deviation from the specified parametric nonlinear regression function in the null hypothesis H_0), the NLS estimator $\widehat{\theta}$ is consistent with its convergence rate depending on the functional form of $g(\cdot, \cdot)$ (c.f., Park and Phillips, 2001). Lemma A.2 in Appendix A will further show that under a given sequence of local alternatives H_1^L, $\widehat{\theta}$ is still consistent with its convergence rate depending on the functional forms of both $g(\cdot, \cdot)$ and $\Delta_n(\cdot)$, where $\Delta_n(\cdot)$ is the distance function defined in (1.4).

2.2. Basic definitions of regular functions

In order for the paper to be more self contained, we next give some basic definitions for regular functions, some of which are introduced in the papers by Park and Phillips (1999, 2001). A function $T(x)$ on $\mathcal{R} \equiv (-\infty, \infty)$ is said to be regular if (i) it is continuous in a neighborhood of infinity; and (ii) for any compact subset $\mathcal{C}$ of $\mathcal{R}$ and each $\epsilon > 0$, there exist continuous functions $\underline{T}_\epsilon$ and $\overline{T}_\epsilon$ as well as a constant $\delta_\epsilon > 0$ such that $\underline{T}_\epsilon(x) \leq T(y) \leq \overline{T}_\epsilon(x)$ for all $|x - y| < \delta_\epsilon$ on $\mathcal{C}$, and $\int_{\mathcal{C}} (\overline{T}_\epsilon - \underline{T}_\epsilon)(x) dx \to 0$ as $\epsilon \to 0$. It is obvious that any continuous or piece-wise continuous function is regular. Before giving the definitions of I-regular and H-regular functions on Θ, we first introduce two preliminary definitions.

Definition 2.1. *We say that a function $m(x, \theta)$ is regular on Θ if*
(i) for all $\theta \in \Theta$, $m(\cdot, \theta)$ is regular;
(ii) for all $x \in \mathcal{R}$, $m(x, \cdot)$ is equicontinuous in a neighborhood of x.

Definition 2.2. *We say that a function $Z_1(x, \lambda, \theta)$ is of order smaller than $Z_2(\lambda, \theta)$ on Θ if*

$$Z_1(x, \lambda, \theta) = a(\lambda, \theta) A(x, \theta) \quad \text{or} \quad Z_1(x, \lambda, \theta) = b(\lambda, \theta) A(x, \theta) B(\lambda x, \theta)$$

where $a(\lambda, \theta) = o(Z_2(\lambda, \theta))$ and $b(\lambda, \theta) = O(Z_2(\lambda, \theta))$ as $\lambda \to \infty$ for all $\theta \in \Theta$, $\sup_{\theta \in \Theta} A(x, \theta) = O(e^{c|x|})$ as $|x| \to \infty$ for some $c > 0$, and $\sup_{\theta \in \Theta} B(x, \theta)$ is bounded and tends to 0 as $|x| \to \infty$.

We are now ready to introduce two classes of functions: I-regular on Θ and H-regular on Θ, which are used in the present paper to classify parametric nonlinear regression functions.

Definition 2.3. *A function $m(x,\theta)$ is said to be I-regular on Θ if the following two conditions are satisfied:*
(i) for each $\theta \in \Theta$, there exist a neighborhood $\mathcal{N}(\theta)$ of θ and a bounded integrable function $M : \mathcal{R} \to \mathcal{R}$ such that $|m(x,\theta') - m(x,\theta)| \leq \|\theta' - \theta\| M(x)$ for all $\theta' \in \mathcal{N}(\theta)$ where $\|\cdot\|$ is the Euclidean norm;
(ii) for some $C > 0$, $|m(x,\theta) - m(y,\theta)| \leq C|x-y|$ for all $\theta \in \Theta$ on each piece $\mathcal{S}_i$ of their common support $\mathcal{S} = \cup_{i=1}^{p} \mathcal{S}_i \subset \mathcal{R}$.

Definition 2.4. *Let $m(\lambda x, \theta) = \kappa(\lambda) H(x,\theta) + R(x,\lambda,\theta)$, where $\kappa(\cdot)$ is nonsingular. The function $m(x,\theta)$ is said to be H-regular on Θ if the following two conditions are satisfied:*
(i) $H(x,\theta)$ is regular on Θ;
(ii) $R(x,\lambda,\theta)$ is of order smaller than $\kappa(\lambda)$ as $\lambda \to \infty$ for all $\theta \in \Theta$.
We call $\kappa(\cdot)$ the asymptotic order of $m(x,\theta)$ and $H(x,\theta)$ the limit homogeneous function.

Remark 2.1. The two conditions in Definition 2.3 are called the I-regularity conditions (c.f., Section 3.2(a) in Park and Phillips, 2001). The first condition indicates that for all x, $m(x,\cdot)$ is continuous on Θ, and the second condition indicates that the function $m(\cdot,\theta)$ is piecewise Lipschitz continuous on a common support which does not rely on the value of θ. Let $m_*(v) = \sum_{i=1}^{p} m_i(v) I(v \in \mathcal{S}_i)$, where $m_i(\cdot)$ is some continuous and integrable function on the set $\mathcal{S}_i$, $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for $1 \leq i \neq j \leq p$, $\cup_{i=1}^{p} \mathcal{S}_i$ is the support of $m_*(\cdot)$ and $I(\cdot)$ is the indicator function. It is easy to check that the two I-regularity conditions in Definition 2.3 are satisfied when $m(v,\theta) = m_*(v)\theta$ and the parameter space Θ is bounded. We can further show that $m(v, \alpha_0, \alpha_1) = \alpha_0 \exp\{-\alpha_1 v^2\}$ with $\theta = (\alpha_0, \alpha_1)^T \in \Theta$ and $m(v,\theta) = \frac{1}{1+\theta v^2}$ also belong to the class of I-regular functions on their respective Θ. The two conditions in Definition 2.4 are called the H-regularity conditions (c.f., Section 3.2(b) in Park and Phillips, 2001). In particular, the second H-regularity condition indicates that the function $m(\cdot,\theta)$ can be asymptotically approximated by the product of its asymptotic order and a limit homogeneous function. This approximation would facilitate the establishment of the asymptotic theory involving H-regular functions. If $m(v, \alpha_0, \alpha_1, \cdots, \alpha_p) = \alpha_0 x^p + \alpha_1 x^{p-1} + \cdots + \alpha_{p-1} x + \alpha_p$ with $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_p)^T \in \Theta$, it is easy to check that the two H-regularity conditions in Definition 2.4 are satisfied with $\kappa(\lambda) = \lambda^p$ and $H(x,\theta) = \alpha_0 x^p$. Furthermore, letting $m_\diamond(\cdot)$ be the distribution function of a random variable which may not be continuous, we can also show that the two H-regularity conditions in Definition 2.4 are satisfied if $m(v,\theta) = m_\diamond(v)\theta$. More discussions and examples on I-regular and H-regular functions can be found in Park and Phillips (1999, 2001).

Under the local alternatives H_1^L, $\widehat{e}_t = Y_t - g(V_t, \widehat{\theta})$ will contain a deviation from the null specification. Hence, the asymptotic behaviour of the test statistic $Q_n(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_t K\left(\frac{V_t - V_s}{h}\right) \widehat{e}_s$ may depend on that of the deviation $\Delta_n(\cdot)$. We next consider two families of functions for the distance $\Delta_n(v)$, i.e., δ_n-integrable functions and δ_n-asymptotically homogeneous functions with $\delta_n \to 0$ as $n \to \infty$.

Definition 2.5. *The distance function $\Delta_n(x)$ is said to be δ_n-integrable if*

$$\delta_n\Gamma_1(x) \leq \Delta_n(x) \leq \delta_n\Gamma_2(x)$$

for all $x \in \mathcal{R}$, where for each $k = 1,\ 2$, $\Gamma_k(x)$ is square integrable.

