ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

A New Test in Parametric Linear Models against Nonparametric Autoregressive Errors

Jiti Gao and Maxwell King

September 2011

Working Paper 20/11

A New Test in Parametric Linear Models against Nonparametric Autoregressive Errors

By Jiti Gao[1] and Maxwell King

The University of Adelaide 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: Autoregressive process; nonlinear time series; nonparametric method; random walk; semiparametric model; unit root test.

JEL Classification: C12, C14, C22

[1]Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East Victoria 3145, Australia. Email: jiti.gao@monash.edu.

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 $\{Y_t : 1 \le t \le T\}$, $\{X_t\}$ is a vector of known deterministic 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). Other 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).

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

$$v_t = g(v_{t-1}) + u_t, \ t = 1, 2, \cdots, T, \tag{1.2}$$

where $g(\cdot)$ is an unknown function defined over $R^1 = (-\infty, \infty)$, $\{u_t\}$ is a sequence of stationary errors with mean zero and finite variance $\sigma_u^2 = E[u_1^2]$, $\{v_t : t \ge 1\}$ is also a sequence of errors with $E[v_t] = 0$, and v_0 is an initial value. Note that v_0 can be either a given initial value or any $O_P(1)$ random variable. We however set $v_0 = 0$ to avoid some unnecessary complications in exposition.

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

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

Existing studies (see, for example, Masry and Tjøstheim 1995) already discuss the case where $\beta \equiv 0$ and $\{v_t\}$ is strictly stationary when certain technical conditions are imposed on the form of $g(\cdot)$. Meanwhile, various existing studies (see, for example, Koul and Stute 1999; Gao 2007 and the references therein) focus on nonparametric estimation and specification testing for the case where $\{v_t\}$ is stationary since the publication of Robinson (1983).

To the best of our knowledge, semiparametric estimation of β and $g(\cdot)$ for the case where $\{v_t\}$ is stationary has only been discussed in Hidalgo (1992). Nonparametric estimation of $g(\cdot)$ for the case where $v_t = v_{t-1} + u_t$ has been done in Phillips and Park (1998), Karlsen and Tjøstheim (2001), and Wang and Phillips (2009).

Model (1.3) is quite general and covers many important cases. For example, in order to test whether $\{Y_t\}$ follows a nonstationary model of the form

$$Y_t = \sum_{i=0}^{d} \gamma_i \ t^i + Y_{t-1} + u_t, \tag{1.4}$$

it suffices to test whether $\mathcal{H}_0: \ v_t = v_{t-1} + u_t$ holds in a $(d+1)$–order polynomial trend model of the form

$$Y_t = \sum_{j=0}^{d+1} \beta_j \ t^j + v_t. \tag{1.5}$$

This paper is then concerned with testing

$$\mathcal{H}_0: \ g(v) = f_0(v, \theta_0) \quad \text{versus} \quad \mathcal{H}_1: \ g(v) = f_1(v, \theta_1) \tag{1.6}$$

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 an unknown semiparametric function indexed by a vector of unknown parameters θ_1.

Forms of $f_0(v, \theta_0)$ include the case of $f_0(v, \theta_0) \equiv 0$. In this case, $v_t = u_t$ and thus $\{v_t\}$ is a sequence of stationary errors. When $\theta_0 = 1$ 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 partially linear AR(1) model of the form $v_t = \theta_1 v_{t-1} + \psi(v_{t-1}) + u_t$, where $\psi(\cdot)$ is an unknown function such that $\min_{\alpha,\beta} E\left[\psi(v_1) - \alpha - \beta v_1\right]^2 \geq M$ for some positive constant M. This is needed to ensure that both θ_1 and $\psi(\cdot)$ are identifiable and estimable.

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 $f_0(v, \theta_0) = \theta_0 v$ with $\theta_0 = 1$. This implies $v_t = \theta_0 v_{t-1} + u_t$ with $\theta_0 = 1$ 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 $\mathcal{H}_1$.

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 $v_t = \rho v_{t-1} + u_t$ for example. 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 Appendices A and B.

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

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

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

Note that $p(\cdot)$ may depend on t when $\{v_t\}$ is nonstationary. Note also that the sample analogue of $E\left[u_t E\left(u_t|v_{t-1}\right) p(v_{t-1})\right]$ is $\frac{1}{T}\sum_{t=1}^{T} u_t E[u_t|v_{t-1}]p(v_{t-1})$. When $E[u_t|v_{t-1}]p(v_{t-1})$ is replaced by a kind of kernel–based sample analogue of the form $\frac{1}{Th}\sum_{s=1}^{n} K_h(v_{s-1} - v_{t-1})u_s$, a kernel–based sample analogue of (2.2) is of the form

$$M_T = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{Th}\sum_{s=1}^{T} u_s\ K_h(v_{s-1} - v_{t-1})\right)\ u_t, \tag{2.3}$$

where $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$ with $K(\cdot)$ being a probability kernel function and h a bandwidth parameter. This suggests using a centralized and then normalized kernel–based sample analogue of (2.2) of the form

$$L_T = L_T(h) = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\ K_h(v_{s-1} - v_{t-1})\ u_t}{\widetilde{\sigma}_T}, \tag{2.4}$$

where $\widetilde{\sigma}_T^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s^2\ K_h^2(v_{s-1} - v_{t-1})\ u_t^2$.

Note that existing literature (see, for example, Chapter 3 of Gao 2007) shows that $\frac{\widetilde{\sigma}_T^2}{\sigma_T^2} \to_P 1$ as $T \to \infty$ when both v_t and u_t are stationary, where $\sigma_T^2 = E\left[\widetilde{\sigma}_T^2\right]$. In the case where $\{v_t\}$ is nonstationary, however, we have only been able to show that $\frac{\widetilde{\sigma}_T^2}{\sigma_T^2} \to \xi^2$ in distribution for some random variable ξ. This is why the proposed test is based on a stochastically normalized version. As a consequence, the proposed test is asymptotically normal regardless of whether or not the errors are stationary, mainly due to the applicability of Lemma B.1 in Appendix B below.

The form of $L_T(h)$ 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). Such tests are extensions of existing tests proposed in Zheng (1996), Li and Wang (1998), Li (1999), and Fan and Linton (2003).

In the original working paper, Gao *et al.* (2006) propose using a version similar to (2.4) for parametric specification in both the nonparametric autoregression model of the form $X_t = g(X_{t-1}) + u_t$ and the nonparametric time series regression model of the form $Y_t = m(X_t) + e_t$ with $X_t = X_{t-1} + u_t$, where $\{u_t\}$ is assumed to be a sequence of independent and normally distributed errors. In the recent published papers, Gao *et al* (2009a, 2009b) consider the specification testing problems for the case where $\{u_t\}$ is assumed to be a sequence of independent and identically distributed errors.

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

$$\widehat{L}_T = \widehat{L}_T(h) = \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s\ K_h(\widehat{v}_{s-1} - \widehat{v}_{t-1})\ \widehat{u}_t}{\widehat{\widehat{\sigma}}_T}, \tag{2.5}$$

where $\widehat{\widehat{\sigma}}_T^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s^2\ K_h^2(\widehat{v}_{s-1} - \widehat{v}_{t-1})\ \widehat{u}_t^2$ and $\widehat{u}_t = \widehat{v}_t - f_0(\widehat{v}_{t-1}, \widehat{\theta}_0)$, 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 Assumption 2.1 for Case A.

ASSUMPTION 2.1. (i) *Let $\{u_t\}$ be a stationary ergodic sequence of martingale differences satisfying $E[u_t|\mathcal{F}_{t-1}] = 0$ and $E[u_t^4|\mathcal{F}_{t-1}] < \infty$ almost surely, where $\{\mathcal{F}_t\}$ is a sequence of σ–fields generated by $\{u_s : 1 \leq s \leq t\}$. Let $\sigma_u^2 = E[u_1^2]$.*

(ii) *Suppose that $\{u_t\}$ has a symmetric marginal density function $g(x)$. Let $g^{(i)}(x)$ be the ith derivative of $g(x)$ and $g^{(i)}(x)$ be continuous at $x \in (-\infty, \infty)$ for $i = 1$.*

For any $m \geq 2$, let $S_{m,t} = \frac{1}{\sqrt{m}\sigma_u}\sum_{s=t+1}^{t+m} u_s$, $f_{m,t}(x)$ be the marginal density function of $S_{m,t}$ and $f_{m,t}(x|\mathcal{F}_t)$ be the conditional density function of $S_{m,t}$ given $\mathcal{F}_t$. Let $f_{m,t}^{(i)}(x)$ and $f_{m,t}^{(i)}(x|\mathcal{F}_t)$ be the respective ith derivatives of $f_{m,t}(x)$ and $f_{m,t}(x|\mathcal{F}_t)$ with respect to x and both $f_{m,t}^{(i)}(x)$ and $f_{m,t}^{(i)}(x|\mathcal{F}_t)$ be continuous at $x \in (-\infty, \infty)$. Suppose that for $i = 0, 1$,

$$\inf_{\delta>0} \lim_{m\to\infty} \sup_{t\geq 1} \sup_{|x|\leq\delta} f_{m,t}^{(i)}(x) < \infty \quad \textit{and} \tag{2.6}$$

$$\inf_{\delta>0} \lim_{m\to\infty} \sup_{t\geq 1} \sup_{|x|\leq\delta} f_{m,t}^{(i)}(x|\mathcal{F}_t) < \infty \quad \textit{with probability one.} \tag{2.7}$$

For Case B, we need the following assumption.

ASSUMPTION 2.2. (i) *Let Assumption 2.1(i) hold. In addition, the marginal density of $\{u_t\}$ is positive and lower–semicontinuous over R^1.*

(ii) *$f_0(v, \theta_0)$ is bounded on any bounded Borel measurable set of R^1. Suppose that there is some constant $|\theta_0| < 1$ such that $f_0(v, \theta_0) = \theta_0 v + o(|v|)$ as $|v| \to \infty$.*

Assumption 2.1(i) assumes that $\{u_t\}$ is a sequence of stationary martingale differences. This is quite general in this kind of problem. Assumption 2.1(ii) imposes a set of general conditions on the marginal and conditional density functions. Similar conditions have been used by Assumption **A4** of Chen, Gao and Li (2007) and Assumption 2.3(ii) of Wang and Phillips (2009). Since $v_t = \sum_{i=1}^{t} u_i$ is a random walk process under H_0, we need to impose certain conditions on the distributional structure of a normalized version of v_t of the form $S_{m,t} = \frac{1}{\sqrt{m}\sigma_u}(v_{t+m} - v_t) = \frac{1}{\sqrt{m}\sigma_u}\sum_{s=t+1}^{t+m} u_s$. Equations (2.6) and (2.7) basically require that the density and conditional density functions and their derivatives are bounded uniformly in $t \geq 1$, $m \to \infty$ and $|x| \leq \delta$ for all small $\delta > 0$.

Equations (2.6) and (2.7) are justifiable. When $\{u_t\}$ is a sequence of independent and identically distributed random variables for example, equation (2.7) reduces to (2.6), which follows from as $m \to \infty$

$$\sup_x |\phi_m(x) - \phi(x)| \ \to \ 0 \ \text{ and } \ \sup_x \left|\phi_m^{(1)}(x) - \phi^{(1)}(x)\right| \to 0, \tag{2.8}$$

under the condition $\int_{-\infty}^{\infty} |v||\psi(v)|dv < \infty$, where $\psi(\cdot)$ is the characteristic function of u_1, $\phi_m^{(1)}(x)$ and $\phi^{(1)}(x)$ are the first derivatives of $\phi_T(x)$, which is the density function of $\frac{1}{\sqrt{m}\sigma_u}\sum_{t=1}^{m} u_t$, and $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ is the density function of the standard normal random variable $N(0,1)$, respectively. The proof of (2.8) is quite standard (see, for example, the proof of Corollary 2.2 of Wang and Phillips 2009).

Assumption 2.2 implies that (see, for example, Tong 1990; Lu 1998; Meitz and Saikkonen 2008) $\{v_t\}$ is strictly stationary and α–mixing with mixing coefficient

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

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 $t \geq 1$.

We now establish the following theorem; its proof is given in Appendix A.

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* $\mathcal{H}_0$

$$\widehat{L}_T(h) \to_D N(0,1) \text{ as } T \to \infty. \tag{2.9}$$

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 small sample properties when using a large sample normal distribution to approximate the small 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 thus propose choosing a simulated α–level critical value, $l_\alpha^*(h)$, by using the following simulation procedure:

(i) Let $Y_0^* = y_0^*$ and $X_0 = x_0$ be the initial values. 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 $\widehat{\beta}$, $\widehat{\theta}_0$ and $\widehat{\sigma}_u^2$ are the respective consistent estimators of β, θ_0 and σ_u^2 based on the original sample $\mathcal{W}_T = \{(X_1, Y_1), \cdots, (X_T, Y_T)\}$, which acts in the resampling as a fixed design, and $\{u_t^*\}$ is generated independently by an existing parametric or nonparametric bootstrap method such that $E\left[u_t^*\right] = 0$, $E\left[u_t^{*2}\right] = 1$ and $E\left[u_t^{*4}\right] < \infty$.

(ii) 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 estimators 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 $\widehat{v}_t^* = Y_t^* - X_t^\tau \widehat{\beta}^*$.

