Nonparametric Estimation and Specification Testing in Nonstationary Time Series Models

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

The University of Adelaide[1] *and Zhejiang University*[2]

Abstract

In this paper, we consider both estimation and testing problems in a nonlinear time series model with nonstationarity. A nonparametric estimation method is proposed to estimate a sequence of nonparametric departure functions. We also propose a test statistic to test whether the regression function is of a known parametric nonlinear form. The power function of the proposed nonparametric test is studied and an asymptotic distribution of the test statistic is shown to depend on the asymptotic behavior of the "distance function" $\Delta_n(\cdot)$ under a sequence of general semiparametric local alternatives. The asymptotic theory developed in this paper differs from existing work on nonparametric estimation and specification testing in the stationary time series case. In order to implement the proposed test in practice, a computer–intensive bootstrap simulation procedure is proposed and asymptotic approximations for both the size and power functions are established. Furthermore, the bandwidth involved in the test statistic is selected by maximizing the power function while the size function is controlled by a significance level. Meanwhile, both simulated and real data examples are provided to illustrate the proposed approach.

Keywords: Asymptotic distribution, Edgeworth expansion, estimation, nonlinear time series, power function, quadratic form, random walk, size function.

Abbreviated Title: Estimation and testing in nonlinear time series

1. Introduction

During the past two decades or so, there exists a rich literature on specification testing for a parametric model versus a nonparametric/semiparametric alternative when the time series satisfy a type of stationarity. Many testing procedures are proposed based on a nonparametric kernel method. Existing tests include Fan and Li (1996), Li and Wang (1998), Li (1999), Fan and Linton (2003), Chen and Gao (2007), and Juhl and Xiao (2005a). It is shown that the leading term of each of many existing nonparametric kernel test statistics is of a quadratic form (see, for example, 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 power function–based selection method for the choice of the bandwidth for optimal test 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, it may be quite restrictive to assume stationarity for time series data in practice. When tackling economic issues from a time perspective, for example, we often deal with nonstationary components. In dealing real–world problems, neither exchange rates nor prices, nor consumption, nor macroeconomic variables follow a stationary distribution. Hence, practitioners might feel more comfortable avoiding restrictions like stationarity for time series data. In this respect, existing literature 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), and Wang and Phillips (2008, 2009).

Meanwhile, there is some existing literature on model specification testing in the nonstationary time series case. Hong and Phillips (2005), and Kasparis (2008) considered model specification testing in cointegration models. 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. Gao and King (2007) considered testing for stationarity in a nonparametric autoregressive error model. More 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 for specifying whether the regression function is of a known parametric form indexed by a vector of unknown parameters and then established an asymptotic distribution for the proposed kernel test statistic under the null hypothesis.

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 some smooth function, $\{e_t\}$ is a sequence of stationary martingale differences, and $\{V_t\}$ is a random walk process of the form

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

in which $\{v_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables. We are then interested in studying a nonparametric test for specifying

$$H_0 : g(v) = g(v, \theta_0) \ \leftrightarrow \ H_1 : g(v) = g(v, \theta_1) + \Delta_n(v), \tag{1.3}$$

where $\theta_0 \in \Theta$ is the true value of the parameter θ under H_0, $\theta_1 \in \Theta$ and $\{\Delta_n(\cdot)\}$ is a sequence of nonparametrically unknown functions.

The choice of this type of semiparametric alternatives is mainly because interest in some cases is to detect whether there is a kind of slight departure from the null hypothesis when there is no sufficient evidence to suggest accepting the null hypothesis. Also in such cases, the level of such departure may be unknown. This is not uncommon when interest is to detect whether there is any slight unknown departure from an existing parametric trend in climatology, economics or finance, for example. To the best of our knowledge, the issue of how to consistently estimate $\Delta_n(\cdot)$ has been mentioned only in Gao *et al* (2009b).

In addition to establishing an asymptotic distribution of the proposed test statistic under H_0 as has been done in Gao *et al* (2009b), the current paper addresses two important issues. The first issue is that we propose using a local–linear kernel estimation method to consistently estimate $\Delta_n(\cdot)$ when the null hypothesis is not true and then establish an asymptotic distribution for a relative difference of the form $\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)} - 1$, where $\widehat{\Delta}_n(v)$ is an estimate to be proposed in Section 2 below. The second issue is that we study asymptotic properties of the proposed test statistic under a sequence of general semiparametric local alternatives. As shown in Theorems 3.1 and 3.2 in Section 3 below, we find that the asymptotic distribution of the proposed test statistic under the alternative hypothesis depends on the asymptotic behavior of the nonparametrically unknown distance function $\Delta_n(\cdot)$. For example, when the distance function is δ_n–integrable as defined in Definition 2.1, the proposed nonparametric kernel test can detect alternatives when $\delta_n n^{1/8} h^{1/4} \to \infty$ as $n \to \infty$. When the distance function is δ_n–asymptotically homogeneous as defined in Definition 2.2, the test can also detect alternatives when $\delta_n n^{3/8} v(\sqrt{n}) h^{1/4} \to \infty$ as $n \to \infty$, where $v(\sqrt{n}) \to \infty$ is as defined in Theorem 3.2 below.

Note that the corresponding order is $n^{-1/2} h^{-1/4}$ when parametrically specifying the mean function of a stationary time series (see, for example, Chapter 3 of Gao 2007). Our results in this paper show that for different distance functions, the nonparametric kernel test in the nonstationary time series case can detect alternatives with either smaller or larger rate than that in the stationary time series case. This is mainly because the rate of convergence of an estimator in the nonstationary case heavily depends the functional form of the function being estimated. Similar observations have been made in Park and Phillips (2001) when the authors dealt with parametric estimation in parametric nonlinear regression models with integrated regressors.

In order to implement the proposed test in practice, we propose a computer–intensive bootstrap simulation procedure for the choice of a suitable bandwidth for optimal testing purposes. The main idea for choosing an optimal bandwidth is to maximize the power function of the proposed test while the size function is controlled by a significance level. Meanwhile, we establish the asymptotic behavior of the bootstrap scheme under mild conditions and obtain an Edgeworth expansion for the asymptotic distribution of the bootstrap test statistic. The rate of the remainder term of the Edgeworth expansion in our paper is $O_P(n^{-1/2})$, which is of an order higher than $O_P(h)$ (since $\sqrt{n}h \to \infty$), the corresponding rate for the stationary time series case as established in Gao and Gijbels (2008). In addition, with the help of an Edgeworth expansion, we obtain some asymptotic approximations for both the size and power functions in Section 4.1.

The rest of the paper is organized as follows. Section 2 introduces a sequence of semiparametric local alternative functions and then proposes a nonparametric estimation method to estimate the "distance function" $\Delta_n(\cdot)$. An asymptotic distribution for the proposed estimate of $\Delta_n(\cdot)$ is then given in Section 2. Section 3 proposes a nonparametric kernel test for testing H_0 and H_1 and then studies asymptotic properties for the proposed nonparametric test. Sections 4 proposes a simulated bootstrap procedure for the bandwidth choice. Section 5 then illustrates the finite sample performance of the proposed test through using both simulated and real data examples. Section 6 concludes this paper with some comments. Appendix A provides some basic definitions for regular functions and necessary assumptions to establish the asymptotic theory. Appendix B gives some useful lemmas and then the proofs of the main results. The proofs of the lemmas are relegated to Appendix C of the supplementary document.

2. Nonparametric kernel estimation

Note that $\Delta_n(v)$ in H_1 of (1.3) can be viewed as the measure of the "distance" between the null hypothesis H_0 and the alternative hypothesis H_1. When the test rejects H_0, how to estimate the the "distance function" $\Delta_n(v)$ is an interesting topic. As far as we know, however, even in the stationary time series case, there is little study on the consistent estimation of $\Delta_n(v)$. In this section, we propose using a semiparametric kernel method to estimate the "distance function".

When H_1 holds, model (1.1) becomes

$$Y_t = g(V_t, \theta_1) + \Delta_n(V_t) + e_t, \quad t = 1, \cdots, n. \tag{2.1}$$

Before we propose our estimation method, we need to impose certain conditions on $g(v, \theta)$. As discussed in Park and Phillips (2001), we consider two classes of parametric nonlinear regression functions: I–regular on Θ and H_0–regular on Θ, whose detailed definitions are given in Appendix A below.

We also need to introduce the following two families of functions $\Delta_n(\cdot)$: δ_n–integrable functions and δ_n–asymptotically homogeneous functions, where $\delta_n \to 0$ as $n \to \infty$. In the sequel, $a(x) \sim b(x)$ denotes that $\frac{a(x)}{b(x)} \to 1$ as $x \to \infty$.

Definition 2.1. $\Delta_n(\cdot)$ *is said to be* δ_n*–integrable if*

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

for all $x \in \mathcal{R} = (-\infty, \infty)$, where $\Gamma_k(x)$, $k = 1, 2$, are integrable.

DEFINITION 2.2. *$\Delta_n(\cdot)$ 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 $x \in \mathcal{R} = (-\infty, \infty)$, 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 $\Delta_n(\cdot)$, $v_1(\lambda) \sim v_2(\lambda)$, $H_k(\cdot)$, $k = 1, 2$, 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 $P_k(\cdot)$ 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 $Q_k(\cdot)$ is locally integrable and vanishes at infinity, i.e., $Q_k(x) \to 0$ as $|x| \to \infty$.

These definitions can be viewed as some extensions of Definitions 4.1 and 4.2 in Park and Phillips (1999) to our asymptotic case.

Since $\Delta_n(v) \to 0$ for each given v by Definitions 2.1 and 2.2, we may construct a consistent estimator $\widehat{\theta}_1$ of θ_1 such that

$$\widehat{\theta}_1 = \arg \min_{\theta \in \Theta} \sum_{t=1}^{n} \left(Y_t - g(V_t, \theta)\right)^2. \tag{2.2}$$

Discussion about the suitability of using (2.2) and alternative estimation methods is given in Remark B.1 of Appendix B below.

We then estimate $\Delta_n(v)$ by a local linear estimate of the form

$$\widehat{\Delta}_n(v) = \sum_{t=1}^{n} \widetilde{w}_{nt}(v)(Y_t - g(V_t, \widehat{\theta}_1)), \tag{2.3}$$

where $\{\widetilde{w}_{nt}(v)\}$ is a sequence of weight functions given by

$$\widetilde{w}_{nt}(v) = \frac{\widetilde{K}_{v,b}(V_t)}{\sum_{k=1}^{n} \widetilde{K}_{v,b}(V_k)} \quad \text{with} \quad \widetilde{K}_{v,b}(V_t) = \frac{1}{b} \widetilde{K}_n\left(\frac{V_t - v}{b}\right), \tag{2.4}$$

in which $\widetilde{K}_n\left(\frac{V_t-v}{b}\right)=K\left(\frac{V_t-v}{b}\right)\left(S_{n,2}(v)-\left(\frac{V_t-v}{b}\right)\ S_{n,1}(v)\right)$, $K(\cdot)$ is a kernel function, b is the bandwidth, $S_{n,j}(v)=\frac{1}{T(n)b}\sum_{s=1}^{n}K\left(\frac{V_s-v}{b}\right)\ \left(\frac{V_s-v}{b}\right)^j$ for $j=0,1,2$, and $T(n)$ is the number of regenerations of $\{V_t\}$ in the time interval $[0,n]$ (see the beginning of Appendix B for more details).

Theorem 2.1 below establishes an asymptotic distribution for a relative difference of the form $\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)}-1$ under certain regularity conditions. Note that it is probably make more sense to use the relative difference than an absolute difference of the form $\widehat{\Delta}_n(v)-\Delta_n(v)$ when $\Delta_n(v)\to 0$ as $n\to\infty$.

We now state the first result of this paper; its proof is given in Appendix B below.

THEOREM 2.1. *Let condition A1 hold. Suppose that A3′ listed in Appendix A is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ has derivatives up to the second order.*

(i) Suppose that $\Delta_n(\cdot)$, $\dot{\Delta}_n(v)$ and $\ddot{\Delta}_n(v):=\frac{\partial^2\Delta_n(v)}{\partial v^2}$ are all δ_n–integrable. If, in addition, A2(i) is satisfied, then

$$\sqrt{T(n)b}\ \Delta_n(v)\left(\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)}-1+\xi_n\right)\stackrel{d}{\longrightarrow}N\left(0,\sigma_e^2\int K^2(v)dv\right),\tag{2.5}$$

where $\xi_n=O_P(b^2+(\sqrt{n}\dot{\kappa}^2(\sqrt{n}))^{-1/2}+(\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1})$.