Definition 2.6. *The distance function $\Delta_n(x)$ is said to be δ_n-asymptotically homogeneous if*

$$\delta_n\Lambda_1(x) \leq \Delta_n(x) \leq \delta_n\Lambda_2(x)$$

for all $x \in \mathcal{R}$, where both $\Lambda_1(x)$ and $\Lambda_2(x)$ are asymptotically homogeneous. That is,

$$\Lambda_k(\lambda x) = v_k(\lambda)H_k(x) + R_k(x, \lambda), \quad k = 1, 2,$$

and $v(\lambda) = v_1(\lambda) + v_2(\lambda)$ is defined as the asymptotically homogeneous order of $\Delta_n(x)$, $v_1(\lambda) \sim v_2(\lambda)$, $H_k(x)$, $k = 1, 2$, are locally integrable and $R_k(\cdot, \cdot)$, $k = 1, 2$, satisfy one of the following conditions:

(i) $|R_k(x, \lambda)| \leq a_k(\lambda)P_k(x)$, where $\limsup_{\lambda\to\infty} \frac{a_k(\lambda)}{v_k(\lambda)} = 0$ and $P_k(x)$ is locally integrable; or

(ii) $|R_k(x, \lambda)| \leq b_k(\lambda)Q_k(\lambda x)$, where $\limsup_{\lambda\to\infty} \frac{b_k(\lambda)}{v_k(\lambda)} < \infty$ and $Q_k(x)$ is locally integrable and vanishes at infinity, i.e., $Q_k(x) \to 0$ as $|x| \to \infty$.

Definitions 2.5 and 2.6 above can be regarded as extensions of Definitions 4.1 and 4.2 in Park and Phillips (1999) to our hypothesis testing setting.

2.3. Assumptions

We next give some regularity conditions, which will be used for establishing our asymptotic theory.

Assumption 1 *(i) Let V_t be generated as $V_t = V_{t-1} + v_t$ with $V_0 = 0$, and $v_t = \sum_{i=0}^{\infty} \phi_i\varepsilon_{t-i}$, where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random errors with $\mathsf{E}[\varepsilon_1] = 0$, $\mathsf{E}[\varepsilon_1^8] < \infty$ and $\sigma_0^2 \equiv \mathsf{E}[\varepsilon_1^2] > 0$, $\{\phi_j\}$ is a sequence of real numbers satisfying $\phi \equiv \sum_{j=0}^{\infty} \phi_j \neq 0$ and $\sum_{j=0}^{\infty} j^{1+\lambda_0}\,|\phi_j| < \infty$ for some $\lambda_0 > 0$. The characteristic function $\varphi(\cdot)$ of ε_1 satisfies $|r|\varphi(r) \to 0$ as $r \to \infty$.*

(ii) Let $\{(e_t, \mathcal{F}_t) : t \geq 1\}$ be a stationary sequence of martingale differences satisfying $\mathsf{E}[e_t^2|\mathcal{F}_{t-1}] = \sigma_e^2 > 0$ and $\sup_{t\geq 1} \mathsf{E}\left[e_t^4|\mathcal{F}_{t-1}\right] < \infty$ almost surely (a.s.), where

$$\mathcal{F}_t = \sigma\left(e_1, \cdots, e_t; \varepsilon_{-\infty}, \cdots, \varepsilon_{t+1}\right)$$

is a σ-field generated by $\{e_i, \varepsilon_j : 1 \leq i \leq t, -\infty < j \leq t+1\}$. Furthermore, ε_t is independent of e_s for all $t \geq s + 1$.

(iii) Letting $E_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t$ and $V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t$, there exists a vector Brownian motion (E, V) such that $(E_n(r), V_n(r)) \Rightarrow (E(r), V(r))$ on $D[0, 1]^2$ as $n \to \infty$, where "$\Rightarrow$" denotes the weak convergence.

Assumption 2 *(i) The kernel $K(\cdot)$ is a continuous, positive and symmetric probability density function with a compact support.*

(ii) For some $0 < \lambda_1 < \frac{1}{4}$, $n^{\lambda_1} h \to 0$ and $n^{\frac{1}{2}-\lambda_1} h \to \infty$ as $n \to \infty$.

Assumption 3 *(i) The nonlinear regression function $g(v, \theta)$ is I-regular on Θ and*

$$\int_{-\infty}^{\infty} \left[g(v, \theta') - g(v, \theta)\right]^2 dv > 0 \quad \text{for all } \theta' \neq \theta.$$

(ii) Both $\dot{g}_\theta(v, \theta)$ and $\ddot{g}_\theta(v, \theta)$ are I-regular on Θ, where $\dot{g}_\theta(v, \theta) = \frac{\partial g(v,\theta)}{\partial \theta}$ and $\ddot{g}_\theta(v, \theta) = \frac{\partial^2 g(v,\theta)}{\partial \theta \partial \theta^T}$. Furthermore, $\int \dot{g}_\theta(v, \theta)\left[\dot{g}_\theta(v, \theta)\right]^T dv$ is positive definite.

Assumption 3′ *(i) The nonlinear regression function $g(v, \theta)$ is H-regular on Θ with limit homogeneous function $h(v, \theta)$ and asymptotic order $\kappa(\cdot)$, which is independent of θ. Furthermore, $\kappa(\lambda)$ is bounded away from zero as $\lambda \to \infty$, and $\int_{|v| \le \delta_0} \left[h(v, \theta) - h(v, \theta')\right]^2 dv > 0$ for all $\theta' \neq \theta$ and all $\delta_0 > 0$.*

(ii) Both $\dot{g}_\theta(v, \theta)$ and $\ddot{g}_\theta(v, \theta)$ are H-regular on Θ with asymptotic orders $\dot{\kappa}(\cdot)$ and $\ddot{\kappa}(\cdot)$, respectively. Furthermore, $\dot{\kappa}(\lambda)$ is bounded away from zero as $\lambda \to \infty$, $\|(\dot{\kappa} \otimes \dot{\kappa})^{-1} \kappa \ddot{\kappa}\| < \infty$, and $\int_{|v| \le \delta_1} \dot{h}(v, \theta)\left[\dot{h}(v, \theta)\right]^T dv$ is a positive definite matrix for all $\delta_1 > 0$, where $\dot{h}(v, \theta)$ is the limit homogeneous function of $\dot{g}_\theta(v, \theta)$ (both $\dot{h}(v, \theta)$ and $\dot{g}_\theta(v, \theta)$ are of the same dimension as θ), and $\otimes$ is the Kronecker product.

Remark 2.2. The condition $V_0 = 0$ on the initial state of the unit root process $\{V_t\}$ in Assumption 1(i) is imposed to simplify the proofs of the main results and it can be relaxed to $V_0 = O_P(1)$. The summability assumption $\sum_{j=0}^{\infty} j^{1+\lambda_0} |\phi_j| < \infty$ in Assumption 1(i) is a commonly used condition for stationary linear processes and indicates that the process $\{v_t\}$ is short-range dependent. Assumption 1(ii) imposes a stationary martingale differences structure on the model errors, which is much weaker than the mutual independence assumption between $\{e_t\}$ and $\{v_t\}$ (c.f., Gao *et al*, 2009b). Wang and Phillips (2012) used a set of distributional assumptions slightly weaker than Assumption 1, but required stronger restrictions on the kernel function and the bandwidth than Assumption 2 (such as requiring $\lim_{n\to\infty} nh^4 \log^2(n) = 0$). Assumption 1(iii) assumes a joint weak convergence result for the two partial sum processes $E_n(r)$ and $V_n(r)$. Such an assumption is not uncommon in the literature on nonstationary regression modelling (c.f., Park and Phillips, 2001; Wang and Phillips, 2012) and can be satisfied under certain restrictions on the stationary processes $\{e_t\}$ and $\{v_t\}$ (c.f., Wu and Min, 2005). Note that the mixing condition is not required in the paper. Assumption 2(i) and (ii) in the present paper are mild conditions on the kernel function $K(\cdot)$ and the bandwidth h.

Assumptions 3 and 3′ are two sets of regularity conditions on the smoothness and functional form of $g(v, \theta)$ so that under either set of conditions, $\widehat{\theta}$ is a consistent estimator of θ_0. Such conditions were initially introduced by Park and Phillips (2001) when they established the asymptotic theory for an NLS estimator in a parametric nonlinear regression model with integrated time series. We have given in Remark 2.1 some examples of I-regular and H-regular functions, which satisfy Assumptions 3 and $3'$, respectively.