(iii) 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 versions 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 $P^*\left(\widehat{L}_T^*(h) \le x\right) = P\left(\widehat{L}_T^*(h) \le x|\mathcal{W}_T\right)$. Let $l_\alpha^*(h)$ satisfy $P^*\left(\widehat{L}_T^*(h) \ge l_\alpha^*(h)\right) = \alpha$ and then estimate $l_\alpha(h)$ by $l_\alpha^*(h)$.

Define the size and power functions by

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

The objective is to choose an optimal bandwidth, $\widehat{h}_{\text{test}}$, such that the power function $\beta^*(h)$ is maximized at $h = \widehat{h}_{\text{test}}$ while the size function $\alpha^*(h)$ is under control.

Let $H_T = \{h : \alpha - \varepsilon_0 < \alpha^*(h) < \alpha + \varepsilon_0\}$ for some $0 < \varepsilon_0 < \alpha$. Choose an optimal bandwidth $\widehat{h}_{\text{test}}$ such that

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

Since $\{v_t\}$ under $\mathcal{H}_1$ is stationary, existing results (§3 of Gao and Gijbels 2008) suggest using an approximate version of the form

$$\widehat{h}_{\text{test}} = \widehat{a}^{-\frac{1}{2}} \widehat{C}_T^{-\frac{3}{2}}, \tag{3.4}$$

where $\widehat{C}_T^2 = \frac{\sum_{t=1}^T \left(\widehat{f}_1(\widehat{v}_{t-1}, \widehat{\theta}_1) - f_0(\widehat{v}_{t-1}, \widehat{\theta}_0)\right)^2 \widehat{p}(\widehat{v}_{t-1})}{\widehat{\mu}_2 \sqrt{2\widehat{\nu}_2} \int K^2(v)dv}$ and $\widehat{a} = \frac{\sqrt{2}K^{(3)}(0)}{3\left(\sqrt{\int K^2(u)du}\right)^3} \widehat{c}(p)$ with $\widehat{c}(p) = \frac{\frac{1}{T}\sum_{t=1}^T \widehat{p}^2(\widehat{v}_{t-1})}{\left(\sqrt{\frac{1}{T}\sum_{t=1}^T \widehat{p}(\widehat{v}_{t-1})}\right)^3}$, in which $\widehat{\mu}_2 = \frac{1}{T}\sum_{t=1}^T \left(\widehat{v}_t - f_0(\widehat{v}_{t-1}, \widehat{\theta}_0)\right)^2$, $\widehat{\nu}_2 = \frac{1}{T}\sum_{t=1}^T \widehat{p}^2(\widehat{v}_{t-1})$, $\widehat{f}_1(v, \widehat{\theta}_1)$ is a consistent estimate of $f_1(v, \theta_1)$, $\widehat{p}(v) = \frac{1}{T\widehat{h}_{\text{cv}}} \sum_{t=1}^T K\left(\frac{\widehat{v}_{t-1} - v}{\widehat{h}_{\text{cv}}}\right)$ with $\widehat{h}_{\text{cv}}$ being chosen by a conventional cross–validation selection method, and $K^{(3)}(\cdot)$ is the three–time convolution of $K(\cdot)$ with itself.

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 block bootstrap (see, for example, Pararoditis and Politis 2003) to generate a sequence of resamples for $\{u_t^*\}$. Since the combination of the proposed simulation procedure with the power–based bandwidth selection method works well in this paper, we use the proposed bootstrap method for both theoretical studies and practical applications.

Under $\mathcal{H}_1$, model (1.3) becomes

$$Y_t = X_t^\tau \beta + v_t \quad \text{with} \quad v_t = f_1(v_{t-1}, \theta_1) + u_t, \tag{3.5}$$

where $f_1(v, \theta_1)$ can be consistently estimated by $\widehat{f}_1(v, \widehat{\theta}_1)$, which depends on the specification of $f_1(v, \theta_1)$. For example, when $f_1(v, \theta_1) = g_1(v, \theta_1) + \psi(v)$ with $g_1(v, \theta_1)$ being parametric and $\psi(v)$ being nonparametric, the form of $\widehat{f}_1(v, \widehat{\theta}_1)$ can be given by

$$\widehat{f}_1(v, \widehat{\theta}_1) = g_1(v, \widehat{\theta}_1) + \widehat{\psi}(v), \tag{3.6}$$

in which

$$\widehat{\psi}(v) = \widehat{\psi}(v, \theta_1) = \frac{\sum_{s=1}^T K_{\widehat{h}_{\rm cv}}(\widehat{v}_{s-1} - v)\left(\widehat{v}_s - g_1(\widehat{v}_{s-1}, \theta_1)\right)}{\sum_{s=1}^T K_{\widehat{h}_{\rm cv}}(\widehat{v}_{s-1} - v)} \quad \text{and} \tag{3.7}$$

$$\widehat{\theta}_1 = \arg\min_{\theta_1} \frac{1}{T}\sum_{t=1}^T \left(\widehat{v}_t - g_1(\widehat{v}_{t-1}, \theta_1) - \widehat{\psi}(\widehat{v}_{t-1}, \theta_1)\right)^2, \tag{3.8}$$

where $\widehat{h}_{\rm cv}$ is chosen by a conventional cross–validation selection method.

To study the power properties of $\widehat{L}_T(\widehat{h}_{\rm test})$, 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 Assumptions 2.1 and 2.2, we need the following assumption.

Assumption 3.1. (i) *Let Assumption 2.2 hold under $\mathcal{H}_1$.*

(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 $\mathcal{H}_1$.

Assumption 3.1(i) is a set of quite general conditions and also standard in this kind of stationary case, as assumed in the literature (see Li 1999 for example). 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 an 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)

$$\mathcal{H}_0: \ v_t = \rho_0 \ v_{t-1} + u_t \quad \text{versus} \quad \mathcal{H}_1: \ v_t = \rho_1 \ v_{t-1} - \frac{v_{t-1}}{1 + v_{t-1}^2} + u_t, \tag{3.9}$$

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 $\rho_0 = 1$. Assumption 2.2 implies that $\{v_t\}$ under $\mathcal{H}_1$ is stationary. In this case, Assumption 3.1(ii) becomes

$$\begin{aligned}\int [f_1(v, \theta_1) - f_0(v, \theta_0)]^2 \pi_1^2(v) dv &= \int \left(2(\rho_0 - \rho_1) + \frac{1}{1+v^2}\right) \frac{v^2}{1+v^2} \pi_1^2(v) \ dv \\ &+ \int (\rho_1 - \rho_0)^2 v^2 \pi_1^2(v) \ dv > 0\end{aligned} \tag{3.10}$$

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 theorem; its proof is given in Appendix A.

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* $\mathcal{H}_0$

$$\lim_{T\to\infty} P\left(\widehat{L}_T(h) > l_\alpha^* | \mathcal{W}_T\right) = \alpha \quad \textit{in probability.}$$

(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* $\mathcal{H}_1$

$$\lim_{T\to\infty} P\left(\widehat{L}_T(h) > l_\alpha^* | \mathcal{W}_T\right) = 1 \quad \textit{in probability.}$$

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.

4. Examples 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

$$Y_t = X_t\ \beta + v_t \quad \text{with} \quad v_t = f_i(v_{t-1}, \theta_i) + u_t,\ 1 \leq t \leq T, \tag{4.1}$$

where $X_t = \sin\left(\frac{2\pi t}{T}\right)$, $\{u_t\}$ 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:

$$f_0(v,\theta_0) = v \text{ and } f_1(v,\theta_1) = v + \theta_1\ v \text{ for Case A, or} \tag{4.2}$$

$$f_0(v,\theta_0) = v \text{ and } f_1(v,\theta_1) = v + \theta_1\ v + \frac{\theta_1\ v}{1+v^2} \text{ for Case A, or} \tag{4.3}$$

$$f_0(v,\theta_0) = 0 \text{ and } f_1(v,\theta_1) = \theta_1\ v \text{ for Case B, or} \tag{4.4}$$

$$f_0(v,\theta_0) = 0.5\ v \text{ and } f_1(v,\theta_1) = 0.5v + \theta_1\ v + \frac{\theta_1\ v}{1+v^2} \text{ for Case B,} \tag{4.5}$$

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), $\theta_1 = -\sqrt{T^{-1}\ \log(\log(T))}$ and $\rho_1 = \rho_0 + \theta_1$. The rate of $\theta_1 = -T^{-\frac{1}{2}}\sqrt{\text{loglog}(T)}$ is chosen because it is an optimal rate of testing in this kind of nonparametric kernel testing problem as discussed in Chapter 3 of Gao (2007). The β parameter is estimated by the conventional semiparametric least squares estimation method (see, for example, Hidalgo 1992). Equations (3.6)–(3.8) are used in the estimation of θ_1. We choose $K(x) = \frac{1}{2} I_{[-1,1]}(x)$ and $\varepsilon_0 = \frac{\alpha}{10}$ involved in (3.3) throughout this section.

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)–(4.5), we first propose using $\widehat{L}_T(h)$ associated with $\widehat{h}_{\text{test}}$ of (3.3). Let

$$L_{1\text{test}} = \widehat{L}_T(\widehat{h}_{\text{test}}). \tag{4.6}$$

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

$$L_{21} = \frac{\sum_{t=2}^T (\widehat{v}_t - \widehat{v}_{t-1})\widehat{v}_{t-1}}{\widehat{\sigma}_{22}\sqrt{\sum_{t=2}^T \widehat{v}_{t-1}^2}}, \tag{4.7}$$

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

$$L_{22} = \frac{\sum_{t=1}^T \sum_{s=1}^T \widehat{v}_s \; a_{st} \; \widehat{v}_t}{\sum_{t=1}^T \sum_{s=1}^T \widehat{v}_s \; b_{st} \; \widehat{v}_t}, \tag{4.8}$$

where $\{a_{st}\}$ 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).

For $i = 1, 2$, let $l_{2i,\alpha}^*$ be the corresponding simulated critical value of L_{2i}. Each of them is computed in the same way as has been proposed in the Simulation Scheme in Section 3. Let z_α be the $1-\alpha$ quantile of the standard Normal distribution. Note that $z_{0.05} = 1.645$ at the $\alpha = 5\%$ level and $z_{0.10} = 1.285$ at the $\alpha = 10\%$ level.

Let $l_{1,\alpha}^* = l_\alpha^*(h_{\text{test}})$ and $L_{1\text{cv}} = \widehat{L}_T(\widehat{h}_{\text{cv}})$, where $\widehat{h}_{\text{cv}}$ is chosen such that

$$\widehat{h}_{\text{cv}} = \arg\min_{h \in H_{\text{cv}}} \frac{1}{T} \sum_{t=1}^T \left(\widehat{v}_t - \widehat{g}_{-t}(\widehat{v}_{t-1}, h)\right)^2, \tag{4.9}$$

in which $\widehat{g}_{-t}(\widehat{v}_{t-1}, h) = \frac{\sum_{s=1,\neq t}^T K\left(\frac{\widehat{v}_{s-1} - \widehat{v}_{t-1}}{h}\right)\widehat{v}_s}{\sum_{u=1,\neq t}^T K\left(\frac{\widehat{v}_{u-1} - \widehat{v}_{t-1}}{h}\right)}$ and $H_{\text{cv}} = \left[T^{-1}, T^{-(1-\delta_0)}\right]$, where $0 < \delta_0 < 1$ is chosen such that $\widehat{h}_{\text{cv}}$ is achievable and unique in each individual case.

We choose $N = 250$ in the Simulation Scheme and use $M = 1000$ replications to compute the two–sized power and size values of the tests in Tables 4.1–4.4 below. Let f_{test} denote the frequency of $L_{1\text{test}} > l_{1,\alpha}^*$, f_{cv} be the frequency of $L_{1\text{cv}} > z_\alpha$, and f_{2i} be the frequency of $L_{2i} > l_{2i,\alpha}^*$ for $i = 1, 2$ and at $\alpha = 5\%$ or 10%.

