Econometric Theory, **25**, 2009, 1869–1892. Printed in the United States of America.
doi:10.1017/S0266466609990363

NONPARAMETRIC SPECIFICATION TESTING FOR NONLINEAR TIME SERIES WITH NONSTATIONARITY

JITI GAO
University of Adelaide

MAXWELL KING
Monash University

ZUDI LU
University of Adelaide

DAG TJØSTHEIM
University of Bergen

This paper considers a nonparametric time series regression model with a nonstationary regressor. We construct a nonparametric test for whether the regression is of a known parametric form indexed by a vector of unknown parameters. We establish the asymptotic distribution of the proposed test statistic. 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 test as well as the choice of simulated critical values. An example of implementation is given to show that the proposed test works in practice.

1. INTRODUCTION

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; 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,

The authors would like to thank Robert Taylor, one of the guest editors, and two referees for their encouragement and constructive comments. The authors also acknowledge useful comments from the conference participants, Bruce Hansen and Peter Phillips in particular, of the Nottingham Conference in Honour of Professor Paul Newbold in September 2007. Thanks also go to Gowry Srianathakumar and Jiying Yin for their excellent computing assistance and the Australian Research Council for its continuing support of the Discovery Grants under grant numbers DP0558602 and DP0879088. Address correspondence to Jiti Gao, School of Economics, University of Adelaide, Adelaide SA 5005, Australia; e-mail: jiti.gao@adelaide.edu.au.

 0266-4666/09 $15.00

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, 1988, 1989; Masry and Tjøstheim, 1995, 1997; Li and Wang, 1998; Li, 1999; Fan and Linton, 2003; Fan and Yao, 2003; Gao, 2007; Li and Racine, 2007; and others).

The stationarity assumption is restrictive because many time series are nonstationary. There is now a large literature on linear modeling of nonstationary time series (see, for example, Dickey and Fuller, 1979; Phillips, 1987, 1997; Phillips and Perron, 1988; Lobato and Robinson, 1998; Phillips and Xiao, 1998; Kapetanios, Shin, and Snell, 2003; Robinson, 2003; and others), but not much has been done in the nonlinear situation. In parametric nonlinear and nonparametric nonlinear time series models with nonstationarity, existing studies include Phillips and Park (PP) (1998), Karlsen and Tjøstheim (KT) (1998, 2001), Park and Phillips (PP) (2001), Wang and Phillips (WP) (2009), Karlsen, Myklebust, and Tjøstheim (KMT) (2007), Phillips (2007), and Chen, Gao, and Li (CGL) (2008). The paper by PP (1998) was among the first to discuss nonparametric kernel estimation in a nonparametric autoregression model with integrated regressors. KT (1998, 2001) independently discuss nonparametric kernel estimation of null recurrent time series. The paper by PP (2001) discusses estimation problems in various parametric nonlinear models with integrated regressors. WP (2006) develop an alternative approach to nonparametric kernel estimation in both autoregression and cointegration models with integrated regressors. The KMT (2007) paper provides a class of nonparametric versions of some of those parametric models proposed in Engle and Granger (1987). Phillips (2007) discusses a nonparametric setting of parametric spurious time series models initially proposed in Granger and Newbold (1974) and then Phillips (1986). More recently, CGL (2008) propose a semiparametric estimation in a partially linear model with nonstationarity.

In the field of model specification with nonstationarity, there are some existing studies (see, for example, Hong and Phillips, 2005; Kasparis, 2007, 2008; and Vadim, 2008). All the cited papers consider specification testing in time series regression with unit roots. The first two papers consider model specification testing in a cointegration setting, while the third paper discusses the applicability of the Bierens test in a class of nonlinear and nonstationary models and establishes some corresponding results. The last paper develops a functional form test in dealing with nonlinearity, nonstationarity, and spurious forecasts. In the original version of this paper, Gao, King, Lu, and Tjøstheim (2007) also propose using a nonparametric kernel test for nonstationarity in an autoregressive model.

In this paper, we are interested in considering a nonlinear time series of the form

$$Y_t = m(X_t) + \sigma_0\, e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \qquad t = 1, 2, \ldots, T, \tag{1.1}$$

where $m(\cdot)$ is an unknown function defined over $R^1 = (-\infty, \infty)$, $\sigma_0 > 0$ is an unknown parameter, $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) errors, and $\{e_t\}$ is a sequence of martingale differences. We are then interested in testing the following null hypothesis:

$$H_0 : P\left(m(X_t) = m_{\theta_0}(X_t)\right) = 1 \quad \text{for all } t \geq 1, \tag{1.2}$$

where $m_{\theta_0}(x)$ is a known parametric function of x indexed by a vector of unknown parameters, $\theta_0 \in \Theta$. Under H_0, model (1.1) becomes a nonlinear parametric model of the form

$$Y_t = m_{\theta_0}(X_t) + \sigma_0\, e_t \quad \text{with} \quad X_t = X_{t-1} + u_t, \qquad t = 1, 2, \ldots, T. \tag{1.3}$$

PP (2001) extensively discuss some estimation problems for this kind of parametric nonlinear time series model.

To the best of our knowledge, the problem of testing (1.2) for the case where $\{X_t\}$ is nonstationary has not been discussed. This paper attempts to derive a simple kernel test for this kind of parametric specification of the conditional mean function when the regressors are integrated. In summary, the main contributions of this paper are:

(i) We propose a new test statistic for model (1.2). Theoretical properties for the proposed test statistic are established.

(ii) In order to implement the proposed test in practice, we develop a new simulation procedure based on the assessment of both the size and power of the proposed test.

The rest of the paper is organized as follows: Section 2 establishes a nonparametric kernel test procedure as well as its asymptotic distribution. A bootstrap simulation scheme is proposed in Section 3. Section 4 shows how to implement the proposed test in practice. Section 5 concludes with some remarks on extensions. Mathematical details are given in the Appendix. Additional details are available from Appendixes B–D of the original version by Gao et al. (2007).

2. NONPARAMETRIC KERNEL TEST

Consider a test problem of the form

$$H_0 : P(m(X_t) = m_{\theta_0}(X_t)) = 1 \quad \text{versus}$$

$$H_1 : P(m(X_t) = m_{\theta_1}(X_t) + \Delta_T(X_t, \theta_1)) = 1 \tag{2.1}$$

for all $t \geq 1$ and some $\theta_1 \in \Theta$ (a parameter space), where $\theta_0 \in \Theta$ denotes the true value of θ if H_0 is true, and $\Delta_T(\cdot, \theta_1)$ is a sequence of unknown functions to ensure that model (1.1) is a semiparametric time series model under H_1.

Note that each $\Delta_T(x, \theta_1)$ behaves like a kind of “distance” function between the null and alternative hypotheses. Such structure allows the inclusion of either a global alternative or a sequence of local alternatives. When $\Delta_T(x, \theta_1) \equiv \Delta(x, \theta_1)$, we are interested in testing a parametric conditional mean function versus a global semiparametric conditional mean function. We are testing a parametric conditional mean function against a sequence of local alternatives when $\{\Delta_T(x,\theta_1)\}$ is a sequence of semiparametric functions.

To construct a nonparametric kernel test, the main idea is to compare a parametric estimator of $m(\cdot)$ under H_0 with a nonparametric kernel estimator. In order to avoid introducing biases associated with nonparametric kernel estimation (Gao and King, 2004), we use a smoothed version of the parametric estimator in the construction.

Similarly to existing studies for the stationary time series case (see, for example, Chap. 3 of Gao, 2007), we propose using a kernel-based test of the form

$$M_T = M_T(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s \; K_h(X_t - X_s)\, \widehat{\epsilon}_t, \tag{2.2}$$

where $K_h(\cdot) = K(\cdot/h)$ with $K(\cdot)$ being a probability kernel function, h is a bandwidth parameter, and $\widehat{\epsilon}_t = Y_t - m_{\widehat{\theta}}(X_t)$, in which $\widehat{\theta}$ is a consistent estimator of θ_0 under H_0. In this paper, we consider $\widehat{\theta}$ as the nonlinear least squares estimator of θ_0 as defined in PP (2001).

In order to establish the asymptotic distribution for M_T, we need to introduce the following assumption.

Assumption 2.1.

(i) The sequence $\{u_t = X_t - X_{t-1}\}$ is a sequence of i.i.d. random errors with $\mathrm{E}[u_t] = 0$, $\mathrm{E}[u_t^2] = \sigma_u^2$, and $\mu_4 = \mathrm{E}[u_t^4] < \infty$. The marginal density function of $\{u_t\}$ is symmetric. The characteristic function $\psi(\cdot)$ of $\{u_t\}$ satisfies $\int_{-\infty}^{\infty} |\psi(v)| dv < \infty$.

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

(iii) The sequences $\{u_s : s \geq 1\}$ and $\{e_t : t \geq 1\}$ are mutually independent.

(iv) The function $K(\cdot)$ is a symmetric and bounded probability density with compact support $C(K)$. In addition, $|K(x+y) - K(x)| \leq \Psi(x)|y|$ for all $x \in C(K)$ and any given y, where $\Psi(x)$ is a nonnegative bounded function for all $x \in C(K)$. Let $K^{(3)}(\cdot)$ denote the three-time convolution of $K(\cdot)$ with itself.