2.4. Asymptotic distributions under local alternatives

Our analysis of the asymptotic properties of the test statistic $Q_n(h)$ will be based on the following decomposition under the local alternative hypotheses H_1^L

$$\begin{aligned} Q_n(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} e_t K_{s,t} e_s + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t} \widetilde{g}_s + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} \Delta_n(V_s) \\ &\quad +2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} \widetilde{g}_s + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t} e_s + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} e_s \\ &\equiv \sum_{j=1}^{6} \overline{Q}_{n,j}(h), \end{aligned} \tag{2.3}$$

where $K_{s,t} = K\left(\frac{V_s - V_t}{h}\right)$, $\widetilde{g}_t = g(V_t, \theta_0) - g(V_t, \widehat{\theta})$.

Define a standardized version of $Q_n(h)$ as

$$\widehat{Q}_n(h) = \frac{Q_n(h)}{\overline{\sigma}_n} \quad \text{with} \quad \overline{\sigma}_n^2 = 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widehat{e}_t^2 K^2\left(\frac{V_t - V_s}{h}\right)\widehat{e}_s^2. \tag{2.4}$$

In Theorems 2.1 and 2.2 below, we give some asymptotic properties for $\widehat{Q}_n(h)$ under the sequence of local alternatives H_1^L. We first consider the case where both $\Delta_n(v)$ and its derivative $\dot{\Delta}_n(v) \equiv \frac{d\Delta_n(v)}{dv}$ are δ_n-integrable.

THEOREM 2.1. *Suppose that Assumptions 1–3 in Section 2.3 hold, and both $\Delta_n(v)$ and $\dot{\Delta}_n(v)$ are δ_n-integrable.*
(i) If $\delta_n = o\left(n^{-1/8}h^{-1/4}\right)$, then under the sequence of local alternative hypotheses H_1^L,

$$\widehat{Q}_n(h) = \frac{Q_n(h)}{\overline{\sigma}_n} \to_D \mathsf{N}(0,1) \qquad \text{as } n \to \infty. \tag{2.5}$$

(ii) If $n^{1/8}h^{1/4}\delta_n \to \infty$ as $n \to \infty$, then under the sequence of local alternative hypotheses H_1^L, $\widehat{Q}_n(h) \to_P \infty$ as $n \to \infty$.

For the case where both $\Delta_n(v)$ and $\dot{\Delta}_n(v)$ are δ_n-asymptotically homogeneous, we have the following asymptotic result.

THEOREM 2.2. *Suppose that Assumptions 1, 2, and $3'$ in Section 2.3 are satisfied, and both $\Delta_n(v)$ and $\dot{\Delta}_n(v)$ are δ_n-asymptotically homogeneous with asymptotically homogeneous orders $v(\cdot)$ and $\dot{v}(\cdot)$, respectively, and $v(n) = O(\dot{\kappa}(n))$ as $n \to \infty$, where $\dot{\kappa}(\cdot)$ is the asymptotic order of $\dot{g}_\theta(v,\theta)$.*
(i) If $\delta_n = o\left(n^{-3/8}v^{-1}(\sqrt{n})h^{-1/4}\right)$, then the asymptotic normal distribution (2.5) holds under the sequence of local alternatives H_1^L.

(ii) If $n^{3/8}v(\sqrt{n})h^{1/4}\delta_n \to \infty$ as $n \to \infty$, then under the sequence of local alternatives H_1^L, $\widehat{Q}_n(h) \to_P \infty$ as $n \to \infty$.

REMARK 2.3. When $\delta_n = O\left(n^{-1/8}h^{-1/4}\right)$, which includes the boundary case of $\delta_n = Cn^{-1/8}h^{-1/4}$ for some $0 < C < \infty$ in Theorem 2.1, or $\delta_n = O\left(n^{-3/8}v^{-1}(\sqrt{n})h^{-1/4}\right)$ which includes the boundary case of $\delta_n = Cn^{-3/8}v^{-1}(\sqrt{n})h^{-1/4}$ in Theorem 2.2, we can show that under the sequence of local alternative hypotheses H_1^L,

$$\widehat{Q}_n(h) - B_n = \frac{Q_n(h)}{\overline{\sigma}_n} - B_n \to_D \mathsf{N}(0,1) \quad \text{as } n \to \infty, \tag{2.6}$$

where $B_n = \frac{1}{\overline{\sigma}_n}\left[\overline{Q}_{n,2}(h) + \overline{Q}_{n,3}(h) + \overline{Q}_{n,4}(h)\right] = O_P(1)$ with $\overline{Q}_{n,i}(h)$, $i = 2, 3, 4$, defined in (2.3).

Theorems 2.1 and 2.2 show that whether the proposed test under H_1^L is asymptotically powerful depends on the rate at which δ_n decays. For example, when the decaying rate of δ_n is faster than that of $n^{-1/8}h^{-1/4} \to 0$, Theorem 2.1(i) shows that $\widehat{Q}_n(h)$ converges in distribution to the standard normal distribution. This implies that the proposed test is not asymptotically powerful. However, when the decaying rate of δ_n is slower than that of $n^{-1/8}h^{-1/4} \to 0$, Theorem 2.1(ii) shows that $\widehat{Q}_n(h) \to \infty$ in probability, which indicates that the test is asymptotically powerful. Similar conclusions can be drawn from Theorem 2.2. Furthermore, as shown in Theorems 2.1 and 2.2, the asymptotic property of the test statistic under the local alternative hypotheses depends on the asymptotic behaviour of the unknown distance function $\Delta_n(\cdot)$. For example, if the distance function is δ_n-integrable, the proposed nonparametric kernel test can detect departures from the specified function $g(v, \theta)$ when $\delta_n n^{1/8}h^{1/4} \to \infty$. In contrast, if the distance function is δ_n-asymptotically homogeneous, the test can detect departures when $\delta_n n^{3/8}v(\sqrt{n})h^{1/4} \to \infty$. It is worth pointing out that Theorem 2.2 also holds in the case where $g(v, \theta)$ is H-regular and $\Delta_n(v)$ is δ_n-integrable (c.f., the earlier version of the paper Chen *et al*, 2011).

Existing literature such as Wang and Phillips (2012) has only established results similar to those in Theorem 2.1(ii) and Theorem 2.2(ii) (c.f., Theorem 3.2 in Wang and Phillips, 2012) when the nonparametric test is asymptotically powerful. In contrast, this paper uses different conditions on the distance function $\Delta_n(\cdot)$ and provides more accurate asymptotic results by also studying the case where the test may not be asymptotically powerful. In the next section, we will introduce a bootstrap procedure for approximating the critical values of the test and then propose a data-driven bandwidth selection method for optimal testing purposes.

3. Edgeworth expansion and optimal bandwidth selection

This section proposes using a bootstrap simulation procedure to approximate the critical values of the test statistic. The main reason for using the bootstrap method is to avoid the possible size distortion which often occurs when the critical values of a test are chosen from asymptotic distribution theory. Some existing literature such as Li and Wang (1998) has shown that the bootstrap method can provide a more accurate approximation of the finite sample distribution of the test statistic than the asymptotic distribution theory. Proposition 3.1 below justifies the use of the bootstrap approximation in our testing setting. Based on the approximate critical values, the size and local power of the test are studied, and the bandwidth for optimal testing purposes is selected so that the local power of the test is maximised while its size is controlled at a given significance level α. For a given significance level α, let l_α be the corresponding critical value of the test, which is the $(1-\alpha)$-quantile of the exact finite sample distribution of $\widehat{Q}_n(h)$ under the null hypothesis. The bootstrap procedure is proposed as follows.

STEP 1: *Generate the bootstrap error terms* $\{e_t^*\}$ *by* $e_t^* = \widehat{\sigma}_e \eta_t^*$, *where*

$$\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}\left[Y_t - g(V_t, \widehat{\theta})\right]^2, \tag{3.1}$$

in which $\{\eta_t^*\}$ *is a sequence of i.i.d. random variables drawn from a pre-specified continuous distribution with* $\mathsf{E}[\eta_1^*] = 0$, $\mathsf{E}[\eta_1^{*2}] = 1$ *and* $\mathsf{E}[|\eta_1^*|^6] < \infty$.