(ii) Suppose that $\Delta_n(\cdot)$, $\dot{\Delta}_n(v)$ and $\ddot{\Delta}_n(v)$ are all δ_n–asymptotically homogeneous with asymptotically homogeneous order $v(\cdot)$, $\dot{v}(\cdot)$ and $\ddot{v}(\cdot)$. If, in addition, A2(ii) is satisfied, then

$$\sqrt{T(n)b}\ \Delta_n(v)\left(\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)}-1+\eta_n\right)\stackrel{d}{\longrightarrow}N\left(0,\sigma_e^2\int K^2(v)dv\right),\tag{2.6}$$

where $\eta_n=O_P\left(b^2+v(\sqrt{n})(\dot{\kappa}(\sqrt{n}))^{-1}+(\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1}\right)$.

Under some additional conditions, Theorem 2.1 implies the following corollary.

COROLLARY 2.1. *(i) Assume that the conditions of Theorem 3.1(i) are satisfied. If, in addition,*

$$\sqrt{\sqrt{n}b}\ \delta_n\xi_n=o(1),\tag{2.7}$$

we have

$$\sqrt{T(n)b}\ \Delta_n(v)\left(\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)}-1\right)\stackrel{d}{\longrightarrow}N\left(0,\sigma_e^2\int K^2(v)dv\right).\tag{2.8}$$

(ii) Assume that the conditions of Theorem 2.1(ii) are satisfied. If, in addition,

$$\sqrt{\sqrt{n}b}\ \delta_n\eta_n = o(1), \tag{2.9}$$

then (2.8) still holds.

REMARK 2.1. (i) Since $T(n)$ is not easy to obtain in practice, it can be replaced by $T_C(n) := \sum_{t=1}^{n} I_C(V_t)$, which is defined as in Karlsen and Tjøstheim (2001). In fact, $T_C(n)$ is the number of times that the process $\{V_t\}$ visits the set C up to the time n. By Lemma 3.2 in Karlsen and Tjøstheim (2001), we have

$$\frac{T_C(n)}{T(n)} = \pi_s(C) + o(1) \ \ a.s.,$$

where π_s is the invariant measure of the random walk $\{V_t\}$.

(ii) As shown in Lemma B.1 below, $T(n)$ is proportional to $\sqrt{n}$. Thus, under the condition $\sqrt{\sqrt{n}b}\ \delta_n \to \infty$, Corollary 2.1 implies

$$\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)} = 1 + O_P\left(\left(\sqrt{\sqrt{n}b}\ \delta_n\right)^{-1}\right).$$

REMARK 2.2. By equation (2.5), the asymptotic mean square error of $\frac{\widehat{\Delta}_n(v)}{\Delta_n(v)} - 1$ equals

$$O_P\left(\frac{1}{T(n)b\delta_n^2} + b^4 + (\sqrt{n}\dot{\kappa}^2(\sqrt{n}))^{-1} + (\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-2}\right) = O_P\left(\frac{1}{T(n)b\delta_n^2} + b^4\right)$$

when $\left(\sqrt{n}\dot{\kappa}^2(\sqrt{n})\right)^{-1} = o(b^4)$ and $(\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-2} = o(b^4)$. This leads to an optimal bandwidth of the form

$$\widetilde{b}_{\text{optimal}} = C_v(T(n)\delta_n^2)^{-1/5}(1 + o_P(1)) = O_P\left(\left(\sqrt{n}\delta_n^2\right)^{-1/5}\right), \tag{2.10}$$

where C_v is some positive constant depending on v. In fact, the above order for an optimal bandwidth is analogous to that in the stationary time series case ($T(n) = n$ and δ_n equals some nonzero constant).

REMARK 2.3. As the leading order for an optimal bandwidth in (2.10) is not sufficient and practically useful in the finite–sample study in Section 5 below, we propose using a semiparametric cross–validation selection method of the form

$$\widehat{b}_{\text{optimal}} = \arg \min_{\{\text{over all possible } b \text{ values}\}} \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - g(V_t, \widehat{\theta}_1) - \widehat{\Delta}_{n,-t}(V_t)\right)^2, \tag{2.11}$$

where $\widehat{\Delta}_{n,-t}(v) = \sum_{s=1,\neq t}^{T} \widetilde{w}_{ns,-t}(v) \left(Y_s - g(V_s, \widehat{\beta}_1)\right)$, in which $\widetilde{w}_{ns,-t}(v) = \frac{\widetilde{K}_{v,b}^{(-t)}(V_s)}{\sum_{s=1,\neq t}^{n} \widetilde{K}_{v,b}^{(-t)}(V_s)}$ and

$$\widetilde{K}_{v,b}^{(-t)}(V_s) = \frac{1}{b} K\left(\frac{V_s - v}{b}\right) \left(S_{n2,-t}(v) - \left(\frac{V_s - v}{b}\right) S_{n1,-t}(v)\right)$$

with $S_{nj,-t}(v) = \frac{1}{\sqrt{nb}} \sum_{s=1,\neq t}^{n} \left(\frac{V_s - v}{h}\right)^j K\left(\frac{V_s - v}{b}\right)$ for $j = 0, 1, 2$.

Theoretical discussion about $\widehat{b}_{\text{optimal}}$ is a difficult and interesting issue in itself and is therefore left for future study.

3. Nonparametric specification testing

This section proposes testing H_0 and H_1. As proposed for both the stationary and nonstationary cases (see Gao 2007; Li and Racine 2007; Gao *et al* 2009b), a nonparametric kernel test of the quadratic form

$$Q_n(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_t K\left(\frac{V_t - V_s}{h}\right) \widehat{e}_s \tag{3.1}$$

has been shown to work well in both the large and small/medium sample cases, where $K(\cdot)$ is some kernel function, h is the bandwidth, and $\widehat{e}_t = Y_t - g(V_t, \widehat{\theta})$, in which $\widehat{\theta}$ is the nonlinear least squares estimator of θ_0 under H_0.

Observe that under H_0:

$$\begin{aligned} Q_n(h) &= \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_t K_{s,t} \widehat{e}_s = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} e_t K_{s,t} e_s \\ &+ \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \overline{g}_t K_{s,t} \overline{g}_s + 2 \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \overline{g}_t K_{s,t} e_s =: \sum_{i=1}^{3} Q_{n,i}(h), \end{aligned}$$

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

As pointed out earlier, Gao *et al* (2009b) already established an asymptotic distribution for $Q_n(h)$ under H_0. In this section, we mainly study the power function of a standardized version of $Q_n(h)$ for the case where the alternative hypothesis H_1 holds.

Before studying the power function of the nonparametric test (1.3), we establish an asymptotic distribution for the test statistic (3.1) under H_0; its proof is given in Appendix B below.

Proposition 3.1. *Assume that conditions A1, A2(iii) and either A3 or A3′ in Appendix A below are satisfied. Then under* H_0*:*

$$\widehat{Q}_n(h) := \frac{Q_n(h)}{\overline{\sigma}_n} \xrightarrow{d} N(0,1), \tag{3.2}$$

where $\overline{\sigma}_n^2 = 2\overline{\sigma}_e^4 \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} K_{s,t}^2$ *with* $\overline{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n} \widehat{e}_t^2$.

While the same asymptotic normality was established by Gao *et al* (2009b), the conditions used are a set of different conditions. The authors basically assumed some high–level conditions on the rate of convergence of $\widehat{\theta}$ to θ_0 for both the establishment and the proof of the asymptotic normality. By contrast, this paper chooses to impose some natural conditions to achieve such a rate of convergence as given in A3 or A3′ in Appendix A below.

As shown in Appendix B below, under H_1 we have

$$\begin{aligned} Q_n(h) &= \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} e_t K_{s,t} e_s + \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t} \widetilde{g}_s + 2\sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t} e_s \\ &+ 2\sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} \widetilde{g}_s + 2\sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} e_s \\ &+ \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} \Delta_n(V_s) \equiv \sum_{j=1}^{6} \overline{Q}_{n,j}(h), \end{aligned} \tag{3.3}$$

where $\widetilde{g}_t = g(V_t, \theta_1) - g(V_t, \widehat{\theta}_1)$.

In view of (3.3), in order to study the asymptotic behavior of $Q_n(h)$, certain conditions on $\{\Delta_n(\cdot)\}$ are needed. We now establish some asymptotic properties for a standardized version of $Q_n(h)$ under H_1 in Theorems 3.1 and 3.2 below; their proofs are given in Appendix B of this paper. We first consider the case where $\Delta_n(\cdot)$ is δ_n–integrable.

Theorem 3.1. *Assume that conditions A1 and A2(iii) in Appendix A are satisfied and that either A3 or A3′ in Appendix A is satisfied when* θ_0 *is replaced by* θ_1*. In addition, both* $\Delta_n(\cdot)$ *and* $\dot{\Delta}_n(v) := \frac{d\Delta_n(v)}{dv}$ *are* δ_n*–integrable.*

(i) If $\delta_n = O\left(n^{-1/8}h^{-1/4}\right)$*, then under* H_1

$$\widehat{Q}_n(h) - \lambda_n = \frac{Q_n(h)}{\overline{\sigma}_n} - \lambda_n \xrightarrow{d} N(0,1) \tag{3.4}$$

as $n \to \infty$*, where* $\lambda_n \equiv \frac{\sum_{j=2}^{6} \overline{Q}_{n,j}(h)}{\overline{\sigma}_n}$ *satisfies* $\lambda_n = O_P(1)$.

Furthermore, if $\delta_n = o\left(n^{-1/8}h^{-1/4}\right)$, *then under* H_1

$$\widehat{Q}_n(h) = \frac{Q_n(h)}{\overline{\sigma}_n} \stackrel{d}{\longrightarrow} N(0,1) \tag{3.5}$$

as $n \to \infty$.

(ii) If $n^{1/8}h^{1/4}\delta_n \to \infty$ *as* $n \to \infty$, *then under* H_1, $\widehat{Q}_n(h) \longrightarrow_P \infty$ *as* $n \to \infty$.

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

THEOREM 3.2. *Assume that conditions A1 and A2(iii) in Appendix A are satisfied and A3′ in Appendix A is satisfied when* θ_0 *is replaced by* θ_1. *Suppose that* $\Delta_n(\cdot)$ *and* $\dot{\Delta}_n(v)$ *are both* δ_n*–asymptotically homogeneous with asymptotically homogeneous order* $v(\cdot)$ *and* $\dot{v}(\cdot)$, *respectively. In addition,* $v(m) = O(\dot{\kappa}(m))$ *as* $m \to \infty$, *where* $\dot{\kappa}(\cdot)$ *is defined in A3′ of Appendix A.*

(i) If $\delta_n = O\left(n^{-3/8}v^{-1}(\sqrt{n})h^{-1/4}\right)$, *then (3.4) still holds under* H_1. *Furthermore, if* $\delta_n = o\left(n^{-3/8}v^{-1}(\sqrt{n})h^{-1/4}\right)$, *then (3.5) still holds under* H_1.

(ii) If $n^{3/8}v(\sqrt{n})h^{1/4}\delta_n \to \infty$ *as* $n \to \infty$, *then under* H_1, $\widehat{Q}_n(h) \longrightarrow_P \infty$ *as* $n \to \infty$.

Theorems 3.1 and 3.2 show that whether the proposed test under H_1 is asymptotically powerful depends on the rate of $\delta_n \to 0$. When the rate of $\delta_n \to 0$ is faster than that of $n^{-1/8}h^{-1/4} \to 0$, for example, Theorem 3.1(i) shows that $\widehat{Q}_n(h)$ converges in distribution to the standard normality. This implies that the proposed test is not asymptotically powerful. When the rate of $\delta_n \to 0$ is slower than that of $n^{-1/8}h^{-1/4} \to 0$, Theorem 3.1(ii) shows that $\widehat{Q}_n(h) \to \infty$ in probability. The same implications can be derived from Theorem 3.2.

In Propositions 4.1 and 4.2 below, we are able to establish more explicit results for both the asymptotic distribution and the power function of a bootstrapping version of the proposed test for the special case where $\{e_t\}$ is a sequence of i.i.d. continuous random variables. This is mainly because the applicability of the Edgeworth expansions involved requires the i.i.d. assumption on $\{e_t\}$. Note also that the use of an Edgeworth expansion for this kind of quadratic form is valid for the case where $\{e_t\}$ is assumed to be a sequence of continuous random variables. As a consequence, the Cramér condition is satisfied (see page 45 of Hall 1992).

4. Bootstrap method and bandwidth selection

Proposition 3.1 shows that the proposed test statistic has an asymptotic standard normal distribution under the null hypothesis H_0. This section proposes combining a bootstrap simulation procedure with an Edgeworth expansion for the test statistic $\widehat{Q}_n(h)$ in order to establish a bandwidth selection method for the choice of an optimal bandwidth h for optimal testing purposes. We first propose using a bootstrap method to construct a simulated critical value in each case.

Let l_α be the α–level critical value, which is the $(1-\alpha)$–quantile of the exact finite sample distribution of $\widehat{Q}_n(h)$.

