ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

An Improved Nonparametric Unit–Root Test

Jiti Gao and Maxwell King

August 2012

Working Paper 16/12

An Improved Nonparametric Unit–Root Test

Jiti Gao and Maxwell King

Department of Econometrics and Business Statistics

Monash University

Melbourne, Australia

Abstract: This paper proposes a simple and improved nonparametric unit–root test. An asymptotic distribution of the proposed test is established. Finite sample comparisons with an existing nonparametric test are discussed. Some issues about possible extensions are outlined.

Key words: Autoregression, nonparametric unit–root test, nonstationary time series, specification testing.

JEL Classification: C12, C14, C22.

1 Introduction

While there is a long literature for the field of parametric unit–root testing, discussion about using non– and semi–parametric tests for unit–root specification in nonlinear time series models has attracted little attention. To the best of our knowledge, the only published work available to us is Gao *et al* (2009a). A follow–up extension is given in Gao and King (2011). As discussed in both papers, the main advantage of estimating and specification testing the mean function simultaneously is that one may avoid bringing possible mis–specification issues through pre–specifying a parametric linear model before testing a linear unit–root problem. As the finite sample studies show in both papers, a nonparametric test is directly applicable when there is some unknown nonlinearity and nonstationarity involved in the mean function. In such a case, existing parametric unit–root tests proposed for the linear unit–root case are not valid. This is basically the main motivation for us to address some nonparametric unit–root testing issues.

In the related literature for the field of nonparametric specification for nonstationary time series, the same type of nonparametric tests have been proposed and studied in Gao *et al* (2009b), and Wang and Phillips (2012) for parametric model specification in nonlinear co–integrating regression models. Discussion about other related estimation and testing issues may be found from two recent survey papers by Gao (2012), and Sun and Li (2012). As may be seen from the relevant literature, existing test statistics are of the same type of standardised quadratic forms of nonstationary time series. While such test statistics all converge in distribution to a standard normality as the limiting distribution, both the establishment and the implementation of such an asymptotically normal test involve all sorts of unnecessary complexities and technicalities, particularly in the case where an autoregressive structure is involved as in the papers by Gao *et al* (2009a), and Gao and King (2011).

In order to address both theoretical and computational issues, this paper develops a simple and improved nonparametric test. As shown in Section 2 below, a functional of the standard Brownian motion is the limiting distribution of the proposed test and its proof is quite concise. This is not unnatural considering the fact that existing parametric unit–root test statistics all have functionals of the standard Brownian motion as their limiting distributions. Section 3 then compares the finite–sample performance of the proposed test with its natural competitor proposed in Gao *et al* (2009a). Our conclusion is that it is easy to implement the proposed test and it is also more powerful than the natural competitor. Additionally, the proposed test is able to test a nonlinear unit–root structure against a sequence of asymptotically localised alternatives.

The organisation of this paper is summarised as follows. Section 2 establishes the proposed test and then develops its asymptotic theory. Section 3 discusses possible extensions of the proposed test for the univariate case to a multivariate case. Simulated examples are used in Section 4 to evaluate the finite–sample performance of the proposed test. Some concluding remarks are given in Section 5. Mathematical assumptions and proofs are all given in Appendices A and B.

2 Test statistic and theory

2.1 Asymptotic theory

Consider a nonlinear time series model of the form

$$\begin{aligned} y_t &= m(x_t) + e_t, \ t = 1, 2, \cdots, n, \\ x_t &= y_{t-1} \ \text{ and } \ E\left[e_t | x_t\right] = 0, \end{aligned} \tag{2.1}$$

where $y_0 = 0$, $m(\cdot)$ is an unknown function defined on $R^1 = (-\infty, \infty)$, and $\{e_t\}$ is a sequence of martingale differences satisfying Assumption A.1 listed in Appendix A below.

Our interest in this paper is to test

$$H_0: \ P\left(m(x_t) = x_t\right) = 1 \ \text{ versus } \ H_1: \ P\left(m(x_t) = x_t + \Delta_n(x_t)\right) = 1, \tag{2.2}$$

where $\Delta_n(x)$ is a local 'departure' function such that $\min_{n\ge 1}\inf_{x\in R^1}|\Delta_n(x)| > 0$. In other words, we are only interested in a kind of local departure from the null hypothesis, because of the explosive nature of the integrated structure of $\{x_t\}$.

When $m(x)$ is parametrically specified as $m(x) = \theta x$, the literature focuses on testing $H_0: \ \theta = 1$. Before we construct our test, we estimate $m(\cdot)$ by minimising

$$\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \beta\right)^2 K\left(\frac{x_t - x}{h}\right) \tag{2.3}$$

over $\beta = m(x)$, where $K(\cdot)$ is a probability kernel function, and h is a bandwidth parameter.

Function $m(x)$ is then estimated by

$$\widehat{m}(x) = \frac{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right) y_t}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{h}\right)}. \tag{2.4}$$

To test H_0, the main idea is to compare $m(x)$ and x through using some distance function, such as

$$L_2(m) = \int_{-\infty}^{\infty}\left(m(x) - x\right)^2 \pi(x)dx, \tag{2.5}$$

where $\pi(x)$ is a known probability weight function satisfying $0 < \int_{-\infty}^{\infty} \pi^2(x)dx < \infty$.

Recall that $m(x)$ is estimated by

$$\widehat{m}(x) = \frac{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right) y_t}{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right)}. \tag{2.6}$$

In order to construct our test, we introduce a smoothed version of x of the form

$$\widetilde{m}(x) = \frac{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right) x_t}{\sum_{t=1}^n K\left(\frac{x_t-x}{h}\right)}. \tag{2.7}$$

We then define the following quantities:

$$\begin{aligned}
\widehat{p}(x) &= \frac{1}{\sqrt{n}h} \sum_{t=1}^n K\left(\frac{x_t - x}{h}\right), \\
\widehat{q}(x) &= \frac{1}{\sqrt{n}h} \sum_{t=1}^n K\left(\frac{x_t - x}{h}\right) y_t, \\
\widetilde{q}(x) &= \frac{1}{\sqrt{n}h} \sum_{t=1}^n K\left(\frac{x_t - x}{h}\right) x_t. \qquad (2.8)
\end{aligned}$$

This paper now proposes using a test statistic of the form

$$\begin{aligned}
L_n(h) &\equiv \sqrt{nh^2} \int_{-\infty}^{\infty} \left(\widehat{q}(x) - \widetilde{q}\,(x)\right)^2 \pi(x)dx \\
&= \sqrt{nh^2} \int_{-\infty}^{\infty} \left(\widehat{m}(x) - \widetilde{m}\,(x)\right)^2 \widehat{p}^2(x)\ \pi(x)dx, \qquad (2.9)
\end{aligned}$$

which is in a similar fashion to the original proposal discussed in Härdle and Mammen (1993) for the independent sample case.

As shown in Appendix A below, we have as $n \to \infty$

$$\begin{aligned}
L_n(h) = &\ \frac{1}{\sqrt{n}} \sum_{t=1}^n \widehat{e}_t^2\ \pi(x_t) \cdot \int_{-\infty}^{\infty} K^2(u)du + \frac{2}{\sqrt{n}} \sum_{t=2}^n \sum_{s=1}^{t-1} \widehat{e}_t\ \widehat{e}_s\ \pi(x_s) \cdot L\left(\frac{x_t - x_s}{h}\right) \\
&+ o_P(1) \equiv \widetilde{L}_n(h) + o_P(1), \qquad (2.10)
\end{aligned}$$

where $\widehat{e}_t = y_t - y_{t-1}$ under H_0, $L(u) = \int_{-\infty}^{\infty} K(v)K(u+v)dv$, and

$$\begin{aligned}
\widetilde{L}_n(h) &= \frac{\sigma^2(K)}{\sqrt{n}} \sum_{t=1}^n \widehat{e}_t^2\ \pi(x_t) + \frac{2}{\sqrt{n}} \sum_{t=2}^n \sum_{s=1}^{t-1} \widehat{e}_t\ \widehat{e}_s\ \pi(x_s) \cdot L\left(\frac{x_t - x_s}{h}\right), \\
&\equiv S_{1n} + S_{2n}, \qquad (2.11)
\end{aligned}$$

where $\sigma^2(K) = \int_{-\infty}^{\infty} K^2(u)du$. Since $L_n(h)$ involves an integral in $R^1 = (-\infty, \infty)$, it is not computationally easy to use $\widetilde{L}_n(h)$ in both simulated and real data examples.

As also shown in Appendix A below, S_{1n} converges in distribution to a random variable and S_{2n} converges to zero in probability. Mainly because of these facts, there is no need to standardise $L_n(h)$ to establish an asymptotic normality as the limiting distribution of the standardised version of $L_n(h)$.