STEP 2: *Obtain* $Y_t^* = g(V_t, \widehat{\theta}) + e_t^*$. *The resulting sample* $\{(Y_t^*, V_t), 1 \le t \le n\}$ *is called a bootstrap sample.*

STEP 3: *Use the sample* $\{(Y_t^*, V_t), 1 \le t \le n\}$ *to re-estimate the parameter vector* θ_0 *and denote the resulting estimate by* $\widehat{\theta}^*$. *Then calculate the test statistic* $\widehat{Q}_n^*(h)$, *which is the corresponding version of* $\widehat{Q}_n(h)$ *by replacing* $\{(Y_t, V_t)\}$ *and* $\widehat{\theta}$ *with* $\{(Y_t^*, V_t)\}$ *and* $\widehat{\theta}^*$, *respectively.*

STEP 4: *Repeat Steps 1–3* M *times and obtain* M *values of* $\widehat{Q}_n^*(h)$. *Denote these* M *values of* $\widehat{Q}_n^*(h)$ *by* $\widehat{Q}_{n,m}^*(h)$, $m = 1, 2, \cdots, M$. *Then, construct the empirical distribution of the bootstrap test statistic* $\widehat{Q}_{n,m}^*(h)$ *as*

$$P^*\left(\widehat{Q}_n^*(h) \le x\right) \equiv \frac{1}{M}\sum_{m=1}^{M} I\left(\widehat{Q}_{n,m}^*(h) \le x\right).$$

For each fixed h, *calculate the bootstrap critical value* l_α^* *as*

$$l_\alpha^* = \inf_x \left\{x : P^*\left(\widehat{Q}_n^*(h) > x\right) < \alpha\right\}$$

and estimate l_α *by* l_α^*.

We next establish more explicit results for the asymptotic distribution and the local power (as a function of the bandwidth h) of the proposed test statistic than those in the previous section. In order to use the Edgeworth expansion to facilitate our analysis, we need to assume that $\{e_t\}$ is a sequence of i.i.d. continuous random variables independent of $\{\varepsilon_t\}$. Proposition 3.1 below is crucial for justifying the use of the above bootstrap precedure for approximating the critical values the test.

PROPOSITION 3.1. *Suppose that Assumptions 1(i), 2 and either 3 or 3′ are satisfied, and* $\{e_t\}$ *is a sequence of i.i.d. continuous random variables with* $\mathsf{E}[e_1] = 0$, $\sigma_e^2 \equiv \mathsf{E}[e_1^2] > 0$, $\mathsf{E}[|e_1|^6] < \infty$, *and* $\{e_t\}$ *is independent of* $\{\varepsilon_t\}$. *Letting* $\widehat{Q}_{n,1}(h) = \overline{Q}_{n,1}(h)/\sigma_{n,1}$ *with* $\overline{Q}_{n,1}(h)$ *defined in (2.3) and* $\sigma_{n,1}^2 = 2\sigma_e^4 \sum_{t=1}^{n} \sum_{s=1,\neq t} K_{s,t}^2$, *we have*

$$\sup_{x\in\mathcal{R}} \left| \mathsf{P}\left(\widehat{Q}_{n,1}(h) \le x \middle| \mathcal{F}_n(V)\right) - \Phi(x) + \rho_n(h)\Phi^{(3)}(x) \right| = O_P(n^{-1/2}), \tag{3.2}$$

where $\mathcal{F}_n(V) = (V_1, \cdots, V_n)$, $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution,* $\Phi^{(3)}(\cdot)$ *is the third derivative of* $\Phi(\cdot)$, and

$$\rho_n(h) = \frac{\sqrt{2}}{3} \frac{\mathsf{Tr}(A_0^3(h))}{\left(\sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} K_{s,t}^2\right)^{3/2}} = O_P\left(n^{-1/4}h^{1/2}\right), \tag{3.3}$$

in which $\mathsf{Tr}(A)$ *represents the trace of a matrix* A, *and* $A_0(h)$ *is defined as*

$$A_0(h) = \begin{pmatrix} 0 & K_{1,2} & \cdots & K_{1,n} \\ K_{2,1} & 0 & \cdots & K_{2,n} \\ \vdots & \vdots & & \vdots \\ K_{n,1} & K_{n,2} & \cdots & 0 \end{pmatrix}. \tag{3.4}$$

In addition, under the null hypothesis H_0,

$$\sup_{x\in\mathcal{R}} \left| \mathsf{P}^*\left(\widehat{Q}_n^*(h) \le x\right) - \mathsf{P}\left(\widehat{Q}_n(h) \le x\right) \right| = o_P(1) \tag{3.5}$$

and as a consequence

$$\mathsf{P}\left(\widehat{Q}_n(h) > l_\alpha^*\right) = \alpha + o(1). \tag{3.6}$$

Based on Proposition 3.1, we estimate the critical value l_α by l_α^* and define the following size and local power of the test statistic $\widehat{Q}_n(h)$ based on l_α^*:

$$\alpha_n^*(h) = \mathsf{P}\left(\widehat{Q}_n(h) > l_\alpha^* | H_0\right) \quad \text{and} \quad \beta_n^*(h) = \mathsf{P}\left(\widehat{Q}_n(h) > l_\alpha^* | H_1^L\right). \tag{3.7}$$

Note that both the size and the local power may depend on h, and hence they have been denoted by $\alpha_n^*(h)$ and $\beta_n^*(h)$, respectively.

For the purpose of optimal testing, we propose choosing an optimal bandwidth h_{test} at the significance level α such that the local power of the test is maximised, i.e.,

$$h_{\text{test}} = \arg\max_{h\in\mathcal{H}_n} \beta_n^*(h), \tag{3.8}$$

where $\mathcal{H}_n = \{h : \alpha - \epsilon_\alpha \le \alpha_n^*(h) \le \alpha + \epsilon_\alpha\}$, in which ϵ_α ($\epsilon_\alpha < \alpha$) is a small positive constant (say, $\epsilon_\alpha = \frac{\alpha}{10}$ or $\epsilon_\alpha = \frac{\alpha}{\sqrt{n}}$ which decays to zero as n tends to infinity). The chosen h_{test} thus maximises the local power of the test for the given sequence of local alternatives while controls the size at the specified significance level α. Such a method is motivated by the existing literature in stationary time series case, such as Gao and Gijbels (2008). In the simulation study in Section 4, we show that this method has good finite sample performance for nonstationary time series.

As the exact functional dependence of $\beta_n^*(h)$ on h is unknown, in order to make the bandwidth selection approach in (3.8) feasible, we establish, in Proposition 3.2 below, an asymptotic expression for the leading term of $\beta_n^*(h)$ when the distance function $\Delta_n(v)$ and its derivative $\dot{\Delta}_n(v)$ are δ_n-integrable or δ_n-asymptotically homogeneous.

Proposition 3.2. *Assume that Assumptions 1(i) and 2 are satisfied, and the bootstrap errors e_t^* are generated as in Step 1 of the bootstrap procedure detailed above.*
(i) *Suppose that both $\Delta_n(v)$ and $\dot{\Delta}_n(v)$ are δ_n-integrable and Assumption 3 holds, and $n^{1/8}h^{1/4}\delta_n \to \infty$ as $n \to \infty$. Then for given $\mathcal{F}_n(V) = (V_1, \cdots, V_n)$, we have, under H_1^L,*

$$\beta_n^*(h) \overset{P}{\sim} 1 - \Phi(l_\alpha^* - \varpi_n) - \rho_n(h)[1 - (l_\alpha^* - \varpi_n)^2]\phi(l_\alpha^* - \varpi_n), \tag{3.9}$$

where

$$\varpi_n \equiv \frac{\overline{Q}_{n,3}(h)}{\sigma_{n,1}}(1 + o_P(1)),$$

$\overline{Q}_{n,3}(h)$ is defined in (2.3), $\sigma_{n,1}^2$ and $\rho_n(h)$ are defined in Proposition 3.1, and $\phi(\cdot)$ is the standard normal probability density function.
(ii) Suppose that both $\Delta_n(v)$ and $\dot{\Delta}_n(v)$ are δ_n-asymptotically homogeneous and Assumption 3′ holds, and $n^{3/8}v(\sqrt{n})h^{1/4}\delta_n \to \infty$ as $n \to \infty$, then given $\mathcal{F}_n(V) = (V_1, \cdots, V_n)$, equation (3.9) holds under H_1^L.