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

$$\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - g(V_t, \widehat{\theta})\right)^2, \tag{4.1}$$

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

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 data set $\{(Y_t^*, V_t), 1 \le t \le n\}$ to re–estimate θ_0 and denote its estimator 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 produce M versions of $\widehat{Q}_n^*(h)$. Denote the M versions of $\widehat{Q}_n^*(h)$ by $\widehat{Q}_{n,m}^*(h)$, $m = 1, 2, \cdots, M$. Then, we construct the empirical distributions of $\widehat{Q}_{n,m}^*(h)$. That is, $P^*\left(\widehat{Q}_n^*(h) \le x\right) = P\left(\widehat{Q}_n^*(h) \le x | \mathcal{W}_n\right)$, where $\mathcal{W}_n = \{(Y_t, V_t), 1 \le t \le n\}$. For each fixed h, choose l_α^* such that $P^*\left(\widehat{Q}_n^*(h) > l_\alpha^*\right) = \alpha$ and estimate l_α by l_α^*.

Proposition 4.1 below establishes an Edgeworth expansion for the bootstrap distribution of the test statistic $\widehat{Q}_n^*(h)$. Asymptotic approximations to the bootstrapping version of the power function of $\widehat{Q}_n^*(h)$ are given in Proposition 4.2 below.

Proposition 4.1. *Assume that conditions A1(i)(ii)(iv), A2(iii) and either A3 or A3′ in Appendix A are satisfied. Furthermore, $\{e_t\}$ is a sequence of i.i.d. continuous random variables with $E[e_1] = 0$, $E[e_1^2] = \sigma_e^2$ and $E[e_1^6] < \infty$. Then under H_0, we have* *as $n \to \infty$,*

$$\sup_{x \in \mathcal{R}} \left| P^*\left(\widehat{Q}_n^*(h) \le x\right) - \Phi(x - s_n^*) + \rho_n^*(h)\Phi^{(3)}(x - s_n^*)\right| = O_P(n^{-1/2}), \tag{4.2}$$

where $s_n^* = O_P(n^{-1/8}h^{1/4})$,

$$\rho_n^*(h) = \frac{\sqrt{2}}{3} \cdot \frac{\mathrm{Tr}(A_0^3(h))}{\widetilde{\sigma}_{n,1}^3(h)} = O_P\left(n^{-1/4}h^{1/2}\right), \tag{4.3}$$

in which $\widetilde{\sigma}_{n,1}^2(h) = \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} K_{s,t}^2$, $\Phi(\cdot)$ is the distribution function of the standard normal distribution, $\Phi^{(3)}(\cdot)$ is the third derivative of $\Phi(\cdot)$ and $\mathrm{Tr}(A_0^3(h))$ is the trace of the matrix $A_0(h)$ which is given by

$$A_0(h) = \begin{pmatrix} 0 & K_{1,2} & \cdots & K_{1,n} \\ K_{2,1} & 0 & \cdots & K_{2,n} \\ \vdots & \vdots & & \vdots \\ K_{n,1} & K_{n,2} & \cdots & 0 \end{pmatrix}. \tag{4.4}$$

The above result extends some existing results in the literature (such as, Li and Wang 1998; Fan and Linton 2003; Gao 2007; and Gao and Gijbels 2008) for the stationary case to the nonstationary time series case. The rate of the right hand side of (4.2) is $O_P(n^{-1/2})$, the rate of which going to zero is faster than that of $O_P(h)$, the corresponding rate in the stationary time series case (as $\sqrt{n}h \to \infty$ by A2(iii) in Appendix A).

To study the size and power functions of the proposed test, we introduce the following bootstrapping version of the size and power functions of the form

$$\alpha_n^*(h) = P^*\left(\widehat{Q}_n^*(h) > l_\alpha^* | H_0\right) \quad \text{and} \quad \beta_n^*(h) = P^*\left(\widehat{Q}_n^*(h) > l_\alpha^* | H_1\right). \tag{4.5}$$

For optimal testing purposes, we then propose choosing an optimal bandwidth h_{test} at the α significance level such that

$$h_{\text{test}} = \arg\max_{h \in \mathcal{H}_n} \beta_n^*(h), \tag{4.6}$$

where $\mathcal{H}_n = \{h : \alpha_n^*(h) = \alpha\}$.

The above bandwidth selection method is based on the criterion that the optimal bandwidth is chosen such that while the size of the bootstrapping version of the test is under control, the power of the bootstrapping version of the test is maximized. This is motivated by existing literature for the stationary time series case, such as Gao and Gijbels (2008). In the finite–sample simulation study and empirical analysis, we show in Section 5 below that this method also performs well for the nonstationary case.

In theory, Proposition 4.2 below provides an asymptotic approximation to the leading term of $\beta_n^*(h)$ in each case.

PROPOSITION 4.2. *Assume that conditions A1(i)(ii)(iv) and A2(iii) in Appendix A are satisfied. Furthermore, $\{e_t\}$ is a sequence of i.i.d. continuous random variables with $E[e_1] = 0$, $E[e_1^2] = \sigma_e^2$ and $E[e_1^6] < \infty$.*

(i) Suppose that both $\Delta_n(\cdot)$ and $\dot{\Delta}_n(v)$ are δ_n–integrable. If, in addition, $n^{1/8}h^{1/4}\delta_n \to \infty$ as $n \to \infty$, then,

$$\beta_n^*(h) = 1 - \Phi(l_\alpha^* - \varpi_n^*) - \rho_n^*(h)(1 - (l_\alpha^* - \varpi_n^*)^2)\phi(l_\alpha^* - \varpi_n^*) + O_P(n^{-1/2}), \qquad (4.7)$$

where $\varpi_n^* = \frac{\sum_{t=1}^{n}\sum_{s\neq t} \Delta_n(V_s)\ K_{s,t}\ \Delta_n(V_t)}{\sigma_e^2\sqrt{2\sum_{t=1}^{n}\sum_{s\neq t} K_{s,t}^2}}(1 + o_P(1))$ *and $\phi(\cdot)$ is the standard normal density function.*

(ii) Suppose that both $\Delta_n(\cdot)$ and $\dot{\Delta}_n(v)$ are δ_n–asymptotically homogeneous. If, in addition, $n^{3/8}v(\sqrt{n})h^{1/4}\delta_n \to \infty$ as $n \to \infty$, then equation (4.7) holds.

The proofs of Propositions 4.1 and 4.2 are given in Appendix B below. Equation (4.7) implies that $\beta_n^*(h)$ can be approximated by

$$\widehat{\beta}_n^*(h) \;=\; 1 - \Phi(l_\alpha^* - \widehat{\varpi}_n^*) - \widehat{\rho}_n^*(h)(1 - (l_\alpha^* - \widehat{\varpi}_n^*)^2)\phi(l_\alpha^* - \widehat{\varpi}_n^*), \qquad (4.8)$$

where $\widehat{\rho}_n^*(h)$ is an estimate of $\rho^*(h)$ with σ_e^2 being replaced by its conventional estimator $\widehat{\sigma}_e^2$ and $\widehat{\varpi}_n^*$ is an estimator of ϖ_n^* with $\Delta_n(\cdot)$ and σ_e^2 being replaced by $\widehat{\Delta}_n(\cdot)$ and $\widehat{\sigma}_e^2$, respectively.

Section 5 below employes (4.6) directly in the choice of h_{test}. Our experience with Tables 5.1–5.3 below shows that the use of $\widehat{\beta}_n^*(h)$ in the implementation of (4.6) is computationally less expensive, although the resulting size and power values are slightly less satisfactory.

5. Examples of implementation

This section provides both simulated and real data examples to implement the test proposed in Section 3 in association with the bandwidth selection procedure proposed in Section 4.

Example 5.1. Consider a time series model of the form

$$Y_t = g(V_t, \theta) + e_t \quad \text{with} \quad V_t = V_{t-1} + v_t, \quad t = 1, 2, \cdots, \tag{5.1}$$

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 $V_0 = 0$.

This example then considers three pairs of hypotheses. The first pair named as HI is as follows:

$$H_0: \ g(v, \theta_0) = \theta_0 v \quad \leftrightarrow \quad H_1: \ g(v, \theta_1) = \theta_{11} v + \theta_{12} v^2. \tag{5.2}$$

where the initial values of the parameters are chosen as follows:

$$\text{Case 1}: \ \theta_0 = \theta_{11} = 1, \ \theta_{12} = 0.005; \ \text{Case 2}: \ \theta_0 = \theta_{11} = 1, \ \theta_{12} = 0.01. \tag{5.3}$$

The second pair named as HII is as follows:

$$H_0: \ g(x, \theta) = \theta_0 v \quad \leftrightarrow \quad H_1: \ g(v, \theta) = \theta_{21} v + \log\left(1 + \theta_{22} v^2\right), \tag{5.4}$$

where the initial values of the parameters are chosen as follows:

$$\text{Case 1}: \ \theta_0 = \theta_{21} = 1, \quad \theta_{22} = 0.005; \ \text{Case 2}: \ \theta_0 = \theta_{21} = 1, \quad \theta_{22} = 0.01. \tag{5.5}$$

The third pair named as HIII is as follows:

$$H_0: \ g(x, \theta) = \frac{1}{1 + \theta_0 v^2} \quad \leftrightarrow \quad H_1: \ g(v, \theta) = \frac{1}{1 + \theta_{31} v^2} + \left(1 - e^{-\theta_{32} v^2}\right), \tag{5.6}$$

where the initial values of the parameters are chosen as follows:

$$\text{Case 1}: \ \theta_0 = \theta_{31} = 0.5, \quad \theta_{32} = 0.001; \ \text{Case 2}: \ \theta_0 = \theta_{31} = 0.5, \quad \theta_{32} = 0.005. \tag{5.7}$$

In the following simulation, we estimate the parameters by an ordinary least squares method. We choose $K(x) = \frac{1}{2} I_{[-1,1]}(x)$ as the kernel function throughout the examples in this section. The bandwidth selection method proposed in (4.6) is applied in the simulation study. Meanwhile, we also propose using the cross–validation (CV) based bandwidth given by h_{cv} as chosen by (2.11). Note that both h_{test} and h_{cv} have two different versions under the null and alternative hypotheses.

To use some simple notations, we introduce $h_{i\text{test}} = h_{\text{test}}$ and $h_{i\text{cv}} = h_{\text{cv}}$ for $i = 0, 1, 2$ to represent $h_{0\text{test}}$ and $h_{0\text{cv}}$ under H_0, and $h_{i\text{test}}$ and $h_{i\text{cv}}$ under H_1 for Cases i with $i = 1, 2$. We then define $Q_{i\text{test}} = \widehat{Q}_n(h_{i\text{test}})$ and $Q_{i\text{cv}} = \widehat{Q}_n(h_{i\text{cv}})$ for $i = 0, 1, 2$. For $i = 0, 1, 2$, let $f_{i\text{test}}$ denote the frequency of $Q_{i\text{test}} > l_\alpha^*(h_{i\text{test}})$ and $f_{i\text{cv}}$ denote the frequency of $Q_{i\text{cv}} > l_\alpha^*(h_{i\text{cv}})$.

Table 5.1. Simulated sizes and power values for HI in (5.2)

	Level 1%		Level 5%		Level 10%	
	H_0 holds					
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	0.013	0.010	0.055	0.050	0.099	0.100
500	0.013	0.010	0.057	0.050	0.100	0.100
800	0.007	0.010	0.050	0.050	0.091	0.100
	H_1 holds (Case 1)					
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	0.286	0.345	0.385	0.465	0.474	0.540
500	0.769	0.878	0.843	0.922	0.876	0.947
800	0.969	0.982	0.985	0.993	0.990	0.996
	H_1 holds (Case 2)					
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	0.584	0.629	0.684	0.738	0.738	0.806
500	0.957	0.978	0.975	0.991	0.985	0.997
800	0.999	1.000	1.000	1.000	1.000	1.000

In Table 5.1, we consider the case where the number of replications for each of the sample versions of the size and power functions is $R = 1000$, each with $M = 250$ number of bootstrapping resamples $\{e_t^*\}$ (involved in the bootstrap scheme introduced in Section 4) from the standard normal distribution $N(0, 1)$, and the simulations are done for the cases of $n = 200$, 500 and 800.

It follows from Table 5.1 that the simulated sizes for the test based on h_{test} perform better than those based on h_{cv} since h_{test} is chosen to make sure that the size function can be controlled by the significance level. Furthermore, the test based on h_{test} is more powerful than that based on h_{cv}. As θ_{12} in Case 2 is larger than that in Case 1, $f_{2\text{test}}$ and $f_{2\text{cv}}$ are larger than $f_{1\text{test}}$ and $f_{1\text{cv}}$, respectively. In fact, both the sizes and power values for the test based on h_{cv} also perform well in Table 4.1 for moderate sample ($n = 500$ and 800). Meanwhile, the power values of the proposed test improve as the sample size increases.