Let $Q_T(\theta) = 1/T \sum_{t=1}^{T} (Y_t - m_\theta(X_t))^2$. 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 2.2.

(i) There are unknown parameters θ_0 and $\sigma_0 > 0$ such that model (1.3) under H_0 is the true identifiable model.

(ii) Here, $\lim_{T\to\infty} h = 0$ and $\limsup_{T\to\infty} T^{(1/2)-\delta_0} h = \infty$ for some $0 < \delta_0 < 1/5$.

(iii) The function $m_\theta(x)$ is differentiable with respect to θ for each fixed x. In addition, under H_i $(i = 0, 1)$ the following equations hold in probability: for $0 < \delta_0 < 1/5$ (τ denotes the transposed),

$$\lim_{T\to\infty} \frac{R_{ij}(T)h}{\sqrt{\left(T^{(3/2)-2\delta_0}h\right)^j}} \left((\widehat{\theta}-\theta_i)^\tau(\widehat{\theta}-\theta_i)\right)^j = 0, \tag{2.3}$$

where, for $i = 0, 1$, $j = 1, 2$, and $R_{ij}(T) = \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} r_{ij}(s)$ with

$$r_{ij}(s) = \int \left\{ \left(\frac{\partial m_{\theta_i}(x)}{\partial\theta}\right)^\tau \left(\frac{\partial m_{\theta_i}(x)}{\partial\theta}\right) \right\}^j \phi\left(\frac{x}{\sqrt{s}}\right) dx,$$

in which $\phi(\cdot)$ is the density function of the normal random variable $N(0, 1)$.

Remark 2.1.

(i) Assumption 2.1(i) requires $\{u_t\}$ to be i.i.d. in order to ensure that $S_t = \sum_{i=1}^t X_i$ have independent increments for all $t \geq 1$. The last sentence of Assumption 2.1(i) imposes a mild condition on the characteristic function, and it holds in many cases. The condition $\int_{-\infty}^{\infty} |\psi(v)| dv < \infty$ ensures certain convergence results. Let $\phi_T(x)$ be the density function of $1/\sqrt{T}\sigma_u \sum_{t=1}^T u_t$. Then Assumption 2.1(i) implies $\sup_x |\phi_T(x) - \phi(x)| \to 0$ as $T \to \infty$, where $\phi(x) = 1/\sqrt{2\pi} e^{-x^2/2}$ is the density function of the standard normal random variable $N(0, 1)$. The proof is standard (see, for example, Chaps. 8 and 9 of Chow and Teicher, 1988).

Assumption 2.1(ii) is quite standard in this kind of problem (see, for example, Ass. 2.1 of PP, 2001). Obviously, Assumption 2.1(ii) covers the case where $\{e_t\}$ is a sequence of i.i.d. errors. Assumption 2.1(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 in this kind of nonstationary problem. Assumption 2.1(iv) is also quite standard in this kind of nonstationary situation.

(ii) Assumption 2.2(i) ensures that the true model (1.3) under H_0 is identifiable. Assumption 2.2(ii) imposes some minimum conditions on the bandwidth. Assumption 2.2(iii) imposes some technical conditions involving both the form of $m_{\theta_0}(\cdot)$ and the rate of convergence of $\widehat{\theta}$ to θ_0. For example, when $m_{\theta_0}(x) = \alpha_0 + \beta_0 x$ and the rate of convergence of $\widehat{\theta}$ to θ_0 is of $o_P(T^{3/8+\delta_0/2}h^{1/4})^{-1}$, Assumption 2.2(iii) holds with $i = 0$. In the case where $m_{\theta_1}(x) = \alpha_1 + \beta_1 x + \gamma_1 x(1 - \exp(-\lambda_1 x^2))$ and the rate of convergence of $\widehat{\theta}$ to θ_1 is of $o_P(T^{7/8+\delta_0/2}h^{1/4})^{-1}$, Assumption 2.2(iii) holds with $i = 1$.

We state the main theorem of this section; its proof is given in the Appendix.

THEOREM 2.1. *Consider model (1.1). Suppose that Assumptions 2.1–2.2 hold with $i = 0$ in Assumption 2.2(iii). Then, under H_0,*

$$\widehat{L}_T = \widehat{L}_T(h) = \frac{M_T(h)}{\widehat{\sigma}_T} \to_D N(0,1) \quad as\ T \to \infty,$$

where $\widehat{\sigma}_T^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s^2\ K_h^2(X_s - X_t)\ \widehat{\epsilon}_t^2$.

Theorem 2.1 shows that $\widehat{L}_T(h)$ converges in distribution to standard normality as $T \to \infty$. Existing studies for the stationary time series case already discuss the small sample performance of this type of nonparametric kernel-based test. When using a normal distribution to approximate the exact finite-sample distribution of this kind of test, the performance of both the size and power functions is not good. In order to improve the finite sample performance of $\widehat{L}_T(h)$, we propose using a bootstrap simulation method. Such a method is known to work quite well in the stationary case. For each given bandwidth satisfying certain theoretical conditions, instead of using an asymptotic value of $l_{0.05} = 1.645$ at the 5% level, for example, we use a simulated critical value for computing the size function and then the power function. An optimal bandwidth is chosen such that the power function is maximized at the optimal bandwidth. Our finite-sample studies show that there is little size distortion when using such a simulated critical value. Such issues are discussed in detail in Section 3.

3. BOOTSTRAP SIMULATION SCHEME

Section 3 discusses how to simulate a critical value for the implementation of $\widehat{L}_T(h)$ in each case. We then examine its finite sample performance using one example in Section 4, below.

Before we look at how to implement $\widehat{L}_T(h)$ in practice, we propose a simulation scheme.

Simulation Scheme 3.1

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 there are unknown quantities, such as parameters and functions, we cannot evaluate $l_\alpha(h)$ in practice. We therefore suggest choosing an approximate α-level critical value, $l_\alpha^*(h)$ by using the following simulation procedure:

(i) For each $t = 1, 2, \ldots, T$, generate $Y_t^* = m_{\widehat{\theta}}(X_t) + \widehat{\sigma}_0\ e_t^*$, where the original sample $(X_1, \ldots, X_T)$ acts in the resampling as a fixed design, $\{e_t^*\}$ is sampled independently either from a prespecified distribution or using a nonparametric bootstrap method, $\widehat{\sigma}_0$ is an initial consistent estimator of σ_0, and $\widehat{\theta}$ is the nonlinear least squares estimator of θ_0 based on the original sample.

(ii) Use the data set $\{(Y_t^*, X_t) : t = 1, 2, \ldots, T\}$ to reestimate (θ_0, σ_0). Denote the resulting estimate by $(\widehat{\theta}^*, \widehat{\sigma}^*)$. Compute the statistic $\widehat{L}_T^*(h)$ that is the corresponding version of $\widehat{L}_T(h)$ by replacing $(\widehat{\theta}, \widehat{\sigma})$ and $\{(Y_t, X_t) : 1 \leq$

$t \leq T\}$ with $(\widehat{\theta}^*, \widehat{\sigma}^*)$ and $\{(Y_t^*, X_t) : 1 \leq t \leq T\}$ on the right-hand side of $\widehat{L}_T(h)$.

(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 values of $\widehat{L}_{Tm}^*(h)$ to construct their empirical bootstrap distribution function. The bootstrap distribution of $\widehat{L}_T^*(h)$ given $\mathcal{W}_T = \{(X_t, Y_t) : 1 \leq t \leq T\}$ is defined by $P^*(\widehat{L}_T^*(h) \leq x) = P(\widehat{L}_T^*(h) \leq x | \mathcal{W}_T)$. Let $l_\alpha^*(h)$ satisfy $P^*(\widehat{L}_T^*(h) \geq l_\alpha^*(h)) = \alpha$ and then estimate $l_\alpha(h)$ by $l_\alpha^*(h)$.

(iv) Define the size and power functions by

$$\alpha(h) = P(\widehat{L}_T(h) \geq l_\alpha^*(h) | H_0) \quad \text{and} \quad \beta(h) = P(\widehat{L}_T(h) \geq l_\alpha^*(h) | H_1).$$

In practice, both $\alpha(h)$ and $\beta(h)$ may be approximated using Edgeworth expansions similarly to (3.23) and (3.24) of Gao (2007).

In order to study both the size and power functions, we specify the form of a sequence of alternatives as

$$H_1: \; P\left(m(X_t) = m_{\theta_1}(X_t) + \Delta_T(X_t, \theta_1)\right) = 1, \tag{3.1}$$

where $\{\Delta_T(x, \theta_1)\}$ is a sequence of unknown functions satisfying certain conditions in Assumption 3.2 below. Under H_1, model (1.1) becomes

$$Y_t = m(X_t) + \epsilon_t = m_{\theta_1}(X_t) + \Delta_T(X_t, \theta_1) + \epsilon_t, \tag{3.2}$$

where $\Delta_T(x, \theta_1)$ can be estimated by $\widehat{\Delta}_T(x, \widehat{\theta}_1)$, in which $\widehat{\theta}_1$ minimizes

$$\sum_{t=1}^{T} \left(Y_t - m_{\theta_1}(X_t) - \widehat{\Delta}_T(X_t, \theta_1)\right)^2, \tag{3.3}$$

and $\widehat{\Delta}_T(x, \theta_1) = \left(\sum_{t=1}^T K_{\widehat{b}_{\text{cv}}}(X_t - x)(Y_t - m_{\theta_1}(X_t))\right) / \left(\sum_{t=1}^T K_{\widehat{b}_{\text{cv}}}(X_t - x)\right)$, with $\widehat{b}_{\text{cv}}$ being chosen by a conventional cross-validation estimation method.