As far as we know, Propositions 3.1 and 3.2 are new to the literature, and complement the results obtained in Gao *et al* (2009b) and Wang and Phillips (2012) on nonparametric specification testing with nonstationarity. With the help of the two propositions above, we next give a feasible procedure for estimating the local power function and then obtaining the optimal bandwidth for the nonparametric kernel test.

- Construct the local linear estimate of the distance function $\Delta_n(v)$:

$$\widehat{\Delta}_n(v) = \sum_{t=1}^n \widetilde{w}_{nt}(v)\big[Y_t - g(V_t, \widehat{\theta})\big], \tag{3.10}$$

  where the weights are defined by

$$\widetilde{w}_{nt}(v) = \widetilde{L}_n\Big(\frac{V_t - v}{b}\Big)/\Big[\sum_{s=1}^n \widetilde{L}_n\Big(\frac{V_s - v}{b}\Big)\Big], \quad \widetilde{L}_n\Big(\frac{V_t - v}{b}\Big) = L\Big(\frac{V_t - v}{b}\Big)\big[S_{n2}(v) - \Big(\frac{V_t - v}{b}\Big)S_{n1}(v)\big],$$

  in which $L(\cdot)$ is a kernel function, b is a bandwidth, $S_{nj}(v) = \frac{1}{\sqrt{n}b}\sum_{t=1}^n \big(\frac{V_t - v}{b}\big)^j L\big(\frac{V_t - v}{b}\big)$ for $j = 0, 1, 2$, and $\widehat{\theta}$ is the NLS estimate of θ_0 defined in (2.2).

- Estimate ϖ_n in Proposition 3.2 by

$$\widehat{\varpi}_n = \frac{1}{\widehat{\sigma}_{n,1}} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{\Delta}_n(V_t) K_{s,t} \widehat{\Delta}_n(V_s), \quad \widehat{\sigma}_{n,1}^2 = 2\widehat{\sigma}_e^4 \sum_{t=1}^{n} \sum_{s=1,\neq t} K_{s,t}^2, \tag{3.11}$$

  where $\widehat{\sigma}_e^2$ is defined in (3.1).

- Approximate the local power function $\beta_n^*(h)$ by

$$\widehat{\beta}_n^*(h) \quad = \quad 1 - \Phi(l_\alpha^* - \widehat{\varpi}_n) - \rho_n(h)[1 - (l_\alpha^* - \widehat{\varpi}_n)^2]\phi(l_\alpha^* - \widehat{\varpi}_n), \tag{3.12}$$

  which is motivated by (3.9) in Proposition 3.2. Here, l_α^* is the bootstrap critical value at the significance level α and $\rho_n(h)$ is defined in (3.3) of Proposition 3.1.

- Choose the optimal bandwidth using the rule in (3.8) with $\beta_n^*(h)$ replaced by $\widehat{\beta}_n^*(h)$.

In practice, one usually does not know, a priori, whether the null H_0 or the local alternative H_1^L is true. When the local alternative H_1^L is true, the above procedure chooses a bandwidth that maximises the local power; and when the null H_0 is true, by Proposition 3.1, the right-hand side of (3.12) is close to α (since $\widehat{\varpi}_n$ is close to 0), and the above procedure chooses a bandwidth that ensures the size of the test is controlled within $[\alpha - \epsilon_\alpha, \alpha + \epsilon_\alpha]$ for a chosen small constant ϵ_α. A simple Monte-Carlo simulation study in the next section shows that the above bandwidth selection method works reasonably well in finite samples.

4. Monte-Carlo simulation

This section provides a simulated example to illustrate the proposed nonparametric specification test and the bandwidth selection procedure.

Example 4.1. Consider the following time series model

$$Y_t = g(V_t) + e_t \quad \text{with} \quad V_t = V_{t-1} + v_t, \quad t = 1, 2, \cdots, n, \tag{4.1}$$

where both $\{e_t\}$ and $\{v_t\}$ are sequences of i.i.d. standard normal random variables, $\{v_t\}$ is independent of $\{e_t\}$ and $V_0 = 0$.

We then consider the following testing problem:

$$H_0 : \; g(v) = g(v, \theta) \quad \leftrightarrow \quad H_1^L : \; g(v) = g(v, \theta) + \Delta_n(v), \tag{4.2}$$

where $g(v, \theta) = \frac{1}{1+\theta v^2}$ with the true value of θ being $\theta_0 = 1$, and the distance function $\Delta_n(\cdot)$ in the local alternatives H_1^L is chosen as one of the following:

$$\text{Case 1}: \; \Delta_n(v) = \frac{\delta_{1n} v}{1 + v^2} \quad \text{with} \quad \delta_{1n} = 0.6 n^{-1/4} \log(n);$$
$$\text{Case 2}: \; \Delta_n(v) = e^{-v^2}\left(1 - e^{-\delta_{2n} v^2}\right) \quad \text{with} \quad \delta_{2n} = n^{-1/32} \log(n).$$

We use the NLS method to estimate the unknown parameter θ_0 and choose $K(x) = 0.5I\big(-1 \leq x \leq 1\big)$ as the kernel function throughout this section. We will use h_{test} to denote the bandwidth selected by the procedure detailed in Section 3 and h_{cv} to denote the bandwidth selected by the leave-one-out cross-validation (CV) method, which is commonly used in nonparametric and semi-parametric kernel-based estimation (c.f., Fan and Yao, 2003). We also compare the performance of the test using these two bandwidths. For easy distinction, we let $h_{0\text{test}}$ and $h_{0\text{cv}}$ represent h_{test} and h_{cv} under H_0, $h_{1\text{test}}$ and $h_{1\text{cv}}$ for the two bandwidths under H_1^L in Case 1, and $h_{2\text{test}}$ and $h_{2\text{cv}}$ for the two bandwidths under H_1^L in Case 2. Moreover, for $i = 0, 1, 2$, let $f_{i\text{test}}$ denote the frequency that $\widehat{Q}_n(h_{i\text{test}}) > l_\alpha^*(h_{i\text{test}})$ occurs over the simulation replications and $f_{i\text{cv}}$ the frequency that $\widehat{Q}_n(h_{i\text{cv}}) > l_\alpha^*(h_{i\text{cv}})$ occurs. The simulation is carried out with $R = 1000$ replications, and in each replication $M = 250$ bootstrap samples of $\{e_t^*\}$ are drawn from the standard normal distribution $\mathsf{N}(0, 1)$. Sample sizes are chosen as n = 200, 500, and 800. The simulation results are summarised in Table 4.1.

Table 4.1 near here

Table 4.1 shows that the test associated with h_{test} avoids any size distortion, when compared with the test based on h_{cv}. This is basically because h_{test} is chosen to ensure that the size of the test is controlled at the given significance level. It is also shown in Table 4.1 that while the local power values of the test based on both h_{cv} and h_{test} are satisfactory, the test based on h_{test} is more powerful than that based on h_{cv}. Note also that $f_{2\text{test}}$ and $f_{2\text{cv}}$ are smaller than $f_{1\text{test}}$ and $f_{1\text{cv}}$, respectively, since the distance function $\Delta_n\,(V_n)$ in Case 2 is smaller than that in Case 1 (in the sense that $\Delta_n(v)$ in Case 2 decays faster than that in Case 1). In addition, the local power of the proposed test in each case increases as the sample size increases. Since very similar results are obtained for several other simulation examples, such simulation results are not reported here, but the reader can find them in an early version of the paper (Chen *et al*, 2011).

5. Conclusions and discussions

The main contributions of this paper can be summarized as follows. We have proposed using a nonparametric kernel test for testing whether the regression function in a model with a nonstationary regressor is of a known parametric form. We have established the asymptotic properties of the test statistic under a sequence of local alternative hypotheses and shown that the asymptotic distribution depends on the functional form of the distance function. The asymptotic theory developed in this paper differs from existing work on nonparametric specification testing for stationary time series (c.f., Gao, 2007). In order to implement the proposed test in practice, we have developed a bootstrap simulation procedure for approximating the critical values of the test statistic and then for selecting a bandwidth for optimal testing purposes. We have also established some higher-order asymptotic properties for a nonstationary kernel-weighted quadratic form and the bootstrap version of the nonparametric kernel test statistic. Such results complement existing results on nonparametric specification testing with nonstationarity (c.f., Gao *et al*, 2009b; Wang and Phillips, 2012). The proposed methodologies have been illustrated using a simulated example which shows that the proposed testing method works well in finite samples.

