New Specification Tests in Nonlinear Time Series with Nonstationarity

JITI GAO
The University of Western Australia

This is joint work with Maxwell King, Zudi Lu and Dag Tjøstheim.

Abstract

This paper considers a class of nonparametric autoregressive processes and then a class of nonparametric time series regression models with a nonstationary regressor. For the autoregression case, we propose a nonparametric unit–root test for the conditional mean. For the nonparametric time series regression case, we construct a nonparametric test for testing whether the regression is of a known parametric form indexed by a vector of unknown parameters. We establish asymptotic distributions of the proposed test statistics. Both the setting and the results differ from earlier work on nonparametric time series regression with stationarity. In addition, we develop a bootstrap simulation scheme for the selection of suitable bandwidth parameters involved in the kernel tests as well as the choice of simulated critical values. An example of implementation is given to show that the proposed tests work in practice.

1. Examples and motivation

Example 1.1 Consider a parametric linear model of the form

$$X_t = \theta\ X_{t-1} + u_t,\ \ t = 1, 2, \cdots, T \tag{1.1}$$

where $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) random errors with $E[u_1] = 0$ and $\sigma_0^2 = E[u_1^2]$, and θ is an unknown parameter.

To check whether there is a kind of unit–root structure, existing results propose testing

$$H_{01}:\ \ \theta = 1. \tag{1.2}$$

Example 1.2 Consider a nonparametric nonlinear model of the form

$$X_t = g(X_{t-1}) + u_t,\ \ t = 1, 2, \cdots, T \tag{1.3}$$

where $g(\cdot)$ is an unknown function.

To test whether there is a kind of unit–root structure, we propose to test

$$\begin{aligned} H_{02}: &\quad P\left(g(X_{t-1}) = X_{t-1}\right) = 1 \text{ versus} \\ H_{12}: &\quad \text{a non–/semiparametric alternative.} \end{aligned} \tag{1.4}$$

The main advantage of using (1.12) and (1.4) is as follows:

- No need to assume the parametric form before testing;
- Model mis–specification may be avoided; and
- Estimation and testing of $g(\cdot)$ may be done simultaneously.

Example 1.3 Consider a nonlinear regression model of the form

$$Y_t = m(X_t) + e_t\ \text{ with }\ X_t = X_{t-1} + u_t, \tag{1.5}$$

where $m(\cdot)$ is unknown, both e_t and u_t are independent, and $\{e_s\}$ are independent of $\{u_t\}$.

Estimation of $m(\cdot)$ has been done in Karlsen, Myklebust and Tjøstheim (2007).

We are also interested in testing

$$\begin{aligned} H_{03}: &\quad P\left(m(X_t) = m_{\theta_0}(X_t)\right) = 1 \text{ vs} \\ H_{13}: &\quad \text{a non–/semiparametric alternative,} \end{aligned} \tag{1.6}$$

where $m_{\theta_0}(\cdot)$ is a parametric function of θ_0.

Under H_{03}, model (1.5) becomes a parametric nonlinear model of the form

$$Y_t = m_{\theta_0}(X_t) + e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \tag{1.7}$$

which has been discussed in Park and Phillips (2001).

When $e_t = \sigma(X_t)\epsilon_t$, it is also interested in testing

$$\begin{aligned} H_{04}: &\quad P\left(m(X_t) = m_{\theta_0}(X_t), \sigma(X_t) = \sigma_{\theta_0}(X_t)\right) = 1 \text{ vs} \\ H_{14}: &\quad \text{a non–/semiparametric alternative,} \end{aligned} \tag{1.8}$$

where $\sigma_{\theta_0}(\cdot)$ is also a parametric function of θ_0.

Example 1.4 Consider a parametric linear model with a nonlinear autoregressive error model of the form

$$Y_t = X_t^\tau \theta + u_t \quad \text{with} \quad u_t = g(u_{t-1}) + e_t, \tag{1.9}$$

where $\{X_t\}$ is a vector of regressors, θ is a vector of unknown parameters, $g(\cdot)$ is an unknown function and $\{e_t\}$ is a sequence of errors.

The interest here is to test

$$\begin{aligned} H_{05}: &\quad P\left(g(u_{t-1}) = \theta_0\ u_{t-1}\right) = 1 \text{ vs} \\ H_{15}: &\quad \text{a non–/semiparametric alternative,} \end{aligned} \tag{1.10}$$

where the case of $\theta_0 \equiv 1$ is included.

In the following discussion, we focus on H_{02} and H_{03}.

During the past two decades or so, there has been much interest in both theoretical and empirical analysis of long–run economic and financial time series data. Models and methods used have been based initially on parametric linear autoregressive moving average representations (Granger and Newbold 1977; Brockwell and Davis 1990; Granger and Teräsvirta 1993; and many others) and then on parametric nonlinear time series models (see e.g. Tong 1990; Granger and Teräsvirta 1993; Fan and Yao 2003). Such parametric linear or nonlinear models, as already pointed out in existing studies, may be too restrictive in some cases. This leads to various nonparametric and semiparametric techniques being used to model nonlinear time series data with the focus of attention being on the case where the observed time series satisfies a type of stationarity. Both estimation and specification testing has been systematically examined in this situation (Robinson 1989; Masry and Tjøstheim 1995, 1997; Härdle, Lütkepohl and Chen 1997; Fan and Yao 1998; Li and Wang 1998; Li 1999; Franke, Kreiss and Mammen 2002; Fan and Yao 2003; Gao 2007; Li and Racine 2007 and others).

The stationarity assumption is restrictive because many time series are nonstationary, and there is now a large literature on linear modeling of nonstationary series, but not much has been done in the nonlinear situation. In nonparametric estimation of nonlinear and nonstationary time series models as well as continuous–time financial models, existing studies include Phillips and Park (1998), Karlsen and Tjøstheim (1998, 2001), Park and Phillips (2001), Bandi and Phillips (2002, 2003, 2005), and Karlsen, Myklebust and Tjøstheim (KMT) (2007). The last paper provides a class of nonparametric versions of some of those parametric models proposed in Engle and Granger (1987). In the field of model specification with nonstationarity, there seems to be very little work on testing in a nonlinear and nonstationary framework. Granger, Inoue and Morin (1997) propose a class of parametric nonlinear random walk models and then discuss their applications in economics and finance. As pointed out in their paper, stochastic trends occur in many macroeconomic and financial series. They also conclude that one of the important problems is how to test whether these nonlinear trends occur and whether such nonlinear trends would be adequate to represent actual data. To provide possible answers to these as well as some other related specification testing problems is the main objective of this paper.

The proposed methodologies and technologies in this paper are applicable to a wide variety of nonlinear time series models, which include a class of nonlinear random walk models proposed by Granger, Inoue and Morin (1997). Specifically, we propose a novel unit root test procedure for stationarity in a nonlinear time series setting. Such a test procedure can initially avoid misspecification through the need to specify a linear conditional mean. In other words, we propose estimating the form of the conditional mean and testing for stationarity simultaneously. Such a test procedure may also be viewed as a nonparametric counterpart of those tests proposed in Dickey and Fuller (1979), Phillips (1987), Phillips and Perron (1988), Phillips (1997), Lobato and Robinson (1998), Phillips and Xiao (1998), Robinson (2003) and many others in the literature.

We consider two different classes of nonlinear time series models with nonstationarity. The first is the class of nonlinear autoregressive models of the form

$$X_t = g(X_{t-1}) + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{1.11}$$

where $g(\cdot)$ is an unknown function defined over $R^1 = (-\infty, \infty)$, $\{\epsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) errors with mean zero and finite variance $\sigma_0^2 = E[\epsilon_1^2]$, and T is the number of observations. The initial value X_0 of X_t may be any $O_p(1)$ random variable. However, we set $X_0 = 0$ in this paper to avoid some unnecessary complications in exposition.

When $g(X_{t-1}) = X_{t-1} + g_1(X_{t-1})$ with $g_1(\cdot)$ being an identifiable nonlinear function, model (1.11) becomes a nonlinear random walk model. Granger, Inoue and Morin (1997) discuss some parametric cases for this model, and suggest several estimation procedures. Recently, Kapetanios, Shin and Snell (2003) propose a testing procedure for a unit root in a parametric nonlinear time series model. As $g_1(\cdot)$ usually represents some kind of nonlinear fluctuation in the conditional mean, it would be both theoretically and practically useful to test whether such a nonlinear term is significant before using model (1.11) in practice. We therefore first consider testing the following null hypothesis:

$$H_{02}: \ P\left(g_1(X_{t-1}) = 0\right) = 1 \text{ or } P\left(g(X_{t-1}) = X_{t-1}\right) = 1 \ \text{ for all } t \geq 1. \tag{1.12}$$

To present our main ideas, we consider only testing (1.12). It should be pointed out that we may also consider a generalized form of model (1.11) with σ_0 replaced by a stochastic volatility function $\sigma(X_{t-1})$. In this case, we should be considering a test for $P\left(g(X_{t-1}) = X_{t-1} \text{ and } \sigma(X_{t-1}) = \sigma_0\right) = 1$ for all $t \geq 1$ instead of (1.12).

Our second class of nonlinear time series regression models is considered under the assumption that H_{02} is true. It is given as:

$$Y_t = m(X_t) + \sigma_{\vartheta_0}(X_t)\ e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \ t = 1, 2, \ldots, T, \tag{1.13}$$

where $m(\cdot)$ is an unknown function defined over $R^1 = (-\infty, \infty)$, $\sigma_{\vartheta_0}(\cdot) > 0$ is a known function indexed by a vector function of unknown parameters ϑ_0, $\{u_t\}$ is a sequence of i.i.d. normal errors, and $\{e_t\}$ is a sequence of martingale differences. We are then interested in testing the following null hypothesis:

$$H_{03}: \ P\left(m(X_t) = m_{\theta_0}(X_t)\right) = 1 \ \text{ for all } t \geq 1, \tag{1.14}$$

where $m_{\theta_0}(x)$ is a known parametric function of x indexed by a vector of unknown parameters, $\theta_0 \in \Theta$. Note that θ_0 is different from ϑ_0 involved in the conditional variance function. Under H_{02}, model (1.13) becomes a nonlinear parametric model of the form

$$Y_t = m_{\theta_0}(X_t) + \sigma_{\vartheta_0}(X_t)\ e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \ t = 1, 2, \ldots, T. \tag{1.15}$$

Park and Phillips (2001) extensively discuss some estimation problems for a form of model (1.15).

