A New Test in Parametric Linear Models with Nonparametric Autoregressive Errors

By Jiti Gao [1] and Maxwell King

The University of Western Australia and Monash University

Abstract: This paper considers a class of parametric models with nonparametric autoregressive errors. A new test is proposed and studied to deal with the parametric specification of the nonparametric autoregressive errors with either stationarity or nonstationarity. Such a test procedure can initially avoid misspecification through the need to parametrically specify the form of the errors. In other words, we propose estimating the form of the errors and testing for stationarity or nonstationarity simultaneously. We establish asymptotic distributions of the proposed test. Both the setting and the results differ from earlier work on testing for unit roots in parametric time series regression. We provide both simulated and real–data examples to show that the proposed nonparametric unit–root test works in practice.

Key words: Error process, nonparametric method, nonlinear time series, random walk, unit root test.

[1] Jiti Gao, School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia. Email: jiti.gao@uwa.edu.au.

1. Introduction

Consider a parametric linear model of the form

$$Y_t = X_t^\tau \beta + v_t, \quad t = 1, 2, \cdots, T, \tag{1.1}$$

where T is the sample size of the time series data $\{Y_t : 1 \leq t \leq T\}$, $X_t = (f_1(t), \cdots, f_p(t))^\tau$ is a vector of parametric deterministic trend functions, $\beta = (\beta_1, \cdots, \beta_p)^\tau$ is a vector of unknown parameters, $\{v_t\}$ is a sequence of time series residuals. Existing studies mainly discuss tests for the case where $\{v_t\}$ satisfies the first–order autoregressive (AR(1)) model of the form $v_t = \rho v_{t-1} + u_t$ with $\{u_t\}$ being a sequence of independent and identically distributed (i.i.d.) errors. Discussion about tests for $|\rho| < 1$ may be found in the survey papers by King (1987), King and Wu (1997) and King (2001).

For the case of $\rho = 1$, there has been much interest in both theoretical and empirical analysis of economic and financial time series with unit roots during the past three decades or so. Various tests for unit roots have been proposed and studied both theoretically and empirically. Models and methods used have been based initially on parametric linear autoregressive moving average representations with or without trend components. Existing studies may be found in the survey paper by Phillips and Xiao (1998).[2]

As pointed out in the literature (Vogelsang 1998; Zheng and Basher 1999), there are cases where there is no priori knowledge about either the form of the residuals or whether the residuals are I(0) or I(1). This motives us to consider using a nonparametric autoregressive error model of the form

$$v_t = g(v_{t-1}) + u_t, \tag{1.2}$$

where $g(\cdot)$ is an unknown function defined over $R^1 = (-\infty, \infty)$, $\{u_t\}$ may be a sequence of strictly stationary errors with mean zero and finite variance $\sigma_u^2 = E[u_1^2]$, and v_s and u_t are mutually independent for all $t > s \geq 1$ and $v_0 = 0$.

Combining model (1.2) into model (1.1) produces a semiparametric time series model of the form

$$Y_t = X_t^\tau \beta + v_t \quad \text{with} \quad v_t = g(v_{t-1}) + u_t. \tag{1.3}$$

[2] Earlier studies include Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Phillips (1987), Phillips and Perron (1988), Dufour and King (1991), Kwiatkowski *et al* (1992), Phillips (1997), Lobato and Robinson (1998), and Robinson (2003).

To the best of our knowledge, estimation problems for model (1.3) have only been discussed in Hidalgo (1992).

This paper is interested in testing

$$\mathcal{H}_0: \ g(v) = f_0(v, \theta_0) \quad \text{versus} \quad \mathcal{H}_1: \ g(v) = f_1(v, \theta_1) \tag{1.4}$$

for all $v \in R^1$, where $f_0(v, \theta_0)$ is a known parametric function indexed by a vector of unknown parameters θ_0, and $f_1(v, \theta_1)$ is either an unknown nonparametric function or a semiparametric function indexed by a vector of unknown parameters θ_1.

Forms of $f_0(v, \theta_0)$ include the case of $f_0(v, \theta_0) \equiv 0$. This implies that $\{v_t\}$ is a sequence of i.i.d. errors. When a suitable θ_0 is chosen such that $f_0(v, \theta_0) = v$, it means that there is a unit root in $\{v_t\}$. Forms of $f_i(v, \theta_i)$ may be chosen suitably to include various existing cases such as a parametric AR(1) model of the form $v_t = \theta_0 v_{t-1} + u_t$ against a semparametric AR(1) model of the form $v_t = \theta_1 v_{t-1} + \psi(v_{t-1}) + u_t$ with $\psi(\cdot)$ being unknown nonparametrically.

In addition, forms of $f_1(v, \theta_1)$ include existing parametric nonlinear functions, such as $f_1(v, \theta_1) = \rho_1 v + \gamma_1 v \ \left(1 - \exp\left\{-\eta_1 v^2\right\}\right)$ as discussed in Kapetanios, Shin and Snell (2003), where $\theta_1 = (\rho_1, \gamma_1, \eta_1)$ is a vector of unknown parameters.

Our discussion in this paper focuses on the following two cases:

Case A: $f_0(v, \theta_0)$ is chosen as $f_0(v, \theta_0) = v$ implying $v_t = v_{t-1} + u_t$ under $\mathcal{H}_0$ while the form of $f_1(v, \theta_1)$ is chosen such that $\{v_t\}$ is a sequence of strictly stationary errors under $\mathcal{H}_1$; and

Case B: The form of each of $f_i(v, \theta_i)$ for $i = 0, 1$ is suitably chosen such that $\{v_t\}$ is a sequence of strictly stationary errors under either $\mathcal{H}_0$ or $\mathcal{H}_1$.

This paper then proposes a nonparametric test for $\mathcal{H}_0$ versus $\mathcal{H}_1$. Unlike existing parametric tests, the proposed test has an asymptotically normal distribution even when $\{v_t\}$ is a sequence of random walk errors. The main advantage of the proposed nonparametric unit root test over existing tests in the parametric case is that it can initially avoid misspecification through the need to parametrically specify the form of $\{v_t\}$ as $v_t = \rho v_{t-1} + u_t$. Such a test may be viewed as a nonparametric counterpart of existing parametric tests proposed in the literature.

Theoretical properties for the proposed nonparametric test are established. Our finite sample results show that the conventional Dickey–Fuller type test is more powerful than the proposed nonparametric unit root test when the alternative model is an AR(1) model of the form $v_t = \rho v_{t-1} + u_t$. When the alternative is a parametric nonlinear autoregressive model, however, the conventional parametric unit root test seems to be inferior to the proposed nonparametric unit root test in the sense that it is less powerful than the proposed nonparametric unit root test.

The rest of the paper is organised as follows. Section 2 establishes a nonparametric test as well as its asymptotic distributional results. A bootstrap simulation procedure is proposed in Section 3. Section 4 presents two examples to show how to implement the proposed test in practice. Section 5 gives some extensions. Mathematical details are relegated to the appendix.

2. A nonparametric test

In the parametric linear case where $v_t = \rho v_{t-1} + u_t$, existing tests for $\rho = 0$ include various versions of the DW test proposed in Durbin and Watson (1950, 1951) as reviewed in King (1987), King and Wu (1997), King (2001) and others. Various extensions of the DF test for $\rho = 1$ proposed in Dickey and Fuller (1979, 1981) have been discussed in Phillips and Xiao (2003), and others.

In order to deal with the nonparametric case where $v_t = g(v_{t-1}) + u_t$, we propose using a nonparametric version of some existing parametric tests. Assuming that $\{v_t\}$ were observable, we would have a parametric autoregressive model of the form

$$v_t = f_0(v_{t-1}, \theta_0) + u_t \tag{2.1}$$

with $E[u_t|v_{t-1}] = 0$ under $\mathcal{H}_0$. We thus have

$$E\left[u_t E\left(u_t|v_{t-1}\right)\pi_t(v_{t-1})\right] = E\left[\left(E^2(u_t|v_{t-1})\right)\pi_t(v_{t-1})\right] = 0 \tag{2.2}$$

under $\mathcal{H}_0$, where $\{\pi_t(\cdot)\}$ is the marginal density function of $\{v_{t-1}\}$.

This would suggest using a normalized kernel–based sample analogue of (2.2) of the form

$$L_T = L_T(h) = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\; K_h(v_{s-1} - v_{t-1})\; u_t}{\sqrt{2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s^2\; K_h^2(v_{s-1} - v_{t-1})\; u_t^2}}, \tag{2.3}$$

where $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$ with $K(\cdot)$ being a probability kernel function and h a bandwidth parameter. Obviously, $L_T(h)$ is invariant to $\sigma_u^2 = E[u_1^2]$.