In the stationary case where $\{x_t\}$ is stationary, however, this kind of standardisation is needed, because of

$$\frac{1}{nh}\sum_{t=1}^{n}\left(\int_{-\infty}^{\infty} K^2\left(\frac{x_t-x}{h}\right)\pi(x)dx\right) e_t^2 = \frac{1}{n}\sum_{t=1}^{n}\pi(x_t)e_t^2\cdot\int_{-\infty}^{\infty}K^2(u)du + o_P(1)$$
$$\to_P \ C_2(K,\pi,\sigma_e^2), \tag{2.12}$$

where $C_2(K,\pi,\sigma_e^2) = \sigma_e^2\cdot E[\pi(x_1)]\cdot\int_{-\infty}^{\infty}K^2(u)du$ is a non–random quantity.

We now state the main theorem of this paper; its proof is given in Appendix A.

Theorem 2.1. *Suppose that Assumptions A.1 and A.2 listed in Appendix A below hold. Then under* H_0

$$L_n(h) = \sqrt{nh^2}\int_{-\infty}^{\infty}\left(\widehat{q}(x)-\widetilde{q}(x)\right)^2\pi(x)dx \to_D \sigma^2(K)\cdot\sigma_e^2\cdot L_B(1,0) \tag{2.13}$$

as $n\to\infty$, *where* $\sigma^2(K)=\int_{-\infty}^{\infty}K^2(v)dv$, $\sigma_e^2 = E[e_1^2]$, *and* $L_B(1,0)$ *is a random variable and its cumulative distribution function is given by*

$$F_L(x) = P\left(L_B(1,0)\le x\right) = \begin{cases} 2\Phi(x)-1, & x\ge 0,\\ 0, & x<0,\end{cases}$$

in which $\Phi(x)$ *is the cdf of* $N(0,1)$.

Remark 2.1. (i) σ_e^2 can be estimated by $\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}\left(y_t-y_{t-1}\right)^2$ under H_0 when it is unknown.

(ii) Note that it is quite common in the parametric case to have a functional of Brownian motion as a limiting distribution of a unit–root test statistic.

2.2 Discussion of power properties

As pointed out in the introductory section, an existing test for the autoregressive case is the test proposed in Gao *et al* (2009a) as follows:

$$M_n(h) = \frac{\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}K\left(\frac{x_t-x_s}{h}\right)\ \widehat{e}_s\ \widehat{e}_t}{\sqrt{2\sum_{t=1}^{n}\sum_{s=1}^{n}K^2\left(\frac{x_t-x_s}{h}\right)\ \widehat{e}_s^2\ \widehat{e}_t^2}}. \tag{2.14}$$

Let $\widehat{\sigma}_n^2 = 2\sum_{t=1}^{n}\sum_{s=1}^{n}K^2\left(\frac{x_t-x_s}{h}\right)\ \widehat{e}_s^2\ \widehat{e}_t^2$. We then have under H_0:

$$\widehat{\sigma}_n^2 = 2\sum_{t=2}^{n}\sum_{s=1}^{n}K^2\left(\frac{x_t-x_s}{h}\right)u_s^2\ u_t^2,$$
$$\sigma_n^2 = E\left[\widehat{\sigma}_n^2\right] = C(1+o(1))\cdot n^{\frac{3}{2}}h, \tag{2.15}$$

where $u_t = y_t - y_{t-1} = e_t$ under H_0, and $C > 0$ is a constant.

In a similar fashion to the derivations used in either the proof of Lemma B.5 of Li *et al* (2011) or the proof of Theorem 3.2 of Wang and Phillips (2012), we may show that there are constants $0 < C_1 < C_2 < \infty$ such that under H_0:

$$\lim_{n\to\infty} P\left(C_1 n^{\frac{3}{2}} h \le \widehat{\sigma}_n^2 \le C_2 n^{\frac{3}{2}} h\right) = 1. \tag{2.16}$$

This section is then interested in a sequence of local departure functions of the form:

$$\Delta_n(x) = \delta_n \cdot \Delta(x), \tag{2.17}$$

where $\delta_n \to 0$ as $n \to \infty$, and $\Delta(x)$ is chosen such that for $j = 1, 2$

$$\int_{-\infty}^{\infty} \Delta^2(x)dx < \infty \ \text{ and } \ \int_{-\infty}^{\infty} \Delta^2(x)\pi^j(x)dx < \infty. \tag{2.18}$$

Let $M_{1n} = \sum_{t=1}^n \sum_{s=1}^n K\left(\frac{x_t - x_s}{h}\right)\ \widehat{e}_s\ \widehat{e}_t$. Note that $\widehat{e}_t = y_t - x_t - \Delta_n(x_t) = u_t - \Delta_n(x_t)$ under H_1. It may be shown under H_1 that

$$\begin{aligned} M_{1n} &= \sum_{t=1}^n \sum_{s=1}^n K\left(\frac{x_t - x_s}{h}\right)\ \widehat{e}_s\ \widehat{e}_t \\ &= \sum_{t=1}^n \sum_{s=1}^n K\left(\frac{x_t - x_s}{h}\right)\ u_s\ u_t + \sum_{t=1}^n \sum_{s=1}^n K\left(\frac{x_t - x_s}{h}\right)\ \Delta_n(x_s)\ \Delta_n(x_t) + o_P(1) \\ &\ge \sum_{t=1}^n \sum_{s=1}^n K\left(\frac{x_t - x_s}{h}\right)\ \Delta_n(x_s)\ \Delta_n(x_t) + o_P(1) \equiv\ M_{2n} + o_P(1), \end{aligned}$$

where, as shown in Appendix B, we have that as $n \to \infty$

$$R_{1n} \equiv \frac{E\left[M_{2n}\right]}{\sigma_n} = C\delta_n^2 \cdot \frac{nh}{\sqrt{n^{\frac{3}{2}}h}} \cdot (1 + o(1)) = C\delta_n^2\ \sqrt{\sqrt{n}h} \cdot (1 + o(1)) \tag{2.19}$$

when $\int_{-\infty}^{\infty} \Delta^2(x)dx < \infty$.

Meanwhile, under H_1 Lemma A.2 in Appendix A below shows that as $n \to \infty$

$$\begin{aligned} L_n(h) &= \frac{1}{\sqrt{n}h} \sum_{t=1}^n \sum_{s=1}^n \widehat{e}_s\ \widehat{e}_t \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x)dx \\ &\ge \frac{\delta_n^2}{\sqrt{n}h} \sum_{t=1}^n \sum_{s=1}^n \Delta(x_s)\Delta(x_t) \int_{-\infty}^{\infty} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x)dx + o_P(1) \\ &\equiv\ L_{1n} + o_P(1), \end{aligned}$$

where we have as $n \to \infty$

$$E[L_{1n}] =\ C(1 + o(1)) \cdot \delta_n^2 \cdot \sqrt{n}h \tag{2.20}$$

when $\int_{-\infty}^{\infty} \Delta^2(x)\pi(x)dx < \infty$.

It follows from equations (2.19) and (2.20) that there is some $C_0 > 0$ such that

$$\frac{E[L_{1n}]}{R_{1n}} = \frac{E[L_{1n}]\ \sigma_{2n}}{E[M_{1n}]} = C_0\sqrt{\sqrt{n}h} \to \infty, \tag{2.21}$$

which implies that $L_n(h)$ is more powerful than $M_n(h)$ under a sequence of local departure functions of the forms (2.17) and (2.18).

Some detailed derivations of equations (2.19)–(2.21) are given in Appendix B below. Section 4 below evaluates the finite sample performance of $L_n(h)$ and $M_n(h)$.

3 Extensions

This section discusses possible extensions of model (2.1) as well as the proposed test to the case where a set of stationary time series regressors may also involve in a nonlinear time series model of the form

$$\begin{aligned} y_t &= m(x_t, z_t) + e_t,\ \ t = 1, 2, \cdots, n, \\ x_t &= y_{t-1} \text{ and } E[e_t|x_t, z_t] = 0, \end{aligned} \tag{3.1}$$

where $y_0 = 0$, $\{z_t\}$ is a vector of stationary regressors, and $m(\cdot, \cdot)$ is an unknown function.

The interest here is in a kind of specification testing of the form

$$\begin{aligned} H_0 &: \ P\left(m(x_t, z_t) = x_t + g(z_t; \theta_0)\right) = 1 \ \text{ versus} \\ H_1 &: \ P\left(m(x_t, z_t) = x_t + g(z_t; \theta_0) + \Delta_n(x_t, z_t)\right) = 1, \end{aligned} \tag{3.2}$$

where $g(\cdot, \theta_0)$ is a parametrically known function indexed by θ_0, a vector of unknown parameters, and $\Delta_n(x, z)$ is a sequence of departure functions.