Similarly to the proof of Proposition 3.1 of Gao and Gijbels (2008) for the stationary case, it may be shown that $\lim_{T\to\infty} (\widehat{\Delta}_T(x, \widehat{\theta}_1))/(\Delta_T(x, \theta_1)) = 1$ in probability for each given x. Since both the establishment and the proof of such a consistency result require more detailed discussion, we wish to leave such details for future research.

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

$$\widehat{h}_{\text{test}} = \arg\max_{h \in H_T} \beta(h). \tag{3.4}$$

Since $\{e_t\}$ is stationary, existing results (Sect. 3 of Gao and Gijbels, 2008) suggest using an approximate version of the form

$$\widehat{h}_{\text{test}} = \widehat{a}^{-1/2} \widehat{C}_T^{-3/2}, \tag{3.5}$$

where $\widehat{C}_T^2 = \frac{\sum_{t=1}^T \widehat{\Delta}^2(X_t,\widehat{\theta}_1)\widehat{p}(X_t)}{\widehat{\mu}_2\sqrt{2\widehat{\nu}_2}\ \int K^2(v)dv}$ and $\widehat{a} = \frac{\sqrt{2}K^{(3)}(0)}{3\left(\sqrt{\int K^2(u)du}\right)^3}\ \widehat{c}(p)$ with $\widehat{c}(p) = \frac{\frac{1}{T}\sum_{t=1}^T \widehat{p}^2(X_t)}{\left(\sqrt{\frac{1}{T}\sum_{t=1}^T \widehat{p}(X_t)}\right)^3}$, in which $\widehat{\mu}_2 = \frac{1}{T}\sum_{t=1}^T (Y_t - m_{\widehat{\theta}}(X_t))^2$, $\widehat{\nu}_2 = \frac{1}{T}\sum_{t=1}^T \widehat{p}^2(X_t)$, $\widehat{p}(x) = \frac{1}{\sqrt{T}\widehat{h}_{\text{cv}}}\sum_{t=1}^T K(\frac{X_t - x}{\widehat{h}_{\text{cv}}})$ 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, as shown in the Appendix, the leading term of $\widehat{L}_T(h)$ is given by

$$\widetilde{L}_T(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \epsilon_s\ K_h(X_t - X_s)\ \epsilon_t}{\sigma_{T1}} + \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T \Delta_T(X_s)\ K_h(X_t - X_s)\ \Delta_T(X_t)}{\sigma_{T1}}, \tag{3.6}$$

where σ_{T1}^2 is proportional to $T^{3/2}h$ as explicitly given in Lemma A.1 of the Appendix.

Equation (3.6) shows that the first term contributes to the asymptotic normality under H_0 and the second term contributes to the asymptotic consistency of the test under H_1. Thus, in order to ensure that the test statistic is asymptotically consistent, we need to impose Assumptions 3.1 and 3.2 below.

Assumption 3.1.

(i) There are consistent estimators $\widehat{\sigma}^*$ and $\widehat{\sigma}$ such that, as $T \to \infty$, $\widehat{\sigma} - \sigma_0 \to_P 0$ and $\widehat{\sigma}^* - \widehat{\sigma} \to_P 0$.

(ii) Let H_0 be true. Then the following equation holds in probability: for $0 < \delta_0 < 1/5$,

$$\lim_{T\to\infty} \frac{\widehat{R}_j(T)h}{\sqrt{\left(T^{(3/2)-2\delta_0}h\right)^j}} \left(\widehat{\theta}^* - \widehat{\theta}\right)^\tau \left(\widehat{\theta}^* - \widehat{\theta}\right)^j = 0, \tag{3.7}$$

where, for $j = 1, 2$, $\widehat{R}_j(T) = \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \widehat{r}_j(s)$ with

$$\widehat{r}_j(s) = \int \left\{ \left(\frac{\partial m_{\widehat{\theta}}(x)}{\partial \theta}\right)^\tau \left(\frac{\partial m_{\widehat{\theta}}(x)}{\partial \theta}\right) \right\}^j \phi\left(\frac{x}{\sqrt{s}}\right) dx.$$

Assumption 3.2. Let H_1 be true. Suppose that Assumption 2.2(iii) holds with $i = 1$. In addition, the following equation holds for $0 < \delta_0 < 1/5$:

$$\lim_{T\to\infty} \frac{D(T)\sqrt{h}}{T^{(3/4)-\delta_0}} = \infty, \tag{3.8}$$

where $D(T) = \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} C_T(s)$ with $C_T(s) = \int \Delta_T^2(x,\theta_1)\phi\left(\frac{x}{\sqrt{s}}\right)dx$.

Assumption 3.1(i) imposes only mild consistency conditions on $\widehat{\sigma}^*$ and $\widehat{\sigma}$ to ensure that the bootstrap critical value $l_\alpha^*(h)$ is an asymptotically correct α-level critical value under any model in H_0. Similarly to Corollary 4.4 of PP (2001), one may impose conditions on the local integrability or the integrability of $m_\theta(\cdot)$ to ensure that Assumption 3.1(i) holds. Assumption 3.1(ii) corresponds to Assumption 2.2(iii) with $i = 0$. Similarly to Remark 2.1(iii), it can be verified that Assumption 3.1(ii) holds when $m_\theta(x)$ belongs to a class of parametric functions.

Assumption 3.2 requires that the "distance" between H_0 and H_1 is large enough to ensure that the test is consistent under H_1. Similarly to Assumption 2.2(iii), Assumption 3.2 involves both the form of $m_{\theta_1}(\cdot)$ under H_1 and the rate of convergence of $\widehat{\theta}$ to θ_1 when the form of $m(x)$ is chosen as $m(x) = m_{\theta_1}(x) + \Delta_T(x, \theta_1)$. In both theory and practice, various forms may be considered for $m_{\theta_0}(\cdot)$ and $m(\cdot)$. For example, we consider the following forms:

$$H_0: \; m_{\theta_0}(x) = \alpha_0 + \beta_0 x \quad \text{versus}$$

$$H_1: \; m(x) = m_{\theta_1}(x) + \Delta_T(x, \theta_1) = \alpha_1 + \beta_1 x + \gamma_1 x^2, \tag{3.9}$$

where $\theta_0 = (\alpha_0, \beta_0)$ is estimated by $\widehat{\theta}$, and $-\infty < \alpha_1, \beta_1, \gamma_1 < \infty$ are unknown parameters. In this case, in order to verify Assumption 3.2, it suffices to show that, as $T \to \infty$,

$$\frac{\mathrm{E}\left[\sum_{t=2}^{T} \sum_{s=1}^{t-1} X_s^2 \; K_h(X_t - X_s) \; X_t^2\right]}{T^{3/4-\delta_0}} \to \infty, \tag{3.10}$$

which follows from (letting $X_{st} = X_t - X_s$)

$$\sum_{t=2}^{T} \sum_{s=1}^{t-1} \mathrm{E}\left[X_s^2 K\left(\frac{X_t - X_s}{h}\right) X_t^2\right]$$

$$= \sum_{t=2}^{T} \sum_{s=1}^{t-1} \mathrm{E}\left[X_s^2 K\left(\frac{X_t - X_s}{h}\right) (X_s + X_t - X_s)^2\right]$$

$$= \sum_{t=2}^{T} \sum_{s=1}^{t-1} \int \int x_s^2 K\left(\frac{x_{st}}{h}\right) (x_s + x_{st})^2 f_s(x_s) f_{st}(x_{st}) \, dx_s \, dx_{st}$$

(letting $y_s = x_s$ and $y_{st} = x_{st}/h$)

$$= h \sum_{t=2}^{T} \sum_{s=1}^{t-1} \int \int y_s^2 K(y_{st})(y_s + y_{st} h)^2 f_s(y_s) f_{st}(y_{st} h) \, dy_s \, dy_{st}$$

$$= h(1 + o(1)) \sum_{t=2}^{T} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \int \int x^4 K(y) g_s\left(\frac{x}{\sqrt{s}}\right) g_{st}\left(\frac{yh}{\sqrt{t-s}}\right) dx \, dy$$

$$= \phi(0)\, h(1+o(1)) \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}} \left(\int x^4 \phi\left(\frac{x}{\sqrt{s}}\right) dx\right) \int K(y)\,dy$$

$$= \phi(0)\, h(1+o(1)) \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}} \int x^4 \phi\left(\frac{x}{\sqrt{s}}\right) dx$$

$$= C\,T^{7/2} h(1+o(1)), \tag{3.11}$$

which follows by using the normal distribution approximation method as outlined in the proof of Lemma A.1 in the Appendix, where $f_s(\cdot)$ denotes the density function of $X_s\, f_{st}(\cdot)$ denotes the density function of $X_t - X_s$, $g_s(\cdot)$ denotes the density function of $X_s/\sqrt{s}$, and $g_{st}(\cdot)$ denotes the density function of $(X_t - X_s)/\sqrt{t-s}$. This shows that Assumption 3.2 holds.