Table 5.2. Simulated sizes and power values for HII in (5.4)

	Level 1%		Level 5%		Level 10%	
	H_0 holds					
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	0.016	0.011	0.068	0.052	0.113	0.100
500	0.018	0.012	0.054	0.050	0.108	0.100
800	0.018	0.011	0.052	0.051	0.103	0.100
	H_1 holds (Case 1)					
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	0.058	0.077	0.119	0.152	0.181	0.241
500	0.241	0.329	0.361	0.491	0.452	0.584
800	0.463	0.557	0.585	0.713	0.664	0.778
	H_1 holds (Case 2)					
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	0.108	0.146	0.198	0.256	0.278	0.355
500	0.380	0.484	0.503	0.622	0.585	0.690
800	0.628	0.712	0.716	0.801	0.767	0.850

We find from Table 5.2 that the test based on h_{test} not only avoids any size distortion, but also is more powerful when the null hypothesis does not hold. Furthermore, $f_{2\text{test}}$ and $f_{2\text{cv}}$ are larger than $f_{1\text{test}}$ and $f_{1\text{cv}}$ in Table 5.2 as θ_{22} in Case 2 is larger than that in Case 1. Meanwhile, as $\log(1 + x_n) < x_n$ for $x_n > 0$, $f_{i\text{test}}$ and $f_{i\text{cv}}$ in Table 5.2 are smaller than corresponding results in Table 5.1 for $i = 1, 2$. As in Table 5.1, the power values in Table 5.2 improve as the sample size increases. The simulation results show that the proposed test performs well even when the distance function is nonlinear under H_1.

Table 5.3. Simulated sizes and power values for HIII in (5.6)

	Level 1%		Level 5%		Level 10%	
	H_0 holds					
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	0.01	0.01	0.043	0.05	0.092	0.100
500	0.014	0.011	0.064	0.050	0.112	0.101
800	0.016	0.011	0.053	0.050	0.100	0.100
	H_1 holds (Case 1)					
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	0.052	0.089	0.108	0.196	0.167	0.269
500	0.427	0.447	0.532	0.548	0.607	0.621
800	0.701	0.719	0.780	0.785	0.825	0.826
	H_1 holds (Case 2)					
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	0.195	0.318	0.316	0.562	0.415	0.644
500	0.852	0.861	0.909	0.912	0.928	0.941
800	0.954	0.959	0.975	0.977	0.985	0.988

The simulation results in Table 5.3 show that the proposed test in this paper performs well when both the regression function and the "distance function" are nonlinear. Analogously, we find that the test based on h_{test} not only avoids the size distortion, but also is more powerful than that based on h_{cv}.

Example 5.2. In this example, we consider the 2–year (X_{1t}) and 30–year (X_{2t}) Australian government bonds, which represent short–term and long–term series in the term structure of interest rates.

Our aim is to analyze the relationship between the long–term data $\{X_{2t}\}$ and short–term data $\{X_{1t}\}$. We first apply the transformed versions defined by $Y_t = \log(X_{2t})$ and $V_t = \log(X_{1t})$. The time frame of the study is during January 1971 to December 2000, with 360 observations for each of $\{Y_t\}$ and $\{V_t\}$.

Consider the null hypothesis defined by

$$H_0: \; Y_t = \alpha_1 + \beta_1 V_t + \gamma_1 V_t^2 + e_t, \tag{5.8}$$

where $\{e_t\}$ is an unobserved error process.

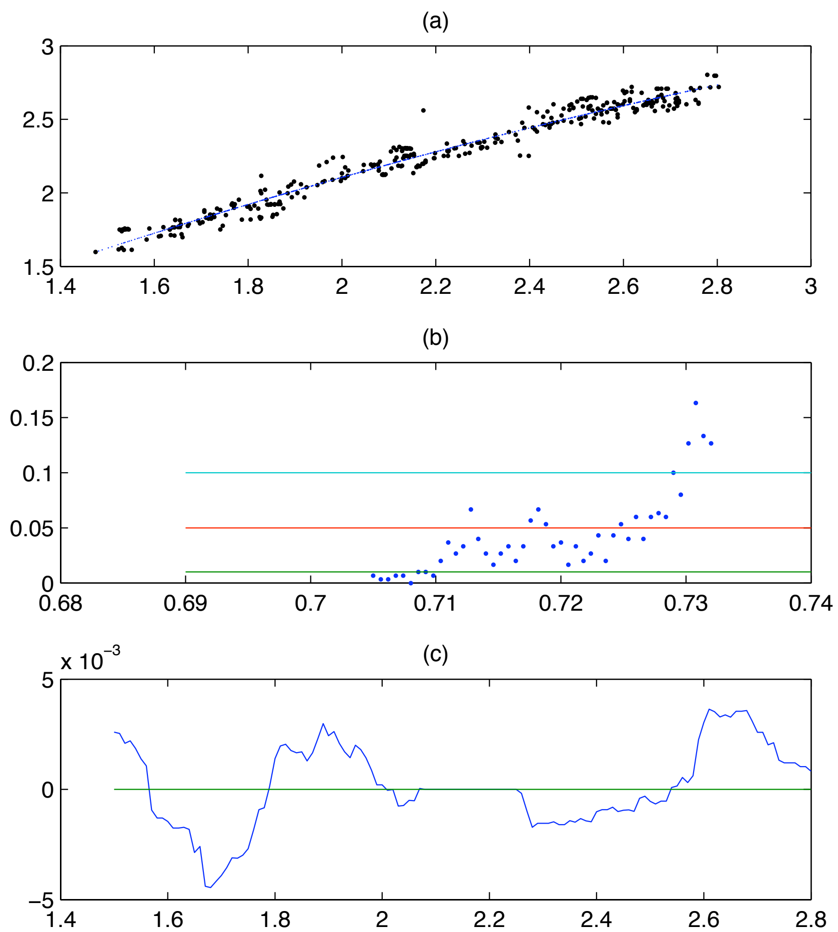

Figure 5.1: (a) provides the scatter chart of (Y_t, V_t) and the binomial regression plot $y = -0.2338 + 1.4446\ v - 0.1374\ v^2$, (b) gives p–values of the test for different bandwidths and (c) is the plot of $\widehat{\Delta}_n(v)$ when the null hypothesis (5.8) does not hold (the values of $\widehat{\Delta}_n(v)$ are between -5×10^{-3} and 5×10^{-3}).

In Figure 5.1(a), we provide the scatter chart of (Y_t, V_t) and a regression estimator of the form $y = -0.2338 + 1.4446\ v - 0.1374\ v^2$. In Figure 5.1(b), we give the p–values of the test for different bandwidths. The upper line in Figure 5.1(b) corresponds to the 10% significance level, the intermediate one corresponds to the 5% significance level and the lower one corresponds to the 1% significance level. The p–value plot suggests that we should apply the second–order polynomial regression model when h is large, since a large bandwidth h will produce a very smooth nonparametric estimator which is close to the second–order polynomial regression function. When h is small, there is only a few data used in the procedure, which implies that the variability of the nonparametric estimate is larger. Hence, for small h, the p–value suggests rejecting the null hypothesis defined in (5.8).

In order to check whether there is any departure from the second–order polynomial fitting, we propose using a semiparametric estimate of the form (2.3) given by

$$\widehat{\Delta}_n(v) = \sum_{t=1}^{n} \widetilde{w}_{nt}(v) \left(Y_t - \widehat{\alpha}_1 - \widehat{\beta}_1 V_t - \widehat{\gamma}_1 V_t^2 \right), \tag{5.9}$$

where $\widehat{\alpha}_1 = -0.2338$, $\widehat{\beta}_1 = 1.4446$, $\widehat{\gamma}_1 = -0.1374$, and $\{\widetilde{w}_{nt}(v)\}$ is as defined in (2.4), in which $K(x) = \frac{1}{2} I_{[-1,1]}(x)$ and the optimal bandwidth $\widehat{b}_{\text{optimal}}$ is chosen by (2.11).

In Figure 5.1(c), we provide the plot of $\widehat{\Delta}_n(v)$. Figure 5.1(c) suggests that the values of $\widehat{\Delta}_n(v)$ are between -5×10^{-3} and 5×10^{-3}. This thus shows that the relationship between Y_t and V_t may be approximately modeled by a second–order polynomial function of the form $y = -0.2338 + 1.4446\ v - 0.1374\ v^2$.

6. Conclusions

The main contributions of this paper can be summarized as follows. We have proposed a nonparametric estimation method to estimate the unknown "distance function". We have also proposed a nonparametric kernel test for specifying whether the regression function of a nonstationary regressor is of a known parametric form. We have first discussed asymptotic properties of the nonparametric estimation method and then shown that the asymptotic distribution of the proposed test statistic under the alternative hypothesis depends on the smoothness of the functional form of the distance function. The asymptotic theory developed in this paper differs from existing work on nonparametric estimation and specification testing in the stationary time series case.

In order to implement the proposed kernel test in practice, we have developed a computer–intensive bootstrap simulation procedure to select a suitable bandwidth for optimal testing purposes. We have also established some higher–order asymptotic properties for the bootstrap version of the proposed test as well as both the size and the power functions. The proposed theory and methodology has been illustrated using both simulated and real data examples.

The finite sample studies have shown that the proposed estimation and testing methods work well numerically.

7. Acknowledgments

This work is part of the ongoing project "Estimation and Specification Testing in Nonlinear Time Series Econometrics" being supported financially by two Australian Research Council Discovery Grants under Grant Numbers: DP0558602 and DP1096374. The authors acknowledge useful comments from various scholars in the field.

Appendix A

In this section, we provide some basic definitions for regular functions and necessary assumptions to establish the asymptotic theory in Sections 2–4. As introduced by Park and Phillips (2001), we consider two classes of parametric nonlinear regression functions: I–regular on Θ and H_0–regular on Θ.

Definition A.1. *We say that $m(\cdot,\cdot)$ 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.

Conditions (i) and (ii) are called regularity conditions. We then introduce two classes of nonlinear regression functions: I–regular on Θ and H_0–regular on Θ.

Definition A.2. *$m(\cdot,\cdot)$ 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 N_0 of θ and $M:\ \mathcal{R} \to \mathcal{R}$ bounded integrable such that $\|m(x,\theta') - m(x,\theta)\| \leq \|\theta' - \theta\| M(x)$ for all $\theta' \in N_0$;

(ii) for some $C > 0$ and $k > 6/(p-2)$ with $p > 4$ defined as in A1 (ii), $\|m(x,\theta) - m(y,\theta)\| \leq C|x-y|^k$ for all $\theta \in \Theta$, on each piece S_i of their common support $S = \cup_{i=1}^m S_i \subset \mathcal{R}$.

Definition A.3. *Let $m(\lambda x,\theta) = \kappa(\lambda,\theta)H(x,\theta) + R(x,\lambda,\theta)$, where κ is nonsingular. $m(\cdot,\cdot)$ is said to be H–regular on Θ if the following two conditions are satisfied:*

(i) $H(\cdot,\cdot)$ is regular on Θ;

(ii) $R(x,\lambda,\theta)$ is of order smaller than $\kappa(\lambda,\theta)$ as $\lambda \to \infty$ for all $\theta \in \Theta$.

Note that *κ is said to be the asymptotic order of $m(\cdot,\cdot)$ and $H(\cdot,\cdot)$ is the limit homogeneous function.* When κ does not depend upon the parameter θ, then $m(\cdot,\cdot)$ is said to be *H_0–regular.*

Throughout the paper, we only consider the case that κ is independent of the parameter θ. For detailed discussion about such function families, we refer to Park and Phillips (2001).

To establish and then prove the main results in Sections 2–4, we need to introduce the following assumptions.

A1 **(i)** $K(\cdot)$ is a continuous and symmetrical probability kernel function with compact support.

(ii) $\{v_t, t \geq 1\}$ is a sequence of i.i.d. random variables with $E[v_1] = 0$, $E[v_1^2] = 1$ and $E\left[|v_1|^p\right] < \infty$ for some $p > 4$. Furthermore, the characteristic function $\psi(\cdot)$ of $\{v_t\}$ satisfies $\int_{-\infty}^{\infty} |\psi(v)|dv < \infty$.

(iii) $\{e_t\}$ is a sequence of stationary martingale differences such that

$$E[e_t|\mathcal{B}_{t-1}] = 0, \quad E[e_t^2|\mathcal{B}_{t-1}] = \sigma_e^2, \quad E[e_t^3|\mathcal{B}_{t-1}] = 0, \quad 0 < E[e_t^4|\mathcal{B}_{t-1}] < \infty, \quad a.s.,$$

where $\mathcal{B}_t$ is the σ–field generated by $\{e_s,\ s \leq t\}$.

(iv) The errors $\{e_t\}$ and $\{v_t\}$ are assumed to be mutually independent.

A2 **(i)** As $n \to \infty$, $b \to 0$, $\sqrt{\sqrt{n}b}\ \delta_n \to \infty$, $\sqrt{n}\dot{\kappa}^2(\sqrt{n}) \to \infty$ and $\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}) \to \infty$.