To the best of our knowledge, the problem of testing both (1.12) and (1.14) for the case where $\{X_t\}$ is nonstationary has not been discussed. This paper proposes two tests and establishes their asymptotic distributions. As the discussion of the two test problems is very different, we will discuss them separately. For model (1.11), we consider testing H_{02} in a nonparametric setting. For model (1.13), we are also able to establish a novel test of H_{03} for the case where $\{X_t\}$ is a classical random walk model of the form (1.13). In other words, this paper discusses separately how to test H_{02} for model (1.11) and then considers testing H_{03} for model (1.13). For both H_{02} and H_{03}, we construct two kernel based test statistics indexed by a pair of bandwidths and establish their asymptotic distributions.

The rest of the paper is organised as follows. Section 2 establishes two test procedures as well as some asymptotic distributional results. One simulation procedure for implementing the proposed tests is established in Section 3. Section 4 shows how to implement the proposed tests in practice. Section 5 concludes the paper with some remarks on extensions. Mathematical details are relegated to Appendix A.

2. Establishment of the tests and asymptotic theory

Let $\widehat{g}(\cdot)$ be a nonparametric estimator of $g(\cdot)$. The idea is to establish a test based on

$$M_T = \frac{1}{T}\sum_{t=1}^{T}\left[\widehat{g}(X_{t-1}) - X_{t-1}\right]^2, \tag{2.1}$$

which is similar to the linear case where

$$\widehat{\theta} - 1 = \frac{\sum_{t=1}^{T} X_{t-1}(X_t - X_{t-1})}{\sum_{t=1}^{T} X_{t-1}^2}. \tag{2.2}$$

A suitably normalized version of M_T suggests using

$$\widehat{L}_T(h_1) = \frac{\sum_{s=1}^{T}\sum_{t=1,\neq s}^{T} \widehat{u}_s\ K\left(\frac{X_{s-1}-X_{t-1}}{h_1}\right)\ \widehat{u}_t}{\sqrt{2\sum_{s=1}^{T}\sum_{t=1,\neq s}^{T} \widehat{u}_s^2\ K^2\left(\frac{X_{s-1}-X_{t-1}}{h_1}\right)\ \widehat{u}_t^2}}, \tag{2.3}$$

where $\widehat{u}_t = X_t - \widehat{g}(X_{t-1})$, $K(\cdot)$ is a probability kernel function and h_1 is a bandwidth parameter.

When $\widehat{g}(x) = \widehat{\theta}\ x$, a Dickey–Fuller type of test is as follows:

$$\mathrm{DF}_T = \frac{\sum_{t=2}^T (X_t - X_{t-1}) X_{t-1}}{\widehat{\sigma}_T \sqrt{\sum_{t=2}^T X_{t-1}^2}} \tag{2.4}$$

where $\widehat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T \left(X_t - \widehat{\theta}_T X_{t-1} \right)^2$ with $\widehat{\theta}_T = \frac{\sum_{t=2}^T (X_t - X_{t-1}) X_{t-1}}{\sum_{t=2}^T X_{t-1}^2}$.

Note that under H_{03}, the true model is a parametric nonlinear model of the form

$$Y_t = f(X_t, \theta_0) + e_t. \tag{2.5}$$

To test H_{03}, we thus propose using a test statistic of the form

$$\widehat{N}_T(h_2) = \frac{\sum_{s=1}^T \sum_{t=1, \neq s}^T \widehat{e}_s\ G\left(\frac{X_s - X_t}{h_2} \right)\ \widehat{e}_t}{\sqrt{2 \sum_{s=1}^T \sum_{t=1, \neq s}^T \widehat{e}_s^2\ G^2 \left(\frac{X_s - X_t}{h_2} \right)\ \widehat{e}_t^2}}, \tag{2.6}$$

where $G(\cdot)$ is a probability kernel function, h_2 is a bandwidth parameter, $\widehat{e}_t = Y_t - f(X_t, \widehat{\theta})$, in which $\widehat{\theta}$ is a consistent estimator of θ_0 under H_{03}.

To establish asymptotic distributions of $\widehat{L}_T(h_1)$ and $\widehat{N}_T(h_2)$, we need to impose the following assumptions.

Assumption 2.1: (i) *Assume that* $\{u_t = X_t - X_{t-1}\}$ *is a sequence of independent and identically distributed (i.i.d.) normal errors with* $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ *and* $0 < E[u_t^4] = \mu_4 < \infty$.

(ii) *Let* $K(\cdot)$ *be a symmetric probability density function with compact support* $C(K)$. *In addition,* $\int K^2(u) du < \infty$.

(iii) *Assume that* $g(x)$ *is twice differentiable in* $x \in R^1 = (-\infty, \infty)$. *In addition,* $\sup_{x \in C(K)} |g'(x)| < \infty$.

(iv) *Assume that* h_1 *satisfies* $\lim_{T \to \infty} h_1 = 0$ *and* $\limsup_{T \to \infty} T^{\frac{1}{2} - \delta_1} h_1 = \infty$ *for some* $0 < \delta_1 < \frac{1}{2}$.

Remark 2.1. Assumption 2.1(i) implies that u_t and X_{t-1} are independent for all $t \geq 1$. In addition, this paper needs to assume that $\{u_t\}$ is an independent $N(0, \sigma_u^2)$ error. As a result, $X_t = \sum_{s=1}^t u_s \sim N(0, t\sigma_u^2)$. However, we believe that the normality assumption could be removed if the so–called "local–time approach" developed by Phillips and Park (1998) or the Markov splitting technique of Karlsen and Tjøstheim (1998, 2001) could be employed in establishing Theorem 2.1 below. As the potential of the two alternative approaches requires further study, we thus assume normality throughout this paper to establish our main results. Assumption 2.1(ii) holds in many cases. For example, when $K(x) = |x|\ I_{[-1,1]}(x)$, Assumption 2.1(ii) holds automatically. In addition, Assumption 2.1(iii) is a very mild condition.

Assumption 2.1(iv) does not look unnatural in the nonstationary case, although it looks more restrictive than for the stationary case. In addition, the conditions of Theorems 5.1 and 5.2 of Karlsen and Tjøstheim (2001) imposed on h_1 become simplified since we are interested in the special case of random walk with a tail index $\beta = \frac{1}{2}$ involved in the conditions. As also pointed out in Remark 3.1 of Karlsen, Myklebust and Tjøstheim (2007), the conditions on h_1 required to establish Theorems 5.1 and 5.2 of Karlsen and Tjøstheim (2001) may be weakened to $\lim_{T\to\infty} h_1 = 0$ and $\lim_{T\to\infty} T^{\frac{1}{2}-\delta} h_1 = \infty$ for some $0 < \delta < \frac{1}{2}$. Such conditions on the bandwidth for nonparametric testing in the nonstationary case are equivalent to the minimal conditions: $\lim_{T\to\infty} h_1 = 0$ and $\lim_{T\to\infty} T h_1 = \infty$ required in nonparametric kernel testing for both the independence and the stationary time series cases (see Zheng 1996; Li and Wang 1998; Fan and Linton 2003; Gao and King 2005).

Assumption 2.2. (i) *Assume that* $\{u_t = X_t - X_{t-1}\}$ *is a sequence of independent and identically distributed normal errors with* $E[u_t] = 0$, $E[u_t^2] = \sigma_u^2$ *and* $0 < E[u_t^4] = \mu_4 < \infty$.

(ii) *Assume that* $\{e_t\}$ *is a sequence of martingale differences satisfying* $E[e_t|\mathcal{B}_{t-1}] = 0$, $E[e_t^2|\mathcal{B}_{t-1}] = 1$ *a.s.*, $E[e_t^3|\mathcal{B}_{t-1}] = 0$ *a.s. and* $0 < \nu_4 = E[e_t^4|\mathcal{B}_{t-1}] < \infty$ *a.s., where* $\mathcal{B}_{t-1} = \sigma\{e_s : 1 \leq s \leq t-1\}$ *is a σ–field generated by* $\{e_s : 1 \leq s \leq t-1\}$.

(iii) *Assume that* u_s *and* e_t *are mutually independent for all* $s, t \geq 1$.

(iv) *Let* $G(\cdot)$ *be a symmetric probability density function with compact support* $C(G)$. *In addition,* $\int G^2(u)du < \infty$.

(v) *Assume that* $m(x)$ *is twice differentiable in* $x \in R^1 = (-\infty, \infty)$. *In addition,* $\sup_{x \in C(G)} |m'(x)| < \infty$.

In addition to Assumption 2.2, we need more conditions on $m_\theta(\cdot)$ under H_{03}. Let

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \left(Y_t - m_\theta(X_t)\right)^2 .$$

Define the nonlinear least squares estimator of θ_0 as the minimizer of $Q_T(\theta)$ over $\theta \in \Theta$:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} Q_T(\theta).$$

Assumption 2.3. (i) *There is a vector,* $\vartheta_0 = (\sigma_0, \rho_0)$, *of unknown parameters such that the conditional variance function can be specified as* $\sigma^2_{\vartheta_0}(x) = \sigma_0^2 |x|^{2\rho_0}$, *where both* $\sigma_0 > 0$ *and* $0 \le \rho_0 < \infty$ *are some constants.*
(ii) *Assume that* h *satisfies* $\lim_{T\to\infty} h_2 = 0$ *and* $\limsup_{T\to\infty} T^{\frac{1}{2}+\rho_0-\delta_2} h_2 = \infty$ *for some* $0 < \delta_2 < \frac{1}{2} + \rho_0$.
(iii) *Furthermore, suppose under* H_{02} *that the following holds in probability:*

$$\lim_{T\to\infty} \frac{D_T}{T} \sum_{t=1}^{T} \left(m_{\theta_0}(X_t) - m_{\widehat{\theta}}(X_t)\right)^2 = 0,$$

where $D_T = T^{\frac{3}{4}-2\rho_0}\sqrt{h_2}$.