When the regressor is multivariate, certain dimension-reduction technique needs to be employed to avoid the curse of dimensionality problem in nonparametric estimation and specification testing. For example, Chen, Gao and Li (2012), and Gao and Phillips (2013) considered the semiparametric estimation of partially linear models with nonstationary multivariate regressors; Cai, Li and Park (2009) studied the nonparametric estimation of functional-coefficient models with nonstationarity. It is possible to extend the testing procedure developed in this paper to the partially linear or functional-coefficient models and investigate the asymptotic property and finite sample behaviour of the test in these settings. However, we will leave this in our future study.

6. Acknowledgments

The authors are grateful to the Co-Editor Professor Michael Jansson and two anonymous referees for their valuable comments and suggestions, which have helped substantially improve the former version of the paper. Thanks also go to Dr Qiying Wang for some constructive discussions and making two relevant papers available to the authors. The second author's research is supported by two Australian Research Council Discovery Grants under Grant Numbers: DP1096374 and DP130104229. The fourth author's research is supported by the National Natural Science Foundation of China.

Appendix A: A sketch of the proofs

In this appendix, we sketch the proofs of the main results in this paper. The full versions of the proofs can be found in the supplementary appendix Chen *et al* (2014). We first introduce several lemmas whose proofs are relegated to Appendix C in the supplementary appendix Chen *et al* (2014). Let $p_t(\cdot)$ be the marginal density function of $V_t = \sum_{s=1}^t v_s$ and $p_{st}(\cdot,\cdot)$ be the joint density function of (V_t, V_s) for $t > s \geq 1$. The following lemma gives the asymptotic orders of $p_t(\cdot)$ and $p_{st}(\cdot,\cdot)$, which will be used, in the supplementary appendix Chen *et al* (2014), to prove the main results and the technical lemmas such as Theorem 2.1(i) and Lemma A.4(ii) below.

Lemma A.1. *Suppose that Assumption 1(i) holds. Then, there exists a sufficiently large* $c_0 > 0$ *such that as* $t \to \infty$ *and* $t - s \to \infty$

$$\sup_{x\in\mathcal{R}} \big|p_t(x)\big| \leq c_0 t^{-1/2}, \qquad \sup_{(x_1,x_2)\in\mathcal{R}^2} \big|p_{st}(x_1,x_2)\big| \leq c_0 (t-s)^{-1/2} s^{-1/2}. \tag{A.1}$$

Park and Phillips (2001) studied the asymptotic properties of the NLS estimator $\widehat{\theta}$ under the null hypothesis H_0. The following lemma establishes the rates of convergence of $\widehat{\theta}$ under the local alternatives H_1^L.

LEMMA A.2. *Suppose that Assumption 1 holds. Let $\widehat{\theta}$ be the NLS estimator defined in (2.2). (i) If Assumption 3 is satisfied and $\Delta_n(v)$ is δ_n-integrable, then under H_1^L, we have*

$$\widehat{\theta} - \theta_0 = O_P\big(\delta_n + n^{-1/4}\big). \tag{A.2}$$

(ii) If Assumption 3′ is satisfied and $\Delta_n(v)$ is δ_n-asymptotically homogeneous with order $v(\cdot)$, then under H_1^L, we have

$$\widehat{\theta} - \theta_0 = O_P\left(\delta_n v(\sqrt{n})[\dot{\kappa}(\sqrt{n})]^{-1} + [\sqrt{n}\dot{\kappa}(\sqrt{n})]^{-1}\right), \tag{A.3}$$

where $\dot{\kappa}(\cdot)$ is defined in Theorem 2.2.

Define

$$\widehat{p}_n(v) = \frac{\sigma_\phi\sqrt{n}}{nh}\sum_{t=1}^{n} K\big(\frac{V_t - v}{h}\big) = \frac{\sigma_\phi}{\sqrt{n}h}\sum_{t=1}^{n} K\big(\frac{V_t - v}{h}\big) \tag{A.4}$$

where $\sigma_\phi^2 = \phi^2\sigma_0^2$ with ϕ and σ_0^2 defined in Assumption 1(i), $K(\cdot)$ is the kernel function and h is the bandwidth, and let $L_B(t,s)$ be the local time process of the standard Brownian motion $B(\cdot)$ defined by

$$L_B(t,s) = \lim_{\delta\to 0}\frac{1}{2\delta}\int_0^t I\big(|B(r) - s| < \delta\big)dr.$$

The following lemma can be obtained as a special case of Theorem 2.1 in Wang and Phillips (2009a).

LEMMA A.3. *Suppose that Assumptions 1(i) and 2 hold. Then, as $n \to \infty$*

$$\widehat{p}_n(v) \to_D L_B(1,0). \tag{A.5}$$

Furthermore, by re-defining $\{\varepsilon_t\}$ on a richer probability space such that

$$\sup_{0\le r\le 1}\big|\frac{1}{\sigma_\phi\sqrt{n}}V_{\lfloor nr\rfloor} - B(r)\big| = o_P(1),$$

where $\lfloor\cdot\rfloor$ denotes the integer part, (A.5) can be strengthened to

$$\widehat{p}_n(v) \to_P L_B(1,0). \tag{A.6}$$

We next give a lemma which is crucial for proving the asymptotic distribution of the proposed test statistic in Sections 2.4.

LEMMA A.4. *(i) Suppose that Assumptions 1 and 2 are satisfied and let $\sigma_{n,1}^2$ be defined as in Proposition 3.1. Then, as $n \to \infty$,*

$$c_1 n^{3/2}h(1 + o_P(1)) \le \sigma_{n,1}^2 \le c_2 n^{3/2}h(1 + o_P(1)), \tag{A.7}$$

where $0 < c_1 < c_2 < \infty$ *are two positive constants.*

(ii) Suppose that Assumptions 1 and 2 are satisfied and let $\overline{Q}_{n,1}(h)$ *be defined as in (2.3). Then, as* $n \to \infty$,

$$\widehat{Q}_{n,1}(h) = \frac{\overline{Q}_{n,1}(h)}{\sigma_{n,1}} \to_D \mathsf{N}(0,1). \tag{A.8}$$

With the above lemmas, we next sketch the proofs of the main results. The full versions of the proofs are given in Appendix B of the supplementary appendix Chen *et al* (2014).

Proof of Theorem 2.1. By (2.3), we have, under H_1^L

$$Q_n(h) = \sum_{i=1}^{6} \overline{Q}_{n,i}(h). \tag{A.9}$$

Following arguments similar to those used in the proof of Theorem 2.1 in Gao *et al* (2009b), we have, by Lemma A.4(ii),

$$\frac{\overline{Q}_{n,1}(h)}{\widetilde{\sigma}_{n,1}} = \frac{\overline{Q}_{n,1}(h)}{\sigma_{n,1}} + o_P(1) \to_D \mathsf{N}(0,1) \quad \text{and} \quad \frac{\overline{\sigma}_n^2 - \widetilde{\sigma}_{n,1}^2}{\widetilde{\sigma}_{n,1}^2} = o_P(1), \tag{A.10}$$

where $\sigma_{n,1}^2 = 2\sigma_e^4 \sum\limits_{t=1}^{n} \sum\limits_{s=1,\neq t} K_{s,t}^2$, $\overline{\sigma}_n^2 = 2\sum_{t=1}^n \sum_{s=1,\neq t}^n \widehat{e}_t^2 K^2\left(\frac{V_t - V_s}{h}\right)\widehat{e}_s^2$, $\widetilde{\sigma}_{n,1}^2 = 2\widetilde{\sigma}_e^4 \sum\limits_{t=1}^{n} \sum\limits_{s=1,\neq t}^{n} K_{s,t}^2$ and $\widetilde{\sigma}_e^2 = \frac{1}{n}\sum\limits_{t=1}^{n} e_t^2$.