In general, we may consider testing various classes of parametric functions under H_0 against nonparametric and/or semiparametric alternatives under H_1. This is both theoretically justifiable and practically implementable, because, as demonstrated by PP (2001, Thms. 5.1 and 5.2), the rate of convergence for one class can be different from that for another class.

We state the following results of this section.

THEOREM 3.1.

(i) *Assume that the conditions of Theorem 2.1 hold. In addition, if Assumption 3.1 holds, then under H_0, we have* $\lim_{T\to\infty} P(\widehat{L}_T(h) \geq l_\alpha^*(h)) = \alpha$.

(ii) *Assume that the conditions of Theorem 2.1 hold. In addition, if Assumptions 3.1 and 3.2 hold, then under H_1, we have* $\lim_{T\to\infty} P(\widehat{L}_T(h) \geq l_\alpha^*(h)) = 1$.

The proof of Theorem 3.1 is given in the Appendix. Theorem 3.1(i) implies that each $l_\alpha^*(h)$ is an asymptotically correct α-level critical value under any model in H_0, and Theorem 3.1(ii) shows that $\widehat{L}_T$ is asymptotically consistent. In Section 4 we illustrate Theorem 3.1 using a simulated example.

4. AN EXAMPLE OF IMPLEMENTATION

This section studies the finite-sample properties of the size and power functions of the proposed test.

Example 4.1

Consider a nonlinear time series model of the form

$$Y_t = m(X_t,\theta) + e_t \quad \text{and} \quad X_t = X_{t-1} + u_t, \qquad t = 1,2,\ldots, \tag{4.1}$$

where $\{e_t\}$ is a sequence of i.i.d. N(0,1), $\{u_t\}$ is also a sequence of i.i.d. N(0,1), $X_0 = 0$, and the forms of $m(x,\theta)$ are given as follows:

$$H_0 : m(x,\theta_0) = \theta_0 x \quad \text{versus} \quad H_1 : m(x,\theta_1) = \theta_{11}x + \theta_{12}x^2 \quad \text{and} \tag{4.2}$$

$$H_0 : m(x,\theta_0) = \theta_0 x \quad \text{versus} \quad H_1 : m(x,\theta_1) = \theta_{21}x + \theta_{22}x\,(1 - e^{-\theta_{23}x^2}), \tag{4.3}$$

where the θ's are chosen as follows: Case 1: $\theta_0 = \theta_{11} = \theta_{21} = 1$ and $\theta_{12} = \theta_{22} = \theta_{23} = 0.08$; Case 2: $\theta_0 = \theta_{11} = \theta_{21} = 1$ and $\theta_{12} = \theta_{22} = 0.05$. Note that Assumptions 2.2 and 3.2 both hold in this case. The form of $m(x, \theta_1)$ in (4.3) has been used in Kapetanios et al. (2003).

In this section, we use an ordinary least squares (OLS) method to estimate the unknown parameters involved in the models under H_0 and the proposed estimation method in (3.3) for the unknown parameters and functions under H_1. In order to compare the performance of the proposed test based on different bandwidths, we evaluate the finite-sample performance of the proposed test associated with both the power-based optimal bandwidth $\widehat{h}_{\text{test}}$ in (3.4) and an estimation-based optimal bandwidth of the form $\widehat{h}_{\text{cv}} = \arg\min_{h \in \mathcal{H}_T} \frac{1}{T} \sum_{i=1}^{T} (Y_i - \widehat{m}_{-i}(X_i; h))^2$, in which $\widehat{m}_{-i}(X_i; h) = \left(\sum_{j=1,\neq i}^{T} K\left(\frac{X_j - X_i}{h}\right) Y_j\right) / \left(\sum_{l=1,\neq i}^{T} K\left(\frac{X_l - X_i}{h}\right)\right)$ with $K(x) = |x| I_{[-1,1]}(x)$ and $\mathcal{H}_T = [T^{-1}, T^{-(1-\delta)}]$ is chosen such that both small and relatively large bandwidth values may be selected, where $0 < \delta < 1$.

Note that $\widehat{h}_{\text{test}}$ and $\widehat{h}_{\text{cv}}$ each has one version under H_0, but both have two versions for Cases 1 and 2 under H_1. To use some simple notation, we introduce $h_{i\text{test}} = \widehat{h}_{\text{test}}$ and $h_{i\text{cv}} = \widehat{h}_{\text{cv}}$ for $i = 0, 1, 2$ to represent $h_{0\text{test}}$ and $h_{0\text{cv}}$ under H_0, and $h_{i\text{test}}$ and $h_{i\text{cv}}$ under H_1 for Cases i with $i = 1, 2$. We then define $L_{i\text{test}} = \widehat{L}_T(h_{i\text{test}})$ and $L_{i\text{cv}} = \widehat{L}_T(h_{i\text{cv}})$ for $i = 0, 1, 2$. For $i = 0, 1, 2$, let $f_{i\text{test}}$ denote the frequency of $L_{i\text{test}} > l_\alpha^*(h_{i\text{test}})$ and $f_{i\text{cv}}$ denote the frequency of $L_{i\text{cv}} > l_\alpha^*(h_{i\text{cv}})$. We consider cases where the number of replications in each of the sample versions of the size and power functions was $M = 1,000$, with $B = 250$ bootstrapping resamples $\{e_t^*\}$ (involved in Simulation Scheme 3.1) from the standard normal distribution $N(0, 1)$; the simulations were done for the cases of $T = 80$, 200, 500, and 800.

Tables 4.1 and 4.2 show that both the proposed test and the proposed simulation scheme are implementable and work well numerically for the cointegration case. First, the augmented test based on $\widehat{h}_{\text{test}}$ is more powerful than that associated with $\widehat{h}_{\text{cv}}$ in each case. Second, Tables 4.1 and 4.2 show that the proposed test is applicable to both linear and nonlinear alternatives. Third, Table 4.2 shows that the proposed test still has power even when the "distance" between the null and an alternative is made deliberately close. For example, when θ_{12} and θ_{22} are made

TABLE 4.1. Simulated sizes at the 1%, 5%, and 10% levels

	1% level		5% level		10% level	
T	$f_{0\text{cv}}$ &	$f_{0\text{test}}$	$f_{0\text{cv}}$ &	$f_{0\text{test}}$	$f_{0\text{cv}}$ &	$f_{0\text{test}}$
80	0.0090	0.0080	0.0400	0.0420	0.0930	0.0920
200	0.0080	0.0060	0.0560	0.0530	0.1050	0.1060
500	0.0110	0.0160	0.0480	0.0540	0.0930	0.0970
800	0.0130	0.0090	0.0520	0.0460	0.1090	0.0990

TABLE 4.2. Simulated power values at the 1%, 5%, and 10% levels

T	Model (4.2)				Model (4.3)			
	f_{1cv}	f_{2cv}	f_{1test}	f_{2test}	f_{1cv}	f_{2cv}	f_{1test}	f_{2test}
				1% level				
80	0.0120	0.0100	0.0090	0.0090	0.0230	0.0200	0.0400	0.0290
200	0.0460	0.0380	0.0580	0.0410	0.2300	0.1370	0.2680	0.1580
500	0.3180	0.2300	0.3940	0.2960	0.7740	0.6600	0.8290	0.7320
800	0.6160	0.5230	0.7120	0.6300	0.9350	0.8880	0.9610	0.9290
				5% level				
80	0.0580	0.0550	0.0590	0.0540	0.1190	0.1010	0.1240	0.1030
200	0.1230	0.0990	0.1320	0.1120	0.4220	0.3200	0.4580	0.3520
500	0.4990	0.4070	0.5850	0.4920	0.8830	0.8070	0.9070	0.8530
800	0.7520	0.6740	0.8360	0.7700	0.9660	0.9380	0.9810	0.9680
				10% level				
80	0.1110	0.1030	0.1150	0.1120	0.1880	0.1660	0.1970	0.1610
200	0.2130	0.1730	0.2240	0.1790	0.5340	0.4410	0.5630	0.4690
500	0.6150	0.5300	0.6850	0.6050	0.9120	0.8620	0.9320	0.9010
800	0.8220	0.7610	0.8860	0.8390	0.9790	0.9630	0.9920	0.9780

as small as 5% and the sample is as medium-sized as $T = 80$, the proposed test still has a power value greater than the nominal level in each case. Finally, Tables 4.1 and 4.2 also show that the power increases when the "distance" between the null hypothesis and an alternative increases.

5. CONCLUSION AND EXTENSIONS

We have proposed a new nonparametric test for the conditional mean function when the regressors are integrated. The asymptotic normal distribution of the proposed test statistic has been established. In addition, we have proposed a 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.

As briefly mentioned in Section 1, we may also consider testing the conditional variance nonparametrically. Furthermore, both the conditional mean and the conditional variance functions may be specified simultaneously. The main idea is that to test

$$H_{01}: \ P(m(X_t) = m_{\theta_0}(X_t) \quad \text{and} \quad \sigma(X_t) = \sigma_{\theta_0}(X_t)) = 1, \tag{5.1}$$

we may use a kernel-based test of the form

$$L_T(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \left[U_s \ K_{h_1}(X_s - X_t) \ U_t + V_s \ G_{h_2}(X_s - X_t) \ V_t\right], \tag{5.2}$$

where $h = (h_1, h_2)$ is a pair of bandwidth parameters, $K(\cdot)$ and $G(\cdot)$ are both probability kernel functions, $U_t = Y_t - m_{\widehat{\theta}}(X_t)$, $V_t = U_t^2 - \sigma_{\widehat{\theta}}^2(X_t)$, and $\widehat{\theta}$ is an estimator of θ_0 under H_{01}. Analogously to Theorem 2.1, we may establish a corresponding theorem for $L_T(h)$. As the detail for this case is extremely lengthy and technical, we leave this issue for future study.