Remark 3.1. (i) Assumption 2.2(i) is the same as Assumption 2.1(i). The key difference is however that Assumption 2.1 imposes the normality condition on the process $\{X_t\}$ under H_{01} while Assumption 3.1 requires $\{u_t\}$ to be standard normal under both H_{03} and H_{13}. This is also because the current section considers testing the parametric conditional mean under the assumption that the explanatory time series $\{X_t\}$ is normally distributed as $N(0, t\sigma_u^2)$, which is the same as assuming that H_{01} is true. When H_{01} is not true, but $\{X_t\}$ belongs to a class of null recurrent processes (as discussed in Karlsen, Myklebust and Tjøstheim 2007), we believe that conditions can be found such that the conclusions of Theorem 2.2 remain true. This case, along with other cases where $\{u_t\}$ is only a sequence of stationary errors, will be left for future research.

(ii) Assumption 2.2(ii) is quite standard in this kind of problem. See, for example, Assumption 2.1 of Park and Phillips (2001). Obviously, Assumption 3.1(ii) covers the case where $\{e_t\}$ is a sequence of independent and standard normally distributed errors.

(iii) Assumption 2.2(iii) imposes the independence between $\{e_s\}$ and $\{u_t\}$ for all $s, t \geq 1$. Such an independence assumption is somewhat restrictive but may not be too unreasonable, since the conditional volatility function $\sigma_{\vartheta_0}(X_t)$ has already been extracted from the error process component. Assumption 2.2(iv)(v) is equivalent to Assumption 2.1(ii)(iii). Such conditions on both the kernel and mean functions are needed in this type of nonstationary cases.

(iv) Assumption 2.3(i) imposes some specific conditions on the form of the conditional variance function, which covers some important models. It is possible that such a specific form may be relaxed to cover some more general parametric functions as may be seen from the derivation in the proof of Lemma B.1. Since the specification of the conditional variance function is not the main interest of this paper, we wish to leave such discussion for future study.

(v) Assumption 2.3(ii) is equivalent to Assumption 2.1(iv), and required to ensure the proofs in Appendix B. Unlike the stationary case, Assumption 2.3(iii) involves both the form of $m_{\theta_0}(\cdot)$ and the rate of convergence of $\widehat{\theta}$ to θ_0. This is because, as discussed extensively by Park and Phillips (2001), the rate of convergence depends on which class $m_{\theta_0}(\cdot)$ belongs to. For example, when $m_{\theta_0}(x) = \alpha_0 + \beta_0 x$, the rate of convergence of $\widehat{\theta}$ to θ_0 is of an order of T^{-1}, faster than the usual rate of $T^{-1/2}$. In this case, Assumption 3.2(ii) reduces to $\lim_{T \to \infty} h = 0$, which is just the first part of Assumption 2.3(ii). In other cases, the rate of convergence of $\widehat{\theta}$ to θ_0 as shown in Theorem 5.1 of Park and Phillips (2001) may be slower than the usual rate of $T^{-1/2}$. Thus, in general the bandwidth h needs to satisfy Assumption 2.3(iii). Assumption 2.3(iii) is needed to establish the asymptotic consistency of the proposed test under H_{02}.

Theorem 2.1: (i) *If Assumption 2.1 holds, then as* $T \to \infty$

$$\widehat{L}_T(h) \to_D N(0,1) \quad \text{under } H_{02}. \tag{2.7}$$

(ii) *If Assumptions 2.2 and 2.3 hold, then as* $T \to \infty$

$$\widehat{N}_T(h) \to_D N(0,1) \quad \text{under } H_{03}. \tag{2.8}$$

The proof of (i) depends on

$$\begin{aligned} L_T(h) &= \frac{\sum_{s=1}^T \sum_{t=1,\neq s}^T u_s\ K\left(\frac{X_{s-1}-X_{t-1}}{h}\right)\ u_t}{\sqrt{2\sum_{s=1}^T \sum_{t=1,\neq s}^T u_s^2\ K^2\left(\frac{X_{s-1}-X_{t-1}}{h}\right)\ u_t^2}} \\ &= \frac{\sum_{s=2}^T \sum_{t=1}^{s-1} u_s\ K\left(\frac{\sum_{i=t}^{s-1} u_i}{h}\right)\ u_t}{\sqrt{\sum_{s=2}^T \sum_{t=1}^{s-1} u_s^2\ K^2\left(\frac{\sum_{i=t}^{s-1} u_i}{h}\right)\ u_t^2}} \\ &\to_D \ N(0,1) \end{aligned} \tag{2.9}$$

as $T \to \infty$, using $u_t \sim N(0, \sigma_0^2)$ and $X_t = \sum_{i=1}^t u_i \sim N(0, t\sigma^2)$.

The proof of (ii) depends on

$$\begin{aligned} N_T(h) &= \frac{\sum_{s=1}^T \sum_{t=1,\neq s}^T e_s\ K\left(\frac{X_s-X_t}{h}\right)\ e_t}{\sqrt{2\sum_{s=1}^T \sum_{t=1,\neq s}^T e_s^2\ K^2\left(\frac{X_s-X_t}{h}\right)\ e_t^2}} \\ &= \frac{\sum_{s=2}^T \sum_{t=1}^{s-1} e_s\ K\left(\frac{\sum_{i=t+1}^{s} u_i}{h}\right)\ e_t}{\sqrt{\sum_{s=2}^T \sum_{t=1}^{s-1} e_s^2\ K^2\left(\frac{\sum_{i=t+1}^{s} u_i}{h}\right)\ e_t^2}} \\ &\to_D \ N(0,1) \end{aligned} \tag{2.10}$$

as $T \to \infty$, using $u_t \sim N(0, \sigma_0^2)$ and $X_t = \sum_{i=1}^t u_i \sim N(0, t\sigma^2)$ as well as the fact that $\{e_s\}$ is assumed to be independent of $\{u_t\}$ for all s, t.

3. Simulation scheme

In this section, we focus on the implementation of Theorem 2.1(i). For ease of expressions, we use $h = h_1$ throughout the rest of this paper.

The exact α–level critical value, $l_\alpha(h)$ $(0 < \alpha < 1)$ is the $1-\alpha$ quantile of the exact finite–sample distribution of $\widehat{L}_T(h)$. We therefore suggest choosing an approximate α–level critical value, $l_\alpha^*(h)$, by using the following simulation procedure:

- Let $X_0^* = 0$. For each $t = 1, 2, \ldots, T$, generate $X_t^* = X_{t-1}^* + \widehat{\sigma}_0 \epsilon_t^*$, $\{\epsilon_t^*\}$ is sampled independently from the Normal distribution: $N(0,1)$, and $\widehat{\sigma}_0^2$ is an initial estimator of σ_0^2 based on the original sample.

- Use the data set $\{X_t^* : t = 1, 2, \ldots, T\}$ to re–estimate σ_0^2. Denote the resulting estimate by $\widehat{\sigma}_0^*$. Compute the test statistic $\widehat{L}_T^*(h)$ that is the corresponding version of $\widehat{L}_T(h)$ by replacing $\widehat{\sigma}_0$ and $\{X_t : 1 \le t \le T\}$ with $\widehat{\sigma}_0^*$ and $\{X_t^* : 1 \le t \le T\}$ on the right–hand side of $\widehat{L}_T(h)$.

- Repeat the above steps M times and produce M versions of $\widehat{L}_T^*(h)$ denoted by $\widehat{L}_{Tm}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{L}_{Tm}^*(h)$ to construct their empirical bootstrap distribution function. The bootstrap distribution of $\widehat{L}_T^*(h)$ given the full sample $\mathcal{Y}_T = \{Y_t : 1 \le t \le T\}$ is defined by $P^*\left(\widehat{L}_T^*(h) \le x\right) = P\left(\widehat{L}_T^*(h) \le x | \mathcal{Y}_T\right)$.

  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

$$\begin{aligned} \alpha(h) &= P\left(\widehat{L}_T(h) \ge l_\alpha^*(h) | H_{02}\right) \text{ and} \\ \beta(h) &= P\left(\widehat{L}_T(h) \ge l_\alpha^*(h) | H_{12}\right). \end{aligned}$$

Let $\mathcal{H} = \{h : \alpha(h) \le \alpha\}$. Choose an optimal bandwidth h_0 such that

$$\widehat{h}_{\text{test}} = \arg\max_{h \in \mathcal{H}} \beta(h).$$

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.

4. Examples of implementation

Example 4.1. Consider a nonlinear time series model of the form

$$X_t = X_{t-1} + g_1(X_{t-1}) + u_t, \tag{4.1}$$

where $g_1(\cdot)$ is an unknown function, $X_0 = 0$, and $\{u_t\}$ is a sequence of independent Normal random errors with $E[u_1] = 0$ and $E[u_1^2] = \sigma_0^2 < \infty$.

In this example, we consider two different alternatives for $g_1(\cdot)$. In the first case, we consider a linear alternative of the form $g_1(x) = \beta x$ with $-2 < \beta < 0$ being estimated by the conventional least squares estimation.