Under H_0, model (3.1) suggests estimating θ_0 by $\widehat{\theta}$ that minimises

$$\frac{1}{n}\sum_{t=1}^{n} [y_t - x_t - g(z_t; \theta)]^2 \quad \text{over all possible } \theta. \tag{3.3}$$

Meanwhile, model (3.1) also suggests estimating $m(\cdot, \cdot)$ by

$$\widehat{m}(x, z) = \frac{\sum_{t=1}^{n} K_1\left(\frac{x_t - x}{h_1}\right) K_2\left(\frac{z_t - z}{h_2}\right) y_t}{\sum_{t=1}^{n} K_1\left(\frac{x_t - x}{h_1}\right) K_2\left(\frac{z_t - z}{h_2}\right)}, \tag{3.4}$$

where $K_i(\cdot)$ for $i = 1, 2$ are probability kernel functions and h_i for $i = 1, 2$ are bandwidth parameters.

To test H_0, discussion in Section 2 suggests constructing a test based on a kind of distance between $\widehat{m}(x,z)$ and $x+g(z;\widehat{\theta})$. In order to construct our test, we introduce a smoothed version of $x+g(z;\theta_0)$ of the form

$$\widetilde{m}(x,z;\theta_0) = \frac{\sum_{t=1}^n K_1\left(\frac{x_t-x}{h_1}\right) K_2\left(\frac{z_t-z}{h_2}\right)(x_t+g(z_t;\theta_0))}{\sum_{t=1}^n K_1\left(\frac{x_t-x}{h_1}\right) K_2\left(\frac{z_t-z}{h_2}\right)}. \tag{3.5}$$

We may then introduce a distance function between $\widehat{m}(x,z)$ and $\widetilde{m}(x,z;\widehat{\theta})$. To avoid introducing some random denominator problems, we propose using a modified distance function by comparing the following quantities:

$$\begin{aligned} \widehat{q}(x,z) &= \frac{1}{\sqrt{n}h_1h_2}\sum_{t=1}^n K_1\left(\frac{x_t-x}{h_1}\right) K_2\left(\frac{z_t-z}{h_2}\right) y_t \quad \text{and} \\ \widetilde{q}(x,\theta_0) &= \frac{1}{\sqrt{n}h_1h_2}\sum_{t=1}^n K_1\left(\frac{x_t-x}{h_1}\right) K_2\left(\frac{z_t-z}{h_2}\right)(x_t+g(z_t;\theta_0)). \end{aligned} \tag{3.6}$$

We then propose using a test statistic of the form

$$\begin{aligned} L_n(h_1,h_2) &= \sqrt{nh_1^2h_2^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\widehat{q}(x,z)-\widetilde{q}\left(x,z;\widehat{\theta}\right)\right)^2\pi_1(x)\pi_2(z)dzdx \\ &= \sqrt{nh_1^2h_2^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\widehat{m}(x,z)-\widetilde{m}\left(x,z;\widehat{\theta}\right)\right)^2\widehat{p}^2(x,z)\ \pi_1(x)\pi_2(z)dzdx, \end{aligned} \tag{3.7}$$

where $\pi_i(u)$ are both known probability weight functions satisfying $0 < \int_{-\infty}^{\infty}\pi_i^2(u)du < \infty$ for $i=1,2$, and $\widehat{p}(x,z) = \frac{1}{\sqrt{n}h_1h_2}\sum_{t=1}^n K_1\left(\frac{x_t-x}{h_1}\right) K_2\left(\frac{z_t-z}{h_2}\right)$.

Since there is an autoregressive structure involved in model (3.1) and y_t and z_t are highly correlated and dependent on each other, it is not so clear whether Theorem 2.1 for the univariate case could be extended for $L_n(h_1,h_2)$ in the multivariate case. Section 4 below however shows that $L_n(h_1,h_2)$ works well numerically.

4 Examples of implementation

4.1 Computational aspects

This section introduces an approximate version of $L_n(h_1,h_2)$ and then a natural extension of the test proposed in Gao *et al* (2009a) before a bandwidth selection method is discussed.

Similarly to the derivation of $\widetilde{L}_n(h)$ in Section 2 above, we may approximate $L_n(h_1,h_2)$

by

$$\widetilde{L}_n(h_1,h_2) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{e}_t^2\ \pi_1(x_t)\pi_2(z_t)\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_1^2(u)K_2^2(v)dvdu$$
$$+\ \frac{2}{\sqrt{n}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\widehat{e}_t\ \widehat{e}_s\ \pi_1(x_s)\pi_2(z_s)\ L_1\left(\frac{x_t-x_s}{h_1}\right)L_2\left(\frac{z_t-z_s}{h_2}\right), \quad (4.1)$$

where $\pi_1(x) = \frac{1}{\pi\ (1+x^2)}$, $\pi_2(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$, $L_i(u) = \int_{-\infty}^{\infty}K_i(v)K_i(u+v)dv$, $\int_{-\infty}^{\infty}K_i^2(u)du = \frac{1}{2\sqrt{\pi}}$ and $L_i(u) = \frac{1}{2\sqrt{\pi}}e^{-\frac{u^2}{4}}$ when $K_i(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ for $i=1,2$, and $\widehat{e}_t = \left(y_t - x_t - g(z_t;\widehat{\theta})\right)$, in which $\widehat{\theta}$ is defined by (3.3).

Meanwhile, a natural extension of the test proposed in Gao *et al* (2009a) for the univariate case can be defined by

$$M_n(h_1,h_2) = \frac{\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widehat{e}_sK_1\left(\frac{x_s-x_t}{h_1}\right)K_2\left(\frac{z_s-z_t}{h_2}\right)\widehat{e}_t}{\sqrt{\sum_{t=1}^{n}\sum_{s=1,\neq t}^{n}\widehat{e}_s^2K_1^2\left(\frac{x_s-x_t}{h_1}\right)K_2^2\left(\frac{z_s-z_t}{h_2}\right)\widehat{e}_t^2}}, \quad (4.2)$$

where $\widehat{e}_t = \left(y_t - x_t - g(z_t;\widehat{\theta})\right)$ and $\widehat{\theta}$ is the same as in (3.3).

As may be seen from the proposed tests, certain bandwidth parameters are involved. In Table 4.1a below, a fixed bandwidth is used. In general, we propose using a cross–validation based method to choose suitable bandwidth parameters.

Because Edgeworth expansions for the distributions of $\widetilde{L}_n(h_1,h_2)$ and $M_n(h_1,h_2)$ are not readily available, we are therefore unable to adopt the power–function approach for the choice of optimal bandwidths (as has been discussed in Li *et al* 2011 for the univariate case). Instead, we propose using an estimation–based optimal bandwidths of the form:

$$\left(\widehat{h}_{1cv},\widehat{h}_{2cv}\right) = \arg\min_{(h_1,h_2)\in H_{\rm cv}}\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \widehat{m}_{-t}(x_t,z_t;h_1,h_2)\right)^2, \quad (4.3)$$

where $\widehat{m}_{-t}(x_t,z_t;h_1,h_2) = \frac{\sum_{s=1,\neq t}^{n}K_1\left(\frac{x_t-x_s}{h_1}\right)K_2\left(\frac{z_t-z_s}{h_2}\right)y_s}{\sum_{u=1,\neq t}^{n}K_1\left(\frac{x_t-x_u}{h_1}\right)K_2\left(\frac{z_t-z_s}{h_2}\right)}$ and

$$H_{\rm cv} = \left[c_1\ n^{-\frac{1}{12}-c_0}, c_2\ n^{-\frac{1}{12}+c_0}\right]\times\left[d_1\ n^{-\frac{1}{6}-d_0}, d_2\ n^{-\frac{1}{6}+d_0}\right]$$

for some $0 < c_1 < c_2 < \infty$, $0 < c_0 < \frac{1}{48}$, $0 < d_1 < d_2 < \infty$ and $0 < d_0 < \frac{1}{24}$. Before selecting $H_{\rm cv}$, we actually calculated equation (4.3) over all possible intervals. Our computation indicates that $H_{\rm cv}$ is the smallest possible interval on which the cross–validation function attains its smallest value.

Let l_r be the asymptotic critical valueI of the sample distribution of the proposed test in each case. In both Examples 4.2 and 4.3 below, we then use the chosen bandwidths involved in a regression bootstrap method to select a simulated critical value l_r^* in each case. Let $Q_n(h_1, h_2)$ denote either $\widetilde{L}_n(h_1, h_2)$ or $M_n(h_1, h_2)$.

Our experience with Examples 4.2 and 4.3 shows that $Q_n\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$ already has some stable sizes and good power values under the choice of $\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$. This may be because this pair of bandwidths may be either exactly identical or extremely close to such bandwidth values that maximise the power function while controlling the size function. In the stationary time series case, the theory developed in Chapter 3 of Gao (2007) shows that such estimation–based optimal bandwidth values may also be optimal for testing purposes.

We then propose using the following bootstrap method to approximate l_r by l_r^* in each case.

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

$$\widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t - g\left(z_t;\widehat{\theta}\right)\right)^2, \tag{4.4}$$

in which $\{\eta_t^*, 1 \leq t \leq n\}$ is a sequence of i.i.d. random variables drawn from

$$P\left(\eta_t^* = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \quad \text{and} \quad P\left(\eta_t^* = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}. \tag{4.5}$$

Step 2: Obtain $y_t^* = x_t + g(z_t;\widehat{\theta}) + e_t^*$. The resulting sample $\{(y_t^*, x_t, z_t), 1 \leq t \leq n\}$ is called a bootstrap sample.