(ii) As $n \to \infty$, $b \to 0$, $\sqrt{\sqrt{n}b}\ \delta_n \to \infty$, $v(\sqrt{n})(\dot{\kappa}(\sqrt{n}))^{-1} \to 0$ and $\delta_n\sqrt{n}\dot{\kappa}(\sqrt{n}) \to \infty$.

(iii) For some $0 < \varepsilon_0 < 1/2$, we have $n^{\varepsilon_0}h \to 0$ and $n^{1/2-\varepsilon_0}h \to \infty$ as $n \to \infty$.

A3 **(i)** The nonlinear regression function $g(\cdot, \theta)$ is I–regular on Θ. Let

$$\int_{-\infty}^{\infty} (g(v, \theta) - g(v, \theta_0))^2 dv > 0 \quad \text{for all } \theta \neq \theta_0.$$

(ii) Both $\dot{g}(\cdot, \theta)$ and $\ddot{g}(\cdot, \theta)$ are I–regular on Θ, where $\dot{g}(\cdot, \theta) = \left(\frac{\partial g}{\partial \theta}\right)$ and $\ddot{g}(\cdot, \theta) = \left(\frac{\partial^2 g}{\partial \theta \partial \theta'}\right)$. Furthermore, $\int \dot{g}(v, \theta_0)\dot{g}(v, \theta_0)'dv$ is some positive definite matrix.

A3′ **(i)** The nonlinear regression function $g(\cdot, \theta)$ is H_0–regular on Θ with asymptotic order $\kappa(\cdot)$ and limit homogeneous function $h(\cdot, \cdot)$. And $\kappa(\lambda)$ is bounded away from zero as $\lambda \to \infty$. Furthermore, $\int_{|v| \leq \varepsilon_0} (h(v, \theta) - h(v, \theta_0))^2 dv > 0$ for all $\theta \neq \theta_0$ and $\varepsilon_0 > 0$.

(ii) Both $\dot{g}(\cdot, \theta)$ and $\ddot{g}(\cdot, \theta)$ are H_0–regular on Θ with asymptotic order $\dot{\kappa}(\cdot)$ and $\ddot{\kappa}(\cdot)$. Furthermore, $\|(\dot{\kappa} \otimes \dot{\kappa})^{-1}\kappa\ddot{\kappa}\| < \infty$ and $\int_{|v| \leq \varepsilon_1} \dot{h}(v, \theta_0)\dot{h}(v, \theta_0)'dv$ is some positive definite matrix for some $\varepsilon_1 > 0$, where $\dot{h}(v, \theta)$ is the limit homogeneous function of $\dot{g}(\cdot, \theta)$ and $\otimes$ is the Kronecker product. In addition, $\dot{\kappa}(\lambda)$ is bounded away from zero as $\lambda \to \infty$.

The above assumptions are quite mild and justifiable. A1(i) is a quite standard condition. If the characteristic function $\psi(v)$ of $\{v_t\}$ satisfies $\int |\psi(v)|dv < \infty$ as in A1 (ii), by the standard convergence results in Chow and Teicher (1988), we have $\sup_{v \in \mathcal{R}} |\phi_t(v) - \phi(v)| = o(1)$ as $t \to \infty$, where $\phi(\cdot)$ is the density function of the standard normal distribution and $\phi_t(\cdot)$ is the density function of $\frac{V_t}{\sqrt{t}}$. This implies that $\phi_t(\cdot)$ can be replaced by $\phi(\cdot)$ when t is large enough. The conditions on the errors in A1(iii) are similar to those used in Gao *et al* (2009b). The independence assumption between $\{e_t\}$ and $\{v_t\}$ in A1(iv) may be somewhat restrictive, but this is what we need to evaluate the orders of various moments for quadratic forms before we can complete the proofs of the main results. It is not clear at the moment whether it is possible to relax this to a set of weak conditions as in Park and Phillips (1999, 2001). A2(i) and A2(ii) are both satisfied when $b = O\left((\sqrt{n}\delta_n^2)^{-\frac{1}{5}}\right)$ and the functional form of $g(v, \theta)$ belongs to a family of either polynomial functions, or trigonometric functions or logarithm functions (see, for example, Park and Phillips 1999, 2001). Condition A2(iii) on the bandwidth corresponds to $h \to 0$ and $nh \to \infty$ as $n \to \infty$ in the stationary time series case.

Conditions A3 and A3$'$ impose some assumptions on the smoothness and functional form of $g(v, \theta_0)$ such that $\widehat{\theta}$ is a consistent estimator of θ_0 under the null hypothesis. Such conditions were initially introduced by Park and Phillips (2001) when they established asymptotic theory for their nonlinear least squares estimator in the parametric nonlinear regression model with integrated time series. It is easy to check that A3 is satisfied for all nonzero I–regular linear–in–parameter regression function ($g(v, \theta) = g(v)\theta$) and it is also satisfied for $g(v, \alpha_0, \alpha_1) = \alpha_0 \exp\{-\alpha_1 v^2\}$ with $\theta = (\alpha_0, \alpha_1) \in \Theta$. Meanwhile, A3$'$ is satisfied for all nonzero H_0–regular linear–in–parameter regression function and it is also satisfied for $g(v, \theta) = \frac{v}{1+\theta v} I_{\{v \geq 0\}}$.

Appendix B

Before we prove the main results in the second part of this appendix, we introduce several lemmas with their proofs being given in Appendix C in the supplementary document. Let $T(n)$ be the regeneration times for the random walk process $\{V_t\}$ defined by (1.2). We first provide an asymptotic rate of $T(n)$ in probability. A detailed definition similar to that given in Karlsen and Tjøstheim (2001) is provided in the first part of Appendix C.

Lemma B.1. *Assume that A1(ii) and A2(iii) hold. Then there are some constants* $0 < C_1 < C_2 < \infty$ *such that*

$$\lim_{n \to \infty} P\left(C_1 < \frac{T(n)}{\sqrt{n}} \leq C_2\right) = 1. \tag{B.1}$$

Park and Phillips (2001) studied the asymptotic properties for the estimator $\widehat{\theta}$ under the null hypothesis. The following lemma establishes rates of convergence for $\widehat{\theta}_1$ under H_1.

Lemma B.2. *Assume that A1(ii)(iii)(iv) holds. Let* $\widehat{\theta}_1 = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} (Y_t - g(V_t, \theta))^2$ *be the nonlinear least squares estimator of* θ_1 *under* H_1.

(i) If A3 is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–integrable, we have

$$\widehat{\theta}_1 - \theta_1 = O_P\left(\delta_n + (\sqrt[4]{n})^{-1}\right). \tag{B.2}$$

(ii) If A3′ is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–integrable, we have

$$\widehat{\theta}_1 - \theta_1 = O_P\left(\delta_n(\sqrt[4]{n}\dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1}\right). \tag{B.3}$$

(iii) If A3′ is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–asymptotically homogeneous with order $v(\cdot)$, we have

$$\widehat{\theta}_1 - \theta_1 = O_P\left(\delta_n v(\sqrt{n})(\dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1}\right). \tag{B.4}$$

LEMMA B.3. *Suppose that A1(ii)(iv) holds and that $\{e_t\}$ is a sequence of i.i.d. continuous random variables with $E[e_1] = 0$, $E[e_1^2] = \sigma_e^2$ and $E[e_1^6] < \infty$.*

Let $\widehat{\theta}_1^* = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} (Y_t^* - g(V_t,\theta))^2$ be the bootstrap nonlinear least squares estimator of θ_1 under H_1.

(i) If A3 is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–integrable, we have

$$\widehat{\theta}_1^* - \theta_1 = O_P\left(\delta_n + (\sqrt[4]{n})^{-1}\right). \tag{B.5}$$

(ii) If A3′ is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–integrable, we have

$$\widehat{\theta}_1^* - \theta_1 = O_P\left(\delta_n(\sqrt[4]{n}\dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1}\right). \tag{B.6}$$

(iii) If A3′ is satisfied when θ_0 is replaced by θ_1 and $\Delta_n(\cdot)$ is δ_n–asymptotically homogeneous with order $v(\cdot)$, we have

$$\widehat{\theta}_1^* - \theta_1 = O_P\left(\delta_n v(\sqrt{n})(\dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n}\dot{\kappa}(\sqrt{n}))^{-1}\right). \tag{B.7}$$

REMARK B.1. (i) To justify the suitability of the definitions of $\widehat{\theta}_1$ and $\widehat{\theta}_1^*$, we provide the following discussion. Let us just focus on the definition of $\widehat{\theta}_1$. Similarly to the nonlinear least squares estimation in the stationary case, in order to choose $\widehat{\theta}_1$ such that $\sum_{t=1}^{n} (Y_t - g(V_t,\theta_1))^2$ is minimized over all θ_1, it suffices to choose such θ_1 such that

$$\frac{1}{d_n}\sum_{t=1}^{n} (Y_t - g(V_t,\theta_1))\,\frac{\partial g(V_t,\theta_1)}{\partial\theta_1} = o_P(1), \tag{B.8}$$

where d_n is a sequence of positive real numbers such that $d_n \to \infty$. Note that equation (B.8) is equivalent to requiring $E\left[(Y_t - g(V_t,\theta_1))\,\frac{\partial g(V_t,\theta_1)}{\partial\theta_1}\right] = 0$ in the case where both $\{Y_t\}$ and $\{V_t\}$ are stationary.

Equation (B.8) follows from

$$\frac{1}{d_n}\sum_{t=1}^n e_t \frac{\partial g(V_t,\theta_1)}{\partial \theta_1} = o_P(1) \quad \text{and} \quad \frac{1}{d_n}\sum_{t=1}^n \Delta_n(V_t)\frac{\partial g(V_t,\theta_1)}{\partial \theta_1} = o_P(1), \tag{B.9}$$

which follow from the proof of Lemma B.2 in the cases of (i) with $d_n = \sqrt{n}$, (ii) with $d_n = n^{3/4}\dot{\kappa}(\sqrt{n})$ and (iii) with $d_n = n\dot{\kappa}(\sqrt{n})v(\sqrt{n})$.

(ii) One alternative estimation method for θ_1 is to choose $\widetilde{\theta}_1$ such that

$$\begin{aligned} \widetilde{\theta}_1 &= \arg \min_{\text{over all } \theta_1} \sum_{t=1}^n \left(Y_t - g(V_t,\theta_1) - \widetilde{\Delta}_n(V_t,\theta_1)\right)^2 \\ &= \arg \min_{\text{over all } \theta_1} \sum_{t=1}^n \left(\widetilde{Y}_t - \widetilde{g}(V_t,\theta_1)\right)^2 \end{aligned} \tag{B.10}$$

where $\widetilde{\Delta}_n(v,\theta_1) = \sum_{s=1}^n \widetilde{w}_{ns}(v)\left(Y_s - g(V_s,\theta_1)\right)$, $\widetilde{Y}_t = Y_t - \sum_{s=1}^n \widetilde{w}_{ns}(V_t) Y_s$ and $\widetilde{g}(V_t,\theta_1) = g(V_t,\theta_1) - \sum_{s=1}^n \widetilde{w}_{ns}(V_t) g(V_s,\theta_1)$, in which $\widetilde{w}_{ns}(v)$ is as defined in (2.4).

Due to the local linear method, similarly to the proof of Theorem 2.1, one may show that as $n \to \infty$

$$\widetilde{g}(V_t,\theta_1) = (1+o_P(1))\ c' g_{20}(V_t,\theta_1)\ b^2 \ \text{ and } \ \widetilde{\Delta}_n(V_t) = (1+o_P(1))\ \dot{\Delta}_n(V_t)\ b^2, \tag{B.11}$$

where $g_{20}(v,\theta_1) = \frac{\partial^2 g(v,\theta_1)}{\partial v^2}$, c is a constant vector and $\dot{\Delta}_n(v) = \frac{d\Delta_n(v)}{dv}$.

As a result, the proof of Lemma B.2 implies that as $n \to \infty$

$$\begin{aligned} b^2\left(\widetilde{\theta}_1 - \theta_1\right) &= c\ (1+o_P(1)) \left(\sum_{t=1}^n \dot{g}_{20}(V_t,\theta_1)\dot{g}_{20}(V_t,\theta_1)'\right)^{-1} \sum_{t=1}^n \dot{g}_{20}(V_t,\theta_1)\ e_t \\ &+ c\ b^2(1+o_P(1)) \left(\sum_{t=1}^n \dot{g}_{20}(V_t,\theta_1)\dot{g}_{20}(V_t,\theta_1)'\right)^{-1} \sum_{t=1}^n \dot{g}_{20}(V_t,\theta_1)\dot{\Delta}_n(V_t), \end{aligned} \tag{B.12}$$

where c is some constant. This implies that the rate of convergence of $b^2\left(\widetilde{\theta}_1 - \theta_1\right)$ is only proportional to the corresponding rate of $\widehat{\theta}_1 - \theta_1$ given on the right–hand side of each of the equations (B.2), (B.3) and (B.4). Therefore, this shows that $\widetilde{\theta}_1$ has a rate slower than that for $\widehat{\theta}_1$, because of $b \to 0$. This is the main reason we propose using $\widehat{\theta}_1$ rather than $\widetilde{\theta}_1$ in this paper.