In the second case, the form of $g_1(\cdot)$ is given by $g_1(x) = \beta x \ \left(1 - e^{\beta x^2}\right)$, where $-2 < \beta < 0$ and $0 < \sigma < \infty$ are estimated using a maximized likelihood estimation procedure.

Since we are interested in assessing the performance of the proposed test for a number of different values for β, the true value of $\sigma_0^2 = 0.05$ was used in generating the data in both cases. In addition to the case of $\sigma_0^2 = 0.05$, we have also tried some other values of σ_0. Probably because the test $\widehat{L}_T(h)$ does not depend on the choice of σ_0, the resulting finite sample results are very similar.

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

$$h_i = \frac{1}{2^{5-i}}\, \widehat{h}_{\text{test}} \quad \text{for} \quad i = 1, \cdots, 5. \tag{4.2}$$

To simplify the notation, we introduce

$$L_{1i} = \widehat{L}_T(h_i) \quad \text{for} \quad i = 1, \cdots, 5. \tag{4.3}$$

Let $L_{05} = \text{DF}_T$. The corresponding simulated sizes and power values with 1000 replications for model (4.4) below are given in Table 4.1.

Consider a linear model of the form

$$\begin{aligned} H_{01} : X_t &= X_{t-1} + u_t \text{ versus} \\ H_{11} : X_t &= X_{t-1} + \beta X_{t-1} + u_t, \end{aligned} \tag{4.4}$$

where $-2 < \beta < 0$.

Table 4.1. Simulated sizes and power values at the 5% level

	$T = 250$		$T = 500$		$T = 750$	
β	L_{05}	L_{15}	L_{05}	L_{15}	L_{05}	L_{15}
0.00	0.037	0.041	0.059	0.039	0.054	0.051
-0.05	0.718	0.464	1.000	0.679	1.000	0.804
-0.10	0.999	0.811	1.000	0.966	1.000	0.986
-0.20	1.000	0.993	1.000	1.000	1.000	1.000

Table 4.1 shows that while the sizes are comparable, the conventional test L_{05} is more powerful than the proposed test L_{15} as expected when the alternative model is a linear autoregressive model. However, the biggest power reduction is only about 36% at the case of $T = 250$ and $\beta = -0.05$. This may suggest that we should use the proposed test for nonstationarity in the conditional mean when there is no priori information about the form of the conditional mean.

When the alternative is a nonlinear parametric form as in (4.5), our studies show that L_{05} is basically inferior to our test in the sense that it is much less powerful than the proposed test. We now give the corresponding simulated sizes and power values with 1000 replications for model (4.5) below for the tests in Tables 4.2 and 4.3.

Consider a nonlinear model of the form

$$\begin{aligned} H_{02} : X_t &= X_{t-1} + u_t \text{ versus} \\ H_{12} : X_t &= X_{t-1} + \beta X_{t-1}\left(1 - e^{\beta X_{t-1}^2}\right) + u_t, \end{aligned} \tag{4.5}$$

where $-2 < \beta < 0$.

Table 4.2. Simulated sizes for $T = 250$ at the 5% level

T	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}	L_{05}
250	0.003	0.010	0.034	0.047	0.039	0.038
500	0.007	0.017	0.026	0.041	0.037	0.061
750	0.005	0.014	0.038	0.050	0.049	0.056

Table 4.3. Power values for $T = 250$ at the 5% level

β	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}	L_{05}
-0.05	0.105	0.123	0.132	0.149	0.218	0.084
-0.10	0.212	0.271	0.353	0.439	0.652	0.123
-0.20	0.562	0.734	0.889	0.978	0.997	0.415
-0.40	0.989	1.000	1.000	1.000	1.000	0.671

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

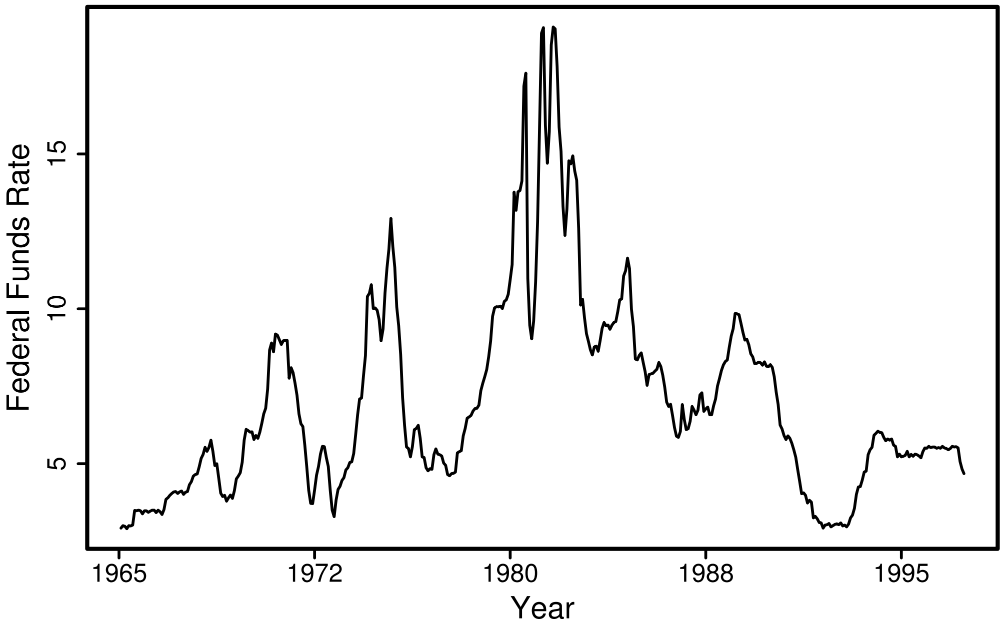
Let $\{X_t : t = 1, 2, \cdots, 432\}$ be the set of Treasury Bill rate data. We assume that $\{X_t\}$ satisfies a nonlinear model of the form

$$X_t = g(X_{t-1}) + u_t, \tag{4.6}$$

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

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, construct $\widehat{g}(\cdot)$, $\widehat{h}_{\text{test}}$ and $\widehat{L}_T(\widehat{h}_{\text{test}})$.
- Let $X_1^* = X_1$. Generate a sequence of bootstrap resamples $\{u_t^*\}$ from $N(0, 1)$ and then $X_t^* = X_{t-1}^* + \widehat{\sigma}_0\ u_t^*$ for $2 \leq t \leq 432$.



- Compute the corresponding version $\widehat{L}_T^*(\widehat{h}_{\text{test}})$ of based on $\{X_t^*\}$.
- Repeat the above steps 1000 times to find the bootstrap distribution of $\widehat{L}_T^*(\widehat{h}_{\text{test}})$ and 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 conclusion is as follows:

- Apply DF_T to test H_{01}. Simulation returns the simulated p–value of $\widehat{p}_1 = 0.005$.
- Apply $\widehat{L}_T(h_0)$ to test H_{02}. Simulation returns the simulated p–value of $\widehat{p}_2 = 0.011$.
- While both of the simulated p–values suggest that there is no enough evidence of supporting unit–root at the 5% significance level, there is some evidence of accepting unit–root based on $\widehat{L}_T(h)$ at the 1% significance level.

5. Conclusions and discussion

- We have proposed a new nonparametric test for the conditional mean functions.
- 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 finite–sample results show that both the proposed test and the Simulation Scheme are practically applicable and implementable.
- Meanwhile, we like to mention some possible extensions of the main ideas to some other closely related models.

Appendix A. As the proofs of the theorems and the necessary lemmas are already extremely technical, we give only an outline for each of the proofs. However, any more details are available from the authors upon request.

To avoid notational complication, we use $h = h_1$ throughout the proof of Theorem 2.1. Let $a_{st} = K_h(\sum_{i=s}^{t-1} u_i) = K\left(\frac{\sum_{i=s}^{t-1} u_i}{h}\right)$ and $\eta_t = 2\sum_{s=1}^{t-1} a_{st}\, u_s$.

Observe that under H_{01}

$$\begin{aligned} L_{1T}(h_1) &= \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s\, K_{h_1}(X_{s-1}-X_{t-1})\, \widehat{u}_t = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\, K_{h_1}(X_{s-1}-X_{t-1})\, u_t \\ &+ \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{g}_s\, K_{h_1}(X_{s-1}-X_{t-1})\, \widetilde{g}_t + 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} u_s\, K_{h_1}(X_{s-1}-X_{t-1})\, \widetilde{g}_t \\ &\equiv L_{1T1} + L_{1T2} + L_{1T3}, \qquad \text{(A.1)} \\ \widehat{\sigma}_{1T}^2 &= 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{u}_s^2\, K_{h_1}^2(X_{s-1}-X_{t-1})\, \widehat{u}_t^2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T} u_s^2\, K_{h_1}^2(X_{s-1}-X_{t-1})\, u_t^2 \\ &+ 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{g}_s^2\, K_{h_1}^2(X_{s-1}-X_{t-1})\, \widetilde{g}_t^2 + \widehat{R}_{1T}, \qquad \text{(A.2)} \end{aligned}$$

where $\widetilde{g}_t = \widehat{g}(X_{t-1}) - g(X_{t-1})$ and $\widehat{R}_{1T}$ is the remainder term given by

$$\widehat{R}_{1T} = \widehat{\sigma}_{1T}^2 - 2\sum_{t=1}^{T}\sum_{s=1}^{T} u_s^2\ K_{h_1}^2(X_{s-1} - X_{t-1})\ u_t^2 - 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widetilde{g}_s^2\ K_{h_1}^2(X_{s-1} - X_{t-1})\ \widetilde{g}_t^2.$$

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

$$\frac{L_{1T1}}{\widetilde{\sigma}_{1T}} \quad \to_D \quad N(0,1), \tag{A.3}$$

$$\frac{L_{1Ti}}{\widetilde{\sigma}_{1T}} \quad \to_P \quad 0 \quad \text{for } i = 2, 3, \tag{A.4}$$

$$L_{1T1} \cdot \left(\frac{1}{\widehat{\sigma}_{1T}} - \frac{1}{\widetilde{\sigma}_{1T}}\right) \quad \to_P \quad 0, \tag{A.5}$$

where $\widetilde{\sigma}_{1S}^2 = 2\sum_{t=1}^{S}\sum_{s=1}^{S} u_s^2\ a_{st}^2\ u_t^2$ for $1 \leq S \leq T$.