By (A.10), in order to prove Theorem 2.1(i), we only need to show that

$$\frac{\overline{Q}_{n,i}(h)}{\overline{\sigma}_n} = o_P(1), \quad i = 2, \cdots, 6, \tag{A.11}$$

if $\delta_n = o\left(n^{-1/8}h^{-1/4}\right)$. The proof of (A.11) is given in Appendix B of the supplementary appendix Chen *et al* (2014) and details are omitted here to save the space.

When $\delta_n n^{1/8}h^{1/4} \to \infty$, we can show that $\overline{Q}_{n,3}(h)$ is the leading term of $Q_n(h)$ asymptotically under H_1^L. Hence, by Lemma A.4(i), to complete the proof of Theorem 2.1(ii), we need only to show that

$$\frac{1}{n^{3/4}h^{1/2}}\overline{Q}_{n,3}(h) \to_P \infty, \tag{A.12}$$

when $\delta_n n^{1/8}h^{1/4} \to \infty$. The proof of (A.12) is given in Appendix B of the supplementary appendix Chen *et al* (2014). We thus complete the proof of Theorem 2.1 ■

Proof of Theorem 2.2. The main idea of the proof is similar to that in the proof of Theorem 2.1. In order to save space, we only give a brief proof of (A.12) when $n^{3/8}\nu(\sqrt{n})h^{1/4}\delta_n \to \infty$. The rest of the proof (including the proof of Theorem 2.2(i)) is given in Appendix B of the supplementary appendix Chen *et al* (2014).

By Definition 2.6, there exist an asymptotically homogeneous function $\widetilde{\Lambda}(v)$ such that

$$\Delta_n^2(v) \geq \delta_n^2\widetilde{\Lambda}(v), \quad \widetilde{\Lambda}(\lambda x) = v^2(\lambda)\widetilde{H}(x) + \widetilde{R}(x,\lambda) \quad \text{and} \quad \int_{-\infty}^{\infty} \widetilde{H}(v)L_B(1,v)dv > 0, \tag{A.13}$$

where $\widetilde{H}(x)$ is locally integrable and $\widetilde{R}(x,\lambda)$ satisfies condition (i) or (ii) in Definition 2.6. Note that

$$\begin{aligned}\overline{Q}_{n,3}(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\Delta_n^2(V_s)K_{s,t} + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\left[\Delta_n(V_t)-\Delta_n(V_s)\right]K_{s,t}\Delta_n(V_s) \\ &\equiv \overline{Q}_{n,3,1}(h)+\overline{Q}_{n,3,2}(h). \qquad (A.14)\end{aligned}$$

By (A.13), Lemma A.3 and Theorem 5.3 in Park and Phillips (1999), we have, as $n\to\infty$,

$$\begin{aligned}\frac{\sigma_\phi}{n^{3/2}L_B(1,0)v^2(\sqrt{n})h\delta_n^2}\overline{Q}_{n,3,1}(h) &\geq \frac{1}{nv^2(\sqrt{n})}\sum_{s=1}^{n}\widetilde{\Lambda}(V_s)\left(1+o_P(1)\right) \\ &\to_P \int_{-\infty}^{\infty}\widetilde{H}(v)\ L_B(1,v)dv, \qquad (A.15)\end{aligned}$$

which, together with the condition $n^{3/8}\nu(\sqrt{n})h^{1/4}\delta_n\to\infty$, implies

$$\frac{1}{n^{3/4}h^{1/2}}\overline{Q}_{n,3,1}(h)\to_P\infty. \qquad (A.16)$$

On the other hand, by Definition 2.6 again, there exists an asymptotically homogeneous function $\overline{\Lambda}(\cdot)$ such that

$$\max\left\{\Delta_n^2(v),\left|\Delta_n(v)\dot{\Delta}_n(v)\right|\right\}\leq\delta_n^2\overline{\Lambda}(v), \qquad (A.17)$$

where $\overline{\Lambda}(\lambda x) = v^2(\lambda)\overline{H}(x)+\overline{R}(x,\lambda)$, in which $\overline{H}(\cdot)$ is locally integrable and $\overline{R}(\cdot,\cdot)$ satisfies condition (i) or (ii) in Definition 2.6. Then, by (A.17) and a first order Taylor expansion of $\Delta_n(\cdot)$ at V_s, we have, as $n\to\infty$,

$$\begin{aligned}\left|\overline{Q}_{n,3,2}(h)\right| &\leq \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\left|\Delta_n(V_t)-\Delta_n(V_s)\right|K_{s,t}\left|\Delta_n(V_s)\right| \\ &\leq (1+o_P(1))\ \sqrt{n}h^2\sum_{s=1}^{n}\left|\dot{\Delta}_n(V_s)\Delta_n(V_s)\right|\Big[\frac{1}{\sqrt{n}h}\sum_{t=1,\neq s}^{n}\left|\frac{V_t-V_s}{h}\right|K_{s,t}\Big] \\ &= O_P\big(n^{3/2}v^2(\sqrt{n})h^2\delta_n^2\big), \qquad (A.18)\end{aligned}$$

which, together with (A.15), implies that $\overline{Q}_{n,3,2}(h)$ is of smaller asymptotic order than $\overline{Q}_{n,3,1}(h)$. Hence, the combination of (A.14), (A.16) and (A.18) leads to (A.12). ■

Proof of Proposition 3.1. The proof of (3.2) in Proposition 3.1 relies on the application of Lemma C.3 given in Appendix C of the supplementary appendix Chen *et al* (2014). To save space for the main document, we relegate the detailed proof of (3.2) into Appendix B of the supplementary appendix. We next only prove (3.5) by using (3.2). Observe that under the null hypothesis H_0

$$Q_n(h)=\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widehat{e}_tK_{s,t}\widehat{e}_s=\overline{Q}_{n,1}(h)+\overline{Q}_{n,2}(h)+\overline{Q}_{n,5}(h), \qquad (A.19)$$

where $\overline{Q}_{n,j}(h)$, $j=1,2,5$, are defined in (2.3). Define the event

$$\mathcal{D}_n=\Big\{\Big|\frac{\overline{Q}_{n,2}(h)+\overline{Q}_{n,5}(h)}{\overline{\sigma}_n}\Big|>s_n\Big\}\cup\Big\{\Big|\frac{\sigma_{n,1}-\overline{\sigma}_n}{\sigma_{n,1}}\Big|>s_n\Big\},$$

where $s_n = n^{-1/8}h^{1/4}$, and let $\mathcal{D}_n^c$ be the complement of $\mathcal{D}_n$. Using arguments similar to those used in the proof of Theorem 2.1(i) and the proof of (A.5) in Gao *et al* (2009b), we can show that

$$\mathsf{P}\big(\mathcal{D}_n\big) = o(1). \tag{A.20}$$

Note that

$$\mathsf{P}\Big(\widehat{Q}_n(h) \le x\Big) = \mathsf{P}\Big(\big\{\widehat{Q}_n(h) \le x\big\} \cap \mathcal{D}_n^c\Big) + \mathsf{P}\Big(\big\{\widehat{Q}_n(h) \le x\big\} \cap \mathcal{D}_n\Big),$$

which together with (A.20), indicates that

$$\mathsf{P}\Big(\widehat{Q}_n(h) \le x\Big) = \mathsf{P}\Big(\big\{\widehat{Q}_n(h) \le x\big\} \cap \mathcal{D}_n^c\Big) + o(1). \tag{A.21}$$

On the event $\mathcal{D}_n^c$, both $\frac{\overline{Q}_{n,2}(h)+\overline{Q}_{n,5}(h)}{\overline{\sigma}_n}$ and $\frac{\sigma_{n,1}-\overline{\sigma}_n}{\sigma_{n,1}}$ are bounded by s_n which converges to zero as n tends to infinity. Hence, by (3.2), (A.19) and (A.21), we have, as $n \to \infty$,

$$\left|\mathsf{P}\left(\widehat{Q}_n(h) \le x\right) - \Phi(x)\right| \to 0 \tag{A.22}$$

uniformly in $x \in \mathcal{R}$. Similarly, we have, as $n \to \infty$,

$$\left|\mathsf{P}^*\left(\widehat{Q}_n^*(h) \le x\right) - \Phi(x)\right| \to_P 0 \tag{A.23}$$

uniformly in $x \in \mathcal{R}$. We can then prove (3.5) by (A.22) and (A.23). The result (3.6) follows directly from (3.5). We thus complete the proof of Proposition 3.1. ■