Another important extension would be to the case where $X_t = (X_{t1}, \ldots, X_{td})$ in (1.1) is a vector of d-dimensional nonstationary sequences. In this case, we are interested in testing

$$H_{02}:\ P\left(m(X_t) = \sum_{i=1}^{d} m_{i\theta_0}(X_{ti})\right) = 1 \quad \text{for all } t \geq 1, \tag{5.3}$$

where each $m_{i\theta_0}(\cdot)$ is a known function indexed by θ_0. Detailed construction of such a test would involve some estimation procedures for additive models as used in Gao, Lu, and Tjøstheim (2006) in the stationary spatial case. Since such an extension is not straightforward, we leave it as a future topic.

REFERENCES

Brockwell, P. & R. Davis (1990) *Time Series Theory and Methods*. Springer.

Chen, J., J. Gao, & D. Li (2008) Semiparametric regression estimation in null recurrent time series. Working paper, University of Adelaide; available at http://www.adelaide.edu.au/directory/jiti.gao.

Chow, Y.S. & H. Teicher (1988) *Probability Theory*. Springer-Verlag.

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

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

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

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

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

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

Gao, J. & M.L. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–883.

Gao, J., M.L. King, Z. Lu, & D. Tjøstheim (2007) Specification Testing in Nonlinear Time Series with Nonstationarity. Working paper, University of Adelaide; available at http://www.adelaide.edu.au/directory/jiti.gao.

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

Granger, C.W.J. & P. Newbold (1974) Spurious regressions in econometrics. *Journal of Econometrics* 2, 111–120.

Granger, C.W.J. & P. Newbold (1977) *Forecasting Economic Time Series*. Academic Press.

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

Hall, P. & C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.

Hong, S.H. & P.C.B. Phillips (2005) Testing Linearity in Cointegrating Relations with an Application to PPP. Cowles Foundation Discussion Paper 1541, Yale University.

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

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

Karlsen, H. & D. Tjøstheim (1998) Nonparametric Estimation in Null Recurrent Time Series. Working paper 50, Sonderforschungsbereich series 373, Humboldt University.

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

Kasparis, I. (2007) The Bierens Test for Certain Nonstationary Models. Discussion paper 2007–04, University of Cyprus.

Kasparis, I. (2008) Detection of functional form misspecification in cointegrating relations. *Econometric Theory* 24, 1373–1403.

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

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

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

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

Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.

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

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

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

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

Phillips, P.C.B. (1997) Unit root tests. In S. Klotz (ed.), *Encyclopedia of Statistical Sciences*, vol. 1, 531–542.

Phillips, P.C.B. (2007) Local limit theory and spurious regressions. Cowles Foundation Discussion Paper, Yale University.

Phillips, P.C.B. & J. Park (1998) Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper 1181, Yale University.

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

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

Robinson, P.M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–964.

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. In *Recent Developments in Time Series*, vol. 1, pp. 526–43. Elgar Reference Collection. International Library of Critical Writings in Econometrics.

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

Vadim, M. (2008) Nonlinearity nonstationarity and spurious forecasts. *Journal of Econometrics* 142, 1–27.

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

APPENDIX

This appendix provides mathematical details for the proofs of the main theorems and their associated lemmas. Additional derivations are available from Appendixes B–D of the original version by Gao et al. (2007).

To avoid notational complication, we introduce the following notation: Let $a_{st} = K_h(X_t - X_s)$, $\epsilon_t = \sigma_0 e_t$, and $\eta_t = 2\sum_{s=1}^{t-1} a_{st}\epsilon_s$. Recall $\lambda_t(\theta_0) = m_{\theta_0}(X_t) - m_{\widehat{\theta}}(X_t)$.

Observe that, under H_0,

$$M_T(h) = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s\ K_h(X_t - X_s)\,\widehat{\epsilon}_t = \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \epsilon_s\ K_h(X_s - X_t)\,\epsilon_t$$

$$+ \sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \lambda_s(\theta_0)\ K_h(X_t - X_s)\ \lambda_t(\theta_0) + 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \epsilon_s\ K_h(X_t - X_s)\ \lambda_t(\theta_0)$$

$$\equiv M_{T1} + M_{T2} + M_{T3}, \tag{A.1}$$

$$\widehat{\sigma}_T^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s^2\ K_h^2(X_t - X_s)\,\widehat{\epsilon}_t^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \epsilon_s^2\ K_h^2(X_t - X_s)\,\epsilon_t^2$$

$$+ 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \lambda_s^2(\theta_0)\ K_h^2(X_t - X_s)\ \lambda_t^2(\theta_0) + \widehat{R}_T, \tag{A.2}$$

where $\widehat{R}_T$ is the remainder term given by

$$\widehat{R}_T = \widehat{\sigma}_T^2 - 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \epsilon_s^2\ K_h^2(X_t - X_s)\,\epsilon_t^2 - 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \lambda_s^2(\theta_0)\ K_h^2(X_t - X_s)\ \lambda_t^2(\theta_0).$$

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

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

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

$$\frac{\widehat{\sigma}_T^2 - \widetilde{\sigma}_T^2}{\widetilde{\sigma}_T^2} \to_P \quad 0, \tag{A.5}$$

where $\widetilde{\sigma}_T^2 = 2\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T} \epsilon_s^2 a_{st}^2 \epsilon_t^2$.

We will return to the proof of (A.4) and (A.5) in the second half of this Appendix after having proved Lemmas A.1–A.3. In order to prove (A.3), we need to introduce a stochastic normalization procedure before we may apply Corollary 3.1 of Hall and Heyde (1980, p. 58) to our case.

Let $C_{10} = 2\sigma_0^4 \int K^2(u)\,du$ and define a random variable of the form

$$\sigma_{10}^2 = C_{10}\ N(T)\ Th, \tag{A.6}$$

in which $N(T)$ has the same definition as $T(n)$ in Karlsen and Tjøstheim (2001). It is the number of regenerations for the Markov chain $\{X_t\}$. Note that we use σ_{10}^2 to express the explicit function of the random variable $N(T)$ for notational simplicity. More details about the definition of $N(T)$ are available from Appendix B of the Gao et al. (2008). In addition, it follows from Appendix B that the inequality

$$T^{1/2-\delta_0} \leq N(T) \leq T^{1/2+\delta_0} \tag{A.7}$$

holds almost surely for large enough T and all $0 < \delta_0 < 1/5$.

As shown in Lemma A.3 below, we have, as $T \to \infty$,

$$\frac{\widetilde{\sigma}_T^2}{\sigma_{10}^2} \to_P 1. \tag{A.8}$$

In view of (A.8), to prove (A.3) it suffices to show that, as $T \to \infty$,

$$\frac{M_{T1}}{\sigma_{10}} \to_D N(0,1). \tag{A.9}$$

We now start to prove (A.9). Before verifying the conditions of Corollary 3.1 of Hall and Heyde (1980), we introduce some notation.

Let $U_{Tt} = \eta_t \epsilon_t / \sigma_{10}$ and $\Omega_{T,s} = \sigma\{U_{Tt} : 1 \leq t \leq s\}$ be the σ–field generated by $\{U_{Tt} : 1 \leq t \leq s\}$. Since $N(T)$ is independent of $\{e_t : 1 \leq t \leq T\}$ by construction, $\mathrm{E}[U_{Tt}|\Omega_{T,t-1}] = 0$. By Corollary 3.1 of Hall and Heyde (1980), in order to prove (A.9), it suffices to show that for all $\delta > 0$,

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

$$\sum_{t=2}^{T} \mathrm{E}\left[U_{Tt}^2 |\Omega_{T,t-1}\right] \to_P \ 1. \tag{A.11}$$

Given the definition of $\{U_{Tt}\}$, in order to verify (A.10) and (A.11) it suffices to show that, as $T \to \infty$,

$$\frac{1}{\sigma_{10}^4} \sum_{t=2}^{T} \eta_t^4 \to_P \ 0, \tag{A.12}$$

$$\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^2 \to_P \ 1. \tag{A.13}$$

The proofs of (A.12) and (A.13) are given in Lemmas A.2 and A.3, respectively.

A.1. Lemmas. Assumption 2.1(i) already assumes that $\{u_i\}$ is a sequence of i.i.d. random variables and has a symmetric probability density function. Now we let $f(x)$ and $f_{st}(x)$ be the density functions of u_i and $X_{st} = X_t - X_s$, respectively, and $g_{st}(x)$ be the density function of $V_{st} = X_{st}/\sqrt{t-s}$. Clearly, $f_{st}(x) = g_{st}(x/\sqrt{t-s}) 1/\sqrt{t-s}$, and by utilizing the usual normal approximation of $V_{st} \to_D N(0,1)$ as $t - s \to \infty$ under the conventional central limit theorem conditions, it follows from Assumption 2.1(i) that

$\sup_{x\in R^1}|g_{st}(x)-\phi(x)|\to 0$ as $t-s\to\infty$. Thus, $\sup_{x\in R^1}|g_{st}(x/\sqrt{t-s})-\phi(x/\sqrt{t-s})|\to 0$ as $t-s\to\infty$, where $\phi(x)=1/\sqrt{2\pi}\exp\{-x^2/2\}$.