We will return to the proof of (A.5) and (A.4) in the second half of this appendix after having proved Lemmas A.1–A.5. In order to prove (A.3), we apply Theorem 3.4 of Hall and Heyde (1980, p.67) to our case. We now start to prove (A.3). The proof of (A.4) is given in the proof of Theorem 2.1 below. Before verifying the conditions of their Theorem 3.4, we introduce the following notation.

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

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

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

$$\frac{\widetilde{\sigma}_{1T}^2}{\sigma_{1T}^2} - \widetilde{U}_{M(T)}^2 \to_P 0. \tag{A.7}$$

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

Before we apply Theorem 3.4 of Hall and Heyde (1980) to our case, we need to state that the conclusion of their Theorem 3.4 remains true if the unconditional assumptions

(3.18) and (3.20) involved in their Theorem 3.4 are replaced by the corresponding conditional assumptions as used in Corollary 3.1 of Hall and Heyde or conditions (A.9) and (A.10) below. Therefore, in order to prove that as $T \to \infty$

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

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

$$\sum_{t=2}^{T} E\left[Y_{Tt}^2 | \Omega_{T,t-1}\right] \quad \to_D \quad \xi^2, \tag{A.9}$$

$$\sum_{t=2}^{T} E\left[Y_{Tt}^2 I_{\{[|Y_{Tt}|>\delta]\}} | \Omega_{T,t-1}\right] \quad \to_P \quad 0, \tag{A.10}$$

$$\sum_{t=2}^{T} E\left[Y_{Tt} | \mathcal{G}_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} + \sum_{t=M(T)+1}^{T} E\left[Y_{Tt} | \Omega_{T,t-1}\right] = \sum_{t=2}^{M(T)} Y_{Tt} \quad \to_P \quad 0, \tag{A.11}$$

$$\sum_{t=2}^{T} \left|E\left[Y_{Tt} | \mathcal{G}_{T,t-1}\right]\right|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 + \sum_{t=M(T)+1}^{T} \left|E\left[Y_{Tt} | \Omega_{T,t-1}\right]\right|^2 = \sum_{t=2}^{M(T)} Y_{Tt}^2 \quad \to_P \quad 0, \tag{A.12}$$

$$\lim_{\delta \to 0} \lim \inf_{T \to \infty} P\left(\frac{\widetilde{\sigma}_{1T}}{\sigma_{1T}} > \delta\right) \quad = \quad 1. \tag{A.13}$$

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

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

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

The proof of (A.13) follows from

$$\frac{\widetilde{\sigma}_{1T}^2}{\sigma_{1T}^2} \to_D \xi^2 > 0. \tag{A.15}$$

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

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

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

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

A.1. Lemmas. The following lemmas are necessary for the complete proof of (A.8).

Lemma A.1. *Assume that the conditions of Theorem 2.1 hold. Then for sufficiently large T*

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

Proof: It follows from the definition that

$$\begin{aligned}
\sigma_{1T}^2 &= \mathrm{E}\left[\sum_{t=1}^{T} \eta_t u_t\right]^2 \\
&= 2\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[a_{st}^2 u_s^2 u_t^2\right] + 4\sum_{t=2}^{T}\sum_{s_1\neq s_2=1}^{t-1} E\left[a_{s_1t}a_{s_2t}u_{s_1}u_{s_2}u_t^2\right] \\
&= 2\sigma_u^2 \sum_{t=1}^{T}\sum_{s=1}^{T} E\left[a_{st}^2 u_s^2\right] + R_{1T}, \qquad (A.18)
\end{aligned}$$

where $R_{1T} = 4\sigma_u^2 \sum_{t=2}^{T}\sum_{s_1\neq s_2=1}^{t-1} E\left[a_{s_1t}a_{s_2t}u_{s_1}u_{s_2}\right]$.

Let $u_{st} = \sum_{i=s+1}^{t-1} u_i$. Throughout this proof, we assume that $\{u_i\}$ is a sequence of independent and normally distributed random variables. Without loss of generality, we also let $\sigma_u^2 = E[u_i^2] \equiv 1$. Let $g_{st}(x) = \frac{1}{\sqrt{2\pi(t-s-1)}} e^{-\frac{x^2}{2(t-s-1)}}$ and $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ be the density functions of $N(0, t-s-1)$ and $N(0,1)$, respectively. A simple calculation implies

$$\begin{aligned}
E[a_{st}^2 u_s^2] &= \int\int K_h^2(u_{st}+u_s) u_s^2 f(u_s) g_{st}(u_{st}) du_s du_{st} \\
&= h\int\int K^2(y) x^2 f(x) g_{st}(hy - x) dx dy \\
&= h(1+o(1)) \int\int K^2(y) x^2 f(x) g_{st}(x) dx dy \\
&= \frac{h(1+o(1))}{\sqrt{2\pi}} \frac{\int K^2(y)dy}{\sqrt{t-s-1}} \frac{t-s-1}{t-s} \qquad (A.19)
\end{aligned}$$

using the fact that both $f(\cdot)$ and $g_{st}(\cdot)$ are normal density functions.

It can then easily be shown that for sufficiently large T

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} E[a_{st}^2 u_s^2] = \frac{4\int K^2(y)dy}{3\sqrt{2\pi}} T^{3/2} h(1+o(1)). \tag{A.20}$$

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

$$Z_i = u_{s_i}, \quad Z_{11} = \sum_{i=s_1+1}^{t-1} u_i, \quad Z_{22} = \sum_{j=s_2+1}^{s_1-1} u_j, \tag{A.21}$$

ignoring the notational involvement of s, t and others.

Let $f_{ii}(x_{ii})$ be the probability density function of Z_{ii} having $Z_{ii} \sim N(0, \sigma_{ii}^2)$ with $\sigma_{11}^2 = t - s_1 - 1$ and $\sigma_{22}^2 = s_1 - s_2 - 1$. Similarly to (A.19), we can show that for sufficiently large T

$$\begin{aligned}
E\left[a_{s_1t}a_{s_2t}u_{s_1}u_{s_2}\right] &= E\left[K_h\left(\sum_{i=s_1}^{t-1} u_i\right) K_h\left(\sum_{j=s_2}^{t-1} u_j\right) u_{s_1}u_{s_2}\right] \\
&= E\left[Z_1 Z_2 K_h\left(Z_2 + Z_{22}\right) K_h\left(Z_1 + Z_2 + Z_{11} + Z_{22}\right)\right] \\
&= E\left[\prod_{i=1}^{2} Z_i K_h\left(\sum_{j=1}^{i}(Z_j + Z_{jj})\right)\right] \\
&= \int\cdots\int x_1 x_2 K_h(x_1 + x_2 + x_{11} + x_{22}) K_h(x_2 + x_{22}) \\
&\times \ f(x_1)f(x_2)f_{11}(x_{11})f_{22}(x_{22})dx_1dx_2dx_{11}dx_{22} \\
&\quad \text{using } \ y_{ii} = \frac{x_i + x_{ii}}{h} \\
&= h^2 \prod_{j=1}^{2}\left[\int K\left(\sum_{i=1}^{j} y_{ii}\right) x_j f(x_j) f_{jj}(x_j - h y_{jj}) dx_j dy_{jj}\right] \\
&\quad \text{using Taylor expansions and } \int x_j f(x_j) f_{jj}(x_j) dx_j = 0 \\
&= h^4(1+o(1)) \prod_{j=1}^{2}\left[\int\int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) x_j f(x_j) f_{jj}'(x_j) dx_j dy_{jj}\right] \\
&= h^4(1+o(1)) \prod_{j=1}^{2}\left[\int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) dy_{jj}\right] \cdot \prod_{j=1}^{2}\left[\int x_j f(x_j) f_{jj}'(x_j) dx_j\right] \\
&= \frac{C_{11}(K) h^4 (1+o(1))}{\prod_{j=1}^{2}\sigma_{jj}^2} \prod_{j=1}^{2}\left[\int x_j^2 f(x_j) f_{jj}(x_j) dx_j\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{C_{11}(K)h^4(1+o(1))}{2\pi} \prod_{j=1}^{2} \frac{1}{\left(\sqrt{1+\sigma_{jj}^2}\right)^3} \\
&= \frac{C_{11}(K)h^4(1+o(1))}{2\pi} \frac{1}{\left(\sqrt{t-s_1}\right)^3} \frac{1}{\left(\sqrt{s_1-s_2}\right)^3}, \qquad \text{(A.22)}
\end{aligned}
$$

using $f'_{jj}(x_j) = -\frac{x_j}{\sigma_{jj}^2} f_{jj}(x_j)$, where $C_{11}(K) = \prod_{j=1}^{2} \int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) dy_{jj} < \infty$ involved in (A.22).

Thus, as for (A.20),

$$
\sum_{t=2}^{T} \sum_{s_1 \neq s_2 = 1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2}\right] = 2 \sum_{t=3}^{T} \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} E\left[a_{s_1 t} a_{s_2 t} u_{s_1} u_{s_2}\right] = o\left(T^{3/2} h\right) \qquad \text{(A.23)}
$$

using Assumption 2.1.

Both (A.20) and (A.23) show that as $T \to \infty$

$$
\sigma_{1T}^2 = \frac{4 \int K^2(y) dy}{3\sqrt{2\pi}} T^{3/2} h(1+o(1)). \qquad \text{(A.24)}
$$

The proof of Lemma A.1 is therefore finished.