$L_T(h)$ may be regarded as a nonparametric counterpart of the DW test for the stationary case (see (5) of Dufour and King 1991) and the DF test for the unit–root case (see (17) of Dufour and King 1991). For the case where the time series involved is strictly stationary, similar versions have been used for nonparametric testing of serial correlation (Li and Hsiao 1998) and nonparametric specification of time series (Gao 2007[3]). For the nonstationary case, Gao *et al.* (2006) were among the first to suggest using a version similar to (2.3) for a unit–root test of the form $H_0: P(m(Y_{t-1}) = Y_{t-1}) = 1$ in a nonlinear time series model of the form $Y_t = m(Y_{t-1}) + u_t$.

Since $\{v_t\}$ and $\{u_t\}$ are unobservable, in practice $L_T(h)$ will need to be replaced by

$$\widehat{L}_T = \widehat{L}_T(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s \; K_h(\widehat{v}_{s-1} - \widehat{v}_{t-1}) \; \widehat{u}_t}{\sqrt{2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^2 \; K_h^2(\widehat{v}_{s-1} - \widehat{v}_{t-1}) \; \widehat{u}_t^2}}, \tag{2.4}$$

where $\widehat{u}_t = \widehat{v}_t - f_0(\widehat{v}_{t-1}, \widehat{\theta}_0)$, in which $\widehat{v}_t = Y_t - X_t^\tau \widehat{\beta}$, and $\widehat{\theta}_0$ and $\widehat{\beta}$ are consistent estimators of θ_0 and β under $\mathcal{H}_0$, respectively.

To establish the asymptotic distribution of $\widehat{L}_T(h)$, we need to introduce the following assumption, in addition to the more technical conditions–Assumptions A.1–A.3 listed in the appendix below.

Assumption 2.1. *Assume that $\{u_t\}$ is a sequence of independent and identically distributed normal errors with $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ and $0 < E[u_t^4] = \mu_4 < \infty$.*

Assumption 2.1 assumes that $\{u_t\}$ is an independent $N(0, \sigma_u^2)$ error. As a result, $v_t = \sum_{s=1}^t u_s \sim N(0, t\sigma_u^2)$ under $\mathcal{H}_0$ for Case A. However, we believe that the normality assumption could be removed if the so–called "local–time approach" developed by Phillips and Park (1998) or the Markov splitting technique of Karlsen and Tjøstheim (2001) could be employed in establishing Theorem 2.1 below. As the potential of the two alternative approaches requires further study, we use Assumption 2.1 in Case A throughout this paper.

To deal with Case B for the stationary case, we may relax that $\{u_t\}$ is just a sequence of strictly stationary and α–mixing errors as assumed in Assumption 2.2 below.

Assumption 2.2. *Assume that both $\{u_t\}$ and $\{v_t\}$ under $\mathcal{H}_0$ are strictly stationary and α–mixing with mixing coefficient*

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$$

[3]Such tests are extensions of existing tests proposed in Zheng (1996), Li and Wang (1998), Li (1999), and Fan and Linton (2003).

for all $s, t \geq 1$, where $\{\Omega_i^j\}$ is a sequence of σ–fields generated by either $\{u_s : i \leq s \leq j\}$ or $\{v_s : i \leq s \leq j\}$. There exist constants $c_r > 0$ and $r \in [0, 1)$ such that $\alpha(t) \leq c_r r^t$ for $t \geq 1$.

We are now ready to establish the main theorem of this paper; its proof is given in the appendix.

THEOREM 2.1: *Assume that either Assumptions 2.1 and A.1–A.3(i) for Case A or Assumption 2.2, A.1–A.3(i) and A.4 for Case B hold. Then under $\mathcal{H}_0$*

$$\widehat{L}_T(h) \to_D N(0, 1) \text{ as } T \to \infty. \tag{2.5}$$

Theorem 2.1 shows that the standard normality can still be an asymptotic distribution of the proposed test even when nonstationarity is involved. Moreover, Theorem 2.1 shows that the same asymptotically normal test can be used to deal with the stationary and nonstationary cases.

It is our experience that in practice the proposed test $\widehat{L}_T(h)$ may not have good finite sample properties when using a normal distribution to approximate the exact finite–sample distribution of the test under consideration. In order to improve the finite sample performance of $\widehat{L}_T(h)$, we propose using a bootstrap method in Section 3 below. Section 3 below also studies the power performance of $\widehat{L}_T(h)$ under $\mathcal{H}_1$.

3. Bootstrap simulation scheme

This section discusses how to simulate a critical value for the implementation of $\widehat{L}_T(h)$ in practice. Before we look at how to implement $\widehat{L}_T(h)$ in practice, we propose the following simulation scheme.

Simulation Scheme: The exact α–level critical value, $l_\alpha(h)$ $(0 < \alpha < 1)$ is the $1 - \alpha$ quantile of the exact finite–sample distribution of $\widehat{L}_T(h)$. Because $l_\alpha(h)$ may be unknown, it cannot be evaluated in practice. We therefore propose choosing an approximate α–level critical value, $l_\alpha^*(h)$, by using the following simulation procedure:

- Let $Y_0^* = X_0 = 0$. For $t = 1, 2, \cdots, T$, generate $Y_t^* = Y_{t-1}^* + (X_t - X_{t-1})^\tau \widehat{\beta} + \widehat{\sigma}_u u_t^*$ for Case A, and $Y_t^* = X_t^\tau \widehat{\beta} + f_0\left(Y_{t-1}^* - X_{t-1}^\tau \widehat{\beta}, \widehat{\theta}_0\right) + \widehat{\sigma}_u u_t^*$ for Case B, where the original sample $\mathcal{W}_T = \{(X_1, Y_1), \cdots, (X_T, Y_T)\}$ acts in the resampling as a fixed design, $\{u_t^*\}$ is sampled independently from the standard Normal distribution: $N(0, 1)$, and $\widehat{\beta}$, $\widehat{\theta}_0$ and $\widehat{\sigma}_u^2$ are the respective $\sqrt{T}$–consistent estimators of β, θ_0 and σ_u^2 based on the original sample.

- Use the data set $\{(X_t, Y_t^*) : t = 1, 2, \ldots, T\}$ to re–estimate β, θ_0 and σ_u. Denote the resulting estimates by $\widehat{\beta}^*$, $\widehat{\theta}_0^*$ and $\widehat{\sigma}_u^*$. Compute $\widehat{L}_T^*(h)$ that is the corresponding version of $\widehat{L}_T(h)$ by replacing $\{(X_t, Y_t) : t = 1, 2, \cdots, T\}$ and $\widehat{\beta}$, $\widehat{\theta}_0$ and $\widehat{\sigma}_u$ with $\{(X_t, Y_t^*) : t = 1, 2, \cdots, T\}$ and $\widehat{\beta}^*$, $\widehat{\theta}_0^*$ and $\widehat{\sigma}_u^*$. That is

$$\widehat{L}_T^* = \widehat{L}_T^*(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^* \ K_h\left(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*\right) \ \widehat{u}_t^*}{\sqrt{2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^{*2} \ K_h^2\left(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*\right) \ \widehat{u}_t^{*2}}}, \tag{3.1}$$

where $\widehat{u}_t^* = \widehat{v}_t^* - f_0(\widehat{v}_{t-1}^*, \widehat{\theta}_0^*)$, in which $\widehat{v}_t^* = Y_t^* - X_t^\tau \widehat{\beta}^*$.

- Repeat the above steps M times and produce M versions of $\widehat{L}_T^*(h)$ denoted by $\widehat{L}_{Tm}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{L}_{Tm}^*(h)$ to construct their empirical bootstrap distribution function. The bootstrap distribution of $\widehat{L}_T^*(h)$ given the full sample $\mathcal{W}_T$ is defined by $P^*\left(\widehat{L}_T^*(h) \leq x\right) = P\left(\widehat{L}_T^*(h) \leq x|\mathcal{W}_T\right)$. Let $l_\alpha^*(h)$ satisfy $P^*\left(\widehat{L}_T^*(h) \geq l_\alpha^*(h)\right) = \alpha$ and then estimate $l_\alpha(h)$ by $l_\alpha^*(h)$.