Step 3: Use the data set $\{(y_t^*, x_t, z_t), 1 \leq t \leq n\}$ to re–estimate (α, β, γ) and denote their estimators by $(\widehat{\alpha}^*, \widehat{\beta}^*, \widehat{\gamma}^*)$. Then calculate the test statistic $\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$, which is the corresponding version of $\widehat{Q}_n\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$ by replacing $\{(y_t, x_t, z_t)\}$ and $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ with $\{(y_t^*, x_t, z_t)\}$ and $(\widehat{\alpha}^*, \widehat{\beta}^*, \widehat{\gamma}^*)$, respectively.

Step 4: Repeat Steps 1–3 $M_b = 250$ times and produce $M_b = 250$ versions of $\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$. Denote the M versions of $\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$ by $\widehat{Q}_{n,m}^*(h_1, h_2)$, $m = 1, 2, \cdots, M_b$. Then, we construct the empirical distributions of $\widehat{Q}_{n,m}^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$. That is,

$$\mathsf{P}^*\left(\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right) \leq x\right) = \mathsf{P}\left(\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right) \leq x | \mathcal{W}_n\right),$$

where $\mathcal{W}_n = \{(y_t, x_t, z_t), 1 \leq t \leq n\}$.

For each pair $\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right)$, choose l_r^* such that

$$P^*\left(\widehat{Q}_n^*\left(\widehat{h}_{1\text{cv}}, \widehat{h}_{2\text{cv}}\right) > l_r^*\right) = r$$

and estimate l_r by l_r^*.

Equation (4.3) is used for the choice of (h_1, h_2) in the implementation of $\widetilde{L}_n(h_1, h_2)$ and $M_n(h_1, h_2)$ in Tables 4.2 and 4.3 below. A special case of (4.3) associated with the proposed bootstrap method is used for the univariate case to be implemented in Table 4.1b below.

4.2 Simulated examples

This section evaluates the finite sample performance of the proposed test and its competitors. As existing studies (such as, Gao *et al* 2009a; Gao and King 2011), already show that $M_n(h)$ is needed and has better finite–sample performances than those proposed for the linear unit–root case, such as the Dickey–Fuller test and its various versions, this paper focuses on the finite–sample comparison between $M_n(h)$ and $L_n(h)$.

Example 4.1. Consider a linear time series model of the form:

$$H_0: \; y_t = \beta_0 x_t + e_t, \quad t = 1, 2, \cdots, n, \tag{4.6}$$

versus

$$H_1: \; y_t = \beta_1 x_t + \Delta_n(x_t) + e_t, \quad t = 1, 2, \cdots, n, \tag{4.7}$$

where $x_t = y_{t-1}$, $y_t = y_{t-1} + u_t$ with $y_0 = 0$ and $u_t \sim N(0, 1)$, $\beta_i = 1$, and

$$\Delta_n(x) = \frac{\delta_n}{\sqrt{1 + x^2}} \quad \text{with } \delta_n = \frac{\log(n)}{2n^{\frac{1}{8}}}. \tag{4.8}$$

The choice of δ_n can be discussed in the same way as will be done for the general case in (4.14) below. This section then compares the finite sample performance of the following two test statistics:

$$L_{1n}(h) \equiv M_n(h) \quad \text{and} \quad L_{2n}(h) = \widetilde{L}_n(h), \tag{4.9}$$

in which $\pi(x) = \frac{1}{\pi \, (1+x^2)}$, $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $L(u) = \int_{-\infty}^{\infty} K(v)K(u+v)dv$ are chosen for the computation of the two test statistics. Note that $\int_{-\infty}^{\infty} K^2(u)du = \frac{1}{2\sqrt{\pi}}$ and $L(u) = \frac{1}{2\sqrt{\pi}} e^{-\frac{u^2}{4}}$ when $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

In this example, we use an asymptotic critical value (acv) and a fixed bandwidth of $h = n^{-\frac{1}{4}}$ in each case. For L_{1n}, we use $z_{0.01} = 2.33$ at the 1% level and $z_{0.05} = 1.645$ at the 5% level. For L_{2n}, we use the critical value, l_r, of $\sigma^2(K) \; L_B(1, 0)$ at the 1% level and at the 5% level.

We then consider cases where the number of replications was $N = 1,000$ and the simulations were done for data sets of sizes $n = 100$, 300 and 500. Let f_{icv} denote the frequency of $L_{1n}(\widehat{h}_{icv}) > z_r$ for $i = 0, 1$ under H_0 or H_1, and g_{icv} denote the frequency of $L_{2n}(\widehat{h}_{icv}) > l_r$ for $i = 0, 1$ under H_0 or H_1. The simulation results are given in Table 4.1a.

In addition, we also consider using a regression bootstrap method to choose bootstrap critical values (bcv) z_r^* and l_r^*. Let $Q_n(h)$ denote either $L_{1n}(h)$ or $L_{2n}(h)$. Table 4.1b then gives the corresponding results for the bootstrap case.

Table 4.1a: Non–bootstrap with $M = 1000$

H_0	L_{1n}		L_{2n}	
n	1%	5%	1%	5%
100	0.001	0.004	0.000	0.005
300	0.001	0.006	0.009	0.026
500	0.009	0.028	0.018	0.051
H_1	L_{1n}		L_{2n}	
n	1%	5%	1%	5%
100	0.466	0.557	0.843	0.861
300	0.863	0.898	0.941	0.944
500	0.966	0.975	0.985	0.990

Table 4.1b: Bootstrap with $M_b = 250$, $M = 1000$

H_0	$L_{1n}(h)$		$L_{2n}(h)$	
n	1%	5%	1%	5%
100	0.009	0.022	0.018	0.068
300	0.005	0.013	0.011	0.068
500	0.009	0.025	0.009	0.056
H_1	$L_{1n}(h)$		$L_{2n}(h)$	
n	1%	5%	1%	5%
100	0.476	0.567	0.961	0.971
300	0.863	0.900	0.995	0.997
500	0.962	0.976	0.994	0.994

Tables 4.1a and 4.1b show that the proposed test works well in the finite sample case. While Table 4.1a shows that the proposed test works well when an asymptotic critical value is

combined with a fixed bandwidth, Table 4.1b shows that both the sizes and power values can be improved when a bootstrap method is used in association with a data–driven bandwidth in each case. Meanwhile, both Tables 4.1a and 4.1b show that L_{2n} is more powerful than L_{1n}.

Similarly to equation (4.9), we introduce the following definitions:

$$L_{1n}(h_1, h_2) \equiv M_n(h_1, h_2) \quad \text{and} \quad L_{2n}(h_1, h_2) = \widetilde{L}_n(h_1, h_2), \tag{4.10}$$

Examples 4.2 and 4.3 below evaluate the finite sample performance of $L_{1n}(h_1, h_2)$ and $L_{2n}(h_1, h_2)$.

Example 4.2. Consider a bivariate linear time series model of the form:

$$H_0: \ y_t = \alpha + \beta y_{t-1} + \gamma z_t + e_t, \quad t = 1, 2, \cdots, n, \tag{4.11}$$

versus

$$H_1: \ y_t = \alpha + \beta y_{t-1} + \gamma z_t + \Delta_n(y_{t-1}, z_t) + e_t, \quad t = 1, 2, \cdots, n, \tag{4.12}$$

where $y_0 = 0$, $\alpha = 0$ and $\beta = \gamma = 1$,

$$\begin{pmatrix} e_t \\ z_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{4.13}$$

with $\rho = 0$ or $\rho = 0.9$,

$$\Delta_n(y, z) = \frac{\delta_n \ z^2}{\sqrt{1 + y^2}} \quad \text{with } \delta_n = \frac{\log(n)}{2n^{\frac{1}{8}}}. \tag{4.14}$$

Note that there is some endogeneity between e_t and z_t when $\rho = E[e_t z_t] = 0.9$. Note also that the choice of δ_n in theory is to ensure that $\delta_n \to 0$ and $\delta_n^2 \sqrt{n} h_1 h_2 \to \infty$. Since the leading orders of h_1 and h_2 are chosen as $n^{-\frac{1}{12}}$ and $n^{-\frac{1}{6}}$, respectively in the cross–validation method in (4.3), the choice of δ_n in (4.8) satisfies the theoretical requirements.

Table 4.2 shows that the extended version, $L_n(h_1, h_2)$, of the proposed test $L_n(h)$ also works well numerically when there is a linear unit–root structure involved in the model under H_0. Meanwhile, Table 4.2 demonstrates that the proposed test is still applicable and even works well in the case where there is some endogeneity between z_t and e_t. This motivates us to develop an asymptotic theory for the proposed test even under certain endogeneity assumptions. In the same pattern as has been seen from Tables 4.1a and 4.1b, additionally, Table 4.2 indicates that L_{2n} is more powerful than L_{1n}.