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

$$
\lim_{T \to \infty} \frac{1}{\sigma_{1T}^4} \sum_{t=2}^{T} E\left[\eta_t^4\right] = 0. \qquad \text{(A.25)}
$$

Proof. Observe that

$$
E\left[\eta_t^4\right] = 16 \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} \sum_{s_3=1}^{t-1} \sum_{s_4=1}^{t-1} E\left[a_{s_1 t} a_{s_2 t} a_{s_3 t} a_{s_4 t} u_{s_1} u_{s_2} u_{s_3} u_{s_4}\right]. \qquad \text{(A.26)}
$$

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.21), let again $Z_i = u_{s_i}$ for $1 \leq i \leq 4$,

$$Z_{11} = \sum_{i=s_1+1}^{t-1} u_i,\ Z_{22} = \sum_{j=s_2+1}^{s_1-1} u_j,\ Z_{33} = \sum_{k=s_3+1}^{s_2-1} u_k,\ Z_{44} = \sum_{l=s_4+1}^{s_3-1} u_l. \tag{A.27}$$

Analogously to (A.22), we may have

$$\begin{aligned}
E\left[\prod_{i=1}^{4} a_{s_i t} u_{s_i}\right] &= E\left[\prod_{j=1}^{4} Z_j K_h\left(\sum_{i=1}^{j}[Z_i + Z_{ii}]\right)\right] \\
&= \int \prod_{j=1}^{4}\left(K_h\left(\sum_{i=1}^{j}[x_i + x_{ii}]\right) x_j f(x_j) f_{jj}(x_{jj}) dx_j dx_{jj}\right) \\
&\quad \text{using } \ y_{ii} = \frac{x_i + x_{ii}}{h} \\
&= h^4 \int \prod_{j=1}^{4}\left[K\left(\sum_{i=1}^{j} y_{ii}\right) x_j f(x_j) f_{jj}(x_j - h y_{jj}) dx_j dy_{jj}\right] \\
&\quad \text{using Taylor expansions and } \int x_j f(x_j) f_{jj}(x_j) dx_j = 0 \\
&= h^8(1 + o(1)) \int \prod_{j=1}^{4}\left[y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) x_j f(x_j) f'_{jj}(x_j) dx_j dy_{jj}\right] \\
&= h^8(1 + o(1)) \prod_{j=1}^{4}\left[\int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) dy_{jj}\right] \cdot \prod_{j=1}^{4}\left[\int x_j f(x_j) f'_{jj}(x_j) dx_j\right] \\
&\quad \text{using } \ f'_{jj}(x_j) = -\frac{x_j}{\sigma_{jj}^2} f_{jj}(x_j) \\
&= \frac{C_{22}(K) h^8 (1 + o(1))}{\prod_{j=1}^{4} \sigma_{jj}^2} \prod_{j=1}^{4}\left[\int x_j^2 f(x_j) f_{jj}(x_j) dx_j\right] \\
&= \frac{C_{22}(K) h^8 (1 + o(1))}{4\pi^2} \prod_{j=1}^{4} \frac{1}{\left(\sqrt{1 + \sigma_{jj}^2}\right)^3},
\end{aligned} \tag{A.28}$$

where $C_{22}(K) = \prod_{j=1}^{4} \int y_{jj} K\left(\sum_{i=1}^{j} y_{ii}\right) dy_{jj} < \infty$ involved in (A.28), $\sigma_{11}^2 = t - s_1 - 1$, $\sigma_{22}^2 = s_1 - s_2 - 1$, $\sigma_{33}^2 = s_2 - s_3 - 1$ and $\sigma_{44}^2 = s_3 - s_4 - 1$.

Hence, similarly to (A.20) we have

$$\sum_{t=2}^{T} \sum_{1 \leq s_4 < s_3 < s_2 < s_1 \leq t-1} E\left[a_{s_1 t} a_{s_2 t} a_{s_3 t} a_{s_4 t} u_{s_1} u_{s_2} u_{s_3} u_{s_4}\right] = o\left(T^3 h^2\right) \tag{A.29}$$

using Assumption 2.1.

Analogously, we can deal with the other terms of (A.26) as follows:

$$\sum_{t=2}^{T} \sum_{1 \le s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^2 a_{s_2 t}^2 u_{s_1}^2 u_{s_2}^2\right] = o\left(T^3 h^2\right), \tag{A.30}$$

$$\sum_{t=2}^{T} \sum_{1 \le s_3 \ne s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^2 a_{s_2 t} a_{s_3 t} u_{s_1}^2 u_{s_2} u_{s_3}\right] = o\left(T^3 h^2\right), \tag{A.31}$$

$$\sum_{t=2}^{T} \sum_{1 \le s_2 \ne s_1 \le t-1} E\left[a_{s_1 t}^3 a_{s_2 t} u_{s_1}^3 u_{s_2}\right] = o\left(T^3 h^2\right), \tag{A.32}$$

using Assumption 2.1.

Thus, we can finish the proof of (A.25) using (A.26)–(A.32).

As in the proof of Lemma A.1, we assume without loss of generality that $\sigma_u^2 = 1$. To prove (A.9), we thus need only to show that as $T \to \infty$

$$\frac{1}{\sigma_{1T}^2} \sum_{t=2}^{T} \eta_t^2 \to_D \xi^2. \tag{A.33}$$

Lemma A.3. *Let the conditions of Theorem 2.1 hold. Then as* $T \to \infty$

$$\frac{1}{\sigma_{1T}^2} \sum_{t=2}^{T} \eta_t^2 \to_D \xi^2 \tag{A.34}$$

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

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

$$\sigma_{1T}^2 = \frac{4\sigma_0^3 J_{02}}{3\sqrt{2\pi}} T^{3/2} h \equiv C_{10}\, T^{3/2} h. \tag{A.35}$$

Observe that

$$\sum_{t=2}^{T} \eta_t^2 = \sum_{t=2}^{T} \left(2 \sum_{s=1}^{t-1} a_{st} u_s\right)^2 = 4 \sum_{t=1}^{T} \sum_{s=t+1}^{T} a_{st}^2 u_t^2 + 4 \sum_{t=2}^{T} \sum_{s_1=1}^{t-1} \sum_{s_2=1, \ne s_1}^{t-1} u_{s_1}\, a_{s_1 t} a_{s_2 t}\, u_{s_2}. \tag{A.36}$$

To continue the following proof, we need to strengthen Theorem 5.1 of Karlsen and Tjøstheim (2001, p.404) in which under Assumption 2.1, the conclusion of their

Theorem 5.1 holds uniformly in x. The detailed proof is quite tedious and therefore relegated to Appendix D below.

Let $Q(u) = \frac{K^2(u)}{J_{02}}$. Then $Q(\cdot)$ is a probability kernel. According to the strengthened version of Theorem 5.1 of Karlsen and Tjøstheim (2001, p.404), we have for a small set C and as $T \to \infty$

$$\begin{aligned} \frac{1}{J_{02} N_C(T) h} \sum_{s=2}^{T} K^2\left(\frac{X_{s-1} - x}{h}\right) &= \frac{1}{N_C(T) h} \sum_{s=2}^{T} Q\left(\frac{X_{s-1} - x}{h}\right) \\ &\to p_C(x) = \frac{\pi_s(x)}{\pi_s(C)} \text{ almost surely (a.s.)} \end{aligned} \tag{A.37}$$

uniformly in x, where $N_C(T) = \sum_{t=0}^{T} I_C(X_t)$ is as defined as in $T_C(n)$ in Remark 3.5 of Karlsen and Tjøstheim (2001, p.384), and $\pi_s(\cdot)$ is as defined in (3.7) of Karlsen and Tjøstheim (2001, p.379).

Since the distribution of $\{X_t\}$ is assumed to be Gaussian, Corollary 4.1 of Karlsen and Tjøstheim (2001, p.395) implies that π_s can be chosen to be $\pi_s(x) \equiv 1$ uniformly in x and that as $T \to \infty$

$$\frac{N_C(T)}{\pi_s(C)} \frac{1}{N(T)} \to_{a.s.} 1, \tag{A.38}$$

where $N(T)$ is defined as $T(n)$ by Karlsen and Tjøstheim (2001, p.383). This, together with (A.37), implies

$$\frac{1}{N(T) h} \sum_{s=2}^{T} Q\left(\frac{X_{s-1} - x}{h}\right) \to_{a.s.} 1 \tag{A.39}$$

uniformly in x.

In addition, Theorem 3.2 of Karlsen and Tjøstheim (2001, p.389) can be applied to the current case of $X_t = X_{t-1} + u_t$ to show that as $T \to \infty$

$$\frac{N(T)}{\sqrt{T}} \to_D M_{\frac{1}{2}}(1). \tag{A.40}$$

Therefore, equations (A.37)–(A.40) implies as $T \to \infty$

$$\begin{aligned} \frac{4}{\sigma_{1T}^2} \sum_{t=1}^{T} \left(\sum_{s=t+1}^{T} a_{st}^2\right) u_t^2 &= \frac{2}{T C_{10}} \sum_{t=1}^{T} u_t^2 \left(\frac{1}{\sqrt{T} h} \sum_{s=1}^{T} a_{st}^2\right) \\ &\to_D \frac{2\ J_{02}}{C_{10}} M_{\frac{1}{2}}(1) = \frac{3\sqrt{2\pi}}{2} M_{\frac{1}{2}}(1). \end{aligned} \tag{A.41}$$

Moreover, we may show that as $T \to \infty$

$$\frac{4}{\sigma_{1T}^2}\sum_{t=2}^{T}\sum_{s_1=1}^{t-1}\sum_{s_2=1,\neq s_1}^{t-1} u_{s_1}\ a_{s_1 t}a_{s_2 t}\ u_{s_2} \quad = \quad \frac{4}{\sigma_{1T}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left(\sum_{t=\min s_1,s_2+1}^{T} a_{s_1 t}a_{s_2 t}\right) u_{s_1}u_{s_2}$$
$$\to_P \ 0. \tag{A.42}$$

Finally, equations (A.41) and (A.42) complete the proof of Lemma A.3.