Define the size and power functions by

$$\alpha(h) = P\left(\widehat{L}_T(h) \geq l_\alpha^*(h)|\mathcal{H}_0\right) \text{ and } \beta(h) = P\left(\widehat{L}_T(h) \geq l_\alpha^*(h)|\mathcal{H}_1\right). \tag{3.2}$$

Let $H_T = \{h : \alpha(h) \leq \alpha\}$. Choose an optimal bandwidth $\widehat{h}_{\text{test}}$ such that

$$\widehat{h}_{\text{test}} = \arg\max_{h\in H_T} \beta(h). \tag{3.3}$$

We then use $l_\alpha^*(\widehat{h}_{\text{test}})$ in the computation of both the size and power values of $\widehat{L}_T(\widehat{h}_{\text{test}})$ for each case.

Note that the above simulation is based on the so–called regression bootstrap simulation procedure discussed in the literature, such as Chen and Gao (2007). We may also use a wild bootstrap to generate a sequence of resamples for $\{u_t^*\}$. Since the proposed simulation procedure works well in this paper, we use it for both theoretical studies and practical applications.

In order to study power properties of $\widehat{L}_T(\widehat{h}_{\text{test}})$, we need to impose certain conditions on $f_1(v, \theta)$ under $\mathcal{H}_1$. Since we are only interested in testing nonstationarity versus stationarity for Case A and stationarity versus stationarity for Case B, assumptions under $\mathcal{H}_1$ are more verifiable than those conditions for the nonstationarity case.

In addition to Assumption A.5 listed in the appendix, we need the following condition.

ASSUMPTION 3.1. (i) *Assume that* $\{v_t\}$ *under* $\mathcal{H}_1$ *is a sequence of strictly stationary and* α*–mixing errors with mixing coefficient*

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$$

for all $s, t \geq 1$, *where* $\{\Omega_i^j\}$ *is a sequence of* σ*-fields generated by* $\{v_s : i \leq s \leq j\}$. *There exist constants* $c_r > 0$ *and* $r \in [0, 1)$ *such that* $\alpha(t) \leq c_r r^t$ *for* $t \geq 1$.

(ii) *Let* $\mathcal{H}_1$ *be true. Then there are* θ_0 *and* θ_1 *such that:*

$$\int [f_1(v, \theta_1) - f_0(v, \theta_0)]^2 \, \pi_1^2(v) dv > 0,$$

where $\pi_1(v)$ *denotes the marginal density of* $\{v_t\}$ *under* $\mathcal{H}_1$ *for either Case A or Case B.*

Assumption 3.1(i) is a set of quite general conditions and also quite standard in this kind of stationary case, as assumed in the literature (see Li 1999 for example). Observe that Assumption 3.1 covers the case where $\{v_t\}$ is a sequence of independent and identically distributed (i.i.d.) errors.

Assumption 3.1(ii) assumes that there is some significant 'distance' between $\mathcal{H}_0$ and $\mathcal{H}_1$ in order for the test to have power. It is obvious that there are various ways of choosing the forms of $f_i(v, \theta_i)$ for $i = 0, 1$. For example, we may consider testing the AR(1) error model against a nonlinear error model of the form (Tong 1990; Granger and Teräsvirta 1993; Granger, Inoue and Morin 1997; Gao 2007)

$$\mathcal{H}_0: \ v_t = \rho_0 \, v_{t-1} + u_t \quad \text{versus} \quad \mathcal{H}_1: \ v_t = \rho_1 \, v_{t-1} - \frac{v_{t-1}}{1 + v_{t-1}^2} + u_t, \tag{3.4}$$

where $\{u_t\}$ is a sequence of i.i.d. normal errors with $E[u_t] = 0$ and $E[u_t^2] = \sigma_u^2 < \infty$, and $|\rho_0| \leq 1$ and $|\rho_1| < 1$ are suitable parameters. It is noted that $\{v_t\}$ under $\mathcal{H}_0$ is stationary when $|\rho_0| < 1$ and nonstationary when $\rho_0 = 1$. Existing studies (Masry and Tjøstheim 1995) show that $\{v_t\}$ is stationary under $\mathcal{H}_1$. In this case, Assumption 3.1 becomes

$$\begin{aligned} \int [f_1(v, \theta_1) - f_0(v, \theta_0)]^2 \, \pi_1^2(v) dv &= \int \left(2(\rho_0 - \rho_1) + \frac{1}{1 + v^2}\right) \frac{v^2}{1 + v^2} \pi_1^2(v) \, dv \\ &+ \int (\rho_1 - \rho_0)^2 v^2 \pi_1^2(v) \, dv > 0 \end{aligned} \tag{3.5}$$

when ρ_1 is chosen such that $\rho_1 \leq \rho_0$. This implies that Assumption 3.1(ii) is verifiable.

We state the following results of this section. Their proofs are relegated to the appendix.

THEOREM 3.1. (i) *Assume that either Assumptions 2.1 and A.1–A.3 for Case A or Assumptions 2.2 and A.1–A.5(i) for Case B hold. Then under* $\mathcal{H}_0$

$$\lim_{T\to\infty} P(\widehat{L}_T(h) > l_\alpha^*) = \alpha.$$

(ii) *Assume that either Assumptions 2.1, 3.1 and A.1–A.3 for Case A or Assumptions 2.2, 3.1 and A.1–A.5 for Case B hold. Then under* $\mathcal{H}_1$

$$\lim_{T\to\infty} P(\widehat{L}_T(h) > l_\alpha^*) = 1.$$

Theorem 3.1(i) implies that l_α^* is an asymptotically correct α–level critical value under any model in $\mathcal{H}_0$, while Theorem 3.1(ii) shows that $\widehat{L}_T(h)$ is asymptotically consistent. In Section 4 below, we show how to illustrate Theorem 3.1 through using a simulated example.

4. An example of implementation.

Example 4.1 compares the small and medium–sample performance of our test with two natural competitors using a simulated example. A real–data application is then given in Example 4.2.

EXAMPLE 4.1. Consider a nonlinear trend model of the form

$$Y_t = X_t\ \beta + v_t = \sin\left(\frac{2\pi t}{T}\right)\ \beta + v_t \text{ with } v_t = f_i(v_{t-1}, \theta_i) + u_t,\ 1 \le t \le T, \tag{4.1}$$

where $\{u_t\}$ is a sequence of i.i.d. N(0,1), and the forms of $f_i(v, \theta_i)$ for $i = 0, 1$ are given as follows:

$$\begin{aligned}
f_0(v, \theta_0) &= v \text{ and } f_1(v, \theta_1) = (1 + \theta_1)\ v \text{ for Case A, or} && (4.2)\\
f_0(v, \theta_0) &= v \text{ and } f_1(v, \theta_1) = (1 + \theta_1)\ v + \frac{\theta_1\ v}{1 + v^2} \text{ for Case A, or} && (4.3)\\
f_0(v, \theta_0) &= 0 \text{ and } f_1(v, \theta_1) = \theta_1\ v \text{ for Case B, or} && (4.4)\\
f_0(v, \theta_0) &= 0.5\ v \text{ and } f_1(v, \theta_1) = (0.5 + \theta_1)\ v + \frac{\theta_1\ v}{1 + v^2} \text{ for Case B,} && (4.5)
\end{aligned}$$

where $\rho_0 = 1$ for models (4.2) and (4.3), $\rho_0 = 0$ for model (4.4) and $\rho_0 = 0.5$ for model (4.5), $\theta_1 = -\sqrt{T^{-1}\ \log(\log(T))}$ or $-T^{-\frac{13}{30}}$ and $\rho_1 = \rho_0 + \theta_1$.

The β parameter is estimated by the conventional least squares estimator. We choose $K(x) = \frac{1}{2} I_{[-1,1]}(x)$ as the kernel function throughout this example.

To compute the size of $\widehat{L}_T(h)$ under $\mathcal{H}_0$ and the power of $\widehat{L}_T(h)$ under $\mathcal{H}_1$ for (4.2)–(4.4), our experience (Gao and Gijbels 2006) shows that the leading term of $\widehat{h}_{\text{test}}$ of (3.3) is proportional to

$$\widehat{h}_0 = \left(T\theta_1^2\right)^{-\frac{3}{2}}, \tag{4.6}$$

where $\theta_1 = \theta_{11} = -\sqrt{T^{-1}\log(\log(T))}$ or $\theta_1 = \theta_{12} = -T^{-\frac{13}{30}}$. The reasoning for the choice of θ_1 is as follows. The rate of $\theta_{11} = -T^{-\frac{1}{2}}\sqrt{\text{loglog}(T)}$ is the optimal rate of testing in this kind of nonparametric kernel testing problem as discussed in Chapter 3 of Gao (2007). The rate of $\theta_{12} = -T^{-\frac{13}{30}}$ implies that the optimal bandwidth $\widehat{h}_0$ in (4.6) above is proportional to $T^{-\frac{1}{5}}$, a bandwidth commonly used in the literature.