(iii) Another alternative method is to construct an instrumental–variable (IV) based consistent estimator for θ_1. This is based on the assumption and existence of an IV of the form $\{\Gamma_n(Z_t,\theta_1)\}$ such that

$$\sum_{t=1}^n \Delta_n(V_t)\ \Gamma_n(Z_t,V_1) = 0 \quad \text{and} \quad \sum_{t=1}^n e_t\ \Gamma_n(Z_t,\theta_1) = 0. \tag{B.13}$$

This paper basically employes an asymptotic version of (B.13) in (B.9) with $\Gamma_n(Z_t,\theta_1) = \frac{\partial g(V_t,\theta_1)}{\partial \theta_1}$. Further discussion on this issue is left for future research. This therefore completes the discussion in Remark B.1.

Let

$$\widehat{f}_n(v,Q,l)=\frac{1}{T(n)}\sum_{t=1}^n Q_{v,h,l}(V_t)\equiv\frac{1}{T(n)}\sum_{t=1}^n\left(\frac{V_t-v}{h}\right)^l K_{v,h}(V_t),$$

where $Q_{v,h,l}(V_t)=\left(\frac{V_t-v}{h}\right)^l K_{v,h}(V_t)$ for $0\le l\le k_0$ with k_0 being some positive integer and $K_{v,h}(V_t)=\frac{1}{h}K\left(\frac{V_t-v}{h}\right)$.

LEMMA B.4. *Let A1(i)(ii) and A2(iii) hold. Then as $n\to\infty$*

$$\left|\widehat{f}_n(v,Q,l)-\int u^l K(u)du\right|=o(1)\quad a.s.,\tag{B.14}$$

uniformly for $|v|\le M\sqrt{n}$ and $0\le l\le k_0$, where M is any given positive constant.

REMARK B.2. Analogously to the proof of Lemma B.4, we conclude that as $n\to\infty$

$$\widehat{f}_n(v,Q,l)\to\int u^l K(u)du\tag{B.15}$$

almost surely for each given v. Similar results for point–wise convergence have been given in Karlsen and Tjøstheim (2001), and Wang and Phillips (2009) for example. Lemma B.4 is concerned with the uniform convergence. Some more general discussion is given in Gao, Li and Tjøstheim (2009).

LEMMA B.5. *Let $Q_{n,1}^*(h)=\sum_{t=1}^n\sum_{s\neq t}e_t^*K_{s,t}e_s^*$, where e_t^* is defined in Section 4. Assume that the conditions of Proposition 4.1 are all satisfied. Then*

$$\sup_{x\in\mathcal{R}}\left|P^*\left(\frac{Q_{n,1}^*(h)}{\overline{\sigma}_n^*}\le x\right)-\Phi(x)+\rho_n^*(h)\ \Phi^{(3)}(x)\right|=O_P(n^{-1/2}),\tag{B.16}$$

where $(\overline{\sigma}_n^*)^2=2\overline{\sigma}_{e,*}^4\sum_{t=1}^n\sum_{s=1,\neq t}^n K_{s,t}^2$ and $\overline{\sigma}_{e,*}^2=\frac{1}{n}\sum_{t=1}^n(\widehat{e}_t^*)^2$.

With the help of the useful lemmas, we now give the proofs of the main results.

PROOF OF THEOREM 2.1. Observe that under H_1,

$$\begin{aligned}\widehat{\Delta}_n(v)&=\sum_{t=1}^n\widetilde{w}_{nt}(v)(Y_t-g(V_t,\widehat{\theta}_1))\\&=\sum_{t=1}^n\widetilde{w}_{nt}(v)e_t+\sum_{t=1}^n\widetilde{w}_{nt}(v)\Delta_n(V_t)+\sum_{t=1}^n\widetilde{w}_{nt}(v)(g(V_t,\theta_1)-g(V_t,\widehat{\theta}_1)),\end{aligned}$$

which implies that

$$\begin{aligned}\widehat{\Delta}_n(v)-\Delta_n(v)&=\sum_{t=1}^n\widetilde{w}_{nt}(v)e_t+\left(\sum_{t=1}^n\widetilde{w}_{nt}(v)\Delta_n(V_t)-\Delta_n(v)\right)\\&\quad+\sum_{t=1}^n\widetilde{w}_{nt}(v)(g(V_t,\theta_1)-g(V_t,\widehat{\theta}_1))\\&=:J_{n1}(v)+J_{n2}(v)+J_{n3}(v).\end{aligned}\tag{B.17}$$

By the central limit theorem for martingale differences (cf. Hall and Heyde, 1980), we have as $n \to \infty$

$$\sum_{t=1}^{n} \left(\frac{\widetilde{w}_{nt}(v)}{\sqrt{\sum_{s=1}^{n} \widetilde{w}_{ns}^2(v)}} \right) e_t \xrightarrow{d} N\left(0, \sigma_e^2\right), \tag{B.18}$$

in which the CLT is applied to the case where $\{e_t\}$ is a sequence of martingale differences and $\sum_{t=1}^{n} b_{nt}(v) e_t$ is a partial sum of martingale differences given $(V_1, \cdots, V_n)$, where $b_{nt}(v) = \frac{\widetilde{w}_{nt}(v)}{\sqrt{\sum_{s=1}^{n} \widetilde{w}_{ns}^2(v)}}$ satisfies $\sum_{t=1}^{n} b_{nt}^2(v) = 1$ and equation (B.19) below.

By Remark B.2 above, we have as $n \to \infty$

$$\begin{aligned} T(n)b \left(\sum_{s=1}^{n} \widetilde{w}_{ns}^2(v) \right) &= \frac{\frac{1}{T(n)b} \sum_{s=1}^{n} \widetilde{K}_n^2\left(\frac{V_s - v}{b}\right)}{\left(\frac{1}{T(n)b} \sum_{s=1}^{n} \widetilde{K}_n\left(\frac{V_s - v}{b}\right) \right)^2} \\ &\xrightarrow{P} \frac{\int K^2(v)dv \left(\int v^2 K(v)dv\right)^2}{\left(\int v^2 K(v)dv\right)^2} \\ &= \int K^2(v)dv. \end{aligned} \tag{B.19}$$

Meanwhile, by the definition of local linear estimator, we have

$$J_{n2}(v) = \frac{1}{2} \ddot{\Delta}_n(v) b^2 (1 + o_P(1)) = O_P(\delta_n b^2). \tag{B.20}$$

By Lemma B.2(ii), we have

$$J_{n3}(v) = (\theta_1 - \widehat{\theta}_1)' \dot{g}(v, \theta_1)(1 + o_P(1)) = O_P(\delta_n (\sqrt[4]{n} \dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n} \dot{\kappa}(\sqrt{n}))^{-1}). \tag{B.21}$$

By equations (B.17)–(B.21), we have shown that (4.3) holds. Hence, the proof of Theorem 2.1(i) is completed. The proof of Theorem 2.1(ii) follows from equations (B.17)–(B.20) and

$$J_{n3}(v) = (\theta_1 - \widehat{\theta}_1)' \dot{g}(v, \theta_1)(1 + o_P(1)) = O_P(\delta_n v(\sqrt{n})(\dot{\kappa}(\sqrt{n}))^{-1} + (\sqrt{n} \dot{\kappa}(\sqrt{n}))^{-1}), \tag{B.22}$$

which follows from Lemma B.2(iii).

Proof of Proposition 3.1: While the main steps of the proof are similar to those for the proof of Theorem 2.1 of Gao *et al* (2009b), we still need to prove the corresponding parts under the conditions of Proposition 3.1. We also only give the proof under A3 since the proof is similar when A3$'$ is satisfied. Recall that under H_0

$$\begin{aligned} Q_n(h) &= \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} \widehat{e}_t K_{s,t} \widehat{e}_s = \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} e_t K_{s,t} e_s \\ &+ \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} \overline{g}_t K_{s,t} \overline{g}_s + 2 \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} \overline{g}_t K_{s,t} e_s =: \sum_{i=1}^{3} Q_{n,i}(h). \end{aligned}$$

Similarly to the proof of Theorem 2.1 in Gao *et al* (2009b), we have

$$\frac{Q_{n,1}(h)}{\widetilde{\sigma}_n} \stackrel{d}{\longrightarrow} N(0,1) \quad \text{and} \quad \frac{\overline{\sigma}_n^2 - \widetilde{\sigma}_n^2}{\widetilde{\sigma}_n^2} = o_P(1), \tag{B.23}$$

where $\widetilde{\sigma}_n^2 = 2\widetilde{\sigma}_e^4 \sum_{t=1}^n \sum_{s=1,\neq t}^n K_{s,t}^2$ and $\widetilde{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^n e_t^2$.

By (B.23), it suffices to show that

$$\frac{Q_{n,i}(h)}{\widetilde{\sigma}_n} = o_P(1), \quad i = 2, 3. \tag{B.24}$$

As in the proof of Theorem 2.1 in Gao *et al* (2009b), we have

$$\widetilde{\sigma}_n^2 = C_0 T(n) nh (1 + o_P(1)) \ \text{ with } \ C_0 = 2\sigma_e^4 \int K^2(v) dv. \tag{B.25}$$

By (B.25) and Lemma B.1, we have for large enough n

$$C_1 \ n^{\frac{3}{2}} h (1 + o_P(1)) \leq \widetilde{\sigma}_n^2 \leq C_2 \ n^{\frac{3}{2}} h (1 + o_P(1)) \tag{B.26}$$

for some constants $0 < C_1 < C_2 < \infty$.

To prove (B.24), in view of (B.26), it suffices to show that

$$\frac{Q_{n,2}(h)}{\widetilde{\sigma}_n} = O_P\left(n^{-1/4}\sqrt{h}\right) \quad \text{and} \quad \frac{Q_{n,3}(h)}{\widetilde{\sigma}_n} = O_P\left(n^{-1/8} h^{1/4}\right). \tag{B.27}$$

To prove the first part of (B.27), we first need to deal with the following term:

$$(\widehat{\theta} - \theta_0)' \sum_{t=1}^n \sum_{s=1,\neq t}^n \dot{g}(V_t, \theta_0) K_{s,t} \dot{g}(V_s, \theta_0)' (\widehat{\theta} - \theta_0). \tag{B.28}$$

A Taylor expansion implies for some $0 < \vartheta_s < 1$ and all $s = 1, \cdots, n$,

$$\begin{aligned} & \sum_{t=1}^n \sum_{s=1,\neq t}^n \dot{g}(V_t, \theta_0) K_{s,t} \dot{g}(V_s, \theta_0)' \\ = & \sum_{t=1}^n \dot{g}(V_t, \theta_0) \dot{g}(V_t, \theta_0)' \sum_{s=1,\neq t}^n K_{s,t} + \sum_{t=1}^n \dot{g}(V_t, \theta_0) \sum_{s=1,\neq t}^n K_{s,t} \\ \times & \ (V_t - V_s) \ddot{g}(V_s + \vartheta_s (V_t - V_s), \theta_0) =: I_{n,1} + I_{n,2}. \end{aligned}$$

By a standard argument, we have

$$\begin{aligned} & \max_{1\leq t\leq n} \left\{ \frac{1}{T(n)h} \sum_{s=1,\neq t}^n K\left(\frac{V_t - V_s}{h}\right) \right\} = \max_{1\leq t\leq n} \left\{ \frac{1}{T(n)h} \sum_{s=1}^n K\left(\frac{V_t - V_s}{h}\right) + \frac{K(0)}{T(n)h} \right\} \\ = & \max_{1\leq t\leq n} \left\{ \frac{1}{T(n)h} \sum_{s=1}^n K\left(\frac{V_t - V_s}{h}\right) \right\} + o_P(1). \end{aligned} \tag{B.29}$$

By Lemma B.4, we have for any given $\varepsilon > 0$

$$P\left(\left\{ \max_{1\leq t\leq n} \left| \frac{1}{T(n)h} \sum_{s=1}^n K\left(\frac{V_t - V_s}{h}\right) - 1 \right| > \varepsilon \right\} \cap \left\{ \max_{1\leq t\leq n} |V_t| \leq M_0 \sqrt{n} \right\} \right) = o(1). \tag{B.30}$$

Meanwhile, for any given small $\varepsilon > 0$, applying the Kolmogorov inequality, we have

$$P\left(\max_{1\le t\le n} |V_t| > M_0\sqrt{n}\right) \le \frac{\mathrm{Var}(V_n)}{M_0^2 n} = \frac{1}{M_0^2} < \varepsilon \tag{B.31}$$

by letting $M_0 > \sqrt{\frac{1}{\varepsilon}}$.