We now introduce the following notation:

$$\widehat{U}_{M(T)}^2 = \frac{\widehat{\sigma}_{1,M(T)}^2}{\sigma_{1,M(T)}^2} \quad \text{with} \quad \sigma_{1,M(T)}^2 = \text{Var}\left[\sum_{t=2}^{M(T)} \eta_t u_t\right],$$
$$\widehat{\sigma}_{1,M(T)}^2 = 2\sum_{t=1}^{M(T)}\sum_{s=1}^{M(T)} \widehat{u}_s^2\ a_{st}^2\ \widehat{u}_t^2 \quad \text{with} \quad \widehat{u}_t = X_t - \widehat{g}(X_{t-1}). \tag{A.43}$$

Lemma A.4. *Let the conditions of Theorem 2.1 hold. Then as $T \to \infty$*

$$\frac{\widehat{\sigma}_{1T}^2}{\sigma_{1T}^2} - \widehat{U}_{M(T)}^2 \to_P 0. \tag{A.44}$$

Proof. For $1 \leq S \leq T$, recall $\widetilde{U}_S^2 = \frac{\widetilde{\sigma}_{1S}^2}{\sigma_{1S}^2}$, where $\widetilde{\sigma}_{1S}^2 = 2\sum_{t=1}^{S}\sum_{s=1}^{S} u_s^2\ a_{st}^2\ u_t^2$. In view of the proof of Theorem 2.1 below, in order to prove (A.44), it suffices to show that as $T \to \infty$

$$\frac{\widetilde{\sigma}_{1T}^2}{\sigma_{1T}^2} - \widetilde{U}_{M(T)}^2 \to_P 0. \tag{A.45}$$

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

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

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

Observe that

$$\frac{\widetilde{\sigma}_{1T}^2}{\sigma_{1T}^2} - \widetilde{U}_{M(T)}^2 = \sum_{i=1}^{m} X_{mi} - \sum_{j=1}^{n} X_{nj} + E[u_i^2]\ \left(\frac{1}{\sigma_m^2}\sum_{i=1}^{m} v_i^2 - \frac{1}{\sigma_n^2}\sum_{j=1}^{n} v_j^2\right)$$
$$\equiv I_{mn} + E[u_1^2]\ J_{mn}. \tag{A.48}$$

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

$$I_{mn} \to_P 0 \quad \text{and} \quad J_{mn} \to_P 0. \tag{A.49}$$

We now prove the first part of (A.49). In view of the fact that the independence of $\{u_i\}$ implies for $n+1 \le i \le m$ and $1 \le j \le n$,

$$\begin{aligned} E\left[X_{mi}\left(X_{mj}-X_{nj}\right)\right] &= \frac{\sigma_n^2-\sigma_m^2}{\sigma_m^4\sigma_n^2}\sum_{k=1}^{i-1}\sum_{l=1}^{j-1}E\left[\left(u_i^2-E[u_i^2]\right)\right] \\ &\times\ E\left[\left(u_j^2-E[u_j^2]\right)\ K_h^2\left(\sum_{p=k}^{i-1}u_p\right)u_k^2\ K_h^2\left(\sum_{q=l}^{j-1}u_q\right)u_l^2\right]=0, \end{aligned}$$

we have

$$\begin{aligned} E[I_{mn}^2] &= E\left[\sum_{i=1}^m X_{mi}-\sum_{j=1}^n X_{nj}\right]^2 = E\left[\sum_{i=n+1}^m X_{mi}+\sum_{j=1}^n\left(X_{mj}-X_{nj}\right)\right]^2 \\ &= E\left[\sum_{i=n+1}^m X_{mi}\right]^2+E\left[\sum_{j=1}^n\left(X_{mj}-X_{nj}\right)\right]^2 \\ &= \frac{1}{\sigma_m^4}\sum_{i=n+1}^m E\left(u_i^2-E\left[u_i^2\right]\right)^2\ E\left[v_i^4\right]+\frac{\left(\sigma_m^2-\sigma_n^2\right)^2}{\sigma_m^4\ \sigma_n^4} \\ &\times\ \sum_{j=1}^n E\left(u_j^2-E\left[u_j^2\right]\right)^2\ E\left[v_j^4\right]. \end{aligned} \tag{A.50}$$

We start by looking at $\sum_{i=n+1}^m E\left[v_i^4\right]$ and $\sum_{j=1}^n E\left[v_j^4\right]$ in order to complete the proof of the first part of (A.49). Before we compute the two terms, we have a look at how to prove the second part of (A.49).

Note that

$$\begin{aligned} E\left[J_{mn}^2\right] &= E\left[\frac{1}{\sigma_m^2}\sum_{i=1}^m v_i^2-\frac{1}{\sigma_n^2}\sum_{j=1}^n v_j^2\right]^2 = E\left[\frac{1}{\sigma_m^2}\sum_{i=n+1}^m v_i^2+\frac{\sigma_n^2-\sigma_m^2}{\sigma_m^2\ \sigma_n^2}\sum_{j=1}^n v_j^2\right]^2 \\ &= \frac{1}{\sigma_m^4}E\left[\sum_{i=n+1}^m v_i^2\right]^2+\frac{\left(\sigma_n^2-\sigma_m^2\right)^2}{\sigma_m^4\ \sigma_n^4}E\left[\sum_{j=1}^n v_j^2\right]^2 \\ &-\ 2\frac{\sigma_m^2-\sigma_n^2}{\sigma_m^4\ \sigma_n^2}\sum_{i=n+1}^m\sum_{j=1}^n E\left[v_i^2v_j^2\right]\equiv I_1+I_2-2I_{12}. \end{aligned} \tag{A.51}$$

We first deal with I_1. Recalling $a_{ji} = K_h\left(\sum_{l=j}^{i-1} u_l\right)$, we have

$$E\left(\sum_{i=n+1}^{m} v_i^2\right)^2 = E\left[\sum_{i=n+1}^{m}\sum_{j=n+1}^{m} v_i^2\ v_j^2\right] = \sum_{i=n+1}^{m} E[v_i^4] + \sum_{i=n+1}^{m}\sum_{j=n+1,\neq i}^{m} E[v_i^2\ v_j^2] \equiv J_1 + J_2. \tag{A.52}$$

We now evaluate the orders of $\sum_{i=n+1}^{m} E[v_i^4]$ and $\sum_{i=n+1}^{m}\sum_{j=n+1,\neq i}^{m} E[v_i^2\ v_j^2]$ 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

$$E\left[\sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1} a_{si}^2 u_s^2 a_{tj}^2 u_t^2\right] = \sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1} E\left[a_{si}^2 u_s^2 a_{tj}^2 u_t^2\right] \tag{A.53}$$

$$= \sum_{i=n+2}^{m}\sum_{j=n+1}^{i-1}\sum_{s=2}^{j-1}\sum_{t=1}^{s-1}$$

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

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 for the proofs of Lemmas A.1 and A.2, we have

$$E\left[K_h^2\left(\sum_{i=1}^{2}(Z_i + Z_{ii})\right) K_h^2\left(Z_2 + Z_{22} + Z_{33}\right) Z_1^2 Z_2^2\right] \tag{A.54}$$

$$= \int\cdots\int K_h^2\left(\sum_{i=1}^{2}(x_i + x_{ii})\right) K_h^2\left(x_2 + x_{22} + x_{33}\right)\left(\prod_{i=1}^{2} x_i^2 f(x_i) f_{ii}(x_{ii}) dx_i\ dx_{ii}\right)$$

$$\times\ f_{33}(x_{33})dx_{33}$$

$$\text{using}\quad y_i = x_i \quad\text{and}\quad y_{ii} = \frac{x_i + x_{ii}}{h}$$

$$= h^2\int\cdots\int K^2(y_{11} + y_{22})K^2(y_{22} + y_{33}) y_1^2 y_2^2 f(y_1) f(y_2) f_{11}(y_1 - y_{11}h) f_{22}(y_2 - y_{22}h) f_{33}(y_{33})$$

$$\times dy_1 dy_2 dy_{11} dy_{22} dy_{33}$$

$$= h^2(1+o(1)) \left(\int K^2(u)du\right)^2 \left(\int x_1^2 f(x_1) f_{11}(x_1) dx_1\right) \left(\int x_2^2 f(x_2) f_{22}(x_2) dx_2\right),$$

which is exactly identical to the squared value of $E[a_{ji}^2 u_j^2]$ involved in (A.19), where $f_{ii}(\cdot)$ denotes the marginal density of Z_{ii} and $f(\cdot)$ denotes the density of Z_i.

In view of (A.53) and (A.54), similarly to the calculations of (A.19), (A.20) and (A.24), it can be shown that for sufficiently large m and n,

$$\begin{aligned} E\left[\sum_{i=n+1}^{m} \sum_{j=n+1,\neq i}^{m} v_i^2\ v_j^2\right] &= \sum_{i=n+1}^{m} \sum_{j=n+1,\neq i}^{m} E\left[v_i^2\ v_j^2\right] \\ &= \left(\frac{4\int K^2(y)dy}{3\sqrt{2\pi}}(m-n)^{3/2}h\right)^2\ (1+o(1)) \\ &= \sigma_{m-n}^4(1+o(1)), \end{aligned} \tag{A.55}$$

where σ_m^2 is as defined above (A.46).

Analogously, we may show that for sufficiently large m and n,

$$\sum_{i=n+1}^{m} E[v_i^4] = \sum_{i=n+1}^{m} \sum_{s=1}^{i-1}\sum_{t=1}^{i-1} E\left[a_{si}^2 a_{ti}^2 u_s^2 u_t^2\right] = o\left(\sigma_{m-n}^4\right) \tag{A.56}$$

due to the fact that there is only a triple summation involved in (A.56) while equation (A.55) involves a quadruple summation.