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

Observation	Model (4.2)			Model (4.3)		
Null Hypothesis Is True						
n	f_{cv}	f_{test}	f_{21}	f_{cv}	f_{test}	f_{21}
250	0.007	0.045	0.046	0.005	0.048	0.058
500	0.003	0.042	0.051	0.006	0.054	0.049
750	0.005	0.053	0.057	0.003	0.044	0.052
Null Hypothesis Is False						
n	f_{cv}	f_{test}	f_{21}	f_{cv}	f_{test}	f_{21}
250	0.171	0.302	0.701	0.462	0.521	0.350
500	0.180	0.345	0.734	0.456	0.554	0.376
750	0.192	0.329	0.752	0.474	0.573	0.402

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

Observation	Model (4.2)			Model (4.3)		
Null Hypothesis Is True						
n	f_{cv}	f_{test}	f_{21}	f_{cv}	f_{test}	f_{21}
250	0.023	0.088	0.107	0.019	0.094	0.096
500	0.038	0.092	0.098	0.035	0.103	0.109
750	0.029	0.103	0.102	0.037	0.089	0.094
Null Hypothesis Is False						
n	f_{cv}	f_{test}	f_{21}	f_{cv}	f_{test}	f_{21}
250	0.201	0.432	0.821	0.536	0.631	0.473
500	0.199	0.469	0.847	0.547	0.655	0.489
750	0.234	0.487	0.862	0.561	0.649	0.512

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

Observation	Model (4.4)			Model (4.5)		
Null Hypothesis Is True						
n	f_{cv}	f_{test}	f_{22}	f_{cv}	f_{test}	f_{22}
250	0.005	0.052	0.049	0.003	0.051	0.048
500	0.004	0.048	0.050	0.007	0.047	0.054
750	0.007	0.051	0.047	0.006	0.052	0.051
Null Hypothesis Is False						
n	f_{cv}	f_{test}	f_{22}	f_{cv}	f_{test}	f_{22}
250	0.112	0.164	0.348	0.349	0.423	0.312
500	0.107	0.182	0.387	0.361	0.456	0.331
750	0.132	0.191	0.372	0.358	0.481	0.342

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

Observation	Model (4.4)			Model (4.5)		
Null Hypothesis Is True						
n	f_{cv}	f_{test}	f_{22}	f_{cv}	f_{test}	f_{22}
250	0.031	0.110	0.097	0.023	0.089	0.101
500	0.040	0.097	0.102	0.038	0.101	0.097
750	0.033	0.103	0.096	0.033	0.098	0.095
Null Hypothesis Is False						
n	f_{cv}	f_{test}	f_{22}	f_{cv}	f_{test}	f_{22}
250	0.197	0.271	0.411	0.452	0.552	0.419
500	0.204	0.267	0.431	0.489	0.581	0.441
750	0.226	0.283	0.456	0.516	0.614	0.476

Tables 4.1 and 4.2 (columns 2–3 and 5–6) show that the test coupled with a bootstrap critical value (bcv) is more powerful than that associated with the use of an asymptotic critical value (acv) in each case, in addition to the fact that there is serious size distortion when using an acv rather than a bcv. The main reasons are as follows: (a) the rate of convergence of each $\widehat{L}_T(h)$ to an asymptotic normal distribution is quite slow in this kind of nonparametric setting; and (b) the use of an optimal bandwidth based on the cross–validation selection criterion may not be optimal for testing purposes. By contrast, there is only small size distortion between using a bcv and an acv for L_{21} and L_{22} in each implementation, although the version of the test associated with a bcv has more stable size performance and better power property than that based on an acv. We therefore compare our nonparametric tests with both L_{21} and L_{22} based on a bcv in each case.

Moreover, 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 seven–day Eurodollar deposit spot rate data given in Figure 1 below sampled daily over the period from 1 June 1973 to 25 February 1995, providing 5505 observations.

Let $\{Y_t : t = 1, 2, \cdots, 5505\}$ be the set of the seven–day Eurodollar deposit spot rate data. The data set has been studied extensively in the literature. Recent studies (see, for example, Bandi 2002) are concerned with whether $\{Y_t\}$ follows a random walk model of the form

$$Y_t = \mu_0 + \mu_1 t + Y_{t-1} + u_t, \tag{4.10}$$

where $\{u_t\}$ is a sequence of strictly stationary errors.

We consider a special form of (1.3) with $X_t^\tau \beta = \beta_0 + \beta_1 t + \beta_2 t^2$. In this case, in

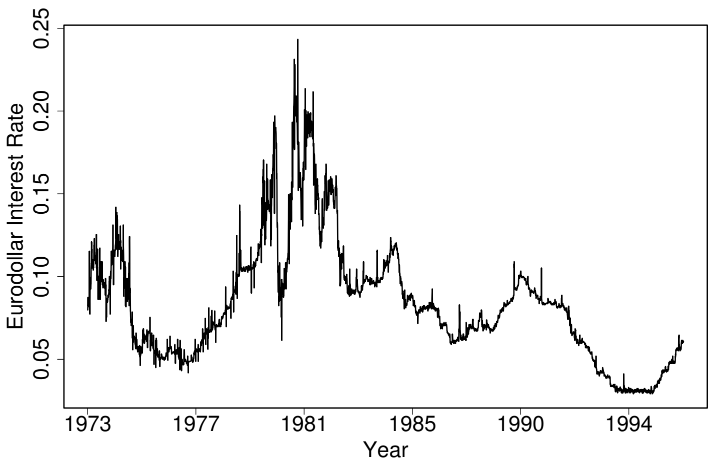

Figure 1: Plot of the seven–day Eurodollar deposit spot rate data

order to apply model (1.3) to test whether $\{Y_t\}$ follows (4.10), it suffices to test

$$\mathcal{H}_0: \ v_t = v_{t-1} + u_t \quad \text{for} \quad Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + v_t. \tag{4.11}$$

To apply the test $\widehat{L}_T(\widehat{h}_{\text{test}})$ to determine whether $\{Y_t\}$ follows a random walk model of the form $Y_t = \mu_0 + \mu_1 t + Y_{t-1} + u_t$, 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 $Y_t^* = Y_{t-1}^* + (X_t - X_{t-1})^\tau \widehat{\beta} + u_t^*$ for $2 \leq t \leq 5505$, where $u_t^* = \widehat{u}_t \eta_t$, in which $\widehat{u}_t = Y_t - Y_{t-1} - (X_t - X_{t-1})^\tau \widehat{\beta}$ and $\{\eta_t\}$ is chosen as a sequence of independent random variables with the following distributional structure: $P\left(\eta_1 = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $P\left(\eta_1 = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}$. Such two–point distributional structure has been commonly used in the literature (see, for example, Li and Wang 1998).
- Compute the corresponding version $\widehat{L}_T^*(\widehat{h}_{\text{test}})$ 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 $\widehat{p}_1 = 0.007$ for L_{22} and $\widehat{p}_2 = 0.013$ 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. This supports the existing conclusions made in Bandi (2002).

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.

This paper has focused on the case where $\{X_t\}$ is a vector of deterministic regressors. The case where $\{X_t\}$ is a vector of stochastic regressors is equally important. Discussion of such a case requires developing new theory and also involves more technicalities. It is therefore left for future research.

6. Acknowledgments. The authors would like to thank Dr Jiying Yin for his excellent computing assistance and the Australian Research Council Discovery Grants under Grant Numbers: DP0558602 and DP0879088 for the financial support.

Appendix A

In this appendix, we introduce several technical conditions and then give some lemmas for the proofs of Theorems 2.1 and 3.1. Assumptions A.1–A.3 are imposed for Case A and Assumptions A.1–A.5 are needed for Case B. To avoid adding some non–essential technicalities, we assume the following initial values $Y_0^* = y_0^* = 0$ and $X_0 = x_0 = 0$, $v_0 = 0$ and $v_0^* = 0$ throughout this appendix.

A.1. *Assumptions*

Assumption A.1. (i) *Let $K(\cdot)$ be a symmetric probability density function with compact support $C(K)$. Let also the existence of $K^{(3)}(\cdot)$, the three–time convolution of $K(\cdot)$ with itself. In addition, there is some positive function $M(\cdot)$ such that*

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

for all $x \in C(K)$ and any small y, where $M(\cdot) \geq 0$ is assumed to satisfy $\int M^2(u)du < \infty$.

(ii) *For Case A, let h satisfy $\lim_{T\to\infty} T^{\frac{3}{10}} h = 0$ and $\limsup_{T\to\infty} T^{\frac{1}{2}-\epsilon_0} h = \infty$ for any $0 < \epsilon_0 < \frac{1}{5}$. Let h satisfy $\lim_{T\to\infty} h = 0$ and $\limsup_{T\to\infty} Th = \infty$ for Case B.*

Assumption A.2. *For $i = 1, 2$, let*

$$\lim_{T\to\infty} \frac{h \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{||Z_t||^i ||Z_s||^i}{\sqrt{t-s}}}{R_T^{2i}\ \lambda_T^i} = 0, \tag{A.1}$$

$$\lim_{T\to\infty} \frac{\sum_{t=2}^{T} \sum_{s=1}^{t-1} ||Z_t||^i \frac{||Z_{t-1}-Z_{s-1}||^i}{\sqrt{t-s}} ||Z_s||^i}{R_T^{3i}\ \lambda_T^i} = 0, \tag{A.2}$$

where $Z_t = X_t - X_{t-1}$ for Case A and $Z_t = X_t$ for Case B, $\lambda_T = T^{\frac{3}{4}}\sqrt{h}$, R_T is chosen such that Assumption A.3 below holds, and $||\cdot||$ denotes the Euclidean norm.

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

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

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*

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

holds with probability one with respect to the distribution of $\mathcal{W}_T$, where $R_T \to \infty$ is the same as in (i).

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. Suppose that $\pi_0(v)$ is continuous and that $f_0(v, \theta)$ is differentiable in both v and θ. In addition,*

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

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. Suppose that* $\pi_1(v)$ *is continuous and that* $f_1(v, \theta)$ *is differentiable in both* v *and* θ*. In addition,*

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

Assumption A.1(i) is a mild condition and holds in many cases. For example, Assumption A.1(i) holds when $K(x) = \frac{1}{2} I_{[-1,1]}(x)$. 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. The corresponding conditions on the bandwidth for nonparametric testing in the stationary case are the same as the minimal conditions: $\lim_{T\to\infty} h = 0$ and $\lim_{T\to\infty} Th = \infty$ that are assumed for nonparametric kernel testing for the case where both the regressors and errors are independent (see, for example, Gao 2007).

Assumption A.2 imposes some minimal conditions on the trend function such that polynomial trends are included. Consider the case where $X_t = t^2$ for Case A, we have for some $0 < C_1, C_2 < \infty$

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} |Z_t| \frac{1}{\sqrt{t-s}} |Z_s| \leq C_1 \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{st}{\sqrt{t-s}} = O\left(T^{\frac{7}{2}}\right),$$
$$R_T^2 = \sum_{t=1}^{T} Z_t^2 = C_2\, T^3 \quad \text{and } T^{\frac{3}{2}}\left(\widehat{\beta} - \beta\right) \to N(0, \sigma_1^2),$$

where $Z_t = X_t - X_{t-1}$, $\widehat{\beta} = \frac{\sum_{t=1}^T Z_t (Y_t - Y_{t-1})}{\sum_{t=1}^T Z_t^2}$ is the ordinary least squares estimator of β based on a model of the form $Y_t - Y_{t-1} = (X_t - X_{t-1})\beta + u_t$, and σ_1 is a positive constant.

In this case, equations (A.1) and (A.2) become respectively

$$\frac{h \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{|Z_t| \cdot |Z_s|}{\sqrt{t-s}}}{R_T^2\, \lambda_T} = O\left(\frac{T^{\frac{7}{2}} h}{T^{3+\frac{3}{4}}\sqrt{h}}\right) = O\left(\frac{\sqrt{h}}{T^{\frac{1}{4}}}\right) = o(1),$$
$$\frac{\sum_{t=2}^T \sum_{s=1}^{t-1} |Z_t| \frac{|Z_{t-1} - Z_{s-1}|}{\sqrt{t-s}} |Z_s|}{R_T^3\, \lambda_T} = O\left(\frac{T^{\frac{9}{2}}}{T^{\frac{9}{2}+\frac{3}{4}}\sqrt{h}}\right) = O\left(\frac{1}{T^{\frac{3}{4}}\sqrt{h}}\right) = o(1).$$

Similarly, in the case where $X_t = t^2$ for Case B, we have for some $0 < D_1, D_2 < \infty$

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} |X_t| \frac{1}{\sqrt{t-s}} |X_s| \leq D_1 \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{s^2 t^2}{\sqrt{t-s}} = O\left(T^{\frac{11}{2}}\right),$$
$$R_T^2 = \sum_{t=1}^{T} X_t^2 = D_2\, T^5 \quad \text{and } T^{\frac{5}{2}}\left(\widehat{\beta} - \beta\right) \to N(0, \sigma_2^2), \tag{A.3}$$

where $\widehat{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}$ is the ordinary least squares estimator of β based on a model of the form $Y_t = X_t \beta + v_t$, and σ_2 is a positive constant.

In this case, equations (A.1) and (A.2) become respectively

$$
\begin{aligned}
\frac{h\sum_{t=2}^T \sum_{s=1}^{t-1} \frac{|X_t|\cdot|X_s|}{\sqrt{t-s}}}{R_T^2\ \lambda_T} &= O\left(\frac{T^{\frac{11}{2}} h}{T^{5+\frac{3}{4}}\sqrt{h}}\right) = O\left(\frac{\sqrt{h}}{T^{\frac{1}{4}}}\right) = o(1),\\
\frac{\sum_{t=2}^T \sum_{s=1}^{t-1} |X_t| \frac{|X_{t-1}-X_{s-1}|}{\sqrt{t-s}} |X_s|}{R_T^3\ \lambda_T} &= O\left(\frac{T^{\frac{15}{2}}}{T^{\frac{15}{2}+\frac{3}{4}}\sqrt{h}}\right) = O\left(\frac{1}{T^{\frac{3}{4}}\sqrt{h}}\right) = o(1).
\end{aligned}
$$

Thus, equations (A.1) and (A.2) hold for $i = 1$. Similarly, we can show that the other cases for (A.1) and (A.2) all hold. In addition, Assumption A.2 is satisfied automatically when the trend functions are all continuous and bounded.

Assumption A.3 requires that the conventional rate of convergence for the parametric case is achievable even when $\{v_t\}$ is nonstationary. When $X_t = t^2$, it has been shown above that the rate of convergence of $\widehat{\beta}$ to β is proportional to $T^{\frac{3}{2}}$ in Case A and $T^{\frac{5}{2}}$ in Case B.