Another key condition used in the following proofs is that $\{e_s\}$ and $\{u_t\}$ are assumed to be mutually independent for all $s,t\geq 1$. In order to complete the proof of Theorem 2.1 we need to evaluate $\sigma_{T1}^2=\text{var}(M_{T1}(h))$. Recall $\epsilon_s=\sigma_0\, e_s$, $X_{st}=X_t-X_s=\sum_{i=s+1}^t u_i$, and define

$$\xi_{st}=K_h(X_{st})\,\epsilon_s\,\epsilon_t \quad \text{with} \quad \lambda_s(\theta_0)=m_{\theta_0}(X_s)-m_{\widehat{\theta}}(X_s). \tag{A.14}$$

We assume without loss of generality that $\sigma_u^2=\text{E}[u_t^2]\equiv 1$ and $\sigma_0^2=\text{E}[e_1^2]\equiv 1$ throughout this Appendix. For $i=1,\ldots,4$, $1\leq s<t\leq T$, and $1\leq s_2<s_1<t\leq T$, we introduce the following notation:

$$B_i(s,t)=\text{E}\left[K_h^i(X_{st})\right]=h\int K^i(x)\,f_{st}(xh)\,dx$$

$$=\frac{h}{\sqrt{t-s}}\int K^i(x)\,g_{st}\left(\frac{xh}{\sqrt{t-s}}\right)dx, \tag{A.15}$$

$$B_i(s_1,s_2,t)=\text{E}\left[K_h^i(X_{s_1t})K_h^i(X_{s_1t}+X_{s_2s_1})\right]$$

$$=h^2\int K^i(x)K^i(x+y)\,f_{s_1t}(xh)f_{s_2,s_1}(hy)\,dx\,dy$$

$$=\frac{h}{\sqrt{t-s_1}}\frac{h}{\sqrt{s_1-s_2}}\int K^i(x)K^i(x+y)$$

$$\times\, g_{s_1t}\left(\frac{xh}{\sqrt{t-s_1}}\right)g_{s_2,s_1}\left(\frac{yh}{\sqrt{s_1-s_2}}\right)dx\,dy, \tag{A.16}$$

$$C_{cdpq}(s,t)=\text{E}\left[e_s^c e_t^d(e_s^2-1)^p(e_t^2-1)^q\right]$$

$$=\begin{cases}1 & \text{if } c=d=2,\ p=q=0,\\ (\nu_4-1)^2 & \text{if } c=d=0,\ p=q=2,\\ 0 & \text{if } c=d=p=q=1,\end{cases} \tag{A.17}$$

where $\nu_4=\text{E}[e_t^4]$.

Since $\{e_t\}$ and $\{u_s\}$ are assumed to be mutually independent for all s,t, we can obtain that, for large enough T,

$$\sigma_{T1}^2=\text{var}\left[M_{T1}(h)\right]=4(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1}\text{E}\left[\xi_{st}^2\right]$$

$$=4\sigma_0^4(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1}B_2(s,t)\,C_{2200}(s,t).$$

Lemma A.1 below derives the order of σ_{T1}^2 and shows that the rate of σ_{T1}^2 diverging to ∞ is slower than T^2h, which is the corresponding rate for the stationary case.

LEMMA A.1. *Assume that the conditions of Theorem 2.1 hold. Then, as* $T \to \infty$,

$$\sigma_{T1}^2 = \text{var}\left[M_{T1}(h)\right] = C_0 T^{(3/2)}\ h\ (1+o(1)), \tag{A.18}$$

where $C_0 = 16\sigma_0^4 \int K^2(u)\,du/3\sqrt{2\pi}$.

Proof. Choose some positive integer $\Pi_T \geq 1$ such that $\Pi_T \to \infty$ and $\frac{\Pi_T}{\sqrt{Th}} \to 0$ as $T \to \infty$. Observe that

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} \mathrm{E}[a_{st}^2] = \sum_{s=1}^{T-1}\ \sum_{t=s+1}^{T} \mathrm{E}[a_{st}^2] = A_{1T} + A_{2T}, \tag{A.19}$$

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

Using (A.15), we have

$$\begin{aligned} A_{2T} &= \sum_{s=1}^{T-1}\ \sum_{\Pi_T+1\leq(t-s)\leq T-1} \mathrm{E}[a_{st}^2] \\ &= \sum_{s=1}^{T-1}\ \sum_{\Pi_T+1\leq(t-s)\leq T-1} \frac{h}{\sqrt{t-s}} \int K^2(x)\, g_{st}\left(\frac{xh}{\sqrt{t-s}}\right) dx \\ &= d_0 h(1+o(1)) \int K^2(x)\,dx \sum_{s=1}^{T-1}\ \sum_{\Pi_T+1\leq(t-s)\leq T-1} \frac{1}{\sqrt{t-s}} \\ &= \frac{4\int K^2(y)\,dy}{3} d_0 T^{3/2} h(1+o(1)), \end{aligned} \tag{A.20}$$

where $d_0 = 1/\sqrt{2\pi}$.

Equations (A.19) and (A.20) imply that, for large enough T,

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} \mathrm{E}[a_{st}^2] = \frac{4\int K^2(y)\,dy}{3\sqrt{2\pi}} T^{3/2} h(1+o(1)). \tag{A.21}$$

Therefore, it follows that, for $T \to \infty$,

$$4\sum_{t=2}^{T}\sum_{s=1}^{t-1} \mathrm{E}\left[\xi_{st}^2\right] = 4\sigma_0^4 \sum_{t=2}^{T}\sum_{s=1}^{t-1} B_2(s,t)\ C_{2200}(s,t) = C_0\ T^{3/2} h(1+o(1)), \tag{A.22}$$

where $C_0 = \frac{16\sigma_0^4 \int K^2(u)\,du}{3\sqrt{2\pi}}$. Thus the proof of Lemma A.1 is completed. ■

For $0 < \delta_0 < \frac{1}{5}$, recall C_{10} as defined in (A.6) and let $\sigma_{20}^2 = C_{10}\ T^{1/2-\delta_0} Th$. We now have the following lemma.

LEMMA A.2. *Under the conditions of Theorem 2.1, we have, as* $T \to \infty$,

$$\frac{1}{\sigma_{10}^4} \sum_{t=2}^{T} \eta_t^4 \to_P 0. \tag{A.23}$$

Proof. In view of (A.7), we have, for large enough T and any given $\delta > 0$,

$$P\left(\frac{1}{\sigma_{10}^4}\sum_{t=2}^{T}\eta_t^4 > \delta\right) \le P\left(\frac{1}{\sigma_{10}^4}\sum_{t=2}^{T}\eta_t^4 > \delta,\ T^{1/2-\delta_0} \le N(T) \le T^{1/2+\delta_0}\right)$$

$$+\ P\left(N(T) < T^{1/2-\delta_0} \text{ or } N(T) > T^{1/2+\delta_0}\right)$$

$$\le P\left(\frac{1}{\sigma_{20}^4}\sum_{t=2}^{T}\eta_t^4 > \delta\right) + o(1) \le \frac{1}{\sigma_{20}^4\,\delta}\sum_{t=2}^{T}\mathrm{E}\left[\eta_t^4\right] + o(1). \qquad \textbf{(A.24)}$$

Thus, in order to prove (A.23), we only need to show that

$$\frac{1}{\sigma_{20}^4}\sum_{t=2}^{T}\mathrm{E}\left[\eta_t^4\right] \to 0. \qquad \textbf{(A.25)}$$

Observe that

$$\mathrm{E}\left[\eta_t^4\right] = 16\sum_{s_1=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{s_3=1}^{t-1}\sum_{s_4=1}^{t-1}\mathrm{E}\left[a_{s_1t}a_{s_2t}a_{s_3t}a_{s_4t}\epsilon_{s_1}\epsilon_{s_2}\epsilon_{s_3}\epsilon_{s_4}\right]. \qquad \textbf{(A.26)}$$

Since Assumption 2.1 imposes mutual independence on $\{u_s\}$ and $\{e_t\}$ for all $s,\ t \ge 1$, in order to prove (A.23), it suffices to show that, as $T \to \infty$,

$$\frac{1}{\sigma_{20}^4}\sum_{t=2}^{T}\sum_{s_1=1}^{t-1}\sum_{s_2=1,\neq s_1}^{t-1}\mathrm{E}\left[a_{s_1t}^2a_{s_2t}^2\right] \to 0, \qquad \textbf{(A.27)}$$

$$\frac{1}{\sigma_{20}^4}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\mathrm{E}\left[a_{st}^4\right] \to 0. \qquad \textbf{(A.28)}$$

To prove (A.27), using (A.16) we have

$$\sum_{t=2}^{T}\sum_{s_1=1}^{t-1}\sum_{s_2=1,\neq s_1}^{t-1}\mathrm{E}\left[a_{s_1t}^2a_{s_2t}^2\right]$$

$$= 4h^2\sum_{t=3}^{T}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\frac{1}{\sqrt{t-s_1}}\frac{1}{\sqrt{s_1-s_2}}\int K^2(x)K^2(x+y)$$

$$\times\, g_{s_1t}\left(\frac{xh}{\sqrt{t-s_1}}\right)g_{s_2,s_1}\left(\frac{yh}{\sqrt{s_1-s_2}}\right)dx\,dy$$

$$= 4h^2(1+o(1))J_{02}^2\,d_0^2\sum_{t=3}^{T}\sum_{s_1=2}^{t-1}\sum_{s_2=1}^{s_1-1}\frac{1}{\sqrt{t-s_1}}\frac{1}{\sqrt{s_1-s_2}}$$

$$= C\,T^2h^2 = o(T^{3-2\delta_0}h^2) = o\left(\sigma_{20}^4\right), \qquad \textbf{(A.29)}$$

using the assumption that $\lim_{T\to\infty} T^{1/2-\delta_0}h = \infty$ for $0 < \delta_0 < 1/5$, where $C > 0$ is some constant, $J_{02} = \int K^2(u)\,du$, and $d_0 = 1/\sqrt{2\pi}$.