In view of the definition of $\{v_i : i \geq 1\}$, we have the following decompositions:

$$\begin{aligned} E\left[v_i^2 v_j^2\right] &= \sum_{\ell=1}^{i-1}\sum_{k=1}^{j-1} E a_{\ell i}^2 a_{kj}^2 u_\ell^2 u_k^2 = \left(\sum_{\ell=1}^{j-1} + \sum_{\ell=j}^{i-1}\right)\sum_{k=1}^{j-1} E\left[a_{\ell i}^2 a_{kj}^2 u_\ell^2 u_k^2\right] \\ &= \sum_{\ell=1}^{j-1}\sum_{k=1}^{j-1} E\left[a_{\ell i}^2 a_{kj}^2 u_\ell^2 u_k^2\right] + \sum_{\ell=j}^{i-1}\sum_{k=1}^{j-1} E\left[a_{\ell i}^2 u_\ell^2\right] E\left[a_{kj}^2 u_k^2\right] \\ &\equiv C_1(i,j) + C_2(i,j), \end{aligned} \tag{A.57}$$

where the mutual independence of $\{u_l : j \leq l \leq i-1\}$ and $\{u_k : 1 \leq k \leq j-1\}$ has been used in $C_2(i,j)$.

Similarly to (A.55), we also have for sufficiently large m and n,

$$\sum_{i=n+1}^{m}\sum_{j=1}^{n} E\left[v_i^2 v_j^2\right] = \sum_{i=n+1}^{m}\sum_{j=1}^{n} [C_1(i,j) + C_2(i,j)] = \sigma_{m-n}^2\ \sigma_n^2\ (1+o(1)). \tag{A.58}$$

Therefore, equations (A.51)–(A.58) imply that as $m,\ n \to \infty$

$$\begin{aligned}
E\left[J_{mn}^2\right] &= E\left[\frac{1}{\sigma_m^2}\sum_{i=1}^m v_i^2 - \frac{1}{\sigma_n^2}\sum_{j=1}^n v_j^2\right]^2 = E\left[\frac{1}{\sigma_m^2}\sum_{i=n+1}^m v_i^2 + \frac{\sigma_n^2-\sigma_m^2}{\sigma_m^2\ \sigma_n^2}\sum_{j=1}^n v_j^2\right]^2 \\
&= \frac{1}{\sigma_m^4}E\left[\sum_{i=n+1}^m v_i^2\right]^2 + \frac{\left(\sigma_n^2-\sigma_m^2\right)^2}{\sigma_m^4\ \sigma_n^4}E\left[\sum_{j=1}^n v_j^2\right]^2 \\
&- 2\frac{\sigma_m^2-\sigma_n^2}{\sigma_m^4\ \sigma_n^2}\sum_{i=n+1}^m\sum_{j=1}^n E\left[v_i^2 v_j^2\right] \\
&= \left(\frac{(m-n)^3}{m^3} + \frac{\left(m^{3/2}-n^{3/2}\right)^2}{m^3} - 2\ \frac{\left(m^{3/2}-n^{3/2}\right)\ (m-n)^{3/2}}{m^3}\right)\ (1+o(1)) \\
&\to (1-r)^3 + (1-r^{3/2})^2 - 2\ (1-r^{3/2})(1-r)^{3/2} \\
&= \left((1-r)^{3/2} - \left((1-r^{3/2}\right)\right)^2 \geq 0 \qquad \text{(A.59)}
\end{aligned}$$

using $\sigma_m^2 = \frac{4\sigma_0^3 J_{02}}{3\sqrt{2\pi}}m^{3/2}h$, $\sigma_n^2 = \frac{4\sigma_0^3 J_{02}}{3\sqrt{2\pi}}n^{3/2}h$ and $r = \lim_{m,n\to\infty}\frac{n}{m}$. Since $r = 0$ from our assumption, we have therefore shown the second part of (A.49). We now complete the first part of (A.49).

Using the results that $\sum_{i=n+1}^m E[v_i^4] = o\left(\sigma_{m-n}^4\right)$ and $\sum_{j=1}^n E[v_j^4] = o\left(\sigma_n^4\right)$, the proof of the first part of (A.49) follows from (A.50). We therefore have completed the proof of Lemma A.4.

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

$$\frac{\widetilde{\sigma}_{1T}^2}{\sigma_{1T}^2} \to_D \xi^2 > 0. \qquad \text{(A.60)}$$

Proof: In view of the fact that $\{u_s\}$ and $\{u_t\}$ are independent for $s \neq t$ and the definition of $\widetilde{\sigma}_{1T}^2 = 2\sum_{t=1}^T\sum_{s=1}^T u_s^2\ a_{st}^2\ u_t^2 = 4\sum_{t=2}^T\left(\sum_{s=1}^{t-1} a_{st}^2 u_s^2\right)u_t^2$, in order to show (A.60), it suffices to show that

$$\frac{4}{\sigma_{1T}^2}\sum_{t=2}^T\left(\sum_{s=1}^{t-1} a_{st}^2 u_s^2\right) \to_D \xi^2, \qquad \text{(A.61)}$$

which follows exactly from the proof of Lemma A.3 by noting the fact that $\widetilde{\sigma}_{1T}^2$ is the leading term involved in (A.36).

A.2. Proof of Theorem 2.1. In view of (A.3), to complete the proof of Theorem 2.1, it suffices to prove (A.5) and (A.4). We only give the proof of (A.4), since the same technique can be used to prove (A.5).

Before we are able to prove (A.5), we need to strengthen Theorem 5.2 of Karlsen and Tjøstheim (2001, p.406) in which under Assumption 2.1, the conclusion of their Theorem 5.2 holds uniformly in x. The detailed proof is similar to that of the strengthened version of their Theorem 5.1, and omitted here but available in Appendix D below.

To prove (A.4) for $i = 3$, in view of the fact that

$$\frac{L_{2T3}}{T\sqrt{T}h} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}h}\sum_{s=t+1}^{T} K\left(\frac{X_{s-1}-X_{t-1}}{h}\right) u_s\right)\widetilde{g}_t, \tag{A.62}$$

by Lemma A.1 and using the uniform convergence of $\widetilde{g}_t$, it suffices to show that

$$\frac{1}{\sqrt{T}h}\sum_{s=t+1}^{T} K\left(\frac{X_{s-1}-X_{t-1}}{h}\right) u_s = O_P(1) \tag{A.63}$$

uniformly in all $t \geq 1$. The proof of (A.63) is very similar to (A.42), and therefore omitted.

Similarly, using the strengthened version of Theorem 5.2 of Karlsen and Tjøstheim (2001, p.406) about uniform convergence, in order to prove (A.4) for $i = 2$, it suffices to show that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}h}\sum_{s=t+1}^{T} K\left(\frac{X_{s-1}-X_{t-1}}{h}\right)\right) = O_P(1). \tag{A.64}$$

But this follows from the strengthened version of Theorem 5.1 of Karlsen and Tjøstheim (2001), which implies that

$$\frac{1}{\sqrt{T}h}\sum_{s=t+1}^{T} K\left(\frac{X_{s-1}-X_{t-1}}{h}\right) = O_P(1) \tag{A.65}$$

uniformly in all $t \geq 1$. This finally completes the proof of Theorem 2.1(i).

The proof of 2.1(ii) follows similarly, and the detail is available from Appendix B of Gao *et al* (2007).

References

Bandi, F., Phillips, P. C. B., 2002. Nonstationary continuous–time processes. In Y. Aït-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics.* Amsterdam, North Holland (forthcoming).

Bandi, F., Phillips, P. C. B., 2003. Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** 241–283.

Bandi, F., Phillips, P. C. B., 2005. A simple approach to the parametric estimation of potentially nonstationary diffusions. Cowles Foundation Discussion Paper No. 1522.

Brockwell, P. and Davis, R., 1990. *Time Series Theory and Methods.* Springer, New York.

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.

Engle, R. F., Granger, C. W. J., 1987. Co–integration and error correction: representation, estimation and testing. *Econometrica* **55** 251–276.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660.

Fan, J., Yao, Q., 2003. *Nonlinear Time Series*: nonparametric and parametric methods. Springer, New York.

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

Franke, J., Kreiss, J. P., Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8** 1–38.

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

Gao, J., King, M. L., 2005. Estimation and specification testing in nonparametric regression. Available from http://www.maths.uwa.edu.au/˜jiti/jems.pdf.

Gao, J., King, M. L., Lu, Z., Tjøstheim, D., 2007. Specification testing in nonlinear time series with nonstationarity. Available from http://www.maths.uwa.edu.au/˜jiti/gklt.pdf.

Gao, J., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. *Annals of Statistics* **34** 1395–1435.

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

Granger, C. W. J., Newbold, P., 1977. *Forecasting Economic Time Series.* Academic Press, New York.

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.

Härdle, W., Lütkepohl, H., Chen, R., 1997. A review of nonparametric time series analysis. *International Statistical Review* **65** 49–72.

Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* **68** 295–321.

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., 1998. Nonparametric estimation in null recurrent time series. Sonderforschungsbereich 373, **50**. Humboldt Univ., Berlin.

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

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

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

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

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

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

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

Masry, E., Tjøstheim, D., 1997. Additive nonlinear ARX time series and projection estimates. *Econometric Theory* **13** 214–252.

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

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

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

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., 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* **56** 511–534.

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.

Schienle, M., 2006. Reaching for econometric generality: nonparametric nonstationary regression. Unpublished manuscript, Department of Economics, The University of Mannheim, Germany.

Tong, H., 1990. *Nonlinear Time Series*: a dynamical system approach. Oxford University Press, Oxford.

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