PROOF OF PROPOSITION 3.2. Since the proof of Proposition 3.2(ii) is similar to that of Proposition 3.2(i), we only provide the proof of Proposition 3.2(i). Observe that

$$\begin{aligned}\beta_n^*(h) &= \mathsf{P}\left(\widehat{Q}_n(h) > l_\alpha^* | H_1^L\right) = 1 - \mathsf{P}\left(\widehat{Q}_n(h) \le l_\alpha^* | H_1^L\right) \\ &= 1 - \mathsf{P}\Big(\widehat{Q}_{n,1}(h) \le l_\alpha^* - \frac{\overline{Q}_{n,3}(h)}{\overline{\sigma}_n} - \frac{\sum_{j=2,\neq 3}^{6} \overline{Q}_{n,j}(h)}{\overline{\sigma}_n} | H_1^L\Big).\end{aligned} \tag{A.24}$$

Following the detailed proof of Theorem 2.1 in Appendix B of the supplementary appendix Chen *et al* (2014), we have, as $n \to \infty$,

$$\overline{Q}_{n,i}(h) = o_P\left(nh\delta_n^2\right), \qquad i = 2, 4, 5, 6 \tag{A.25}$$

and

$$\mathsf{P}\Big(c_3^* nh\delta_n^2 < \big|\overline{Q}_{n,3}(h)\big| < c_4^* nh\delta_n^2\Big) \to 1, \tag{A.26}$$

for some $0 < c_3^* < c_4^* < \infty$. By (A.25) and (A.26), we may show that

$$\frac{\overline{Q}_{n,i}(h)}{\overline{\sigma}_n} = o_P\Big(\frac{\overline{Q}_{n,3}(h)}{\overline{\sigma}_n}\Big) \quad \text{for } i = 2, 4, 5, 6. \tag{A.27}$$

By (A.27), $\frac{\sum_{j=2,\neq 3}^{6} \overline{Q}_{n,j}(h)}{\overline{\sigma}_n}$ is dominated by $\frac{\overline{Q}_{n,3}(h)}{\overline{\sigma}_n}$ which is asymptotically equivalent to $\frac{\overline{Q}_{n,3}(h)}{\sigma_{n,1}}$ by (B.54) in Appendix B of the supplementary appendix. Hence, we have

$$\beta_n^*(h) \overset{P}{\sim} 1 - \mathsf{P}\Big(\widehat{Q}_{n,1}(h) \le l_\alpha^* - \frac{\overline{Q}_{n,3}(h)}{\sigma_{n,1}}(1 + o_P(1))\Big). \tag{A.28}$$

Noting that $l_\alpha^* - \frac{\overline{Q}_{n,3}(h)}{\sigma_{n,1}}$ is independent of $\{e_t\}$, by (3.2) in Proposition 3.1 and (A.28), we can complete the proof of Proposition 3.2 (i). ■

REFERENCES

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

Chen, J., Gao, J., Li, D., 2012. Estimation in semiparametric regression with nonstationary regressors. Bernoulli 18, 678–702.

Chen, J., Gao, J., Li, D. and Lin, Z., 2014. Supplementary appendix for 'Specification testing in nonstationary time series models'. Available at http://www.york.ac.uk/economics/research/discussion-papers/2014/

Chen, J., Gao, J., Li, D. and Lin, Z., 2011. Nonparametric specification testing in nonlinear and nonstationary time series models: theory and practice. Available at http://ssrn.com/abstract=2235356.

Chen, J., Li, D., Zhang, L., 2010. Robust estimator in a nonlinear cointegration model. Journal of Multivariate Analysis 101, 706–717.

Chen, S. X., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series models. Journal of Econometrics 141, 950–972.

Fan, J. and Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York.

Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Linton, O., 2003. Some higher–order theory for a consistent nonparametric model specification test. Journal of Statistical Planning and Inference 109, 125-154.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall, London.

Gao, J., Gijbels, I., 2008. Bandwidth selection in nonparametric kernel testing. Journal of the American Statistical Association 484, 1584-1594.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2009a. Specification testing in nonlinear and nonstationary time series autoregression. Annals of Statistics 37, 3893–3928.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2009b. Nonparametric specification testing for nonlinear time series with nonstationarity. Econometric Theory 25, 1869–1892.

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

Hong, S. H., Phillips, P. C. B., 2010. Testing linearity in cointegrating relations with an application to purchasing power parity. Journal of Business and Economic Statistics 28, 96–114.

Juhl, T., Xiao, Z., 2005a. A nonparametric test for changing trends. Journal of Econometrics 127, 179-199.

Juhl, T., Xiao, Z., 2005b. Testing for cointegration using partially linear models. Journal of Econometrics 124, 363–394.

Karlsen, H. A., Tjøstheim, D., 2001. Nonparametric estimation in null recurrent time series. Annals of Statistics 29, 372–416.

Karlsen, H. A., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. Annals of Statistics 35, 252–299.

Kasparis, I., 2008. Detection of functional form misspecification in cointegrating relations. Econometric Theory 24, 1373-1403

Li, Q., 1999. Consistent model specification tests for time series econometrics models. Journal of Econometrics 92, 101-147.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. Journal of Econometrics 87, 145-165.

Lin, Z., Li, D., Chen, J., 2009. Local linear M-estimator in null recurrent time series. Statistica Sinica 19, 1683–1703.

Marmer, V., 2008. Nonlinearity, nonstationarity, and spurious forecasts. Journal of Econometrics 142, 1-27.

Park, J., Phillips, P. C. B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269-298.

Park, J., Phillips, P. C. B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117-162.

Phillips, P. C. B., Park, J., 1998. Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper, No. 1181, Yale University.

Wang, Q. Y., Phillips, P. C. B., 2009a. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory 25, 710-738.

Wang, Q. Y., Phillips, P. C. B., 2009b. Structural Nonparametric Cointegrating Regression. Econometrica 77, 1901–1948.

Wang, Q. Y., Phillips, P. C. B., 2011. Asymptotic theory for zero energy functionals with nonparametric regression application. Econometric Theory 27, 235–259.

Wang, Q. Y., Phillips, P. C. B., 2012. A Specification test for nonlinear nonstationary models. Annals of Statistics 40, 727–758.

Wu, W., Min, W., 2005. On linear processes with dependent innovations. Stochastic Processes and Their Applications 115, 939–958.

Table 4.1. Size and local power of the test for the simulated example in Section 4

	Level 1%		Level 5%		Level 10%	
	size of test with bandwidths h_{cv} and h_{test}					
n	$f_{0\text{cv}}$	$f_{0\text{test}}$	$f_{0\text{cv}}$	$f_{0\text{test}}$	$f_{0\text{cv}}$	$f_{0\text{test}}$
200	0.0160	0.0110	0.0580	0.0510	0.1110	0.0990
500	0.0150	0.0100	0.0560	0.0500	0.0960	0.1000
800	0.0100	0.0100	0.0550	0.0500	0.0950	0.1000
	local power of test in Case 1 with bandwidths h_{cv} and h_{test}					
n	$f_{1\text{cv}}$	$f_{1\text{test}}$	$f_{1\text{cv}}$	$f_{1\text{test}}$	$f_{1\text{cv}}$	$f_{1\text{test}}$
200	0.1640	0.2390	0.3370	0.3980	0.4520	0.5140
500	0.3190	0.3740	0.4750	0.5140	0.5690	0.5870
800	0.3900	0.4250	0.5350	0.5600	0.6050	0.6650
	local power of test in Case 2 with bandwidths h_{cv} and h_{test}					
n	$f_{2\text{cv}}$	$f_{2\text{test}}$	$f_{2\text{cv}}$	$f_{2\text{test}}$	$f_{2\text{cv}}$	$f_{2\text{test}}$
200	0.1060	0.1150	0.2240	0.2400	0.3210	0.3240
500	0.1610	0.1580	0.2850	0.2860	0.3800	0.3930
800	0.2030	0.2230	0.3410	0.3410	0.4430	0.4500