Equations (B.29), (B.30) and (B.31) imply that for the given small $\varepsilon > 0$

$$\begin{aligned}
& P\left(\max_{1\le t\le n}\left|\frac{1}{T(n)h}\sum_{s=1}^{n} K\left(\frac{V_t - V_s}{h}\right) - 1\right| > \varepsilon\right) \qquad\qquad (B.32)\\
&= P\left(\left\{\max_{1\le t\le n}\left|\frac{1}{T(n)h}\sum_{s=1}^{n} K\left(\frac{V_t - V_s}{h}\right) - 1\right| > \varepsilon\right\} \cap \left\{\max_{1\le t\le n} |V_t| \le M_0\sqrt{n}\right\}\right)\\
&+ P\left(\left\{\max_{1\le t\le n}\left|\frac{1}{T(n)h}\sum_{s=1}^{n} K\left(\frac{V_t - V_s}{h}\right) - 1\right| > \varepsilon\right\} \cap \left\{\max_{1\le t\le n} |V_t| > M_0\sqrt{n}\right\}\right)\\
&\le P\left(\left\{\max_{1\le t\le n}\left|\frac{1}{T(n)h}\sum_{s=1}^{n} K\left(\frac{V_t - V_s}{h}\right) - 1\right| > \varepsilon\right\} \cap \left\{\max_{1\le t\le n} |V_t| \le M_0\sqrt{n}\right\}\right)\\
&+ P\left(\max_{1\le t\le n} |V_t| > M_0\sqrt{n}\right) = o(1).
\end{aligned}$$

By (B.29)–(B.32), we thus have

$$I_{n,1} = T(n)h(1 + o_P(1))\sum_{t=1}^{n} \dot{g}(V_t, \theta_0)\dot{g}(V_t, \theta_0)'. \tag{B.33}$$

Furthermore, when A3 is satisfied, following the proof of Lemma B.2, we have

$$(\widehat{\theta} - \theta_0)'\sum_{t=1}^{n} \dot{g}(V_t, \theta_0)\dot{g}(V_t, \theta_0)'(\widehat{\theta} - \theta_0) = O_P(1). \tag{B.34}$$

Similarly, we can show that $I_{n,2}$ is of an order smaller than $I_{n,1}$. Hence, by (B.33), (B.34) and Lemma B.1, we have shown that

$$\frac{Q_{n,2}(h)}{\widetilde{\sigma}_n} = O_P\left(\frac{\sqrt{n}h}{n^{3/4}\sqrt{h}}\right) = O_P\left(n^{-1/4}\sqrt{h}\right), \tag{B.35}$$

which shows that the first part of (B.27) holds. Meanwhile, note that

$$\begin{aligned}
Q_{n,3}^2(h) &= 4\left(\sum_{t=1}^{n}\sum_{s=1,\ne t}^{n} \overline{g}_t\sqrt{K_{s,t}}\sqrt{K_{s,t}}e_s\right)^2\\
&\le 4\left(\sum_{t=1}^{n}\sum_{s=1,\ne t}^{n} \overline{g}_t K_{s,t}\overline{g}_s\right)\left(\sum_{t=1}^{n}\sum_{s=1,\ne t}^{n} e_t K_{s,t} e_s\right) \qquad (B.36)\\
&= O\left(Q_{n,1}(h)Q_{n,2}(h)\right) = O_P\left(\sqrt{nh}n^{3/4}h\right),
\end{aligned}$$

which implies the second part of (B.27). The proof of Proposition 3.1 is therefore completed.

Proof of Theorem 3.1(i): Note that under H_1, we have

$$\begin{aligned} Q_n(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widehat{e}_t K_{s,t}\widehat{e}_s \\ &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} e_t K_{s,t} e_s + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t}\widetilde{g}_s + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widetilde{g}_t K_{s,t} e_s \\ &+ 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t}\widetilde{g}_s + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t} e_s \\ &+ \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \Delta_n(V_t) K_{s,t}\Delta_n(V_s) =: \sum_{i=1}^{6} \overline{Q}_{n,i}(h). \end{aligned} \tag{B.37}$$

where $\widetilde{g}_t = g(V_t,\theta_1) - g(V_t,\widehat{\theta}_1)$.

Following the proof of Proposition 3.1 as above, we have

$$\frac{1}{\overline{\sigma}_n}\overline{Q}_{n,1}(h) \stackrel{d}{\longrightarrow} N(0,1). \tag{B.38}$$

When A3 is satisfied with θ_0 being replaced by θ_1, by the proof of Lemma B.2(i), we have

$$\begin{aligned} & (\widehat{\theta}_1-\theta_1)'\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\dot{g}(V_t,\theta_1)'(\widehat{\theta}_1-\theta_1) \\ \leq\ & O_P\left((\delta_n^2 + (\sqrt{n})^{-1})\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\dot{g}(V_t,\theta_1)'\right) \\ =\ & O_P\left(\delta_n^2\sqrt{n}+1\right). \end{aligned}$$

Letting $\overline{I}_{n,1} = \sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\dot{g}(V_t,\theta_1)'\sum_{s=1,\neq t}^{n} K_{s,t}$, we then have

$$\overline{I}_{n,1} = O_P(n\delta_n^2 h + \sqrt{n}h) = O_P\left(n^{3/4}\sqrt{h}\right), \tag{B.39}$$

by the condition of $\delta_n = O\left(n^{-1/8}h^{-1/4}\right)$.

Hence, we have

$$\overline{Q}_{n,2}(h) = O_P(\overline{\sigma}_n). \tag{B.40}$$

By Definition 2.1, we can find an integrable function $\overline{\Gamma}(\cdot)$ such that

$$\max\{\Delta_n^2(v), \left|\Delta_n(v)\dot{\Delta}_n(v)\right|\} \leq \delta_n^2\overline{\Gamma}(v). \tag{B.41}$$

Note that

$$\begin{aligned} \overline{Q}_{n,4}(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \left(\Delta_n^2(V_s)K_{s,t} + \left(\Delta_n(V_t)-\Delta_n(V_s)\right)K_{s,t}\Delta_n(V_s)\right) \\ &=: \overline{Q}_{n,4,1}(h) + \overline{Q}_{n,4,2}(h). \end{aligned} \tag{B.42}$$

By the condition of $\delta_n = O\left(n^{-1/8}h^{-1/4}\right)$, (B.41), Lemma B.1, Lemma B.4, Theorem 5.1 in Park and Phillips (1999) and following the calculation of $I_{n,1}$,

$$\begin{aligned}
\overline{Q}_{n,4,1}(h) &= \sum_{s=1}^{n} \Delta_n^2(V_s) \sum_{t=1,\neq s}^{n} K_{s,t} \\
&= T(n)h \sum_{s=1}^{n} \Delta_n^2(V_s) \left(\frac{1}{T(n)h} \sum_{t=1}^{n} K_{s,t} - \frac{K(0)}{T(n)h} \right) \\
&\leq (T(n)\sqrt{n}h\delta_n^2) \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \overline{\Gamma}(V_s)\,(1 + o_P(1)) \\
&= O_P(nh\delta_n^2) = O_P(n^{3/4}h^{1/2}). \qquad \text{(B.43)}
\end{aligned}$$

Meanwhile, note that $\dot{\Delta}_n(\cdot)$ is also δ_n–integrable. By Taylor's expansion, we can show that

$$\begin{aligned}
\left|\overline{Q}_{n,4,2}(h)\right| &\leq \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \left|\Delta_n(V_t) - \Delta_n(V_s)\right| K_{s,t} \left|\Delta_n(V_s)\right| \\
&\overset{P}{\sim} T(n)h^2 \sum_{s=1}^{n} \left|\dot{\Delta}_n(V_s)\Delta_n(V_s)\right| \left(\frac{1}{T(n)h} \sum_{t=1,\neq s}^{n} \left|\frac{V_t - V_s}{h}\right| K_{s,t} \right) \\
&\leq \left(T(n)\sqrt{n}h^2\delta_n^2 \int |u|K(u)du \right) \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \overline{\Gamma}(V_s)\,(1 + o_P(1)) \\
&= O_P(nh^2\delta_n^2) = o_P(n^{3/4}h^{1/2}), \qquad \text{(B.44)}
\end{aligned}$$

where $a_n \overset{P}{\sim} b_n$ means that $\frac{a_n}{b_n} = 1 + o_P(1)$.

Thus, we have

$$\overline{Q}_{n,4}(h) = O_P(\overline{\sigma}_n). \qquad \text{(B.45)}$$

Similarly to the proof of (B.36), we can also show that

$$\overline{Q}_{n,i}(h) = O_P(\overline{\sigma}_n), \quad i = 3, 5, 6. \qquad \text{(B.46)}$$

When A3$'$ is satisfied for θ_1, the proof is analogous with the help of Lemma B.2(ii). Therefore, the proof of Theorem 3.1(i) is completed.

To complete the proof for the case of $\delta_n = o\left(n^{-1/8}h^{-1/4}\right)$, in view of the proof for the case of $\delta_n = O\left(n^{-1/8}h^{-1/4}\right)$above, it suffices to verify

$$\begin{aligned}
\overline{Q}_{n,2}(h) &= o_P(\overline{\sigma}_n), \qquad \text{(B.47)} \\
\overline{Q}_{n,4,1}(h) &= O_P(nh\delta_n^2) = o_P(n^{3/4}h^{1/2}). \qquad \text{(B.48)} \\
\left|\overline{Q}_{n,4,2}(h)\right| &= O_P(nh^2\delta_n^2) = o_P(n^{3/4}h^{1/2}), \qquad \text{(B.49)} \\
\overline{Q}_{n,4}(h) &= o_P(\overline{\sigma}_n). \qquad \text{(B.50)} \\
\overline{Q}_{n,i}(h) &= o_P(\overline{\sigma}_n), \quad i = 3, 5, 6, \qquad \text{(B.51)}
\end{aligned}$$

which follows from (B.40) and (B.43)–(B.46), respectively, when $\delta_n = o\left(n^{-1/8}h^{-1/4}\right)$. The proof of Theorem 3.1(i) is therefore completed.

Proof of Theorem 3.1(ii): Following the proof of Theorem 3.1(i), in order to prove Theorem 3.1(ii), we need only to show that

$$\frac{1}{n^{3/4}h^{1/2}}\overline{Q}_{n,4,1}(h) \xrightarrow{P} \infty, \tag{B.52}$$

when the condition of $\delta_n\ n^{1/8}h^{1/4} \to \infty$ is satisfied. By Definition 2.1, we can find some integrable function $\widetilde{\Gamma}(\cdot)$ such that

$$\Delta_n^2(v) \geq \delta_n^2\widetilde{\Gamma}(v) \quad \text{and} \quad \int_{-\infty}^{\infty}\widetilde{\Gamma}(v)dv > 0. \tag{B.53}$$

By (B.53), Lemma B.4, Theorem 5.1 in Park and Phillips (1999) and following the calculation of $I_{n,1}$,

$$\begin{aligned} \frac{1}{T(n)\sqrt{n}h\delta_n^2}\overline{Q}_{n,4,1}(h) &= \frac{1}{T(n)\sqrt{n}h\delta_n^2}\sum_{s=1}^{n}\Delta_n^2(V_s)\sum_{t=1,\neq s}^{n}K_{s,t} \\ &= \frac{1}{\sqrt{n}\delta_n^2}\sum_{s=1}^{n}\Delta_n^2(V_s)\left(\frac{1}{T(n)h}\sum_{t=1}^{n}K_{s,t} - \frac{K(0)}{T(n)h}\right) \\ &\geq \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\widetilde{\Gamma}(V_s)\left(1+o_P(1)\right) \\ &\xrightarrow{P} \int_{-\infty}^{\infty}\widetilde{\Gamma}(v)dvL_W(1,0)(1+o_P(1)), \end{aligned} \tag{B.54}$$

where $L_W(1,0)$ is the local time of the standard Wiener process at the origin. Hence, we can show that (B.52) holds by $\delta_n\ n^{1/8}h^{1/4} \to \infty$ and Lemma B.1. Then, the proof of Theorem 3.1(ii) is completed.

Proof of Theorem 3.2(i): By Definition 2.2, we can find some asymptotically homogeneous function $\overline{\Lambda}(\cdot)$ such that

$$\max\{\Delta_n^2(v), \left|\Delta_n(v)\dot{\Delta}_n(v)\right|\} \leq \delta_n^2\overline{\Lambda}(v), \tag{B.55}$$

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 Definition 2.2(i)(ii).