In order to assess the variability of both the size and power with respect to various bandwidth values, we then use a set of bandwidth values of the form

$$h_i = \widehat{h}_0 \quad \text{corresponding to the two values of } \theta_{1i}, \quad i = 1, 2. \tag{4.7}$$

We then introduce

$$L_{1i} = \widehat{L}_T(h_i) \quad \text{for} \quad i = 1, 2. \tag{4.8}$$

For model (4.2) and (4.3), we compare our test with the conventional DF (Dickey and Fuller 1979) test of the form

$$L_{21} = \frac{\sum_{t=2}^{T}(\widehat{v}_t - \widehat{\rho}_0\widehat{v}_{t-1})\widehat{v}_{t-1}}{\widehat{\sigma}_{22}\sqrt{\sum_{t=2}^{T}\widehat{v}_{t-1}^2}}, \tag{4.9}$$

where $\widehat{\sigma}_{22}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\left(\widehat{v}_t - \widehat{\rho}_0\widehat{v}_{t-1}\right)^2$ with $\widehat{\rho}_0 = \frac{\sum_{t=2}^{T}\widehat{v}_t\widehat{v}_{t-1}}{\sum_{t=2}^{T}\widehat{v}_{t-1}^2}$.

For models (4.4) and (4.5), we also compare our test with the DK test (Dufour and King 1991) of the form

$$L_{22} = \frac{\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{v}_s\ a_{st}\ \widehat{v}_t}{\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{v}_s\ b_{st}\ \widehat{v}_t}, \tag{4.10}$$

where $\widehat{v}_t = Y_t - X_t\ \widehat{\beta}$, $\{a_{st}\}$ is the (s,t)–th element of A_0 given by

$$A_0 = -2(1-\rho_0)\ I_T + A_1 - 2\rho_0\ C_1$$

with I_T being the $T \times T$ identity matrix, A_1 and C_1 being given in (6) and (7) of Dufour and King (1991, p.120), and $\{b_{st}\}$ is the (s,t)–th element of Σ_0^{-1}, in which $\Sigma_0 = \Sigma(\rho_0)$ with $\Sigma(\rho)$ being given above (G1) of Dufour and King (1991, p.118).

Since the power of L_{21} and L_{22} in each case depends on the choice of θ_{1i} for $i = 1, 2$, we define L_{21i} and L_{22i} as the corresponding versions of L_{21} and L_{22} according to θ_{1i}. We choose $N = 250$ in the Simulation Scheme and use $M = 1000$ replications to compute the double–sized power and size values of the tests in Tables 4.1–4.4 below.

Table 4.1. Sizes and power values for models (4.2) and (4.3) at the $\alpha = 5\%$ significance level

Test	Model (4.2)			Model (4.3)		
	Null Hypothesis Is True					
T	250	500	750	250	500	750
L_{11}	0.043	0.041	0.055	0.046	0.060	0.046
L_{12}	0.049	0.047	0.055	0.046	0.053	0.044
L_{21}	0.045	0.059	0.062	0.060	0.059	0.055
	Null Hypothesis Is False					
T	250	500	750	250	500	750
L_{11}	0.242	0.307	0.335	0.519	0.593	0.575
L_{12}	0.241	0.252	0.276	0.565	0.610	0.613
L_{211}	0.723	0.746	0.785	0.255	0.359	0.371
L_{212}	0.725	0.748	0.787	0.228	0.296	0.328

Table 4.2. Sizes and power values for models (4.2) and (4.3) at the $\alpha = 10\%$ significance level

Test	Model (4.2)			Model (4.3)		
	Null Hypothesis Is True					
T	250	500	750	250	500	750
L_{11}	0.087	0.085	0.101	0.092	0.106	0.083
L_{12}	0.108	0.088	0.092	0.094	0.116	0.106
L_{21}	0.116	0.118	0.121	0.100	0.117	0.090
	Null Hypothesis Is False					
T	250	500	750	250	500	750
L_{11}	0.302	0.373	0.394	0.616	0.619	0.650
L_{12}	0.278	0.320	0.357	0.658	0.657	0.674
L_{211}	0.803	0.851	0.850	0.342	0.416	0.490
L_{212}	0.808	0.852	0.853	0.303	0.400	0.453

Table 4.3. Sizes and power values for models (4.4) and (4.5) at the $\alpha = 5\%$ significance level

Test	Model (4.4)			Model (4.5)		
	Null Hypothesis Is True					
T	250	500	750	250	500	750
L_{11}	0.057	0.053	0.047	0.049	0.042	0.054
L_{12}	0.053	0.044	0.052	0.052	0.050	0.050
L_{22}	0.057	0.043	0.062	0.050	0.064	0.053
	Null Hypothesis Is False					
T	250	500	750	250	500	750
L_{11}	0.158	0.156	0.182	0.400	0.461	0.420
L_{12}	0.183	0.192	0.194	0.469	0.534	0.477
L_{221}	0.251	0.288	0.286	0.293	0.304	0.334
L_{222}	0.300	0.338	0.347	0.338	0.411	0.431

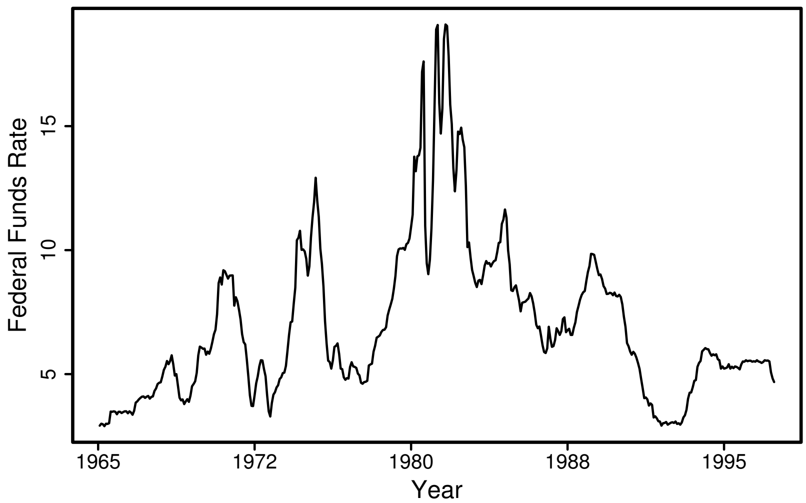
Table 4.4. Sizes and power values for models (4.4) and (4.5) at the $\alpha = 10\%$ significance level

Test	Model (4.4)			Model (4.5)		
	Null Hypothesis Is True					
T	250	500	750	250	500	750
L_{11}	0.131	0.094	0.112	0.106	0.107	0.107
L_{12}	0.110	0.092	0.099	0.094	0.102	0.094
L_{22}	0.107	0.103	0.107	0.113	0.091	0.091
	Null Hypothesis Is False					
T	250	500	750	250	500	750
L_{11}	0.267	0.258	0.280	0.547	0.578	0.607
L_{12}	0.271	0.317	0.297	0.608	0.654	0.692
L_{221}	0.380	0.407	0.403	0.432	0.419	0.442
L_{222}	0.428	0.475	0.471	0.497	0.508	0.525

Tables 4.1 and 4.2 show that the proposed test is less powerful than the conventional DF test when the true model (4.2) is linear. When the true model (4.3)

is nonlinear, however, the DF test is still applicable but is less powerful than the proposed test. Tables 4.3 and 4.4 also show that the proposed test is more powerful than the DK test when the true model (4.5) is nonlinear. When the true model (4.4) is linear, the DK test is more powerful than the proposed test. In summary, Tables 4.1–4.4 show that the proposed test is more powerful in the nonlinear case while the sizes are comparable with the two competitors for the parametric linear case. This supports that the proposed test, which is dedicated to the nonlinear case, is needed to deal with testing stationarity in nonlinear time series models.

Example 4.2. This example examines the three–month Treasury Bill rate data given in Figure 1 below sampled monthly over the period from January 1963 to December 1998, providing 432 observations. Since we consider a monthly data set, this gives $\Delta = \frac{20}{250}$.



Let $\{Y_t : t = 1, 2, \cdots, 432\}$ be the set of Treasury Bill rate data. As the plot does not suggest that there is any significant trend for the data set, it is not unreasonable to assume that $\{Y_t\}$ satisfies a nonlinear model of the form

$$Y_t = f(Y_{t-1}) + u_t \tag{4.11}$$

with the form of $f(\cdot)$ to be unknown.