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 $\mathcal{H}_1$. Note that Assumptions A.4(i) and A.5(i)(ii) may also be satisfied even when $\{u_t\}$ is correlated. In this case, an instrumental–variable method may be used to construct a consistent estimator (see, for example, Frölich 2008)

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

Let $\sigma_u^2 = E[u_1^2] \equiv 1$ throughout the rest of this paper. To avoid notational complication, we introduce

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

Observe that

$$
\begin{aligned}
\widehat{M}_T &\equiv \sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widehat{u}_t = \sum_{t=1}^T \sum_{s=1,\neq t}^T u_s\ K_h(v_{s-1}-v_{t-1})\ u_t\\
&+ \sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t + 2\sum_{t=1}^T \sum_{s=1,\neq t}^T u_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t\\
&+ M_{T4} \equiv M_{T1} + M_{T2} + M_{T3} + M_{T4}, \qquad \text{(A.4)}\\
\widehat{\sigma}_T^2 &\equiv 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widehat{u}_t^2 = 2\sum_{t=1}^T \sum_{s=1,s\neq t}^T u_s^2\ K_h^2(v_{s-1}-v_{t-1})\ u_t^2\\
&+ 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t^2 + 2\sum_{t=1}^T \sum_{s=1,\neq t}^T u_s^2\ K_h^2(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t^2\\
&+ \widetilde{\sigma}_{T4}^2 \equiv \widetilde{\sigma}_{T1}^2 + \widetilde{\sigma}_{T2}^2 + \widetilde{\sigma}_{T3}^2 + \widetilde{\sigma}_{T4}^2, \qquad \text{(A.5)}
\end{aligned}
$$

where for Case A under $\mathcal{H}_0$: $v_t = v_{t-1} + u_t$,

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

In view of (A.4) and (A.5), in order to prove Theorem 2.1 for Case A, it suffices to show that as $T \to \infty$

$$\frac{M_{T1}}{\widetilde{\sigma}_{T1}} \to_D N(0,1), \tag{A.6}$$

$$\frac{M_{Ti}}{\widetilde{\sigma}_{T1}} \to_P 0 \quad \text{for } i = 2,3,4, \tag{A.7}$$

$$\frac{\widetilde{\sigma}_{Tj}}{\widetilde{\sigma}_{T1}} \to_P 0 \quad \text{for } j = 2,3,4. \tag{A.8}$$

We will return to the proof of (A.7) and (A.8) in Lemma A.5 after having proved (A.6) in Lemmas A.1–A.4 below. In order to prove (A.6), we need to apply Lemma B.1 of Appendix B below.

Before verifying the conditions of the Lemma B.1, we introduce the following notation. Let $Y_{Tt} = \frac{\eta_t u_t}{\sigma_{T1}}$, $\Omega_{T,s} = \sigma\{Y_{Tt} : 1 \leq t \leq s\}$ be a σ–field generated by $\{Y_{Tt} : 1 \leq t \leq s\}$, $\mathcal{G}_T = \Omega_{T,M(T)}$ and $\mathcal{G}_{T,s}$ be defined by

$$\mathcal{G}_{T,s} = \begin{cases} \Omega_{T,M(T)}, & 1 \leq s \leq M(T), \\ \Omega_{T,s}, & M(T)+1 \leq s \leq T, \end{cases} \tag{A.9}$$

where $\sigma_{T,1}^2 = \text{var}\left[\sum_{t=2}^T \eta_t u_t\right]$ and $M(T)$ 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 $\sigma_{S,1}^2 = \text{var}\left[\sum_{t=2}^S \eta_t u_t\right]$ for all $1 \leq S \leq T$. We can show that as $T \to \infty$

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} - \widetilde{U}_{M(T)}^2 \to_P 0. \tag{A.10}$$

Thus, condition (B.2) of the Lemma B.1 of Appendix B below can be satisfied. The proof of (A.10) is given in Lemma A.4 below.

Therefore, in view of the Lemma B.1, in order to prove that as $T \to \infty$

$$\frac{M_{T1}}{\widetilde{\sigma}_{T1}} = \frac{1}{\widetilde{\sigma}_{T1}} \sum_{t=2}^T \eta_t u_t \to_D N(0,1), \tag{A.11}$$

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

$$\sum_{t=2}^{T} E\left[Y_{Tt}^2 I_{\{[Y_{Tt}|>\epsilon]\}}(Y_{Tt})|\Omega_{T,t-1}\right] \to_P 0, \tag{A.12}$$

$$\sum_{t=2}^{T} E\left[Y_{Tt}|\mathcal{G}_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} + \sum_{t=M(T)+1}^{T} E\left[Y_{Tt}|\Omega_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} \to_P 0, \tag{A.13}$$

$$\sum_{t=2}^{T} |E\left[Y_{Tt}|\mathcal{G}_{T,t-1}\right]|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 + \sum_{t=M(T)+1}^{T} |E\left[Y_{Tt}|\Omega_{T,t-1}\right]|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 \to_P 0, \tag{A.14}$$

$$\lim_{\delta\to 0} \lim \inf_{T\to\infty} P\left(\frac{\widetilde{\sigma}_{T1}}{\sigma_{T1}} > \epsilon\right) = 1, \tag{A.15}$$

where $I_A(x)$ is the conventional indicator of the form $I_A(x) = 1$ when $x \in A$ and $I_A(x) = 0$ when $x \notin A$. The proof of (A.12) follows from Lemma A.2 below. The proof of (A.13) is similar to that of (A.14), which follows from

$$\sum_{t=2}^{M(T)} E\left[Y_{Tt}^2\right] = O\left(\left(\frac{M(T)}{T}\right)^{\frac{3}{2}}\right) \to 0 \tag{A.16}$$

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

In order to prove (A.12), it suffices to show that

$$\frac{1}{\sigma_{T1}^4} \sum_{t=2}^{T} E\left[\eta_t^4\right] \to 0, \tag{A.17}$$

which is given in Lemma A.2 below.

The proof of (A.15) follows from

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} \to_D \xi^2 > 0, \tag{A.18}$$

which is given in Lemma A.3 below.

Before we establish several lemmas for the proof of Theorem 2.1, we need to introduce the following notation.

For any $t > s \geq 1$ and $\alpha = \frac{1}{2}$, define $v_{st} = \frac{v_{t-1}-v_{s-1}}{C_\alpha (t-s)^\alpha}$, where $0 < C_\alpha < \infty$ is a normalized constant. We assume without loss of generality that $C_\alpha = 1$ in this appendix. Recall that $g(u)$ is the marginal density of the stationary time series $\{u_t\}$. Let $f_{st}(\cdot)$ be the density function of v_{st} and $g_{st}(\cdot)$ be the density function of $u_{st} = v_{t-1} - v_{s-1}$. Then, the i–th derivative of $g_{st}(v)$ satisfies for $i = 0, 1$

$$g_{st}^{(i)}(v) = \frac{1}{C_\alpha (t-s)^{(1+i)\alpha}} f_{st}^{(i)}\left(\frac{v}{(t-s)^\alpha}\right). \tag{A.19}$$

Similarly, let $f(\cdot|\mathcal{F}_s)$ and $g(\cdot|\mathcal{F}_s)$ be the conditional density functions of v_{st} and u_{st} given $\mathcal{F}_{s-1}$, where $\{\mathcal{F}_s\}$ is a sequence of σ–fields such that $\{v_s\}$ is adapted to $\mathcal{F}_s$. Then

$$g_{st}(v|\mathcal{F}_{s-1}) = \frac{1}{C_\alpha (t-s)^\alpha} f_{st}\left(\frac{v}{(t-s)^\alpha}|\mathcal{F}_{s-1}\right), \tag{A.20}$$

and the first derivatives of $g_{st}(\cdot|\mathcal{F}_{s-1})$ and $f_{st}(\cdot|\mathcal{F}_{s-1})$ satisfy

$$g_{st}^{(1)}(v|\mathcal{F}_{s-1}) = \frac{1}{C_\alpha (t-s)^{2\alpha}} f_{st}^{(1)}\left(\frac{v}{(t-s)^\alpha}|\mathcal{F}_{s-1}\right). \tag{A.21}$$

Assumption 2.1(ii) then implies the following useful results: as $t-s \to \infty$

$$\sup_{|x|\le\delta}\left|f_{st}^{(i)}(x)\right| = O(1), \tag{A.22}$$

$$\sup_{|x|\le\delta}\left|f_{st}^{(i)}(x|\mathcal{F}_{s-1})\right| = O_P(1) \tag{A.23}$$

for $i = 0, 1$, where $\delta > 0$ is some small constant. Equations (A.22) and (A.23) are used repeatedly in the proofs of Lemmas A.1–A.5 below.

Lemma A.1. *Let Assumptions 2.1 and A.1 hold. Then for large enough T*

$$\sigma_{T1}^2 = \mathrm{var}\left[\sum_{t=2}^T \eta_t u_t\right] = \frac{16\int K^2(x)dx}{3\sqrt{2\pi}} T^{3/2} h\ (1+o(1)). \tag{A.24}$$

Proof: It follows from the definition that

$$\begin{aligned}
\sigma_{T1}^2 &= E\left[\sum_{t=1}^T \eta_t u_t\right]^2 = 2\sum_{t=1}^T\sum_{s=1}^T E\left[a_{st}^2 u_s^2 u_t^2\right] + 4\sum_{t=2}^T \sum_{s_1\neq s_2=1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2} u_t^2\right] \\
&= 2\sigma_u^2 \sum_{t=1}^T\sum_{s=1}^T E\left[a_{st}^2 u_s^2\right] + R_{1T},
\end{aligned} \tag{A.25}$$

where $R_{1T} = 4\sigma_u^2 \sum_{t=2}^T \sum_{s_1\neq s_2=1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2}\right]$.

Let $w_{st} = \sum_{i=s+1}^{t-1} u_i$ and $g_{st}(\cdot,\cdot)$ be the joint density function of w_{st} and u_s. Assumption 2.1(ii) then implies

$$\begin{aligned}
E[a_{st}^2 u_s^2] &= \int\int K_h^2(w_{st}+u_s) u_s^2 g_{st}(u_{st}, u_s) du_s du_{st} \\
&= \int\int K_h^2(u_{st}+u_s) u_s^2 g_{st}(w_{st}|u_s) f(u_s) du_s du_{st} \\
&= \frac{1}{(t-s-1)^\alpha}\int\int K_h^2(w_{st}+u_s) u_s^2 f_{st}\left(\frac{u_{st}}{(t-s-1)^\alpha}|u_s\right) g(u_s) du_s du_{st} \\
&= \frac{h}{(t-s-1)^\alpha}\int\int K^2(x_{st}) x^2 f_{st}\left(\frac{x_{st}h}{(t-s-1)^\alpha}|u_s\right) g(x) dx dx_{st}.
\end{aligned} \tag{A.26}$$

Choose $m_T \ge 1$ such that $m_T \to \infty$ and $\frac{m_T}{\sqrt{Th}} \to 0$ as $T\to\infty$. Observe that

$$\sum_{t=2}^T\sum_{s=1}^{t-1} E[a_{st}^2 u_s^2] = \sum_{s=1}^{T-1}\sum_{t=s+1}^T E[a_{st}^2 u_s^2] = A_{1T} + A_{2T}, \tag{A.27}$$

where $A_{1T} = \sum_{s=1}^{T-1}\sum_{1\le(t-s)\le m_T} E[a_{st}^2 u_s^2] = O(Tm_T) = o(T^{3/2}h)$ using the fact that $E\left[a_{st}^2 u_s^2\right] \le k_0^2 E\left[u_s^2\right] = k_0^2$ due to the boundedness of the kernel $K(\cdot)$ by a constant $k_0 > 0$.

Using Assumption 2.1, it follows from (A.26) that

$$
\begin{aligned}
A_{2T} &= \sum_{s=1}^{T-1} \sum_{m_T+1\le(t-s)\le T-1} E[a_{st}^2 u_s^2] \\
&= (1+o(1))C_0 \sum_{s=1}^{T-1} \sum_{m_T+1\le(t-s)\le T-1} \frac{h}{(t-s-1)^\alpha} \int\int K^2(y) x^2 g(x) dx dy \\
&= \frac{4\sigma_u^2 \int K^2(y)dy}{3} C_0 T^{3/2} h(1+o(1)). \qquad (A.28)
\end{aligned}
$$

To deal with R_{1T}, we need to introduce the following notation: for $1 \le i \le 2$,

$$
Z_i = u_{s_i}, \quad Z_{11} = \sum_{i=s_1+1}^{t-1} u_i, \quad Z_{22} = \sum_{j=s_2+1}^{s_1-1} u_j, \qquad (A.29)
$$

ignoring the notational involvement of s, t and others.