Table 4.2: Bootstrap with $M_b = 250$, $M = 1000$

	$\rho = 0$				$\rho = 0.9$			
H_0	$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$		$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$	
n	1%	5%	1%	5%	1%	5%	1%	5%
100	0.002	0.012	0.021	0.076	0.004	0.017	0.025	0.063
300	0.006	0.022	0.015	0.060	0.011	0.034	0.015	0.068
500	0.009	0.031	0.011	0.052	0.012	0.033	0.016	0.051
H_1	$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$		$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$	
n	1%	5%	1%	5%	1%	5%	1%	5%
100	0.195	0.322	0.515	0.698	0.200	0.379	0.887	0.944
300	0.436	0.540	0.614	0.797	0.590	0.715	0.960	0.974
500	0.568	0.655	0.736	0.856	0.651	0.739	0.960	0.979

Example 4.3. Consider a nonlinear time series model of the form:

$$H_0: \ y_t = \beta y_{t-1} + \frac{1}{\tau + y_{t-1}^2} + \gamma z_t + e_t, \quad t = 1, 2, \cdots, n, \tag{4.15}$$

versus

$$H_1: \ y_t = \beta y_{t-1} + \frac{1}{\tau + y_{t-1}^2} + \gamma z_t + \Delta_n(y_{t-1}, z_t) + e_t, \quad t = 1, 2, \cdots, n, \tag{4.16}$$

where $y_0 = 0$, and $\beta = \gamma = \tau = 1$,

$$\begin{pmatrix} e_t \\ z_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{4.17}$$

with $\rho = 0$ or $\rho = 0.9$ and $\Delta_n(y, z) = \frac{\delta_n \ z^2}{\sqrt{1+y^2}}$, in which $\delta_n = \frac{\log(n)}{2n^{\frac{1}{8}}}$.

Table 4.3 reveals that the tests are applicable to the case where a nonlinear unit–root structure is involved in model (4.15) under H_0, in addition to supporting the findings reported in Table 4.2 that $L_{2n}(h_1, h_2)$ is more powerful than $L_{1n}(h_1, h_2)$ and that both tests are applicable to the case where there is some endogeneity between z_t and e_t. Note that $\{y_t\}$ generated by (4.15) is a $\frac{1}{2}$–null recurrent Markov chain (see, for example, Gao, Tøstheim and Yin 2011), even though it is not linearly integrated. This may imply that the proposed tests are applicable to the case where $\{y_t\}$ is a nonstationary, but not necessarily a linear unit–root time series.

Table 4.3: Bootstrap with $M_b = 250$ and $M = 1000$

	$\rho = 0$				$\rho = 0.9$			
H_0	$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$		$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$	
n	1%	5%	1%	5%	1%	5%	1%	5%
100	0.003	0.017	0.020	0.080	0.009	0.023	0.020	0.054
300	0.004	0.023	0.016	0.061	0.011	0.039	0.015	0.052
500	0.018	0.048	0.014	0.052	0.011	0.034	0.012	0.048
H_1	$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$		$L_{1n}(h_1, h_2)$		$L_{2n}(h_1, h_2)$	
100	0.142	0.245	0.500	0.614	0.325	0.467	0.884	0.921
300	0.262	0.373	0.655	0.756	0.590	0.701	0.946	0.958
500	0.368	0.484	0.717	0.819	0.672	0.755	0.957	0.974

5 Conclusions and discussions

This paper has proposed a simple and improved nonparametric test for specifying a unit–root structure involved in a nonlinear time series model. The proposed test has been compared with an existing one both theoretically and empirically. Meanwhile, the proposed test has been extended to a multivariate version for the case where both stationary and nonstationary regressors may be involved simultaneously in the same model. The finite sample performance of the proposed tests and their competitors have all been evaluated, while the theory for the multivariate version of the proposed test has not been established in this paper. Examples 4.2 and 4.3 also show that the tests under study are all applicable to the cases where there is some endogeneity between the regressors and the error terms, although we have not developed our theory to cover such an important case. Further discussion is left for future research.

6 Acknowledgments

The authors would like to thank the conference participants of the 3rd WISE–Humboldt Workshop on "High Dimensional and Nonstationary Time Series" held on 19–20 May 2012, Xiamen, China, for their constructive comments, particularly those from Professor Wolfgang Härdle and Professor Yongmiao Hong. The authors would also like to thank Dr Jiying Yin for his excellent computing assistance. Thanks also go to the Australian Research Council Discovery Grants Program under Grant Number: DP1096374.

7 Appendices

7.1 Appendix A

In order to prove Theorem 2.1, we need to introduce the following assumptions.

ASSUMPTION A.1. (i) *Let* $\{\mathcal{F}_t\}$ *be a* σ*–field generated by* $\{e_s : 1 \leq s \leq t\}$. *Let* $\{e_t\}$ *be a sequence of stationary martingale differences satisfying* $E\left[e_t|\mathcal{F}_{t-1}\right] = 0$ *and* $E\left[e_t^2|\mathcal{F}_{t-1}\right] = \sigma_e^2$ *almost surely, where* $0 < \sigma_e^2 < \infty$ *is some constant. In addition,* $\max_{t\geq 1} E\left[|e_t|^{2+\delta_0}|\mathcal{F}_{t-1}\right] < \infty$ *almost surely for some* $\delta_0 > 0$.

(ii) *Let* $p(u)$ *be the marginal density function of* e_1 *and* $p_\tau(v, w)$ *be the joint density of* $(e_1, e_{1+\tau})$ *for any* $\tau \geq 1$. *Suppose that* $p(u)$ *is continuous in* u *and* $p_\tau(v, w)$ *is continuous in* (v, w).

(iii) *For any positive integers* $1 \leq t_1 < t_2 < \cdots < t_n \leq n$, *define* $S_{ij} = \sum_{k=t_i+1}^{t_j} e_k$ *for* $1 \leq i < j \leq n$. *Let* $q_{ij}(w|v_{kl})$ *be the conditional density function of* $\frac{S_{ij}}{\sqrt{t_j - t_i}}$ *given* $S_{kl} = v_{kl}$ *for all* $1 \leq k \leq i-1$ *and* $1 \leq l \leq j-1$. *Suppose that there is some constant* $0 < C_s < \infty$ *such that* $\max_{(i,j,k,l)} \sup_{(w,v_{kl})} q_{ij}(w|v_{kl}) \leq C_s$. *Let* $g(\cdot)$ *be the marginal density of* $\frac{\sum_{s=1}^t e_s}{\sqrt{t}}$ *and then satisfy* $\max_{t\geq 1}\sup_u g\left(\frac{u}{\sqrt{t}}\right) \leq C_g$ *for some constant* $0 < C_g < \infty$.

ASSUMPTION A.2. (i) *Let* $K(\cdot)$ *be a symmetric and continuous probability kernel function satisfying for* $j = 0, 1, 2$,

$$\int_{-\infty}^{\infty} ||u||^j K^2(u)du < \infty \text{ and } \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} ||v||^j K(u+v)K(v)dv\right)^2 du < \infty.$$

(ii) *The bandwidth* h *satisfies* $h \to 0$, $nh^2 \to \infty$ *and* $nh^4 = o(1)$ *as* $n \to \infty$.

(iii) *Let* $\pi(\cdot)$ *be a known probability weight function such that* $\int_{-\infty}^{\infty} \pi^{2+\delta_0}(u)du < \infty$ *for the same* $\delta_0 > 0$ *as in Assumption A.1(i).*

(iv) *In addition, there is some function* $D(x)$ *satisfying* $\int_{-\infty}^{\infty} D(x)dx < \infty$ *such that* $|\pi(y) - \pi(x)| \leq D(x) \cdot |y - x|$ *for any* $(x, y) \in \Omega(\epsilon) = \{(x, y) : |y - x| \leq \epsilon, x, y \in R^1\}$, *where* $\epsilon > 0$ *is some small constant.*

Assumptions A.1 and A.2 are quite reasonable and easily verifiable. Assumption A.1(i) imposes the martingale difference structure to avoid imposing a kind of mixing condition on $\{e_t\}$. In this case, one will need to use some existing inequalities (such as, Lemma A.1 of Gao 2007) to deal with such terms: $E[e_ie_je_ke_l]$ for the case where i, j, k and l are different. As may be seen from the proof of Theorem 2.1 below, the derivations involving $E[e_t|\mathcal{F}_{t-1}] = 0$ may be replaced by $|E[e_se_t] - E[e_s]E[e_t]| \leq C\alpha^{\frac{4}{4+\delta_0}}(|t-s|)$ when an α–mixing condition is used, in which $\alpha(k)$ represents the mixing coefficient (as defined in Lemma A.1 of Gao 2007 for example). In summary, this paper adopts the martingale–difference assumption to avoid dealing with all sorts of technicalities that are part of the consequence of imposing a mixing condition on $\{e_t\}$.