To apply the test $\widehat{L}_T(\widehat{h}_{\text{test}})$ to determine whether $\{Y_t\}$ follows a random walk model of the form $Y_t = Y_{t-1} + u_t$, we need to propose the following procedure for computing the p–value of $\widehat{L}_T(\widehat{h}_{\text{test}})$:

- For the real data set, compute $\widehat{h}_{\text{test}}$ and $\widehat{L}_T(\widehat{h}_{\text{test}})$.
- Let $Y_1^* = Y_1$. Generate a sequence of bootstrap resamples $\{u_t^*\}$ from $N(0,1)$ and then $Y_t^* = Y_{t-1}^* + \widehat{\sigma}_u\ u_t^*$ for $2 \leq t \leq 432$.
- Compute the corresponding version $\widehat{L}_T^*(\widehat{h}_{\text{test}})$ of based on $\{Y_t^*\}$.
- Repeat the above steps N times to find the bootstrap distribution of $\widehat{L}_T^*(\widehat{h}_{\text{test}})$ and then compute the proportion that $\widehat{L}_T(\widehat{h}_{\text{test}}) < \widehat{L}_T^*(\widehat{h}_{\text{test}})$. This proportion is an approximate p–value of $\widehat{L}_T(\widehat{h}_{\text{test}})$.

Our simulation results return the simulated p–values of $\widehat{p}_1 = 0.005$ for L_{22} and $\widehat{p}_2 = 0.011$ for $\widehat{L}_T(\widehat{h}_{\text{test}})$. While both of the simulated p–values suggest that there is no enough evidence of accepting the unit–root structure at the 5% significance level, there is some evidence of accepting the unit–root structure based on $\widehat{L}_T(\widehat{h}_{\text{test}})$ at the 1% significance level.

5. Conclusion. We have proposed a new nonparametric test for the parametric specification of the residuals. An asymptotically normal distribution of the proposed test has been established. In addition, we have also proposed the Simulation Scheme to implement the proposed test in practice. The small and medium–sample results show that both the proposed test and the Simulation Scheme are practically applicable and implementable.

6. Acknowledgments. The authors would like to thank Dr Jiying Yin for his excellent computing assistance and the Australian Research Council Discovery Grants Program for its financial support.

Appendix

In this appendix, we first list some technical conditions and then give the proofs of Theorems 2.1 and 3.1. As the proofs of the theorems and the necessary lemmas are already extremely technical, we give only an outline for each of the proofs. However, further details are available from the authors upon request.

A.1. *Assumptions*

Assumption A.1. (i) *Let $K(\cdot)$ be a symmetric probability density function with compact support $C(K)$. In addition, there is some positive function $M(\cdot)$ such that*

$$|K(y) - K(x)| \leq M(x)\ |y - x|$$

for all $x,\ y \in C(K)$*, where both* $K(\cdot)$ *and* $M(\cdot)$ *satisfy* $\int K^2(u)du < \infty$ *and* $\int M^2(u)du < \infty$.

(ii) *Assume that* h *satisfies* $\lim_{T\to\infty} h = 0$ *and* $\limsup_{T\to\infty} Th^{2+\frac{4\delta}{1-2\delta}} = \infty$ *for some* $\frac{1}{8} \le \delta < \frac{1}{2}$.

Assumption A.2. *Let*

$$\limsup_{T\to\infty} \max_{1\le t\le T} \frac{||X_t||^2}{T^{\frac{1}{4}}} \le C < \infty,$$

for some $0 < C < \infty$*, where* $||\cdot||$ *denotes the Euclidean norm. In addition,*

$$\limsup_{T\to\infty} \frac{1}{T} \left|\left| \sum_{t=1}^{T} X_t X_t^\tau \right|\right| < \infty.$$

Assumption A.3. (i) *Let* $\mathcal{H}_0$ *be true. Then there is some* $\widehat{\beta}$ *such that*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\beta} - \beta|| > B_0\right) < \varepsilon_0$$

for any $\varepsilon_0 > 0$ *and some* $B_0 > 0$.

(ii) *Let* $\mathcal{H}_0$ *be true. There is an estimator* $\widehat{\beta}^*$ *such that for some positive constants* $B_0^* > 0$ *and* ε_0^* *the following inequality*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\beta}^* - \widehat{\beta}|| > B_0^* | \mathcal{W}_T\right) < \varepsilon_0^*$$

holds with probability one with respect to the distribution of $\mathcal{W}_T$.

Assumption A.4. (i) *Let* $\mathcal{H}_0$ *be true. Then there is an estimator* $\widehat{\theta}_0$ *such that*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\theta}_0 - \theta_0|| > C_0\right) < \epsilon_0$$

for any $\epsilon_0 > 0$ *and some* $C_0 > 0$

(ii) *Let* $\pi_0(v)$ *denote the marginal density of* $\{v_t\}$ *under* $\mathcal{H}_0$ *for Case B. Assume that* $\pi_0(v)$ *is continuous and that* $f_0(v,\theta)$ *is differentiable at both* v *and* θ*. In addition,*

$$0 < \int \left[\frac{\partial f_0(v,\theta_0)}{\partial v}\right]^2 \pi_0^2(v)\, dv < \infty \text{ and } 0 < \int \left|\left| \frac{\partial f_0(v,\theta)}{\partial \theta}|_{\theta=\theta_0} \right|\right|^2 \pi_0^2(v)\, dv < \infty.$$

Assumption A.5. (i) *Let* $\mathcal{H}_0$ *be true. Then there is an estimator* $\widehat{\theta}_0^*$ *such that for some positive constants* $C_0^* > 0$ *and* ϵ_0^* *the following inequality*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\theta}_0^* - \widehat{\theta}_0|| > C_0^* | \mathcal{W}_T\right) < \epsilon_0^*$$

holds with probability one with respect to the distribution of $\mathcal{W}_T$.

(ii) *Let* $\mathcal{H}_1$ *be true. There exists an estimator* $\widehat{\theta}_1$ *such that*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\widehat{\theta}_1 - \theta_1|| > C_1\right) < \epsilon_1$$

for any $\epsilon_1 > 0$ and some $C_1 > 0$.

(iii) *Let $\pi_1(v)$ denote the marginal density of $\{v_t\}$ under $\mathcal{H}_1$ for either A or Case B. Assume that $\pi_1(v)$ is continuous and that $f_1(v,\theta)$ is differentiable at both v and θ. In addition,*

$$0 < \int \left[\frac{\partial f_1(v,\theta_1)}{\partial v}\right]^2 \pi_1^2(v)\, dv < \infty \text{ and } 0 < \int \left|\left|\frac{\partial f_1(v,\theta)}{\partial \theta}|_{\theta=\theta_1}\right|\right|^2 \pi_1^2(v)\, dv < \infty.$$

Assumption A.1(i) holds in many cases. For example, $K(x) = \frac{1}{2} I_{[-1,1]}(x)$. In addition, Assumption A.1(i) is a very mild condition. While Assumption A.1(ii) imposes certain conditions, which may look more restrictive than those for the stationary case, they don't look unnatural in the nonstationary case. Such conditions on the bandwidth for nonparametric testing in the nonstationary case are equivalent to the minimal conditions: $\lim_{T\to\infty} h = 0$ and $\lim_{T\to\infty} Th = \infty$ required in nonparametric kernel testing for both the independence and the stationary time series cases (see Gao 2007 for example).

Assumption A.2 imposes some minimal conditions on the trend functions. When the trend functions are all continuous and bounded, Assumption A.2 is satisfied automatically.

Assumption A.3 requires that the conventional rate of convergence for the parametric case is achievable even when $\{v_t\}$ is nonstationary. As a matter of fact, in Case A the rate of convergence of the conventional least squares (LS) estimator of β under $\mathcal{H}_0$ given by

$$\widehat{\beta} = \left(\sum_{t=1}^{T}(X_t - X_{t-1})(X_t - X_{t-1})^\tau\right)^{-1} \sum_{t=1}^{T}(X_t - X_{t-1})(Y_t - Y_{t-1})$$

may be obtained when the inverse matrix does exist and Assumption A.3 holds. In Case B, the rate of convergence is also achievable. For example, the semiparametric LS estimator constructed by Hidalgo (1992) achieves the required rate of convergence.