Let $g(x_{11}, x_1, x_{22}, x_2)$ be the joint density of $(Z_{11}, Z_1, Z_{22}, Z_2)$, $g_{11}(x_{11}|x_1, x_{22}, x_2)$ be the conditional density function of Z_{11} given (Z_1, Z_{22}, Z_2), $g(x_1|x_{22}, x_2)$ be the conditional density function of Z_1 given (Z_{22}, Z_2), and $g_{22}(x_{22}|x_2)$ be the conditional density function of Z_{22} given Z_2. Similarly to (A.26), we have that for large enough T

$$
\begin{aligned}
& E\left[a_{s_1t}a_{s_2t}u_{s_1}u_{s_2}\right] = E\left[K_h\left(\sum_{i=s_1}^{t-1} u_i\right) K_h\left(\sum_{j=s_2}^{t-1} u_j\right) u_{s_1}u_{s_2}\right] \\
&= E\left[Z_1Z_2K_h\left(Z_2+Z_{22}\right)K_h\left(Z_1+Z_2+Z_{11}+Z_{22}\right)\right] \\
&= \int\cdots\int x_1x_2K_h(x_1+x_2+x_{11}+x_{22})K_h(x_2+x_{22}) \\
&\times g(x_{11}, x_1, x_{22}, x_2)dx_1dx_2dx_{11}dx_{22} \\
&= \int\cdots\int x_1x_2K_h(x_1+x_2+x_{11}+x_{22})K_h(x_2+x_{22}) \\
&\times g_{11}(x_{11}|x_1, x_{22}, x_2)g(x_1|x_{22}, x_2)g_{22}(x_{22}|x_2)g(x_2)dx_1dx_2dx_{11}dx_{22} \\
&(\text{using } y_{ii} = \tfrac{x_i+x_{ii}}{h}) \\
&= h^2\int\cdots\int K\left(y_{22}\right)K\left(y_{11}+y_{22}\right)x_1x_2 \\
&\times g_{11}(y_{11}h - x_1|x_1, y_{22}h, x_2)g(x_1|hy_{22}, x_2)g_{22}(hy_{22}-x_2|x_2)g(x_2) \\
&\times dx_1dx_2dy_{11}dy_{22} \\
&(\text{using Taylor expansions}) \\
&= h^2(1+o(1))\int\cdots\int K\left(y_{22}\right)K\left(y_{11}+y_{22}\right)x_1x_2 \\
&\times g_{11}(-x_1|x_1, 0, x_2)g(x_1|0, x_2)g_{22}(-x_2|x_2)g(x_2)dx_1dx_2dy_{11}dy_{22} \\
&+ h^4(1+o(1))\int\cdots\int K\left(y_{22}\right)K\left(y_{11}+y_{22}\right)x_1x_2 \\
&\times g_{11}'(-x_1|x_1, 0, x_2)g(x_1|0, x_2)g_{22}'(-x_2|x_2)g(x_2)dx_1dx_2dy_{11}dy_{22}
\end{aligned}
$$

$$\begin{aligned}
&= h^2(1+o(1)) \int \cdots \int K(y_{22}) K(y_{11}+y_{22}) x_1 x_2 g(x_1|0,x_2) g(x_2) \\
&\times \frac{1}{(t-s_1-1)^\alpha} \frac{1}{(s_1-s_2-1)^\alpha} f_{11}\left(\frac{-x_1}{(t-s_1-1)^\alpha}|x_1,0,x_2\right) \\
&\times f_{22}\left(\frac{-x_2}{(s_1-s_2-1)^\alpha}|x_2\right) dx_1 dx_2 dy_{11} dy_{22} \\
&+ h^4(1+o(1)) \int \cdots \int y_{11} y_{22} K(y_{22}) K(y_{11}+y_{22}) x_1 x_2 g(x_1|0,x_2) g(x_2) \\
&\times \frac{1}{(t-s_1-1)^{2\alpha}} \frac{1}{(s_1-s_2-1)^{2\alpha}} f_{11}'\left(\frac{-x_1}{(t-s_1-1)^\alpha}|x_1,0,x_2\right) \\
&\times f_{22}'\left(\frac{-x_2}{(s_1-s_2-1)^\alpha}|x_2\right) dx_1 dx_2 dy_{11} dy_{22}. \qquad \text{(A.30)}
\end{aligned}$$

Thus, similarly to (A.27) and (A.28), we can show

$$\begin{aligned}
&\sum_{t=2}^{T} \sum_{s_1 \neq s_2 = 1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2}\right] = 2 \sum_{t=3}^{T} \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2}\right] \\
&= o\left(T^{3/2} h\right) + 2C_0^2 h^2 (1+o(1)) \sum_{t=3}^{T} \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \frac{1}{(t-s_1-1)^\alpha} \frac{1}{(s_1-s_2-1)^\alpha} \\
&\times \int \cdots \int K(y_{22}) K(y_{11}+y_{22}) x_1 x_2 g(x_1|0,x_2) g(x_2) dx_1 dx_2 dy_{11} dy_{22} \\
&+ o\left(T^{3/2} h\right) + 2h^4 (1+o(1)) \sum_{t=3}^{T} \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \frac{1}{(t-s_1-1)^{2\alpha}} \frac{1}{(s_1-s_2-1)^{2\alpha}} \\
&\times \int \cdots \int y_{11} y_{22} K(y_{22}) K(y_{11}+y_{22}) x_1 x_2 g(x_1|0,x_2) g(x_2) dx_1 dx_2 dy_{11} dy_{22} \\
&= o\left(T^{3/2} h\right) \qquad \text{(A.31)}
\end{aligned}$$

using Assumption 2.1.

Equations (A.27), (A.28) and (A.31) show that for large enough T

$$\sigma_{T1}^2 = \frac{16 \int K^2(y) dy}{3\sqrt{2\pi}} T^{3/2} h (1+o(1)). \qquad \text{(A.32)}$$

The proof of Lemma A.1 is therefore finished.

Lemma A.2. *Let Assumptions 2.1 and A.1 hold. Then for large enough T*

$$\lim_{T\to\infty} \frac{1}{\sigma_{T1}^4} \sum_{t=2}^{T} E\left[\eta_t^4\right] = 0. \qquad \text{(A.33)}$$

Proof. Observe that

$$E\left[\eta_t^4\right] = 16 \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} \sum_{s_3=1}^{t-1} \sum_{s_4=1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} a_{s_3 t} a_{s_4 t} u_{s_1} u_{s_2} u_{s_3} u_{s_4}\right]. \qquad \text{(A.34)}$$

We mainly consider the cases of $s_i \neq s_j$ for all $i \neq j$ in the following proof. Since the other terms involve at most triple summations, we may deal with such terms similarly. Without loss of generality, we only look at the case of $1 \leq s_4 < s_3 < s_2 < s_1 \leq t-1$ in the following evaluation. Let

$$\begin{aligned}
u_{s_1 t} &= u_{s_1} + \sum_{i=s_1+1}^{t-1} u_i, \quad u_{s_2 t} = u_{s_1} + u_{s_2} + \sum_{i=s_2+1}^{s_1-1} u_i + \sum_{j=s_1+1}^{t-1} u_j, \\
u_{s_3 t} &= u_{s_1} + u_{s_2} + u_{s_3} + \sum_{k=s_3+1}^{s_2-1} u_k + \sum_{i=s_2+1}^{s_1-1} u_i + \sum_{j=s_1+1}^{t-1} u_j, \\
u_{s_4 t} &= u_{s_1} + u_{s_2} + u_{s_3} + u_{s_4} + \sum_{l=s_4+1}^{s_3-1} u_l + \sum_{k=s_3+1}^{s_2-1} u_k + \sum_{i=s_2+1}^{s_1-1} u_i + \sum_{j=s_1+1}^{t-1} u_j.
\end{aligned}$$

Similarly to (A.29), let again $Z_i = u_{s_i}$ for $1 \leq i \leq 4$,

$$Z_{11} = \sum_{i=s_1+1}^{t-1} u_i, \; Z_{22} = \sum_{j=s_2+1}^{s_1-1} u_j, \; Z_{33} = \sum_{k=s_3+1}^{s_2-1} u_k, \; Z_{44} = \sum_{l=s_4+1}^{s_3-1} u_l.$$

Analogously to (A.30), we may have

$$\begin{aligned}
& E\left[\prod_{i=1}^{4} a_{s_i t} u_{s_i}\right] = E\left[\prod_{j=1}^{4} Z_j K_h\left(\sum_{i=1}^{j}[Z_i + Z_{ii}]\right)\right] \\
& = \int \cdots \int g(x_{11}, x_1, \cdots, x_{44}, x_4) \\
& \times \prod_{j=1}^{4}\left(K_h\left(\sum_{i=1}^{j}[x_i + x_{ii}]\right) x_j dx_{jj} dx_j\right) \\
& = \int \cdots \int g_{11}(x_{11}|x_1, \cdots, x_{44}, x_4) g(x_1|x_{22}, \cdots, x_{44}, x_4) \\
& \times g_{22}(x_{22}|x_2, \cdots, x_{44}, x_4) g(x_2|x_{33}, \cdots, x_{44}, x_4) \\
& \times g_{33}(x_{33}|x_3, x_{44}, x_4) g(x_3|x_{44}, x_4) g_{44}(x_{44}|x_4) g(x_4) \\
& \times \prod_{j=1}^{4}\left(K_h\left(\sum_{i=1}^{j}[x_i + x_{ii}]\right) x_j dx_{jj} dx_j\right) \\
& \left(\text{using } y_{ii} = \frac{x_i + x_{ii}}{h} \text{ and } y_i = x_i\right) \\
& = h^4 \int \cdots \int g_{11}(y_{11}h - y_1|y_1, \cdots, hy_{44}, y_4) g(y_1|hy_{22}, \cdots, hy_{44}, y_4) \\
& \times g_{22}(hy_{22} - y_2|y_2, \cdots, hy_{44}, y_4) g(y_2|hy_{33}, \cdots, hy_{44}, y_4) \\
& \times g_{33}(hy_{33} - y_3|y_3, hy_{44}, y_4) g(y_3|hy_{44}, y_4) g_{44}(hy_{44} - y_4|y_4) g(y_4) \\
& \times \prod_{j=1}^{4}\left(K\left(\sum_{i=1}^{j} y_{ii}\right) y_j dy_{jj} dy_j\right) \\
& = h^4(1 + o(1)) \int \cdots \int g_{11}(-y_1|y_1, \cdots, 0, y_4) g(y_1|0, \cdots, 0, y_4) \\
& \times g_{22}(-y_2|y_2, \cdots, 0, y_4) g(y_2|0, \cdots, 0, y_4) \\
& \times g_{33}(-y_3|y_3, 0, y_4) g(y_3|0, y_4) g_{44}(-y_4|y_4) g(y_4) \\
& \times \prod_{j=1}^{4}\left(K\left(\sum_{i=1}^{j} y_{ii}\right) y_j dy_{jj} dy_j\right), \qquad \text{(A.35)}
\end{aligned}$$

where $C_{22}(K) \equiv \prod_{j=1}^{4} \int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) dy_{jj} < \infty$ involved in (A.35).

Hence, similarly to (A.31) we have by Assumption 2.1

$$\begin{aligned}
&\sum_{t=2}^{T} \sum_{1 \le s_4 < s_3 < s_2 < s_1 \le t-1} E\left[a_{s_1 t} a_{s_2 t} a_{s_3 t} a_{s_4 t} u_{s_1} u_{s_2} u_{s_3} u_{s_4}\right] \\
&= O\left(h^4\right) \sum_{t=2}^{T} \sum_{1 \le s_4 < s_3 < s_2 < s_1 \le t-1} \frac{1}{\sqrt{t-s_1}} \frac{1}{\sqrt{s_1-s_2}} \frac{1}{\sqrt{s_2-s_3}} \frac{1}{\sqrt{s_3-s_4}} \\
&= O\left(T^3 h^4\right) = o\left(T^3 h^2\right). \qquad \text{(A.36)}
\end{aligned}$$

Analogously, we can deal with the other terms of (A.34) as follows:

$$\sum_{t=2}^{T} \sum_{1 \le s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^2 a_{s_2 t}^2 u_{s_1}^2 u_{s_2}^2\right] = o\left(T^3 h^2\right), \qquad \text{(A.37)}$$

$$\sum_{t=2}^{T} \sum_{1 \le s_3 \ne s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^2 a_{s_2 t} a_{s_3 t} u_{s_1}^2 u_{s_2} u_{s_3}\right] = o\left(T^3 h^2\right), \qquad \text{(A.38)}$$

$$\sum_{t=2}^{T} \sum_{1 \le s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^3 a_{s_2 t} u_{s_1}^3 u_{s_2}\right] = o\left(T^3 h^2\right). \qquad \text{(A.39)}$$

Thus, the proof of (A.33) is completed using (A.34)–(A.39).

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

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} \to_D \xi^2 > 0 \qquad \text{(A.40)}$$

with $\xi^2 = \frac{\sqrt{\pi}}{2} M_{\frac{1}{2}}(1)$*, 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).*

Proof. To simplify the following proof, ignoring the higher–order term we rewrite

$$\sigma_{T1}^2 = \frac{16\sigma_u^4 J_{02}}{3\sqrt{2\pi}} T^{3/2} h \equiv C_{10}\, T^{3/2} h. \qquad \text{(A.41)}$$

Let $Q(u) = \frac{K^2(u)}{J_{02}}$ and $N(T)$ be the same as $T(n)$ in Karlsen and Tjøstheim (2001). It then follows from Lemma B.2 below that as $T \to \infty$

$$\max_{1 \le t \le T} \left| \frac{1}{N(T)h} \sum_{s=2}^{T} Q\left(\frac{v_{s-1} - v_{t-1}}{h}\right) - 1 \right| = o(1) \quad \text{almost surely.} \qquad \text{(A.42)}$$

Meanwhile, Theorem 3.2 of Karlsen and Tjøstheim (2001, p.389) is applicable to the current case of $v_t = v_{t-1} + u_t$ under H_0 to show that as $T \to \infty$

$$\frac{N(T)}{L_0 \sqrt{T}} \to_D M_{\frac{1}{2}}(1) \qquad \text{(A.43)}$$

when the slowly–varying function in this case is $L_0 = \frac{2\sqrt{2}}{3}$.