Assumption A.1(ii) is standard, and Assumption A.1(iii) basically imposes the boundedness of the conditional density function for all (t_i, t_j). When $t_j - t_i \to \infty$, Lemma B.1 in Appendix B below implies that $q_{ij}(w|v) \to \phi(w)$, where $\phi(\cdot)$ is the density function of the standard normal random variable $U \sim N(0,1)$. When (i,j) is fixed and the support of x_s is compact, the boundedness of $g_{ij}(w|v)$ follows from the common assumption of the continuity of $g_{ij}(w|v)$ in (w,v). When (i,j) is fixed and $g_{ij}(w|v) \to 0$ as $w \to \infty$, the boundedness of $g_{ij}(w|v)$ also follows trivially. The second part of Assumption A.1(iii) follows similarly and trivially. In summary, it is not unreasonable to assume the boundedness in Assumption A.1(iii). Assumption A.2 is also quite standard except that Assumption A.2(ii) imposes a stronger condition than $nh^5 = O(1)$. Existing literature, such as Gao *et al* (2009a), have to assume $nh^{\frac{10}{3}} = 0$.

We now introduce some necessary lemmas before we prove Theorem 2.1.

LEMMA A.1. *Let the conditions of Theorem 2.1 hold. Under H_0, we have as $n \to \infty$*

$$\begin{aligned}
\widehat{S}_{1n} &= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h}\right) \widehat{e}_t^2 \right) \pi(x)dx \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h}\right) \ e_t^2 \right) \pi(x)dx \equiv S_{1n}, \qquad \text{(A.1)} \\
\widehat{S}_{2n} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_s \widehat{e}_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} e_s \ e_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x)dx \equiv S_{2n}. \qquad \text{(A.2)}
\end{aligned}$$

LEMMA A.2. *Let the conditions of Theorem 2.1 hold. Under H_1, we have as $n \to \infty$*

$$\begin{aligned}
\widehat{S}_{1n} &= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h}\right) \widehat{e}_t^2 \right) \pi(x)dx \\
&\geq \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h}\right) \Delta_n^2(x_t) \right) \pi(x)dx + o_P(1), \qquad \text{(A.3)} \\
\widehat{S}_{2n} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} \widehat{e}_s \widehat{e}_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x)dx \qquad \text{(A.4)} \\
&\geq \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \Delta_n(x_s)\Delta_n(x_t)\pi(x)dx + o_P(1).
\end{aligned}$$

LEMMA A.3. *Let Assumptions A.1 and A.2 hold. Then as $n \to \infty$*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \psi(x_t) \to_D L_{B_u}(1,0) \cdot \sigma_e^2 \cdot \int_{-\infty}^{\infty} \psi(x)dx, \qquad \text{(A.5)}$$

where $\psi(\cdot) = \pi(\cdot)$ or $D(\cdot)$.

The proofs of Lemmas A.1–A.3 are given in Appendix B below. We now give the proof of Theorem 2.1.

Proof of Theorem 2.1. Without loss of generality, we let $\sigma_e^2 \equiv 1$ throughout Appendices A and B. In view of Lemma A.1, in order to prove Theorem 2.1, it suffices to show that as $n \to \infty$

$$S_{1n} = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}h} \sum_{t=1}^{n} K^2 \left(\frac{x_t - x}{h} \right) \; e_t^2 \right) \pi(x) dx \to_D L_{B_u}(1,0) \cdot \left(\int_{-\infty}^{\infty} K^2(u) du \right), \tag{A.6}$$

$$S_{2n} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h} \sum_{t=1}^{n} \sum_{s=1,\neq t}^{n} e_s \; e_t \cdot K \left(\frac{x_t - x}{h} \right) K \left(\frac{x_s - x}{h} \right) \pi(x) dx = o_P(1). \tag{A.7}$$

We start with the proof of (A.6). Under Assumptions 2.1 and 2.2, using Lemma A.3, we have as $n \to \infty$

$$\begin{aligned}
S_{1n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \int_{-\infty}^{\infty} K^2(u) \pi(x_t - uh) du = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \int_{-\infty}^{\infty} K^2(u) \pi(x_t) du \\
&+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \int_{-\infty}^{\infty} K^2(u) \left(\pi(x_t - uh) - \pi(x_t) \right) du \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \pi(x_t) \int_{-\infty}^{\infty} K^2(u) du + O(h) \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 D(x_t) \int_{-\infty}^{\infty} |u| K^2(u) du \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t^2 \pi(x_t) \int_{-\infty}^{\infty} K^2(u) du + o_P(1) \to_D L_{B_u}(1,0) \cdot \int_{-\infty}^{\infty} K^2(u) du,
\end{aligned} \tag{A.8}$$

which completes the proof of (A.6).

We then prove (A.7). Let

$$\begin{aligned}
B(s,t) &\equiv B(x_s, x_t) = \int_{-\infty}^{\infty} K \left(\frac{x_t - x}{h} \right) K \left(\frac{x_s - x}{h} \right) \pi(x) dx, \\
A_t &\equiv A(x_1, \cdots, x_t; e_1, \cdots, e_{t-1}) = \frac{2}{h} \sum_{s=1}^{t-1} B(s,t) \; e_s, \\
S_{2n} &= \frac{1}{\sqrt{n}} \sum_{t=2}^{n} A_t \; e_t.
\end{aligned} \tag{A.9}$$

Similarly to the derivations in (A.8), we have

$$\begin{aligned}
B(s,t) &= \int_{-\infty}^{\infty} K \left(\frac{x_t - x_s}{h} + \frac{x_s - x}{h} \right) K \left(\frac{x_s - x}{h} \right) \pi(x) dx \\
&= h \cdot \int_{-\infty}^{\infty} K \left(\frac{x_t - x_s}{h} + u \right) K(u) \pi(x_s - uh) du \\
&= h \; \pi(x_s) \cdot L \left(\frac{x_t - x_s}{h} \right) + h \cdot \int_{-\infty}^{\infty} K \left(\frac{x_t - x_s}{h} + u \right) K(u) \left(\pi(x_s - uh) - \pi(x_s) \right) du \\
&\equiv h \left(B_1(s,t) + B_2(s,t) \right),
\end{aligned} \tag{A.10}$$

where $L(v) = \int_{-\infty}^{\infty} K(u+v) K(u) du$.

We then have

$$\begin{aligned}
S_{2n} = \frac{1}{\sqrt{n}} \sum_{t=2}^{n} A_t \; e_t &= \frac{2}{\sqrt{n}} \sum_{t=2}^{n} \left(\sum_{s=1}^{t-1} B_1(s,t) e_s \right) e_t + \frac{2}{\sqrt{n}} \sum_{t=2}^{n} \left(\sum_{s=1}^{t-1} B_2(s,t) e_s \right) e_t \\
&\equiv S_{2n1} + S_{2n2}.
\end{aligned} \tag{A.11}$$

We first deal with $E\left[S_{2n1}^2\right]$. Observe that

$$
\begin{aligned}
E\left[S_{2n1}^2\right] &= \frac{8}{n}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[\pi^2(x_s)L^2\left(\frac{x_t-x_s}{h}\right)e_s^2e_t^2\right] \\
&+ \frac{8}{n}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{v=1}^{s-1} E\left[\pi(x_s)\pi(x_v)L\left(\frac{x_t-x_s}{h}\right)L\left(\frac{x_t-x_v}{h}\right)e_se_ve_t^2\right] \\
&+ \frac{8}{n}\sum_{t_1=3}^{n} E\left[\left(\sum_{t_2=2}^{t_1-1}\left(\sum_{s_1=1}^{t_1-1}\sum_{s_2=1}^{t_2-1} B_1(s_1,t_1)B_1(s_2,t_2)e_{s_1}e_{s_2}\right)e_{t_2}\right)e_{t_1}\right] \\
&\equiv J_{1n}+J_{2n}+J_{3n}. \qquad \text{(A.12)}
\end{aligned}
$$

Before we evaluate the order of $E\left[J_{1n}^2\right]$, we introduce the following definitions. Let $u_t = y_t - x_t = y_t - y_{t-1}$. Note that $u_t = e_t$ under H_0. Let $p_t(x)$ and $q_t(x)$ be the probability density functions of x_t and $\frac{x_t}{\sqrt{t}}$, respectively. We then have $p_t(x) = \frac{1}{\sqrt{t}}q_t\left(\frac{x}{\sqrt{t}}\right)$. For $s < t$, let $q_{st}(u|v)$ be the conditional density function of $\frac{x_t}{\sqrt{t}}$ given $\frac{x_s}{\sqrt{s}}$.