Similarly to (A.20), using (A.15) we have

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} \mathrm{E}[a_{st}^4] = \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{h}{\sqrt{t-s}} \int K^4(x)\, g_{st}\left(\frac{xh}{\sqrt{t-s}}\right) dx$$

$$= C_0\, h(1+o(1)) \int K^4(x)\,dx \cdot \sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}}$$

$$= C\, T^{3/2} h\, (1+o(1)) = o\left(T^{3-2\delta_0} h^2\right) = o\left(\sigma_{20}^4\right), \tag{A.30}$$

using the assumption that $\lim_{T\to\infty} T^{1/2-\delta_0} h = \infty$, where $C > 0$ is some constant.

Equations (A.29) and (A.30) complete the proofs of (A.27) and (A.28). This completes the proof of Lemma A.2. ■

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

$$\frac{1}{\sigma_{10}^2} \sum_{t=2}^{T} \eta_t^2 \to_P 1. \tag{A.31}$$

Proof. Observe that

$$\sum_{t=2}^{T} \eta_t^2 = \sum_{t=2}^{T} \left(2 \sum_{s=1}^{t-1} a_{st}\epsilon_s\right)^2 = 4 \sum_{t=2}^{T}\sum_{s=1}^{t-1} a_{st}^2 \epsilon_s^2 + 4 \sum_{t=2}^{T}\sum_{s_1=1}^{t-1} \sum_{s_2=1,\neq s_1}^{t-1} \epsilon_{s_1}\, a_{s_1 t} a_{s_2 t}\, \epsilon_{s_2}. \tag{A.32}$$

We first show that, as $T \to \infty$,

$$\frac{4}{\sigma_{10}^2} \sum_{t=2}^{T}\sum_{s=1}^{t-1} a_{st}^2 \epsilon_s^2 \to_P 1. \tag{A.33}$$

Similarly to the proofs of Lemmas A.1 and A.2, it can be shown that

$$\sum_{t=2}^{T} \left(\sum_{s=1}^{t-1} a_{st}^2 \left(\epsilon_s^2 - 1\right)\right) = o_P\left(\sigma_{10}^2\right), \tag{A.34}$$

using the assumption that $\{\epsilon_t\}$ is independent of $\{u_s\}$ for all s, t, and $\mathrm{E}[\epsilon_1^2] = 1$.

In view of (A.34), in order to prove (A.33), it suffices to show that, as $T \to \infty$,

$$\frac{4}{\sigma_{10}^2} \sum_{t=2}^{T}\sum_{s=1}^{t-1} a_{st}^2 = \frac{2}{\sigma_{10}^2} \sum_{t=1}^{T}\sum_{s=1}^{T} a_{st}^2 \to_P 1. \tag{A.35}$$

Let $Q(u) = K^2(u)/\int K^2(u)\,du$. Then $Q(\cdot)$ is a probability kernel. According to Lemma C.1 in Appendix C of Gao et al. (2007), we have that, as $T \to \infty$,

$$\frac{1}{N(T)h} \sum_{s=1}^{T} Q\left(\frac{X_s - x}{h}\right) \to_P 1 \tag{A.36}$$

uniformly in $x \in R^1$, where we have used the result that the invariant measure of the random walk $\{X_t\}$ can be taken to be a Lebesgue measure with corresponding density

$p(x) \equiv 1$. The uniform convergence in (A.36) strengthens the pointwise convergence of Theorem 5.1 of Karlsen and Tjøstheim (2001) in the random walk case. For more details, refer to Gao et al. (2007) and its Appendixes B–D.

Thus, the proof of (A.35) follows from (A.36) and

$$\frac{2}{\sigma_{10}^2}\sum_{t=1}^T\sum_{s=1}^T a_{st}^2 = \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{N(T)h\ J_{02}}\sum_{s=1}^T K^2\left(\frac{X_s-X_t}{h}\right)\right)$$

$$= \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{N(T)h}\sum_{s=1}^T Q\left(\frac{X_s-X_t}{h}\right)\right) \to_P 1 \qquad \textbf{(A.37)}$$

as $T \to \infty$.

In view of (A.31) and (A.32), in order to complete the proof of (A.31) we need to show that

$$\frac{1}{\sigma_{10}^2}\sum_{t=2}^T\sum_{s_1=1}^{t-1}\sum_{s_2=1,\neq s_1}^{t-1} \epsilon_{s_1}\ a_{s_1t}a_{s_2t}\ \epsilon_{s_2} \to_P 0 \quad \text{as } T\to\infty. \qquad \textbf{(A.38)}$$

Similarly to (A.24), the proof of (A.38) follows from

$$\frac{1}{\sigma_{20}^4}\mathrm{E}\left[\sum_{t=2}^T\sum_{s_1=1}^{t-1}\sum_{s_2=1,\neq s_1}^{t-1} \epsilon_{s_1}\ a_{s_1t}a_{s_2t}\ \epsilon_{s_2}\right]^2 \to 0, \qquad \textbf{(A.39)}$$

which, using the same arguments as in (A.25)–(A.30) and the fact that $\{\epsilon_s\}$ is a sequence of martingale differences and also independent of $\{u_t\}$, follows from

$$\sum_{t_1=2}^T\sum_{t_2=1}^T\sum_{s_1=1}^T\sum_{s_2=1}^T \mathrm{E}\left[a_{s_1t_1}a_{s_2t_1}a_{s_1t_2}a_{s_2t_2}\right] = O\left(T^{5/2}h^3\right) = o\left(\sigma_{20}^4\right). \qquad \textbf{(A.40)}$$

This therefore completes the proof of Lemma A.3. ■

A.2. Proofs of Theorems.

Proof of Theorem 2.1. In view of (A.3), to complete the proof of Theorem 2.1 it suffices to prove (A.4) and (A.5). We only give the proof of (A.4), since the proof of (A.5) is very similar.

Taylor expansions of $m_\theta(x)$ with respect to θ at θ_0 imply

$$m_\theta(x) - m_{\theta_0}(x) = \left(\frac{\partial m_{\theta_0}(x)}{\partial\theta}\right)^\tau (\theta-\theta_0) + o_P\left(||\theta-\theta_0||\right) \qquad \textbf{(A.41)}$$

for each given x. Thus, in order to prove (A.4), using the same arguments as in (A.24), it suffices to show that

$$(\widehat{\theta}-\theta_0)^\tau \sum_{t=1}^T\sum_{s=1}^T \left(\frac{\partial m_{\theta_0}(X_s)}{\partial\theta}\right) K\left(\frac{X_t-X_s}{h}\right)\left(\frac{\partial m_{\theta_0}(X_t)}{\partial\theta}\right)^\tau (\widehat{\theta}-\theta_0) = o_P(\sigma_{10}). \qquad \textbf{(A.42)}$$

Note that using the same arguments as in (A.24), the proof of (A.42) follows when (A.42) holds with σ_{10} replaced by σ_{20}.

To do so, we first evaluate the following quantity. Straightforward calculations imply that, for large enough T (letting $Y_1 = X_s$ and $Y_{12} = X_t - X_s$ and then $x_1 = y_1$ and $x_2 = y_2/h$),

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1} \mathrm{E}\left[\left(\frac{\partial m_{\theta_0}(X_s)}{\partial\theta}\right)^{\tau} K\left(\frac{X_t - X_s}{h}\right)\left(\frac{\partial m_{\theta_0}(X_s + X_t - X_s)}{\partial\theta}\right)\right]$$

$$= \sum_{t=2}^{T}\sum_{s=1}^{t-1} \int\int \left(\frac{\partial m_{\theta_0}(y_1)}{\partial\theta}\right)^{\tau} K\left(\frac{y_{12}}{h}\right)\left(\frac{\partial m_{\theta_0}(y_1 + y_{12})}{\partial\theta}\right) f_s(y_1) f_{st}(y_{12})\, dy_1\, dy_{12}$$

$$= h\sum_{t=2}^{T}\sum_{s=1}^{t-1} \int\int \left(\frac{\partial m_{\theta_0}(x_1)}{\partial\theta}\right)^{\tau} K(x_2)\left(\frac{\partial m_{\theta_0}(x_1 + x_2 h)}{\partial\theta}\right) f_s(x_1) f_{st}(x_2 h)\, dx_1\, dx_2$$

$$= h(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}} \int\int \left(\frac{\partial m_{\theta_0}(x_1)}{\partial\theta}\right)^{\tau} K(x_2)\left(\frac{\partial m_{\theta_0}(x_1)}{\partial\theta}\right)$$

$$\times\ g_s\left(\frac{x_1}{\sqrt{s}}\right) g_{st}\left(\frac{x_2 h}{\sqrt{t-s}}\right) dx_1\, dx_2$$

$$= d_0\, h(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1} \frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}} \int \left(\frac{\partial m_{\theta_0}(x)}{\partial\theta}\right)^{\tau}\left(\frac{\partial m_{\theta_0}(x)}{\partial\theta}\right)\phi\left(\frac{x}{\sqrt{s}}\right) dx. \tag{A.43}$$

This, along with Assumption 2.2(iii) with $j = 1$ and the Markov inequality, implies that (A.42) holds, with σ_{10} replaced by σ_{20}. This therefore proves (A.4) for $i = 2$.