Therefore, equations (A.42) and (A.43) imply as $T \to \infty$

$$\begin{aligned}
&\frac{4}{\sigma_{T1}^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{T} a_{st}^2\right) u_t^2 = \frac{2}{TC_{10}}\sum_{t=1}^{T} u_t^2 \left(\frac{1}{\sqrt{T}h}\sum_{s=1}^{T} a_{st}^2\right) \\
&= \frac{2L_0 J_{02}}{C_{10}} \frac{N(T)}{L_0\sqrt{T}} \; \frac{1}{T}\sum_{t=1}^{T} u_t^2 \left(\frac{1}{N(T)h}\sum_{s=1}^{T} Q\left(\frac{v_{s-1}-v_{t-1}}{h}\right) - 1\right) \\
&+ \frac{2L_0 J_{02}}{C_{10}} \frac{N(T)}{L_0\sqrt{T}} \; \frac{1}{T}\sum_{t=1}^{T} u_t^2 \to_D \frac{2\ J_{02} L_0}{C_{10}} \; M_{\frac{1}{2}}(1) = \frac{\sqrt{\pi}}{2} \; M_{\frac{1}{2}}(1) \equiv \xi^2. \qquad \text{(A.44)}
\end{aligned}$$

Therefore, equation (A.44) completes the proof of Lemma A.3.

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

$$\frac{\widetilde{\sigma}_{T1}^2}{\sigma_{T1}^2} - \frac{\widetilde{\sigma}_{M(T),1}^2}{\sigma_{M(T),1}^2} \to_P 0. \qquad \text{(A.45)}$$

Proof. To simplify our proofs, we introduce the following lower case notation: $m = T$, $n = M(T)$, $\sigma_m^2 = \sigma_{T1}^2$, $\sigma_n^2 = \sigma_{M(T),1}^2$, and for $1 \le i \le n$, $1 \le j \le i-1$,

$$e_{ij} = \left(u_i^2 - E[u_i^2]\right) \; K_h^2\left(\sum_{l=j}^{i-1} u_l\right) u_j^2 \quad \text{and} \quad W_{mi} = \frac{1}{\sigma_m^2}\sum_{j=1}^{i-1} e_{ij}. \qquad \text{(A.46)}$$

$$w_i^2 = \sum_{j=1}^{i-1} K_h^2\left(\sum_{l=j}^{i-1} u_l\right) u_j^2 = \sum_{j=1}^{i-1} K_h^2\left(\sum_{l=j+1}^{i-1} u_l + u_j\right) u_j^2. \qquad \text{(A.47)}$$

Note that $W_{mi} = \frac{1}{\sigma_m^2}\left(u_i^2 - E[u_1^2]\right) w_i^2$.

Observe that

$$\begin{aligned}
\frac{\widetilde{\sigma}_{m1}^2}{\sigma_{m1}^2} - \frac{\widetilde{\sigma}_{n1}^2}{\sigma_{n1}^2} &= \sum_{i=1}^{m} W_{mi} - \sum_{j=1}^{n} W_{nj} + E[u_1^2] \; \left(\frac{1}{\sigma_m^2}\sum_{i=1}^{m} w_i^2 - \frac{1}{\sigma_n^2}\sum_{j=1}^{n} w_j^2\right) \\
&\equiv I_{mn} + E[u_1^2] \; J_{mn}. \qquad \text{(A.48)}
\end{aligned}$$

In view of (A.47), in order to prove (A.45), it suffices to show that as $m,\ n \to \infty$

$$I_{mn} \to_P 0 \quad \text{and} \quad J_{mn} \to_P 0. \qquad \text{(A.49)}$$

We start by proving the second part of (A.49). Observe also that

$$\begin{aligned}
E\left[J_{mn}^2\right] &= E\left[\frac{1}{\sigma_m^2}\sum_{i=1}^{m} w_i^2 - \frac{1}{\sigma_n^2}\sum_{j=1}^{n} w_j^2\right]^2 = E\left[\frac{1}{\sigma_m^2}\sum_{i=n+1}^{m} w_i^2 + \frac{\sigma_n^2 - \sigma_m^2}{\sigma_m^2\ \sigma_n^2}\sum_{j=1}^{n} w_j^2\right]^2 \\
&= \frac{1}{\sigma_m^4}\sum_{i=n+1}^{m}\sum_{k=n+1}^{m} E\left[w_k^2 w_i^2\right] + \frac{\left(\sigma_n^2 - \sigma_m^2\right)^2}{\sigma_m^4\ \sigma_n^4}\sum_{j=1}^{n}\sum_{k=1}^{n} E\left[w_k^2 w_j^2\right] \\
&- 2\frac{\sigma_m^2 - \sigma_n^2}{\sigma_m^4\ \sigma_n^2}\sum_{i=n+1}^{m}\sum_{j=1}^{n} E\left[w_i^2 w_j^2\right]. \qquad \text{(A.50)}
\end{aligned}$$

We first deal with the first term. Recalling $a_{ji} = K_h\left(\sum_{l=j}^{i-1} u_l\right)$, we have

$$E\left(\sum_{i=n+1}^{m} w_i^2\right)^2 = E\left[\sum_{i=n+1}^{m}\sum_{j=n+1}^{m} w_i^2\ w_j^2\right] = \sum_{i=n+1}^{m} E[w_i^4] + \sum_{i=n+1}^{m}\sum_{j=n+1,\neq i}^{m} E[w_i^2\ w_j^2]. \quad (A.51)$$

We now evaluate the orders of $\sum_{i=n+1}^{m} E[w_i^4]$ and $\sum_{i=n+1}^{m}\sum_{j=n+1,\neq i}^{m} E[w_i^2\ w_j^2]$ respectively. To do so, we now consider one of the cases: $1 \leq t \leq s-1; 2 \leq s \leq j-1; n+1 \leq j \leq i-1; n+2 \leq i \leq m$ for the following term

$$E\left[\sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1} a_{si}^2 u_s^2 a_{tj}^2 u_t^2\right] = \sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1} E\left[a_{si}^2 u_s^2 a_{tj}^2 u_t^2\right]$$
$$= \sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1} E\left[K_h^2\left(\sum_{c=s+1}^{j-1} u_c + \sum_{c=j}^{i-1} u_c + u_s\right) u_s^2\right.$$
$$\left.\times\ K_h^2\left(\sum_{d=t+1}^{s-1} u_d + \sum_{d=s+1}^{j-1} u_d + u_s + u_t\right) u_t^2\right].$$

Other terms may be dealt with similarly. To simplify our calculation, we now introduce the following simplistic symbols: $Z_{11} = \sum_{d=t+1}^{s-1} u_d$, $Z_{22} = \sum_{c=s+1}^{j-1} u_c$, $Z_{33} = \sum_{c=j}^{i-1} u_c$, $Z_1 = u_t$ and $Z_2 = u_s$.

As in the proof of Lemma A.1, using the same techniques as in (A.35) we have

$$E\left[K_h^2\left(\sum_{i=1}^{2}(Z_i + Z_{ii})\right) K_h^2\left(Z_2 + Z_{22} + Z_{33}\right) Z_1^2 Z_2^2\right]$$
$$= \int \cdots \int K_h^2\left(\sum_{i=1}^{2}(x_i + x_{ii})\right) K_h^2\left(x_2 + x_{22} + x_{33}\right)\ x_1^2\ x_2^2$$
$$\times\ g(x_{33}, x_{22}, x_2, x_{11}, x_1)\ dx_{33}dx_{22}dx_{11}dx_1dx_2$$
$$= \int \cdots \int K_h^2\left(\sum_{i=1}^{2}(x_i + x_{ii})\right) K_h^2\left(x_2 + x_{22} + x_{33}\right)\ x_1^2\ x_2^2$$
$$\times\ g_{33}(x_{33}|x_{22}, x_2, x_{11}, x_1) g_{22}(x_{22}|x_2, x_{11}, x_1) g(x_2|x_{11}, x_1) g_{11}(x_{11}|x_1) g(x_1)$$
$$\times\ dx_{33}dx_{22}dx_{11}dx_1dx_2$$

(using $y_i = x_i$ and $y_{ii} = \frac{x_i + x_{ii}}{h}$ for $i = 1, 2$ and $y_{33} = \frac{x_{33}}{h}$)

$$= h^3 \int \cdots \int K^2(y_{11} + y_{22}) K^2(y_{22} + y_{33})\ y_1^2 y_2^2$$
$$\times\ g_{33}(y_{33}h|y_{22}h - y_2, y_2, y_{11}h - y_1, y_1) g_{22}(y_{22}h - y_2|y_2, y_{11}h - y_1, y_1)$$
$$\times\ g_{11}(y_{11}h - y_1|y_1) g(y_2|y_{11}h - y_1, y_1) g(y_1)\ dy_{33}dy_{22}dy_{11}dy_1dy_2$$
$$= h^3(1 + o(1)) \int \cdots \int K^2(y_{11} + y_{22}) K^2(y_{22} + y_{33})\ y_1^2 y_2^2$$
$$\times\ g_{33}(0| - y_2, y_2, -y_1, y_1) g_{22}(-y_2|y_2, -y_1, y_1) g_{11}(-y_1|y_1)$$
$$\times\ g(y_2| - y_1, y_1) g(y_1)\ dy_{33}dy_{22}dy_{11}dy_1dy_2$$

$$
\begin{aligned}
&= h^3(1+o(1)) \int \cdots \int K^2(y_{11}+y_{22}) K^2(y_{22}+y_{33})\ y_1^2 y_2^2 \\
&\times\ f_{33}(0|-y_2, y_2, -y_1, y_1) f_{22}\left(\frac{-y_2}{(j-s-1)^\alpha}|y_2, -y_1, y_1\right) \\
&\times\ f_{11}\left(\frac{-y_1}{(s-t-1)^\alpha}|y_1\right) g(y_2|-y_1, y_1) g(y_1)\ dy_{33} dy_{22} dy_{11} dy_1 dy_2. \qquad \text{(A.52)}
\end{aligned}
$$

In view of (A.52) and (A.52), similarly to the calculations of (A.26), (A.27) and (A.28), it can be shown that for large enough m and n,

$$
\begin{aligned}
& E\left[\sum_{i=n+1}^{m} \sum_{j=n+1,\neq i}^{m} w_i^2\ w_j^2\right] = \sum_{i=n+1}^{m} \sum_{j=n+1,\neq i}^{m} E\left[w_i^2\ w_j^2\right] \\
&= Ch^3(1+o(1)) \sum_{i=n+2}^{m} \sum_{j=n+1}^{i-1} \sum_{s=2}^{j-1} \sum_{t=1}^{s-1} \frac{1}{(i-j)^\alpha} \frac{1}{(j-s-1)^\alpha} \frac{1}{(s-t-1)^\alpha} \\
&= Ch^3(1+o(1))(m-n)^{\frac{5}{2}}. \qquad \text{(A.53)}
\end{aligned}
$$

Similarly to (A.53), we may have for sufficiently large m and n,

$$
\begin{aligned}
\sum_{i=n+1}^{m} E[w_i^4] &= Ch^2(1+o(1))(m-n)^{\frac{3}{2}}. \qquad \text{(A.54)} \\
E\left[\sum_{i=n+1}^{m} \sum_{j=1}^{n} w_i^2\ w_j^2\right] &= o\left(h^3(m-n)^{\frac{5}{2}}\right), \qquad \text{(A.55)} \\
E\left[\sum_{i=}^{m} \sum_{j=1}^{n} w_i^2\ w_j^2\right] &= o\left(h^3(m-n)^{\frac{5}{2}}\right) \qquad \text{(A.56)}
\end{aligned}
$$

using $\lim_{m,n\to\infty} \frac{n}{m} = 0$.

Thus, equations (A.50)–(A.56) imply that for large enough m and n,

$$
\begin{aligned}
E\left[J_{mn}^2\right] &= E\left[\frac{1}{\sigma_m^2} \sum_{i=1}^{m} w_i^2 - \frac{1}{\sigma_n^2} \sum_{j=1}^{n} w_j^2\right]^2 \\
&= \frac{1}{\sigma_m^4} \sum_{i=n+1}^{m} \sum_{k=n+1}^{m} E\left[w_k^2 w_i^2\right] + \frac{\left(\sigma_n^2 - \sigma_m^2\right)^2}{\sigma_m^4\ \sigma_n^4} \sum_{j=1}^{n} \sum_{k=1}^{n} E\left[w_k^2 w_j^2\right] \\
&- 2\frac{\sigma_m^2 - \sigma_n^2}{\sigma_m^4\ \sigma_n^2} \sum_{i=n+1}^{m} \sum_{j=1}^{n} E\left[w_i^2 w_j^2\right] \\
&= Ch\left(1 - \frac{n}{m}\right)^{\frac{3}{2}}\ (1+o(1)) = o(1) \qquad \text{(A.57)}
\end{aligned}
$$

using again $\lim_{m,n\to\infty} \frac{n}{m} = 0$. We thus complete the second part of (A.49).

Let $z_i = u_i^2 - E[u_1^2]$. We now come back to prove the first part of (A.49). Note that for $n+1 \le i \le m$ and $1 \le j \le n$,

$$
I_{mn} = \frac{1}{\sigma_m^2} \sum_{i=1}^{m} \left(u_i^2 - E[u_1^2]\right) w_i^2 - \frac{1}{\sigma_n^2} \sum_{j=1}^{n} \left(u_j^2 - E[u_1^2]\right) w_j^2 = \frac{1}{\sigma_m^2} \sum_{i=1}^{m} z_i\ w_i^2 - \frac{1}{\sigma_n^2} \sum_{j=1}^{n} z_j\ w_j^2. \qquad \text{(A.58)}
$$

Note that $\{w_i^2\}$ is a function of $\{u_j : 1 \leq j \leq i-1\}$ while $\{z_i\}$ is a function of $\{u_i\}$. Let $g_{zw}(\cdot,\cdot)$ be the joint density function of (z_i, w_i^2), $g_{z|w}(\cdot|\cdot)$ be the conditional density function of z_i given w_i, and $g_w(\cdot)$ be the marginal density function of w_i^2. Obviously, $g_{zw}(z,w) = g_z(z)g_w(w)$ when $\{u_i\}$ is assumed to be a sequence of independent random variables.