Assumption A.4 imposes the differentiability conditions as well as the moment conditions on $f_0(\cdot,\cdot)$. As $\{v_t\}$ is strictly stationary, it is possible to verify Assumption A.4 in many cases. Assumption A.5(i) is the bootstrap version of Assumption A.4(i). Assumption A.5(ii)(iii) is a kind of corresponding version of Assumption A.4 under $\mathcal{H}_1$.

A.2. *Proof of Theorem 2.1 in Case A*

To avoid notational complication, we introduce

$$a_{st} = K_h\left(\sum_{i=s}^{t-1} u_i\right) \quad \text{and} \quad \eta_t = 2\sum_{s=1}^{t-1} a_{st}\, u_s.$$

Note that under $\mathcal{H}_0$

$$
\begin{aligned}
\widehat{M}_T &\equiv \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widehat{u}_t = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\ K_h(v_{s-1}-v_{t-1})\ u_t \\
&+ \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{u}_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t + 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t \\
&\equiv M_{T1}+M_{T2}+M_{T3}+M_{T4}, \qquad \text{(A.1)}\\
\widehat{\sigma}_T^2 &\equiv 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widehat{u}_t^2 = 2\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} u_s^2\ K_h^2(v_{s-1}-v_{t-1})\ u_t^2 \\
&+ 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{u}_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t^2 + 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t^2 \\
&+ \widetilde{\sigma}_{T4}^2 \equiv \widetilde{\sigma}_{T1}^2+\widetilde{\sigma}_{T2}^2+\widetilde{\sigma}_{T3}^2+\widetilde{\sigma}_{T4}^2, \qquad \text{(A.2)}
\end{aligned}
$$

where for Case A under $\mathcal{H}_0$: $f_0(v,\theta_0)\equiv v$,

$$
\begin{aligned}
\widetilde{u}_t &= X_t^\tau(\beta-\widehat{\beta}) + f_0(v_{t-1},\theta_0) - f_0\left(v_{t-1}+X_t^\tau(\beta-\widehat{\beta}),\widehat{\theta}_0\right)\\
&= X_t^\tau(\beta-\widehat{\beta}),\\
\widehat{v}_{s-1}-\widehat{v}_{t-1} &= v_{s-1}-v_{t-1}+(X_{s-1}-X_{t-1})^\tau\,(\beta-\widehat{\beta}),\\
M_{T4} &= \widehat{M}_T - M_{T1}-M_{T2}-M_{T3},\\
\widetilde{\sigma}_{T4}^2 &= \widehat{\sigma}_T^2-\widetilde{\sigma}_{T1}^2-\widetilde{\sigma}_{T2}^2-\widetilde{\sigma}_{T3}^2.
\end{aligned}
$$

In view of (A.1) and (A.2), in order to prove Theorem 2.1 for Case A, it suffices to show that as $T\to\infty$

$$
\begin{aligned}
\frac{M_{T1}}{\widetilde{\sigma}_{T1}} &\to_D \quad N(0,1), \qquad \text{(A.3)}\\
\frac{M_{Ti}}{\widetilde{\sigma}_{T1}} &\to_P \quad 0 \quad \text{for } i=2,3,4, \qquad \text{(A.4)}\\
\frac{\widetilde{\sigma}_{Tj}}{\widetilde{\sigma}_{T1}} &\to_P \quad 0 \quad \text{for } j=2,3,4. \qquad \text{(A.5)}
\end{aligned}
$$

We will return to the proof of (A.4) and (A.5) in Lemma A.6 after having proved (A.3) in Lemmas A.1–A.5. In order to prove (A.3), we apply Theorem 3.4 of Hall and Heyde (1980, p.67) to our case. Before verifying the conditions of their Theorem 3.4, we introduce the following notation.

Let $Y_{Tt}=\frac{\eta_t u_t}{\sigma_{T1}(h)}$, $\Omega_{T,s}=\sigma\{u_t: 1\le t\le s\}$ be a σ–field generated by $\{u_t: 1\le t\le s\}$, $\mathcal{G}_T=\Omega_{T,M(T)}$ and $\mathcal{G}_{T,s}$ be defined by

$$
\mathcal{G}_{T,s} = \begin{cases} \Omega_{T,M(T)}, & 1\le s\le M(T),\\ \Omega_{T,s}, & M(T)+1\le s\le T, \end{cases} \qquad \text{(A.6)}
$$

where $M(T)$ is chosen such that $M(T) \to \infty$ and $\frac{M(T)}{T} \to 0$ as $T \to \infty$. Let $\widetilde{U}_{M(T)}^2 = \frac{\widetilde{\sigma}_{M(T),1}^2}{\sigma_{M(T),1}^2}$, where $\sigma_{S,1}^2 = \mathrm{Var}\left[\sum_{t=2}^S \eta_t u_t\right]$ for all $1 \le S \le T$. We can prove that as $T \to \infty$

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} - \widetilde{U}_{M(T)}^2 \to_P 0. \tag{A.7}$$

Thus, equation (3.28) of Hall and Heyde (1980) can be satisfied. The proof of (A.7) is given in Lemma A.4 below.

Before we apply Theorem 3.4 of Hall and Heyde (1980) to our case, we need to state that the conclusion of their Theorem 3.4 remains true if the unconditional assumptions (3.18) and (3.20) involved in their Theorem 3.4 are replaced by the corresponding conditional assumptions as used in Corollary 3.1 of Hall and Heyde or conditions (A.9) and (A.10) below. Therefore, in order to prove that as $T \to \infty$

$$\frac{M_{T1}}{\widetilde{\sigma}_{T1}} = \frac{1}{\widetilde{\sigma}_{T1}} \sum_{t=2}^T \eta_t u_t \to_D N(0,1), \tag{A.8}$$

it suffices to show that there is an almost surely finite random variable ξ such that for all $\epsilon > 0$,

$$\begin{aligned}
\sum_{t=2}^T E\left[Y_{Tt}^2 | \Omega_{T,t-1}\right] &\to_D \xi^2, && \text{(A.9)}\\
\sum_{t=2}^T E\left[Y_{Tt}^2 I_{\{[|Y_{Tt}|>\epsilon]\}} | \Omega_{T,t-1}\right] &\to_P 0, && \text{(A.10)}\\
\sum_{t=2}^T E\left[Y_{Tt} | \mathcal{G}_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} + \sum_{t=M(T)+1}^T E\left[Y_{Tt}|\Omega_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} &\to_P 0, && \text{(A.11)}\\
\sum_{t=2}^T \left|E\left[Y_{Tt} | \mathcal{G}_{T,t-1}\right]\right|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 + \sum_{t=M(T)+1}^T \left|E\left[Y_{Tt}|\Omega_{T,t-1}\right]\right|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 &\to_P 0, && \text{(A.12)}\\
\lim_{\delta \to 0} \lim \inf_{T \to \infty} P\left(\frac{\widetilde{\sigma}_{T1}}{\sigma_{T1}} > \epsilon\right) &= 1. && \text{(A.13)}
\end{aligned}$$

The proof of (A.9) is given in Lemma A.3, while Lemma A.2 below gives the proof of (A.10). The proof of (A.11) is similar to that of (A.12), which follows from

$$\sum_{t=2}^{M(T)} E\left[Y_{Tt}^2\right] = O\left(\left(\frac{M(T)}{T}\right)^{\frac{3}{2}}\right) \to 0 \tag{A.14}$$

as $T \to \infty$, in which Lemma A.1 below is used.

The proof of (A.13) follows from

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} \to_D \xi^2 > 0. \tag{A.15}$$

An outline of the proof of (A.15) is given in Lemma A.5 below.

In order to prove (A.10), it suffices to show that

$$\frac{1}{\sigma_{T1}^4}\sum_{t=2}^{T} E\left[\eta_t^4\right] \to 0. \tag{A.16}$$

The proof of (A.16) is given in Lemma A.2 below.

The following lemmas are necessary to complete the proof of Theorem 2.1. Since the proofs of the lemmas are extremely technical, they are relegated to those of Lemmas A.1–A.9 of the full version of this paper by Gao and King (2007).