Meanwhile, it follows from (A.3) that

$$\frac{1}{\sigma_{10}}\sum_{t=1}^{T}\sum_{s=1}^{T} \epsilon_s\, K\left(\frac{X_t - X_s}{h}\right)\epsilon_t = O_P(1). \tag{A.44}$$

Thus, the proof of (A.4) for $i = 3$ follows from (A.42)–(A.44) and

$$\left|\sum_{t=1}^{T}\sum_{s=1}^{T} \epsilon_s \sqrt{K\left(\frac{X_t - X_s}{h}\right)}\sqrt{K\left(\frac{X_t - X_s}{h}\right)}\lambda_t(\theta_0)\right|^2$$

$$\leq \sum_{t=1}^{T}\sum_{s=1}^{T} \epsilon_s\, K\left(\frac{X_t - X_s}{h}\right)\epsilon_t \sum_{t=1}^{T}\sum_{s=1}^{T} \lambda_s(\theta_0)\, K\left(\frac{X_t - X_s}{h}\right)\lambda_t(\theta_0)$$

$$= O_P(\sigma_{10}) \cdot o_P(\sigma_{10}) = o_P\left(\sigma_{10}^2\right). \tag{A.45}$$

Similarly to (A.41)–(A.43), using Assumption 2.2(iii) with $j = 2$, one may verify (A.5). ■

Proof of Theorem 3.1. Using

$$\widehat{\epsilon}_t^* \equiv Y_t^* - m_{\widehat{\theta}^*}(X_t) = m_{\widehat{\theta}}(X_t) - m_{\widehat{\theta}^*}(X_t) + \widehat{\sigma}_0\, e_t^*,$$

we have

$$M_T^*(h) \equiv \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s^* \ K_h(X_s - X_t) \ \widehat{\epsilon}_t^* = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\sigma}_0 \ e_s^* \ K_h(X_s - X_t) \ \widehat{\sigma}_0 \ e_t^*$$
$$+ \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \lambda_s^* \ K_h(X_s - X_t) \ \lambda_t^* + 2 \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\sigma}_0 \ e_s^* \ K_h(X_s - X_t) \ \lambda_t^*, \tag{A.46}$$

where $\lambda_t^* = m_{\widehat{\theta}}(X_t) - m_{\widehat{\theta}^*}(X_t)$.

Using Assumptions 2.1, 2.2, and 3.1, in view of the notation of $\widehat{L}_T^*(h)$ introduced in Simulation Scheme 3.1 as well as the proof of Theorem 2.1, we may show that, as $T \to \infty$,

$$P^* \left(\widehat{L}_T^*(h) \leq x \right) \to \Phi(x) \quad \text{for all } x \in (-\infty, \infty) \tag{A.47}$$

holds in probability with respect to the distribution of the original sample $\{(X_t, Y_t) : 1 \leq t \leq T\}$. In detail, in order to prove (A.47), using the fact that $\{e_s^*\}$ and $\{(X_t, Y_t)\}$ are independent for all $s, t \geq 1$, we may show that the proofs of Lemmas A.2 and A.3 remain true by successive conditioning arguments.

Let z_α be the $1 - \alpha$ quantile of $\Phi(\cdot)$ such that $\Phi(z_\alpha) = 1 - \alpha$. Then it follows from (A.47) that, as $T \to \infty$,

$$P^* \left(\widehat{L}_T^*(h) \geq z_\alpha \right) \to 1 - \Phi(z_\alpha) = \alpha \quad \text{in probability.} \tag{A.48}$$

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

$$l_\alpha^* - z_\alpha \to_P 0. \tag{A.49}$$

Using the conclusion of Theorem 2.1 and (A.47) again, we have, as $T \to \infty$,

$$P^* \left(\widehat{L}_T^*(h) \leq x \right) - P \left(\widehat{L}_T(h) \leq x \right) \to_P 0 \quad \text{for all } x \in (-\infty, \infty). \tag{A.50}$$

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

$$\lim_{T \to \infty} P \left(\widehat{L}_T(h) \geq l_\alpha^* \right) = \alpha \tag{A.51}$$

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

To prove Theorem 3.1(ii), we need to decompose $M_T(h)$ as follows:

$$M_T(h) = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \widehat{\epsilon}_s \ K(X_{st}) \ \widehat{\epsilon}_t = \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \epsilon_s(\theta_1) \ K_h(X_{st}) \ \epsilon_t(\theta_1)$$
$$+ \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \lambda_s(\theta_1) \ K_h(X_{st}) \ \lambda_t(\theta_1) + 2 \sum_{t=1}^{T} \sum_{s=1,\neq t}^{T} \lambda_s(\theta_1) \ K_h(X_{st}) \ \epsilon_t(\theta_1),$$

where $\epsilon_t(\theta_1) = Y_t - m(X_t)$ and $\lambda_t(\theta_1) = m(X_t) - m_{\widehat{\theta}}(X_t)$ under H_1.

By the proof of Theorem 2.1, in order to prove Theorem 3.2(ii) it suffices to show that, under H_1,

$$\frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\lambda_s(\theta_1)\ K_h(X_t-X_s)\ \lambda_t(\theta_1)}{\sigma_{10}}\to_P\infty. \tag{A.52}$$

Using Taylor expansions to $m_\theta(\cdot)$ with respect to θ, we have

$$\begin{aligned} m(X_t)-m_{\widehat{\theta}}(X_t) &= \Delta_T(X_t,\theta_1)+m_{\theta_1}(X_t)-m_{\widehat{\theta}}(X_t)\\ &= \Delta_T(X_t,\theta_1)+\left(\theta_1-\widehat{\theta}\right)^\tau \frac{\partial m_\theta(X_t)}{\partial\theta}|_{\theta=\theta_1}. \end{aligned} \tag{A.53}$$

In view of (A.52), using Assumption 2.2(iii) with $i=1$, in order to prove (A.52) it suffices to show that

$$\frac{\sum_{t=1}^{T}\sum_{s=1,\neq t}^{T}\mathrm{E}\left[\Delta_T(X_s,\theta_1)\ K_h(X_t-X_s)\ \Delta_T(X_t,\theta_1)\right]}{\sigma_{20}}\to\infty. \tag{A.54}$$

Note that (letting $X_{st}=X_t-X_s$)

$$\begin{aligned} &\sum_{t=2}^{T}\sum_{s=1}^{t-1}\mathrm{E}\left[\Delta_T(X_s,\theta_1)K\left(\frac{X_t-X_s}{h}\right)\Delta_T(X_t,\theta_1)\right]\\ &=\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left[\Delta_T(X_s,\theta_1)K\left(\frac{X_t-X_s}{h}\right)\Delta_T\left(X_s+X_t-X_s,\theta_1\right)\right]\\ &=\sum_{t=2}^{T}\sum_{s=1}^{t-1}\int\int\Delta_T(x_s,\theta_1)K\left(\frac{x_{st}}{h}\right)\Delta_T(x_s+x_{st},\theta_1)f_s(x_s)f_{st}(x_{st})\,dx_s\,dx_{st}\\ &\qquad\left(\text{letting }\ y_s=x_s\ \text{ and } y_{st}=\frac{x_{st}}{h}\right)\\ &=h\sum_{t=2}^{T}\sum_{s=1}^{t-1}\int\int\Delta_T(y_s,\theta_1)K(y_{st})\Delta_T(y_s+y_{st}h)f_s(y_s)f_{st}(y_{st}h)\,dy_s\,dy_{st}\\ &=h(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}}\int\int\Delta_T^2(x,\theta_1)K(y)g_s\left(\frac{x}{\sqrt{s}}\right)g_{st}\left(\frac{yh}{\sqrt{t-s}}\right)dx\,dy\\ &=\phi(0)\ h(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}}\left(\int\Delta_T^2(x,\theta_1)\phi\left(\frac{x}{\sqrt{s}}\right)dx\right)\int K(y)\,dy\\ &=\phi(0)\ h(1+o(1))\sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}}\ C_T(s). \end{aligned} \tag{A.55}$$

The proof of Theorem 3.1(ii) therefore follows from Assumption 3.2 and equations (A.53)–(A.55). ■