Thus, in view of the relationship $g_{zw}(z,w) = g_{z|w}(z|w)g_w(w)$ and the fact that the conditional moments of z_i given w_i do not affect the order of $E\left[I_{mn}^2\right]$, by using the same arguments as in (A.50)–(A.57), we can show that for large enough m and n,

$$\begin{aligned}
E\left[I_{mn}^2\right] &= E\left[\frac{1}{\sigma_m^2}\sum_{i=1}^m z_i\ w_i^2 - \frac{1}{\sigma_n^2}\sum_{j=1}^n z_j\ w_j^2\right]^2 \\
&= \frac{1}{\sigma_m^4}\sum_{i=n+1}^m\sum_{k=n+1}^m E\left[z_k\ w_k^2\ z_i\ w_i^2\right] + \frac{\left(\sigma_n^2-\sigma_m^2\right)^2}{\sigma_m^4\ \sigma_n^4}\sum_{j=1}^n\sum_{k=1}^n E\left[z_k\ w_k^2\ z_j\ w_j^2\right] \\
&- 2\frac{\sigma_m^2-\sigma_n^2}{\sigma_m^4\ \sigma_n^2}\sum_{i=n+1}^m\sum_{j=1}^n E\left[z_i\ w_i^2\ z_j\ w_j^2\right] \\
&= Ch\left(1-\frac{n}{m}\right)^{\frac{3}{2}}\ (1+o(1)) = o(1). \qquad (A.59)
\end{aligned}$$

We therefore have completed the proof of Lemma A.4.

Lemma A.5. *Let the conditions of Theorem 2.1 hold. Then as $T \to \infty$*

$$\frac{M_{Ti}}{\widetilde{\sigma}_{T1}} \to_P 0 \quad \text{for } i = 2,3,4, \qquad (A.60)$$
$$\frac{\widetilde{\sigma}_{Tj}}{\widetilde{\sigma}_{T1}} \to_P 0 \quad \text{for } j = 2,3,4. \qquad (A.61)$$

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.60) and (A.61), it suffices to show that as $T \to \infty$

$$\frac{M_{Ti}}{\sigma_{T1}} \to_P 0 \quad \text{for } i = 2,3,4, \qquad (A.62)$$
$$\frac{\widetilde{\sigma}_{Tj}}{\sigma_{T1}} \to_P 0 \quad \text{for } j = 2,3,4. \qquad (A.63)$$

Since the details are very similar, we prove only (A.62) for $i = 2$. Observe that

$$\begin{aligned}
M_{T2} &= \sum_{t=1}^T\sum_{s=1,\neq t}^T \widetilde{u}_s\ K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1})\ \widetilde{u}_t = \sum_{t=1}^T\sum_{s=1,\neq t}^T \widetilde{u}_s\ K_h(v_{s-1}-v_{t-1})\ \widetilde{u}_t \\
&+ \sum_{t=1}^T\sum_{s=1,\neq t}^T \widetilde{u}_s\ \left(K_h(\widehat{v}_{s-1}-\widehat{v}_{t-1}) - K_h(v_{s-1}-v_{t-1})\right)\ \widetilde{u}_t \\
&\equiv M_{T21} + M_{T22}. \qquad (A.64)
\end{aligned}$$

For some $B_0 > 0$, let $\Theta(\beta) = \left\{\widehat{\beta}:\ ||\widehat{\beta} - \beta|| \leq B_0 R_T^{-1}\right\}$ and $I_{\Theta(\beta)}(\widehat{\beta})$ be the conventional indicator function. Thus, for sufficiently large T and any given $\epsilon > 0$,

$$
\begin{aligned}
& P\left(\left|M_{T21} I_{\Theta(\beta)}(\widehat{\beta})\right| \geq \epsilon\sigma_{T1}\right) \leq \frac{E\left|M_{T21} I_{\Theta(\beta)}(\widehat{\beta})\right|}{\sigma_{T1}\epsilon} \\
& \leq \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left[|\widetilde{u}_s|\ \ K_h(v_{s-1} - v_{t-1})\ \ |\widetilde{u}_t|\, I_{\Theta(\beta)}(\widehat{\beta})\right]}{\sigma_{T1}\epsilon} \\
& \leq C\ \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} ||X_s - X_{s-1}||\ E\left[K_h(v_{s-1} - v_{t-1})\right]\ ||X_t - X_{t-1}||}{R_T^2 \sigma_{T1}} \\
& \leq C \frac{2h \sum_{s=2}^{T}\sum_{t=1}^{s-1} ||X_s - X_{s-1}||\ \frac{1}{\sqrt{s-t}}\ ||X_t - X_{t-1}||}{R_T^2\sigma_{T1}} = o(1) \qquad \text{(A.65)}
\end{aligned}
$$

using $\widetilde{u}_t = (X_t - X_{t-1})^\tau \left(\beta - \widehat{\beta}\right)$, recalling the definition of $K_h(\cdot) = K\left(\frac{\cdot}{h}\right)$, the first part of Assumption A.2 and for all $s > t$, $E\left[K_h(v_{s-1} - v_{t-1})\right] \leq \frac{Ch}{\sqrt{s-t}}$, which follows from

$$
\begin{aligned}
E\left[K_h(v_{s-1} - v_{t-1})\right] &= E\left[K\left(\frac{v_{s-1} - v_{t-1}}{h}\right)\right] \\
&= \frac{h}{\sqrt{s-t}} \int K(x) f_{st}\left(\frac{xh}{\sqrt{s-t}}\right) dx \\
&\leq C\frac{h}{\sqrt{s-t}},
\end{aligned}
$$

using the same argument as in (A.26) of the proof of Lemma A.1, where $f_{st}(\cdot)$ is the density of $v_{st} = \frac{v_{s-1} - v_{t-1}}{\sqrt{s-t}}$ and $f_{st}\left(\frac{xh}{\sqrt{s-t}}\right)$ is bounded by (2.6) of Assumption 2.1(i).

Therefore, for sufficiently small $\epsilon > 0$

$$
\begin{aligned}
P\left(|M_{T21}| \geq \epsilon\sigma_{T1}\right) &= P\left(\left(|M_{T21}| \geq \epsilon\sigma_{T1}\right) \cap \left(\widehat{\beta} \notin \Theta(\beta)\right)\right) \\
&+ P\left(\left(|M_{T21}| \geq \epsilon\sigma_{T1}\right) \cap \left(\widehat{\beta} \in \Theta(\beta)\right)\right) \\
&\leq P\left(||\widehat{\beta} - \beta|| > B_0 R_T^{-1}\right) + P\left(\left|M_{T21} I_{\Theta(\beta)}(\widehat{\beta})\right| \geq \epsilon\sigma_{T1}\right) \\
&\to 0 \quad \text{as} \quad T \to \infty. \qquad \text{(A.66)}
\end{aligned}
$$

In view of $\widehat{v}_{s-1} - \widehat{v}_{t-1} = v_{s-1} - v_{t-1} + (X_{s-1} - X_{t-1})^\tau (\beta - \widehat{\beta})$ and using Assumption A.1, we have

$$
\begin{aligned}
\overline{K}_h(s,t) &\equiv \left|K\left(\frac{v_{s-1} - v_{t-1} + (X_{s-1} - X_{t-1})^\tau (\beta - \widehat{\beta})}{h}\right) - K\left(\frac{v_{s-1} - v_{t-1}}{h}\right)\right| \\
&\leq M\left(\frac{v_{s-1} - v_{t-1}}{h}\right)\left|\frac{(X_{s-1} - X_{t-1})^\tau (\beta - \widehat{\beta}))}{h}\right|.
\end{aligned}
$$

This implies that for large enough T

$$
\begin{aligned}
P\left(|M_{T22} I_{\Theta(\beta)}(\widehat{\beta})| \geq \epsilon\sigma_{T1}\right) &\leq \frac{E\left|M_{T22} I_{\Theta(\beta)}(\widehat{\beta})\right|}{\sigma_{T1}\epsilon} \\
&\leq \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} E\left[|\widetilde{u}_s|\ M\left(\frac{v_{s-1}-v_{t-1}}{h}\right)\left|\frac{(X_{s-1}-X_{t-1})^{\tau}(\beta-\widehat{\beta}))}{h}\right|\ |\widetilde{u}_t|\, I_{\Theta(\beta)}(\widehat{\beta})\right]}{\sigma_{T1}\epsilon} \\
&\leq \epsilon\ \frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} ||X_s - X_{s-1}||\ ||X_{s-1} - X_{t-1}||\ E\left[M\left(\frac{v_{s-1}-v_{t-1}}{h}\right)\right]\ ||X_t - X_{t-1}||}{R_T^3\ h\ \sigma_{T1}} \\
&\leq C\frac{\sum_{s=2}^{T}\sum_{t=1}^{s-1} ||X_s - X_{s-1}||\ \frac{||X_{s-1}-X_{t-1}||}{\sqrt{s-t}}\ ||X_t - X_{t-1}||}{R_T^3\sigma_{T1}} = o(1) \qquad \text{(A.67)}
\end{aligned}
$$

using the second part of Assumption A.2.

We thus have that for sufficiently small $\epsilon > 0$

$$
\begin{aligned}
P\left(|M_{T22}| \geq \epsilon\sigma_{T1}\right) &= P\left((|M_{T22}| \geq \epsilon\sigma_{T1}) \cap \left(\widehat{\beta} \notin \Theta(\beta)\right)\right) \\
&+ P\left((|M_{T22}| \geq \epsilon\sigma_{T1}) \cap \left(\widehat{\beta} \in \Theta(\beta)\right)\right) \\
&\leq P\left(||\widehat{\beta} - \beta|| > B_0 R_T^{-1}\right) + P\left(\left|M_{T22} I_{\Theta(\beta)}(\widehat{\beta})\right| \geq \epsilon\sigma_{T1}\right) \\
&\to 0 \quad \text{as} \quad T \to \infty. \qquad \text{(A.68)}
\end{aligned}
$$

As the detailed proofs for $i = 3, 4$ are very similar to those for the case of $i = 2$, we need only to mention the proof for the case of $i = 2$. Similarly to (A.64), we can have

$$
\begin{aligned}
\widetilde{\sigma}_{T2}^2 &= 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{u}_s^2\ K_h^2(\widehat{v}_{s-1} - \widehat{v}_{t-1})\ \widetilde{u}_t^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{u}_s^2\ K_h^2(v_{s-1} - v_{t-1})\ \widetilde{u}_t^2 \\
&+ 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{u}_s^2\ \left(K_h^2(\widehat{v}_{s-1} - \widehat{v}_{t-1}) - K_h^2(v_{s-1} - v_{t-1})\right)\ \widetilde{u}_t^2. \qquad \text{(A.69)}
\end{aligned}
$$

Analogously to (A.66) and (A.68), using Assumption A.2 with $i = 2$ we can show that for any given $\epsilon > 0$

$$
P\left(\widetilde{\sigma}_{T2}^2 \geq \epsilon\ \sigma_{T1}^2\right) \to 0 \quad \text{as} \quad T \to \infty. \qquad \text{(A.70)}
$$

This completes the proof of Lemma A.5 and thus the proof of Theorem 2.1 for Case A.

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

In view of (A.4) and (A.5), in order to prove Theorem 2.1 for Case B, it suffices to show that equations (A.6)–(A.8) hold. These proofs are given in Lemmas A.6 and A.7 below.