Lemma A.1. *Assume that Assumptions 2.1 and A.1 hold. Then for sufficiently large* T

$$\sigma_{T1}^2 = \mathrm{Var}\left[\sum_{t=2}^{T}\eta_t u_t\right] = \frac{4\int K^2(x)dx}{3\sqrt{2\pi}} T^{3/2} h\ (1+o(1)). \tag{A.17}$$

Lemma A.2. *Under the conditions of Theorem 2.1, we have*

$$\lim_{T\to\infty}\frac{1}{\sigma_{T1}^4}\sum_{t=2}^{T} E\left[\eta_t^4\right] = 0. \tag{A.18}$$

Lemma A.3. *Let Assumptions 2.1 and A.1 hold. Then as* $T \to \infty$

$$\frac{1}{\sigma_{T1}^2}\sum_{t=2}^{T}\eta_t^2 \to_D \xi^2 \tag{A.19}$$

with $\xi^2 = \frac{3\sqrt{2\pi}}{2}\, M_{\frac{1}{2}}(1)$, *where* $M_{\frac{1}{2}}(\cdot)$ *is a special case of the Mittag–Leffer process* $M_\beta(\cdot)$ *with* $\beta = \frac{1}{2}$ *as described by Karlsen and Tjøstheim (2001, p.388).*

Lemma A.4. *Let Assumptions 2.1 and A.1 hold. Then as* $T \to \infty$, $M(T) \to \infty$ *and* $\frac{M(T)}{T} \to 0$

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} - \widetilde{U}_{M(T)}^2 \to_P 0. \tag{A.20}$$

Lemma A.5. *Let Assumptions 2.1 and A.1 hold. Then as* $T \to \infty$

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} \to_D \xi^2 > 0. \tag{A.21}$$

Lemma A.6. *Let the conditions of Theorem 2.1 hold. Then as* $T \to \infty$

$$\frac{M_{Ti}}{\widetilde{\sigma}_{T1}} \quad \to_P \quad 0 \quad \text{for } i = 2, 3, \tag{A.22}$$

$$\frac{\widetilde{\sigma}_{Tj}}{\widetilde{\sigma}_{T1}} \quad \to_P \quad 0 \quad \text{for } j = 2, 3, 4. \tag{A.23}$$

Proof: Since $\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} \to_D \xi^2$ as shown in Lemma A.3, in order to prove (A.22) and (A.23), it suffices to show that as $T \to \infty$

$$\frac{M_{Ti}}{\sigma_{T1}} \quad \to_P \quad 0 \quad \text{for } i = 2, 3, 4, \tag{A.24}$$

$$\frac{\widetilde{\sigma}_{Tj}}{\sigma_{T1}} \quad \to_P \quad 0 \quad \text{for } j = 2, 3, 4. \tag{A.25}$$

Since the details are quite technical, they are relegated to Lemma A.6 of Gao and King (2007).

A.3. *Proof of Theorem 2.1 in Case B*

In view of (A.1) and (A.2), in order to prove Theorem 2.1 for Case B, it suffices to show that equations (A.3)–(A.5) hold. The proofs are given in Lemmas A.7 and A.8 below.

Lemma A.7. *Let Assumptions 2.2 and A.1 hold. Then under* $\mathcal{H}_0$

$$\frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T u_s \ K_h(v_{s-1} - v_{t-1}) \ u_t}{\sqrt{2 \sum_{t=1}^T \sum_{s=1,\neq t}^T u_s^2 \ K_h^2(v_{s-1} - v_{t-1}) \ u_t^2}} \to_D N(0,1) \quad as \quad T \to \infty, \tag{A.26}$$

where $v_t = f_0(v_{t-1}, \theta_0) + u_t$.

Proof: The asymptotic normality in (A.26) is a standard result for the case where both $\{u_t\}$ and $\{v_t\}$ are strictly stationary and α–mixing time series. The proof follows from Lemma A.1 of Gao and King (2004) or Theorem A.1 of Gao (2007). As the details are very similar to the proof of Theorem 2.1 of Gao and King (2004), they are omitted here.

Lemma A.8. *Let the conditions of Theorem 2.1 hold. Then as* $T \to \infty$

$$\frac{M_{Ti}}{\sigma_{T1}} \quad \to_P \quad 0 \quad \text{for } i = 2, 3, 4, \tag{A.27}$$

$$\frac{\widetilde{\sigma}_{Tj}}{\sigma_{T1}} \quad \to_P \quad 0 \quad \text{for } j = 2, 3, 4. \tag{A.28}$$

Proof: Since both $\{u_t\}$ and $\{v_t\}$ are assumed to be strictly stationary and α–mixing time series in Case B, the proofs of (A.24) and (A.25) remain true, but become more standard through using Assumptions 2.1, A.3(i) and A.4.

A.4. *Proof of Theorem 3.1(i)*

Recall the notation introduced in the Simulation Scheme in Section 3 and let

$$\begin{aligned}
\widetilde{v}_t^* &= Y_t^* - X_t^\tau \widehat{\beta} = f_0(\widetilde{v}_{t-1}, \widehat{\theta}_0) + \widehat{\sigma}_u \ u_t^*, \\
\widehat{v}_t^* &= Y_t^* - X_t^\tau \widehat{\beta}^* = \widetilde{v}_t^* + X_t^\tau \left(\widehat{\beta} - \widehat{\beta}^*\right), \\
\widetilde{u}_t^* &= X_t^\tau (\widehat{\beta} - \widehat{\beta}^*) + f_0(\widetilde{v}_{t-1}^*, \widehat{\theta}_0) - f_0\left(\widetilde{v}_{t-1}^* + X_t^\tau (\widehat{\beta} - \widehat{\beta}^*), \widehat{\theta}_0^*\right), \\
\widehat{u}_t^* &= \widehat{v}_t^* - f_0(\widehat{v}_{t-1}^*, \widehat{\theta}_0^*) = \widehat{\sigma}_u \ u_t^* + \widetilde{u}_t^*, \\
\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^* &= \widetilde{v}_{s-1}^* - \widetilde{v}_{t-1}^* + (X_{s-1} - X_{t-1})^\tau \left(\widehat{\beta} - \widehat{\beta}^*\right).
\end{aligned}$$

We thus have

$$\begin{aligned}
\widehat{M}_T^* &\equiv \sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^* \; K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widehat{u}_t^* = \sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u u_s^* \; K_h(\widetilde{v}_{s-1}^* - \widetilde{v}_{t-1}^*) \; \widehat{\sigma}_u u_t^* \\
&+ \sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s^* \; K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widetilde{u}_t^* + 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u u_s^* \; K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widetilde{u}_t^* \\
&+ M_{T4}^* \equiv M_{T1}^* + M_{T2}^* + M_{T3}^* + M_{T4}^*, \qquad (A.29) \\
\widehat{\sigma}_T^{*2} &\equiv 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^{*2} \; K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widehat{u}_t^{*2} \\
&= 2\sum_{t=1}^T \sum_{s=1,s\neq t}^T \widehat{\sigma}_u^2 u_s^{*2} \; K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widehat{\sigma}_u^2 \; u_t^{*2} \\
&+ 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s^{*2} \; K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widetilde{u}_t^{*2} \\
&+ 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u^2 u_s^{*2} \; K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*) \; \widetilde{u}_t^{*2} + \widetilde{\sigma}_{T4}^{*2} \equiv \sum_{j=1}^4 \widetilde{\sigma}_{Tj}^{*2}, \qquad (A.30)
\end{aligned}$$

where $\widetilde{\sigma}_{T4}^{*2} = \widehat{\sigma}_T^{*2} - \widetilde{\sigma}_{T1}^{*2} - \widetilde{\sigma}_{T2}^{*2} - \widetilde{\sigma}_{T3}^{*2}$ and $M_{T4}^* = \widehat{M}_T^* - M_{T1}^* - M_{T2}^* - M_{T3}^*$.

Using the conditions of Theorem 3.1(i), in view of the notation of $\widehat{L}_T^*(h)$ introduced in the Simulation Scheme in Section 3 as well as the proof of Theorem 2.1, we may show that as $T \to \infty$

$$P^* \left(\widehat{L}_T^*(h) \le x \right) \to \Phi(x) \;\; \text{for all } x \in (-\infty, \infty) \qquad (A.31)$$

holds in probability with respect to the distribution of the original sample $\mathcal{W}_T$. In detail, in order to prove (A.31), in view of the fact that $\{u_s^*\}$ and $\{Y_t\}$ are independent for all $s, t \ge 1$, we may show that the proofs of Lemmas A.1–A.6 all remain true by successive conditioning arguments.