Since $x_t = \sum_{i=1}^{t-1} u_i$, $x_t - x_s = \sum_{j=s}^{t-1} u_j$ and $x_t = x_s + u_s + \sum_{j=s+1}^{t-1} u_j$, additionally, one will need to involve the joint distributions of u_s, $v_s = x_s$ and $w_{st} = \sum_{j=s+1}^{t-1} u_j$ to evaluate the order of J_{1n} in equation (A.12). Let $p_{st}(u,v,w)$ and $q_{st}(u,v,w)$ be the joint distributions of (u_s, v_s, w_{st}) and $\left(u_s, \frac{x_s}{\sqrt{s}}, \frac{w_{st}}{\sqrt{t-s-1}}\right)$, respectively. Then, we have the following expressions involving the joint density functions of $\left(\frac{w_{st}}{\sqrt{t-s-1}}, \frac{x_s}{\sqrt{s}}, u_s\right)$:

$$
\begin{aligned}
p_{st}(u,v,w) &= \frac{1}{\sqrt{s}\sqrt{t-s-1}}q_{st}\left(u, \frac{v}{\sqrt{s}}, \frac{w}{\sqrt{t-s-1}}\right) \\
&= \frac{1}{\sqrt{s}\sqrt{t-s-1}}q_{st}\left(\frac{w}{\sqrt{t-s-1}}|u, \frac{v}{\sqrt{s}}\right) q_s\left(\frac{v}{\sqrt{s}}|u\right) p(u), \qquad \text{(A.13)}
\end{aligned}
$$

where $q_s(\cdot|u)$ denotes the conditional density of $\frac{x_s}{\sqrt{s}}$ given u, and $p(u)$ is the marginal density function of u_s.

In view of the notational introduction just below (A.12) and Assumption A.1(iii), we have as $n \to \infty$

$$
\begin{aligned}
J_{1n} &= \frac{8}{n}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[\pi^2(x_s)L^2\left(\frac{x_t-x_s}{h}\right)e_s^2\right] \\
&= \frac{8}{n}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}\sqrt{t-s}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u^2\pi^2(v)L\left(\frac{u+w}{h}\right) \\
&\times q_{st}\left(\frac{w}{\sqrt{t-s}}|\frac{v}{\sqrt{s}}, u\right) q_s\left(\frac{v}{\sqrt{s}}|u\right) p(u)dwdvdu \\
&= \frac{8h}{n}\sum_{t=2}^{m}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}\sqrt{t-s}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u^2\pi^2(v)L\left(w\right) \\
&\times q_{st}\left(\frac{wh-u}{\sqrt{t-s}}|\frac{v}{\sqrt{s}}, u\right) q_s\left(\frac{v}{\sqrt{s}}|u\right) p(u)dwdvdu
\end{aligned}
$$

$$
\begin{aligned}
&\leq\ C\int_{-\infty}^{\infty}u^2p(u)du\cdot\int_{-\infty}^{\infty}\pi^2(v)dv\int_{-\infty}^{\infty}L(w)dw\cdot\frac{8h}{n}\sum_{t=2}^{m}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}\sqrt{t-s}}\\
&= C(1+o(1))\cdot\frac{nh}{n} = O(h) = o(1).
\end{aligned}
\tag{A.14}
$$

Similarly to equation (A.13), one may deal with J_{2n}. Let $z_{sv} = \sum_{k=v+1}^{s-1} u_k$. Note that $x_t - x_v = x_t - x_s + x_s - x_v = u_s + \sum_{j=s+1}^{t-1} u_j + u_v + \sum_{k=v+1}^{s-1} u_k$. The joint density functions of $\left(\frac{w_{st}}{\sqrt{t-s-1}}, \frac{z_{sv}}{\sqrt{s-v-1}}, u_s, u_v\right)$ are given by

$$
\begin{aligned}
&p_{st}(u_1,u_2,w,z) = \frac{1}{\sqrt{s}\sqrt{t-s-1}}\frac{1}{\sqrt{s-v-1}}q_{st}\left(u_1,u_2,\frac{z}{\sqrt{s-v-1}},\frac{w}{\sqrt{t-s-1}}\right)\\
&=\ \frac{1}{\sqrt{s-v-1}\sqrt{t-s-1}}q_{st}\left(\frac{w}{\sqrt{t-s-1}}|u_1,u_2,\frac{z}{\sqrt{s-v-1}}\right)\\
&\times\ q_{sv}\left(\frac{z}{\sqrt{s-v-1}}|u_1,u_2\right)p(u_1,u_2),
\end{aligned}
\tag{A.15}
$$

where $p(u_1, u_2)$ denotes the joint density of (u_v, u_s).

In a similar fashion to the derivations in (A.14), we have as $n \to \infty$

$$
\begin{aligned}
&E\left[|J_{2n}|\right] \leq \frac{8}{n}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{v=1}^{s-1}E\left[\pi(x_s)\pi(x_v)L\left(\frac{x_t-x_s}{h}\right)L\left(\frac{x_t-x_v}{h}\right)|e_se_v|e_t^2\right]\\
&=\ \frac{8}{n}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{v=1}^{s-1}\frac{1}{\sqrt{v-1}\sqrt{s-v-1}\sqrt{t-s-1}}\\
&\cdots\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|v_1v_2|\pi(v_5)\pi(v_2+v_4+v_5)L\left(\frac{v_1+v_3}{h}\right)L\left(\frac{v_2+v_4+v_1+v_3}{h}\right)\\
&\times\ q_{st}\left(\frac{v_3}{\sqrt{t-s-1}}|\frac{v_4}{\sqrt{s-v-1}},\frac{v_5}{\sqrt{v-1}},v_2,v_1\right)q_{sv}\left(\frac{v_4}{\sqrt{s-v-1}}|\frac{v_5}{\sqrt{v-1}},v_2,v_1\right)\\
&\times\ q_v\left(\frac{v_5}{\sqrt{v-1}}|v_2,v_1\right)p(v_1,v_2)dv_5dv_4dv_3dv_2dv_1\\
&\leq C(1+o(1))\cdot\int_{-\infty}^{\infty}\pi^2(v_5)dv_5\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}L(v_3)L(v_3+v_4)dv_4dv_3\\
&\cdot\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|v_1v_2|p(v_1,v_2)dv_1dv_2\\
&\times\ \frac{8h^2}{n}\sum_{t=3}^{n}\sum_{s=2}^{t-1}\sum_{v=1}^{s-1}\frac{1}{\sqrt{v-1}\sqrt{s-v-1}\sqrt{t-s-1}} = O\left(\sqrt{n}h^2\right) = o(1)
\end{aligned}
\tag{A.16}
$$

when $nh^4 = o(1)$.

Meanwhile, it is obvious that

$$
J_{3n} = 0. \tag{A.17}
$$

Equations (A.12)–(A.16) then imply as $n \to \infty$

$$
E\left[S_{2n1}^2\right] = o(1). \tag{A.18}
$$

In a similar way to the derivations of (A.12)–(A.16), using Assumption 2.2(iv) in particular, we have as $n \to \infty$

$$
E\left[S_{2n2}^2\right] \leq Ch, \quad \text{which deduces } S_{2n2} = o_P(1). \tag{A.19}
$$

The proof of Theorem 2.1 follows from equations (A.6), (A.7), (A.17)–(A.19).

Appendix B

This appendix gives the proofs of Lemmas A.1–A.3 and then the derivations of equations (2.18)–(2.20) are given in the last part of this appendix.

We first introduce a very useful lemma, which has been used in the verification of Assumption A.1(iii). The proof of Lemma B.1 below follows from some standard central limit theorems (see, for example, Awad 1981; Denker and Gordin 2003).

Let $\{e_j\}$ satisfy Assumption A.1(i) and $\widehat{\phi}_k(x)$ be the probability density function of $L_k = \frac{1}{\sqrt{k}\ \sigma_e}\sum_{j=1}^k e_j$ and $\widehat{\phi}_k(x|\mathcal{F}_{k-1})$ be the conditional probability density function of L_k given $\mathcal{F}_{k-1}$, where $\{\mathcal{F}_k\}$ is a sequence of σ–fields generated by $\{e_i, 1 \le i \le k\}$ and σ_e^2 is the same as in Assumption A.1(i).

Lemma B.1. *Under Assumption A.1(i)(ii), we have as $k \to \infty$*

$$\sup_{x\in R^1}\left|\widehat{\phi}_k(x) - \phi(x)\right| \to 0 \quad \textit{and} \quad \sup_{x\in R^1}\left|\widehat{\phi}_k(x|\mathcal{B}_{k-1}) - \phi(x)\right| \to 0 \quad \textit{almost surely}, \tag{B.1}$$

where $\phi(\cdot)$ is the probability density of the standard normal random variable $U \sim N(0,1)$.

Proof of Lemma A.1. Recall that under H_0 : $\widehat{e}_t = y_t - x_t = y_t - y_{t-1} = e_t$. Thus, the verification of Lemma A.1 follows trivially.

Proof of Lemma A.2. We only prove equation (A.4), as the proof of (A.3) follows similarly. Let

$$\widehat{\varepsilon}_t(x) = \widehat{e}_t\ K\left(\frac{x_t - x}{h}\right) \quad \text{and} \quad \widehat{e}_t = e_t + \Delta_n(x_t) \tag{B.2}$$

under H_1, where $\Delta_n(x) = \delta_n \Delta(x)$ is the same as defined in (2.17) and (2.18).