Lemma A.6. *Let Assumptions 2.2 and A.1 hold. Then under* $\mathcal{H}_0:\ v_t = f_0(v_{t-1}, \theta_0) + u_t$

$$
\frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\ K_h(v_{s-1} - v_{t-1})\ u_t}{\sqrt{2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s^2\ K_h^2(v_{s-1} - v_{t-1})\ u_t^2}} \to_D N(0,1) \quad \textit{as} \quad T \to \infty. \qquad \text{(A.71)}
$$

Proof: The asymptotic normality in (A.71) is a standard result for the case where $\{u_t\}$ is a sequence of martingale differences and $\{v_t\}$ is a strictly stationary and α–mixing sequence. 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.7. *Let Assumption 2.2, A.1–A.3(i) and A.4 hold. Then as* $T \to \infty$

$$\frac{M_{Ti}}{\sigma_{T1}} \to_P 0 \quad \text{for } i = 2, 3, 4, \tag{A.72}$$

$$\frac{\widetilde{\sigma}_{Tj}}{\sigma_{T1}} \to_P 0 \quad \text{for } j = 2, 3, 4. \tag{A.73}$$

Proof: Since $\{u_t\}$ is a sequence of martingale differences and $\{v_t\}$ is a strictly stationary and α–mixing time series in Case B, the proofs of (A.60) and (A.61) remain true, but become more standard through using Assumptions 2.2, 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

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

We thus have

$$\begin{aligned}
&\widehat{M}_T^* \equiv \sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^*\ K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widehat{u}_t^* = \sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u u_s^*\ K_h(\widetilde{v}_{s-1}^* - \widetilde{v}_{t-1}^*)\ \widehat{\sigma}_u u_t^* \\
&+ \sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s^*\ K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widetilde{u}_t^* + 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u u_s^*\ K_h(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widetilde{u}_t^* \\
&+ M_{T4}^* \equiv M_{T1}^* + M_{T2}^* + M_{T3}^* + M_{T4}^*,
\end{aligned} \tag{A.74}$$

$$\begin{aligned}
&\widehat{\sigma}_T^{*2} \equiv 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{u}_s^{*2}\ K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widehat{u}_t^{*2} = 2\sum_{t=1}^T \sum_{s=1,s\neq t}^T \widehat{\sigma}_u^2 u_s^{*2}\ K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widehat{\sigma}_u^2\ u_t^{*2} \\
&+ 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widetilde{u}_s^{*2}\ K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widetilde{u}_t^{*2} \\
&+ 2\sum_{t=1}^T \sum_{s=1,\neq t}^T \widehat{\sigma}_u^2 u_s^{*2}\ K_h^2(\widehat{v}_{s-1}^* - \widehat{v}_{t-1}^*)\ \widetilde{u}_t^{*2} + \widetilde{\sigma}_{T4}^{*2} \equiv \sum_{j=1}^4 \widetilde{\sigma}_{Tj}^{*2},
\end{aligned} \tag{A.75}$$

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

In view of (A.74) and (A.75), to prove Theorem 3.1(i), it suffices to show that as $T \to \infty$

$$\frac{M_{T1}^*}{\widetilde{\sigma}_{T1}^*} \to_D N(0,1), \tag{A.76}$$

$$\frac{M_{Ti}^*}{\widetilde{\sigma}_{T1}^*} \to_P 0 \quad \text{for } i = 2,3,4, \tag{A.77}$$

$$\frac{\widetilde{\sigma}_{Tj}^*}{\widetilde{\sigma}_{T1}^*} \to_P 0 \quad \text{for } j = 2,3,4. \tag{A.78}$$

Note that $\widetilde{v}_t^* = \widetilde{v}_{t-1}^* + \widehat{\sigma}_u u_t^* = \widetilde{v}_0^* + \widehat{\sigma}_u \sum_{s=1}^t u_s^* = \widehat{\sigma}_u \sum_{s=1}^t u_s^*$. Note also that $\widehat{\sigma}_u^2 = E[u_1^2] + o_P(1)$. Thus, in order to prove equations (A.76)–(A.78), in view of the fact that $\{u_t^*\}$ is a sequence of independent and identically distributed errors with $E[u_t^*] = 0$ and $E\left[u_t^{*2}\right] = 1$, and also independent of $\{Y_s\}$ for all $s, t \geq 1$, it suffices to complete the proofs of the bootstrapping versions of Lemmas A.1–A.5 by successive conditioning arguments.

As a matter of the fact, the derivations in the proofs of Lemmas A.1–A.5 now become less technical and tedious due to the fact that $\{u_t^*\}$ is a sequence of independent and identically distributed errors. 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, we thus may show that as $T \to \infty$

$$P^*\left(\widehat{L}_T^*(h) \leq x\right) \to \Phi(x) \quad \text{for all } x \in (-\infty, \infty) \tag{A.79}$$

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

Let z_α be the $1-\alpha$ quantile of $\Phi(\cdot)$ such that $\Phi(z_\alpha) = 1-\alpha$. Then it follows from (A.79) that as $T \to \infty$

$$P^*\left(\widehat{L}_T^*(h) \geq z_\alpha\right) \to 1 - \Phi(z_\alpha) = \alpha. \tag{A.80}$$

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

$$l_\alpha^* - z_\alpha \to_P 0. \tag{A.81}$$

Using the conclusion of Theorem 2.1 and (A.79) again, we have that as $T \to \infty$

$$P^*\left(\widehat{L}_T^*(h) \leq x\right) - P\left(\widehat{L}_T(h) \leq x\right) \to_P 0 \quad \text{for all } x \in (-\infty, \infty) \tag{A.82}$$

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

$$\lim_{T\to\infty} P\left(\widehat{L}_T(h) > l_\alpha^*\right) = \alpha \tag{A.83}$$

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

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

Note that under $\mathcal{H}_1: \; v_t = f_1(v_{t-1}, \theta_1) + u_t$

$$
\begin{aligned}
\widehat{u}_t &= \widehat{v}_t - f_0\left(\widehat{v}_{t-1}, \widehat{\theta}_0\right) = X_t^\tau\left(\beta - \widehat{\beta}\right) + v_t - f_0\left(\widehat{v}_{t-1}, \widehat{\theta}_0\right) \\
&= u_t + X_t^\tau\left(\beta - \widehat{\beta}\right) + f_1(v_{t-1}, \theta_1) - f_0\left(\widehat{v}_{t-1}, \widehat{\theta}_0\right) \equiv u_t + \widetilde{u}_t, \\
\widetilde{u}_t &= X_t^\tau(\beta - \widehat{\beta}) + f_1(v_{t-1}, \theta_1) - f_0\left(v_{t-1} + X_t^\tau(\beta - \widehat{\beta}), \widehat{\theta}_0\right) \\
&= X_t^\tau(\beta - \widehat{\beta}) + f_1(v_{t-1}, \theta_1) - f_0(v_{t-1}, \theta_0) \\
&+ f_0(v_{t-1}, \theta_0) - f_0\left(v_{t-1} + X_t^\tau(\beta - \widehat{\beta}), \widehat{\theta}_0\right). \qquad \text{(A.84)}
\end{aligned}
$$

To complete the proof of Theorem 3.1(ii), we need the following lemma.

Let

$$
\begin{aligned}
\Lambda_{T1} &= \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widetilde{u}_s \; K_h(v_{s-1} - v_{t-1}) \; \widetilde{u}_t \quad \text{and} \\
\Lambda_{T2} &= \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} f_{10}(v_{s-1}) \; K_h(v_{s-1} - v_{t-1}) \; f_{10}(v_{t-1}).
\end{aligned}
$$

Then we have the following lemma.

Lemma A.8. *Let the conditions of Theorem 3.1(ii) hold. Then as* $T \to \infty$

$$\sigma_{T1} \; \Lambda_{T1}^{-1} \to_P 0. \qquad \text{(A.85)}$$

Proof: Let $f_{10}(v) = f_1(v, \theta_1) - f_0(v, \theta_0)$. In view of (A.84), using Assumptions A.4 and A.5(ii), in order to prove (A.85), it suffices to show that as $T \to \infty$

$$\sigma_{T1} \; \Lambda_{T2}^{-1} \to_P 0, \qquad \text{(A.86)}$$

which follows from $\sigma_{T1} = O\left(T\sqrt{h}\right)$ and

$$\sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} E\left[f_{10}(v_{s-1}) \; K_h(v_{s-1} - v_{t-1}) \; f_{10}(v_{t-1})\right]$$

$$= T^2 h (1 + o(1)) \cdot \left(\int f_{10}^2(x) \; \pi_1^2(x) \; dx\right) \; \left(\int K(y) \; dy\right) = O\left(T^2 h\right),$$

using Assumption 3.1, where $\pi_1(v)$ denotes the marginal density of $\{v_t\}$ under $\mathcal{H}_1$. Note that in such cases where $\{v_t\}$ is strictly stationary and α–mixing, existing results for the α–mixing case (such as Lemmas A.1 and A.2 of the Appendix of Gao 2007) can be used to show that $E\left[\psi(v_{1+\tau_1}, \ldots, v_{1+\tau_l})\right]$ can be approximated by $E\left[\psi(z_{1+\tau_1}, \ldots, z_{1+\tau_l})\right]$ with certain rate of convergence related to the α–mixing coefficient for all $2 \leq l \leq 4$, where $\{z_i\}$ is a sequence of independent random variables having the same marginal density $\pi_1(\cdot)$ as $\{v_i\}$ and each $\psi(x_1, \cdots, x_l)$ is a symmetric function.

Proof of Theorem 3.1(ii): In view of the definition of $\widehat{L}_T(h)$ and the proofs of Lemmas A.6–A.8, 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) then follows from Lemma A.8.

Appendix B

In this appendix, we give two secondary lemmas for the proofs in Appendix A above.

Lemma B.1. *Assume that the probability space* $(\Omega_n, \mathcal{F}_n, P_n)$ *supports square integrable random variables* $S_{n,1}, S_{n,2}, \cdots, S_{n,k_n}$*, and that the* $S_{n,t}$ *are adapted to* σ*–algebras* $\mathcal{F}_{n,t}$*,* $1 \leq t \leq k_n$*, where*

$$\mathcal{F}_{n,1} \subset \mathcal{F}_{n,2} \subset \cdots \subset \mathcal{F}_{n,k_n} \subset \mathcal{F}_n.$$

Let $X_{n,t} = S_{n,t} - S_{n,t-1}$*,* $S_{n,0} = 0$ *and* $U_{n,t}^2 = \sum_{s=1}^{t} X_{n,s}^2$*. If* $\mathcal{G}_n$ *is a sub–*σ*–algebra of* $\mathcal{F}_n$*, let* $\mathcal{G}_{n,t} = \mathcal{F}_{n,t} \vee \mathcal{G}_n$ *(the* σ*–algebra generated by* $\mathcal{F}_{n,t} \cup \mathcal{G}_n$*) and let* $\mathcal{G}_{n,0} = \{\Omega_n, \phi\}$ *denote the trivial* σ*–algebra. Moreover, suppose that*

$$\sum_{t=1}^{n} E\left(X_{n,t}^2 I_{\{[|X_{n,t}|>\delta]\}}(X_{n,t}) | \mathcal{G}_{n,t-1}\right) \to_P 0 \tag{B.1}$$

for some $\delta > 0$*, and there exists a* $\mathcal{G}_n$*–measurable random variable* u_n^2*, such that*

$$U_{n,k_n}^2 - u_n^2 \to_P 0, \tag{B.2}$$

$$\sum_{t=1}^{n} E\left(X_{n,t} | \mathcal{G}_{n,t-1}\right) \to_P 0, \tag{B.3}$$

and

$$\sum_{t=1}^{n} \left|E\left(X_{n,t} | \mathcal{G}_{n,t-1}\right)\right|^2 \to_P 0. \tag{B.4}$$

If

$$\lim_{\delta \to 0} \lim_{n \to \infty} \inf P\left\{U_{n,k_n} > \delta\right\} = 1, \tag{B.5}$$

then $\frac{S_{n,k_n}}{U_{n,k_n}} \to_D N(0,1)$ *as* $n \to \infty$.

Proof. The proof of Lemma B.1 follows from Corollary 3.1 and Theorem 3.4 of Hall and Heyde (1980).

Lemma B.2 below is concerned with uniform strong convergence of nonparametric kernel density estimate of a nonstationary time series of the form $v_t = v_{t-1} + u_t$. The proof of Lemma B.2 follows from that of Proposition 3.1 of Chen, Gao and Li (2007).

Recall that $N(T)$ is defined in the same way as $T(n)$ in Karlsen and Tjøstheim (2001) and define

$$\widehat{f}(v) = \widehat{f}_s(v) = \frac{1}{N(T)h} \sum_{l=1}^{T} K\left(\frac{v_{l-1} - v}{h}\right). \tag{B.6}$$

Lemma B.2. *Let Assumptions 2.1(i) and A.1 hold. Then under* $\mathcal{H}_0: \ v_t = v_{t-1} + u_t$ *and as* $T \to \infty$

$$\max_{1 \le t \le T} \left| \widehat{f}(v_{t-1}) - 1 \right| = o(1) \quad \textit{almost surely.} \tag{B.7}$$

Note that in the random walk case, the invariant measure π_s of $\{v_t\}$ is proportional to the Lebesgue measure on R^1, i.e., $d\pi_s(x) = c_s\, dx$ with c_s being a proportionality factor. Referring to the uniqueness discussion in Remark 3.1 of Karlsen and Tjøstheim (2001), we can choose s such that $c_s = 1$ and $d\pi_s(x) = dx$, the Lebesgue measure. This means that π_s has a constant density $f(x) = f_s(x) \equiv 1$. This choice shows why the limit in (B.7) is one.

References

Bandi, F., 2002. Short–term interest rate dynamics: a spatial approach. Journal of Financial Economics 65, 73–110.

Chen, J., Gao, J., Li, D., 2007. Semiparametric regression estimation in null recurrent time series. Working paper available from http://www.adelaide.edu.au/directory/jiti.gao.

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

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.

Frölich, M., 2008. Parametric and nonparametric regression in the presence of endogenous control variables. International Statistical Review 76, 214–227.

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

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

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

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2006. Specification testing in nonlinear time series with nonstationarity. Working paper available from http://www.adelaide.edu.au/directory/jiti.gao.

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

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

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.

Koul, H. L., Stute, W., 1999. Nonparametric model checks for time series. Annals of Statistics 27, 204–236.

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.

Lu, Z., 1998. On the geometric ergodicity of a nonlinear autoregressive model with an autoregressive conditional heteroscedastic term. Statistica Sinica 8, 1205–1217.

Meitz, M., Saikkonen, P., 2008. Stability of nonlinear AR–GARCH models. Journal of Time Series Analysis 29, 453–475.

Pararoditis, E., Politis, D. N., 2003. Residual–based block bootstrap for unit root testing. Econometrica 71, 813–856.

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., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185–207.

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.

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

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.