Let z_α be the $1-\alpha$ quantile of $\Phi(\cdot)$ such that $\Phi(z_\alpha) = 1-\alpha$. Then it follows from (A.31) that as $T \to \infty$

$$P^* \left(\widehat{L}_T^*(h) \ge z_\alpha \right) \to 1 - \Phi(z_\alpha) = \alpha. \qquad (A.32)$$

This, together with the construction that $P^* \left(\widehat{L}_T^*(h) \ge l_\alpha^* \right) = \alpha$, implies that as $T \to \infty$

$$l_\alpha^* - z_\alpha \to_P 0. \qquad (A.33)$$

Using the conclusion of Theorem 2.1 and (A.31) again, we have that as $T \to \infty$

$$P^* \left(\widehat{L}_T^*(h) \le x \right) - P \left(\widehat{L}_T(h) \le x \right) \to_P 0 \;\; \text{for all } x \in (-\infty, \infty) \qquad (A.34)$$

holds in probability. This, along with the construction that $P^*\left(\widehat{L}_T^*(h) \geq l_\alpha^*\right) = \alpha$ again, shows that as $T \to \infty$

$$\lim_{T\to\infty} P\left(\widehat{L}_T(h) \geq l_\alpha^*\right) = \alpha \tag{A.35}$$

holds in probability. Therefore the conclusion of Theorem 3.1(i) is proved.

A.4. *Proof of Theorem 3.1(ii)*

Note that under $\mathcal{H}_1$

$$\begin{aligned}
\widetilde{u}_t &= X_t^\tau(\beta - \widehat{\beta}) + f_1(v_{t-1}, \theta_1) - f_0\left(v_{t-1} + X_t^\tau(\beta - \widehat{\beta}), \widehat{\theta}_0\right) \\
&= X_t^\tau(\beta - \widehat{\beta}) + f_1(v_{t-1}, \theta_1) - f_0(v_{t-1}, \theta_0) \\
&+ f_0(v_{t-1}, \theta_0) - f_0\left(v_{t-1} + X_t^\tau(\beta - \widehat{\beta}), \widehat{\theta}_0\right).
\end{aligned} \tag{A.36}$$

Lemma A.9. *Let the conditions of Theorem 3.1(ii) hold. Then as* $T \to \infty$

$$\frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s\ K_h(v_{s-1} - v_{t-1})\ \widetilde{u}_t}{\sigma_{T1}} \to_P \infty. \tag{A.37}$$

Proof: Let $f_{10}(v) = f_1(v, \theta_1) - f_0(v, \theta_0)$. In view of (A.36), using Assumptions A.4 and A.5(ii), in order to prove (A.37), it suffices to show that as $T \to \infty$

$$\frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T f_{10}(v_{s-1})\ K_h(v_{s-1} - v_{t-1})\ f_{10}(v_{t-1})}{\sigma_{T1}} \to_P \infty. \tag{A.38}$$

The proof of (A.38) follows from as $T \to \infty$ and $h \to 0$

$$\begin{aligned}
&\frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T E\left[f_{10}(v_{s-1})\ K_h(v_{s-1} - v_{t-1})\ f_{10}(v_{t-1})\right]}{\sigma_{T1}} \\
&= \frac{T^2 h(1+o(1))}{\sigma_{T1}} \left(\int f_{10}^2(x)\ \pi_1^2(x)\ dx\right) \left(\int K(y)\ dy\right) = O\left(T\sqrt{h}\right) \to \infty
\end{aligned}$$

using $\sigma_{T1} = O\left(T\sqrt{h}\right)$ and Assumption 3.1, where $\pi_1(v)$ denotes the marginal density of $\{v_t\}$ under $\mathcal{H}_1$.

Proof of Theorem 3.1(ii): In view of the definition of $\widehat{L}_T(h)$ and the proofs of Lemmas A.7–A.9, it may be shown that as $T \to \infty$

$$\begin{aligned}
\widehat{L}_T(h) &= \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s\ K_h(\widehat{v}_{s-1} - \widehat{v}_{t-1})\ \widehat{u}_t}{\sqrt{2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^2\ K_h^2(\widehat{v}_{s-1} - \widehat{v}_{t-1})\ \widehat{u}_t^2}} \\
&= (1+o_P(1))\ \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T u_s\ K_h(v_{s-1} - v_{t-1})\ u_t}{\sigma_{T1}} \\
&+ (1+o_P(1))\ \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T f_{10}(v_{s-1})\ K_h(v_{s-1} - v_{t-1})\ f_{10}(v_{t-1})}{\sigma_{T1}}.
\end{aligned}$$

The proof of Theorem 3.1(ii) follows from Lemma A.9.

References

Chen, S. X., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series regression (available at http://www.sciencedirect.com/science/journal/03044076). *Journal of Econometrics* **138** (in press, 23 pages).

Dickey, D. A., Fuller, W. A., 1979. Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74** 427–431.

Dickey, D. A., Fuller, W. A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057–1072.

Dufour, J. M., King, M. L., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *Journal of Econometrics* **47** 115–143.

Durbin, J., Watson, G. S., 1950. Testing for serial correlation in least squares regression I. *Biometrika* **37** 409–428.

Durbin, J., Watson, G. S., 1951. Testing for serial correlation in least squares regression II. *Biometrika* **38** 159–178.

Evans, G. B. A., Savin, N. E., 1981. Testing for unit roots: I. *Econometrica* **49** 753–779.

Evans, G. B. A., Savin, N. E., 1984. Testing for unit roots: II. *Econometrica* **52** 1241–1269.

Fan, Y., Linton, O., 2003. Some higher–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/CRC, London.

Gao, J., Gijbels, I., 2006. Selection of smoothing parameters in nonparametric and semiparametric testing. Working paper available from www.maths.uwa.edu.au/˜jiti/gg51.pdf.

Gao, J., King, M. L., 2004. Adaptive testing in continuous–time diffusion models. *Econometric Theory* **20** 844–882.

Gao, J., King, M. L., 2007. Specification testing in parametric linear models with nonparametric autoregressive errors. Working paper available at www.maths.uwa.edu.au/˜jiti/gaoking65.pdf.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2006. Specification testing in nonlinear time series with nonstationarity. Working paper available from www.maths.uwa.edu.au/˜jiti/gklt.pdf.

Granger, C. W. J., Inoue, T., Morin, N., 1997. Nonlinear stochastic trends. *Journal of Econometrics* **81** 65–92.

Granger, C. W. J., Teräsvirta, T., 1993. *Modelling Nonlinear Dynamic Relationships*. Oxford University Press.

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

Hidalgo, F. J., 1992. Adaptive semiparametric estimation in the presence of autocorrelation of unknown form. *Journal of Time Series Analysis* **13** 47–78.

Kapetanios, G., Shin, Y., Snell, A., 2003. Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics* **112** 359–379.

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

King, M. L., 1987. Testing for autocorrelation in linear regression models: a survey. *Specification Analysis in the Linear Model* (M. L. King and D. E. A. Giles eds), 19–73. Routledge and Kegan Paul, London.

King, M. L., 2001. Serial correlation. In B.H. Baltagi, ed., *A Companion to Theoretical Econometrics*. Blackwell Publishers, 62–81.

King, M. L., Wu, X., 1997. Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* **16** 131–156.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* **54** 159–178.

Li, Q., 1999. Consistent model specification tests for time series econometric models. *Journal of Econometrics* **92** 101–147.

Li, Q., Hsiao, C., 1998. Testing serial correlation in semiparametric panel data models. Journal of Econometrics **87** 207–237.

Li, Q., Wang, S., 1998. A simple consistent bootstrap tests for a parametric regression functional form. *Journal of Econometrics* **87** 145–165.

Lobato, I., Robinson, P. M., 1998. A nonparametric test for I(0). *Review of Economic Studies* **65** 475-95.

Masry, E., Tjøstheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH models. *Econometric Theory* **11** 258–289.

Phillips, P. C. B., 1987. Time series regression with a unit root. *Econometrica* **55** 277–302.

Phillips, P. C. B., 1997. Unit root tests. *Encyclopedia of Statistical Sciences* (Ed. by Samuel Kotz). **1** 531–542.

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

Phillips, P. C. B., Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* **75** 335–346.

Phillips, P. C. B., Xiao, Z., 1998. A primer on unit root testing. *Journal of Economic Surveys* **12** 423–469.

Robinson, P. M., 2003. Efficient Tests of Nonstationary Hypotheses. *Recent Developments in Time Series*. Volume 1, 526–43. Elgar Reference Collection. International Library of Critical Writings in Econometrics. Cheltenham, United Kingdom.

Tong, H., 1990. *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Vogelsang, T., J., 1998. Trend function hypothesis testing in the presence of serial correlation. *Econometrica* **66** 123–148.

Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* **75** 263–289.

Zheng, X., Basher, R. E., 1999. Structural time series models and trend detection in global and regional temperature series. *Journal of Climate* **12** 2347-2358.