Then, we have under H_1:

$$\begin{aligned}
\widehat{T}_n =\ & \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h}\sum_{t=1}^n\sum_{s=1}^n \widehat{e}_s\widehat{e}_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right)\pi(x)dx \\
=\ & \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h}\sum_{t=1}^n\sum_{s=1}^n e_s e_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right)\pi(x)dx \\
+\ & \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h}\sum_{t=1}^n\sum_{s=1}^n \Delta_n(x_t)e_s \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right)\pi(x)dx \\
+\ & \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h}\sum_{t=1}^n\sum_{s=1}^n \Delta_n(x_s)e_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right)\pi(x)dx \\
+\ & \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}h}\sum_{t=1}^n\sum_{s=1}^n \Delta_n(x_s)\Delta_n(x_t) \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right)\pi(x)dx \\
\equiv\ & \sum_{j=1}^4 \widehat{T}_{jn}.
\end{aligned} \tag{B.3}$$

We will show that under H_1:

$$\widehat{T}_{kn} = o_P\left(\delta_n^2 \sqrt{n} h\right) \quad \text{for } k = 2, 3. \tag{B.4}$$

We need only to prove (B.4) for either $k = 2$ or $k = 3$. Meanwhile, similarly to the derivations in (A.14), in order to deal with $\widehat{T}_{3n}$, it suffices to show that as $n \to \infty$

$$E\left[\delta_n \sum_{t=1}^{n} \left(\sum_{s=1}^{n} \Delta(x_s)\pi(x_s) L\left(\frac{x_t - x_s}{h}\right)\right) e_t\right]^2$$
$$= C\ (1 + o(1)\delta_n^2 n h, \tag{B.5}$$

which follows similarly from the proof of (A.14). Equation (B.5) then implies that as $n \to \infty$

$$\widehat{T}_{3n} = O_P\left(\delta_n \sqrt{\sqrt{n} h}\right) = o_P\left(\delta_n^2 \sqrt{n} h\right), \tag{B.6}$$

which, along with the derivations in (B.3)–(B.6), shows that under H_1, as $n \to \infty$

$$\widehat{T}_n = \int_{-\infty}^{\infty} \frac{1}{\sqrt{n} h} \sum_{t=1}^{n} \sum_{s=1}^{n} \widehat{e}_s \widehat{e}_t \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x) dx \tag{B.7}$$
$$\geq \int_{-\infty}^{\infty} \frac{1}{\sqrt{n} h} \sum_{t=1}^{n} \sum_{s=1}^{n} \Delta_n(x_s)\Delta_n(x_t) \cdot K\left(\frac{x_t - x}{h}\right) K\left(\frac{x_s - x}{h}\right) \pi(x) dx + o_P(1),$$

which completes the proof of Lemma A.2.

Proof of Lemma A.3. In view of existing results (such as, Theorem 2.1 of Wang and Phillips 2009), in order to prove Lemma A.3, it suffices to show that as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi(x_t) e_t^2 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi(x_t) E[e_t^2] + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi(x_t)\left(e_t^2 - E\left[e_t^2\right]\right)$$
$$= \frac{\sigma_e^2}{\sqrt{n}} \sum_{t=1}^{n} \psi(x_t) + o_P(1), \tag{B.8}$$

which follows from

$$\frac{1}{n}\left[\sum_{t=1}^{n} \psi(x_t)\left(e_t^2 - E\left[e_t^2\right]\right)\right]^2 = \frac{1}{n} \sum_{t=1}^{n} E\left[\psi(x_t)\left(e_t^2 - E\left[e_t^2\right]\right)\right]^2$$
$$+ \frac{2}{n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E\left[\psi(x_t)\psi(x_s)\left(e_s^2 - E\left[e_s^2\right]\right)\left(e_t^2 - E\left[e_t^2\right]\right)\right]$$
$$= \frac{1}{n} \sum_{t=1}^{n} E\left[\psi(x_t)\left(e_t^2 - E\left[e_t^2\right]\right)\right]^2$$
$$\leq \frac{C}{n} \sum_{t=1}^{n} E\left[\psi^2(x_t)\right] = \frac{C}{n} \sum_{t=1}^{n} \int_{-\infty}^{\infty} \psi^2(x) p_t(x) dx$$
$$= \frac{C}{n} \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} \psi^2(x) q_t\left(\frac{x}{\sqrt{t}}\right) dx = O\left(\frac{1}{\sqrt{n}}\right) = o(1)$$

by Assumptions A.1 and A.2, where $p_t(\cdot)$ and $q_t(\cdot)$ denote the marginal densities of x_t and $\frac{x_t}{\sqrt{t}}$, respectively, and we have used the relationship of $p_t(x) = \frac{1}{\sqrt{t}} q_t\left(\frac{x}{\sqrt{t}}\right)$. Equation (B.8) then completes the proof of Lemma A.3.

Derivations of equations (2.19) and (2.20). Similarly to the proof of Lemma A.2, under H_1, we have as $n \to \infty$

$$\begin{aligned}
&\sum_{t=1}^{n}\sum_{s=1}^{n} K\left(\frac{x_t - x_s}{h}\right)\ \widehat{e}_s\ \widehat{e}_t = \sum_{t=1}^{n}\sum_{s=1}^{n} K\left(\frac{x_t - x_s}{h}\right)\ e_s\ e_t \\
&+\delta_n^2 \sum_{t=1}^{n}\sum_{s=1}^{n} K\left(\frac{x_t - x_s}{h}\right)\ \Delta(x_s)\ \Delta(x_t) + o_P(1) \\
&\geq\ \delta_n^2 \cdot \sum_{t=1}^{n}\sum_{s=1}^{n} K\left(\frac{x_t - x_s}{h}\right)\ \Delta(x_s)\ \Delta(x_t) + o_P(1) \\
&\equiv\ \delta_n^2 \cdot Q_n(h) + o_P(1), \qquad\qquad (B.9)
\end{aligned}$$

where $Q_n(h) = \sum_{t=1}^{n}\sum_{s=1}^{n} K\left(\frac{x_t - x_s}{h}\right)\ \Delta(x_s)\ \Delta(x_t)$.

Straightforward derivations imply that as $n \to \infty$

$$E\left[Q_n(h_1, h_2)\right] = C_1(1 + o(1))\ nh \qquad (B.10)$$

for some $C_1 > 0$.

Similarly, we may show that as $n \to \infty$

$$\sigma_{2n}^2 \equiv E\left[\sum_{t=1}^{n}\sum_{s=1}^{n} K^2\left(\frac{x_t - x_s}{h}\right) e_s^2 e_t^2\right] = C_2(1 + o(1))\ n^{\frac{3}{2}} h. \qquad (B.11)$$

Equations (B.9)–(B.11) thus complete an outline of the derivation of (2.19).

In view of the proof of Lemma A.2, in order to complete the derivation of (2.20), it suffices to show that as $n \to \infty$ and for some $C_2 > 0$

$$\begin{aligned}
&\frac{1}{\sqrt{n}h} \sum_{t=1}^{n}\sum_{s=1}^{n} E\left[\int_{-\infty}^{\infty} K\left(\frac{x_t - x}{h_1}\right) K\left(\frac{x_s - x}{h}\right) \Delta_n(x_s)\Delta_n(x_t)\right] \pi(x)dx \\
&=\ C_2(1 + o(1))\ \sqrt{n}h, \qquad\qquad (B.12)
\end{aligned}$$

which follows similarly to the derivations used elsewhere. Thus, we omit these details. Such details are available upon request, however.

REFERENCES

Awad, A. A., 1981. Conditional central limit theorems for martingales and reversed martingales. The INdian Journal of Statistics Series A 43, 100–106.

Denker, M., Gordin, M., 2003. On conditional central limit theorems for stationary processes, In Lecture Notes–Monograph Series, Vol. 41, *Probability, Statistics and Their Applications: Papers in Honor of Rabi Bhattacharya*, pp. 133–152.

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

Gao, J., 2012. Identification, estimation and specification in a class of semi–linear time series models. Working paper available at http://ideas.repec.org/p/pra/mprapa/39256.html.

Gao, J., King, M. L., 2011. A new test in linear models against nonparametric errors. Working paper available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp20-11.pdf.

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., Tjøstheim, D., Yin, J., 2011. Estimation in threshold autoregressive models with a stationary and a unit root regime. Available at http://ideas.repec.org/p/msh/ebswps/2011-21.html.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Li, D., Gao, J., Chen, J., Lin, Z., 2011. Nonparametric specification testing in nonlinear and nonstationary time series models. Working paper available at http://www.jitigao.com/page1006.aspx.

Sun, Y., Li, Q., 2012. Nonparametric and Semiparametric Estimation and Hypothesis Testing with Nonstationary Time Series. Forthcoming in Handbook on Applied Nonparametric Econometrics and Statistics. Oxford University Press.

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

Wang, Q. Y., Phillips, P. C. B., 2012. Specification testing for nonlinear cointegrating regression. Annals of Statistics 40, 727–758.