By (B.55), Lemma B.1, Lemma B.4, Theorem 5.3 in Park and Phillips (1999) and following the calculation of $I_{n,1}$,

$$\begin{aligned} \overline{Q}_{n,4,1}(h) &= \sum_{s=1}^{n}\Delta_n^2(V_s)\sum_{t=1,\neq s}^{n}K_{s,t} \leq (T(n)nv^2(\sqrt{n})h\delta_n^2) \\ &\times \frac{1}{nv^2(\sqrt{n})}\sum_{s=1}^{n}\overline{\Lambda}(V_s)\left(1+o_P(1)\right) = O_P(n^{3/2}v^2(\sqrt{n})h\delta_n^2), \end{aligned}$$

which implies that either (B.45) or (B.50) holds by either

$$n^{3/8}\nu(\sqrt{n})h^{1/4}\delta_n = O(1) \text{ or } n^{3/8}\nu(\sqrt{n})h^{1/4}\delta_n \to 0.$$

The rest of the proof is the same as the proof of Theorem 3.1(i).

Proof of Theorem 3.2(ii): By Definition 2.2, we can find some asymptotically homogeneous function $\widetilde{\Lambda}(\cdot)$ such that $\Delta_n^2(v) \geq \delta_n^2 \widetilde{\Gamma}(v)$, and

$$\widetilde{\Lambda}(\lambda x) = v^2(\lambda)\widetilde{H}(x) + \widetilde{R}(x,\lambda) \text{ and } \int_{-\infty}^{\infty} \widetilde{H}(v) L_W(1,v)dv > 0. \tag{B.56}$$

By the condition of $n^{3/8}\nu(\sqrt{n})h^{1/4}\delta_n \to \infty$, (B.56), Lemma B.4 and Theorem 5.3 in Park and Phillips (1999) and following the calculation of $I_{n,1}$,

$$\begin{aligned} \frac{1}{T(n)nv^2(\sqrt{n})h\delta_n^2}\overline{Q}_{n,4,1}(h) &= \frac{1}{T(n)nv^2(\sqrt{n})h\delta_n^2}\sum_{s=1}^{n}\Delta_n^2(V_s)\sum_{t=1,\neq s}^{n} K_{s,t} \\ &\geq \frac{1}{nv^2(\sqrt{n})}\sum_{s=1}^{n}\widetilde{\Lambda}(V_s)\left(1+o_P(1)\right) \\ &\xrightarrow{P} \int_{-\infty}^{\infty}\widetilde{H}(v)L_W(1,v)dv(1+o_P(1)), \end{aligned} \tag{B.57}$$

which implies that equation (B.52) holds by Lemma B.1. The proof of Theorem 3.2(ii) is therefore completed.

Proof of Proposition 4.1: Recall the definition of $Q_{n,1}^*(h)$ in Lemma B.5 and observe that under H_0

$$Q_n^*(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \widehat{e}_t^* K_{s,t}\widehat{e}_s^* = Q_{n,1}^*(h) + Q_{n,2}^*(h) + Q_{n,3}^*(h), \tag{B.58}$$

where $Q_{n,2}^*(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \overline{g}_t^* K_{s,t}\overline{g}_s^*$ and $Q_{n,3}^*(h) = 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} \overline{g}_t^* K_{s,t} e_s^*$, in which $\overline{g}_t^* = g(V_t,\widehat{\theta}) - g(V_t,\widehat{\theta}^*)$, $(\overline{\sigma}_n^*)^2 = 2\overline{\sigma}_{e,*}^4 \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} K_{s,t}^2$ and $\overline{\sigma}_{e,*}^2 = \frac{1}{n}\sum_{t=1}^{n}(\widehat{e}_t^*)^2$.

By (B.58) and Lemma B.5, to prove (4.2), we first show that

$$\frac{Q_{n,2}^*(h) + Q_{n,3}^*(h)}{\overline{\sigma}_n^*} = O_P(s_n^*) \text{ and } \frac{\widetilde{\sigma}_n - \overline{\sigma}_n^*}{\widetilde{\sigma}_n} = o_P(s_n^*) \tag{B.59}$$

given $\mathcal{V}_n = (V_1, \cdots, V_n)'$.

The first part of (B.59) can be proved by arguments similar to those used in the proof of (B.27) in the proof of Proposition 3.1. We thus need only to give the proof of the second part of (B.59). Following the proof of Lemma B.3, we have

$$\frac{1}{n}\sum_{t=1}^{n}(\widehat{e}_t^*)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - g(V_t,\widehat{\theta}^*)\right)^2 \tag{B.60}$$
$$= \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - g(V_t,\widehat{\theta}) + g(V_t,\widehat{\theta}) - g(V_t,\widehat{\theta}^*)\right)^2$$
$$= \frac{1}{n}\sum_{t=1}^{n}(e_t^*)^2 + \frac{1}{n}\sum_{t=1}^{n}(\overline{g}_t^*)^2 + \frac{2}{n}\sum_{t=1}^{n}e_t\overline{g}_t^* + o_P(s_n^*).$$

Then, following the same arguments as in the proof Lemma B.5, we have

$$\frac{1}{n}\sum_{t=1}^{n}(\widehat{e}_t^*)^2 = \frac{1}{n}\sum_{t=1}^{n}e_t^2 + o_P(s_n^*). \tag{B.61}$$

This implies that as $n \to \infty$

$$\frac{\frac{1}{n}\sum_{t=1}^{n}(\widehat{e}_t^*)^2}{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \to_P 1, \tag{B.62}$$

which shows that the second part of (B.59) holds.

Meanwhile, equation (B.26) implies for some $0 < C_1 < C_2 < \infty$,

$$C_1\ n^{\frac{3}{2}}h\ (1 + o_P(1)) \le \widetilde{\sigma}_{n,1}^2(h) \le C_2\ n^{\frac{3}{2}}h\ (1 + o_P(1)), \tag{B.63}$$

where $\widetilde{\sigma}_{n,1}^2(h) = \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}K_{s,t}^2$.

To complete the proof of Proposition 4.1, in view of (B.63), we need only to show that

$$E\left[\mathrm{Tr}(A_0^3(h))\right] = O\left(n^2h^2\right). \tag{B.64}$$

Note that

$$E\left[\mathrm{Tr}(A_0^3(h))\right] = \sum_{s=1}^{n}\sum_{\substack{t=1\\t\neq s}}^{n}\sum_{\substack{q=1\\q\neq s,t}}^{n}E\left[K_{s,t}K_{t,q}K_{q,s}\right] \tag{B.65}$$
$$= \sum_{s=1}^{n}\sum_{t<s}\sum_{q<t}E\left[K_{s,t}K_{t,q}K_{q,s}\right] + \sum_{s=1}^{n}\sum_{t<s}\sum_{q=t+1}^{s-1}E\left[K_{s,t}K_{t,q}K_{q,s}\right]$$
$$+ \sum_{s=1}^{n}\sum_{t<s}\sum_{q>s}E\left[K_{s,t}K_{t,q}K_{q,s}\right] + \sum_{s=1}^{n}\sum_{t>s}\sum_{q<s}E\left[K_{s,t}K_{t,q}K_{q,s}\right]$$
$$+ \sum_{s=1}^{n}\sum_{t>s}\sum_{q=s+1}^{t-1}E\left[K_{s,t}K_{t,q}K_{q,s}\right] + \sum_{s=1}^{n}\sum_{t>s}\sum_{q>t}E\left[K_{s,t}K_{t,q}K_{q,s}\right] =: \sum_{i=1}^{6}\widehat{\Pi}_{n,i}.$$

By the definition of the random walk $\{V_t\}$, we have

$$\begin{aligned}
\widehat{\Pi}_{n,1} &= \sum_{s=1}^{n}\sum_{t<s}\sum_{q<t} E\left[K\left(\tfrac{V_s-V_t}{h}\right)K\left(\tfrac{V_t-V_q}{h}\right)K\left(\tfrac{V_q-V_s}{h}\right)\right] \\
&= \sum_{s=1}^{n}\sum_{t<s}\sum_{q<t}\int\int\int K\left(\tfrac{\sqrt{s-t}w}{h}\right)K\left(\tfrac{\sqrt{t-q}v}{h}\right)K\left(\tfrac{\sqrt{s-t}w+\sqrt{t-q}v}{h}\right) \\
&\quad \varphi_q(u)\varphi_{t-q}(v)\varphi_{s-t}(w)dudvdw \qquad \text{(B.66)}\\
&= \sum_{s=1}^{n}\sum_{t<s}\sum_{q<t}\frac{h^2}{\sqrt{(s-t)(t-q)}}\int\int\int K(z_1)K(z_2)K(z_1+z_2) \\
&\quad \varphi_q(z_3)\varphi_{t-q}\left(\tfrac{hz_1}{\sqrt{t-q}}\right)\varphi_{s-t}\left(\tfrac{hz_2}{\sqrt{s-t}}\right)dz_1dz_2dz_3 = O(n^2h^2).
\end{aligned}$$

Equation (B.66) also holds for $\widehat{\Pi}_{n,i}$ for $i=2,\cdots,6$. Hence, equation (B.64) holds with the help of Lemma B.1.

PROOF OF PROPOSITION 4.2: Observe that

$$\begin{aligned}
Q_n^*(h) &= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widehat{e}_t^*K_{s,t}\widehat{e}_s^* \\
&= \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}e_t^*K_{s,t}e_s^* + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widetilde{g}_t^*K_{s,t}\widetilde{g}_s^* + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widetilde{g}_t^*K_{s,t}e_s^* \\
&+ 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\Delta_n(V_t)K_{s,t}\widetilde{g}_s^* + 2\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\Delta_n(V_t)K_{s,t}e_s^* \\
&+ \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\Delta_n(V_t)K_{s,t}\Delta_n(V_s) =: \sum_{i=1}^{6}\overline{Q}_{n,i}^*(h). \qquad \text{(B.67)}
\end{aligned}$$

where $\widetilde{g}_t^* = g(V_t,\theta_1) - g(V_t,\widehat{\theta}_1^*)$.

Let $\widehat{Q}_{n,1}^*(h) = \frac{\overline{Q}_{n,1}^*(h)}{\overline{\sigma}_n^*}$. Observe that

$$\begin{aligned}
\beta_n^* &= P^*\left(\widehat{Q}_n^*(h) > l_\alpha^*|H_1\right) = 1 - P^*\left(\widehat{Q}_n^*(h) \leq l_\alpha^*|H_1\right) \\
&= 1 - P^*\left(\widehat{Q}_{n,1}^*(h) \leq l_\alpha^* - \frac{\overline{Q}_{n,6}^*(h)}{\overline{\sigma}_n^*} - \frac{\sum_{j=2}^5\overline{Q}_{n,j}^*(h)}{\overline{\sigma}_n^*}|H_1\right). \qquad \text{(B.68)}
\end{aligned}$$

Since the proof of Proposition 4.2(ii) is very similar to that for Proposition 4.2(i), we then need only to provide the proof of Proposition 4.2(i). To prove Proposition 4.2(i), in view of Lemma B.5, (B.59) and (B.68), it suffices to show that as $n \to \infty$

$$\frac{\overline{Q}_{n,i}^*(h)}{\overline{\sigma}_n^*} = o_P\left(\frac{\overline{Q}_{n,6}^*(h)}{\overline{\sigma}_n^*}\right) \quad \text{for } i=2,\cdots,5. \qquad \text{(B.69)}$$

In view of the corresponding derivations of (B.69) in the proofs of Theorems 3.1(i)(ii) and 3.2(i)(ii) as well as the second part of (B.59), replacing Lemma B.2 by Lemma B.3 in the derivations of (B.39) and (B.40) as well as (B.47) and (B.51), we can obtain for $i=2,\cdots,5$

$$\overline{Q}_{n,i}^*(h) = O_P\left(\overline{\sigma}_n^*\right) = o_P\left(nh\delta_n^2\right) = O_P\left(\overline{Q}_{n,6}^*\right) \qquad \text{(B.70)}$$

using equations (B.54) and (B.57) as well as the conditions of either Proposition 4.2(i) or Proposition 4.2(ii).

Equation (B.70) completes the proof of (B.69). We therefore have completed all the proofs.

REFERENCES

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

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

Chow, Y., Teicher, H., 1988. *Probability Theory.* Springer–Verlag, New York.

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., 2007. A new test in parametric linear models with nonparametric autoregressive errors. Working paper is available at http://www.adelaide.edu.au/directory/jiti.gao.

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

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

Gao, J., Li, D., Tjøstheim, D., 2009. Uniform consistency for nonparametric estimators in null recurrent time series. Working paper is available at http://www.adelaide.edu.au/directory/jiti.gao.

Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and Its Applications.* Academic Press, New York.

Hong, S. H., Phillips, P. C. B., 2005. Testing linearity in cointegrating relations with an application to purchasing power parity. Cowles Foundation Discussion Paper 1541, Yale University.

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 tests for a parametric regression models. *Journal of Econometrics* **87**, 145-165.

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., 2008. Structural Nonparametric Cointegrating Regression. Cowles Foundation Discussion Paper No. 1657. Forthcoming in *Econometrica*.

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