THE UNIVERSITY OF ADELAIDE
AUSTRALIA

The University of Adelaide
School of Economics

Research Paper No. 2009-04
January 2009

Specification Testing for Nonlinear Time Series with Long-Range Dependence

Jiti Gao, Qiying Wang and Jiying Yin

Specification Testing in Nonlinear Time Series with Long–Rang Dependence[1]

By Jiti Gao[†,∗],[2] Qiying Wang[‡] and Jiying Yin[†,∗]

The University of Adelaide[†], The University of Western Australia[∗] and The University of Sydney[‡]

Abstract

This paper proposes a model specification testing procedure for parametric specification of the conditional mean function in a nonlinear time series model with long–range dependence. An asymptotically normal test is established even when long–range dependence is involved. In order to implement the proposed test in practice using a simulated example, a bootstrap simulation procedure is established to find a simulated critical value to compute both the size and power values of the proposed test.

Keywords: Asymptotic theory, Gaussian process, nonlinear time series, long–range dependence, parametric specification.

1 Introduction

Consider a nonlinear time series model of the form

$$Y_t = m(X_t) + e_t, \ t = 1, 2, \cdots, n, \tag{1.1}$$

where $m(\cdot)$ is an unknown function over $I\!R = (-\infty, \infty)$, $\{X_t\}$ is a sequence of strictly stationary time series regressors, and $\{e_t\}$ is a sequence of strictly stationary time series errors with $E[e_1] = 0$ and $0 < E[e_1^2] < \infty$.

[1]The authors also acknowledge the financial support from the Australian Research Council Discovery Grants Program.

[2]Correspondence: Jiti Gao, School of Economics, The University of Adelaide, Adelaide SA 5005, Australia. Email: jiti.gao@adelaide.edu.au.

Both nonparametric estimation and parametric specification of $m(\cdot)$ have been discussed extensively in the literature for the case where both X_t and e_t are strictly stationary and short–range dependent time series. Such results may be found in the recent monographs by Fan and Yao (2003), Gao (2007), and Li and Racine (2007) for example. For the case where both X_t and e_t are strictly stationary and long–range dependent time series, estimation of $m(\cdot)$ has also been quite active during the last ten years or so. See for example, Beran (1994), Cheng and Robinson (1994), Hidalgo (1997), Robinson (1997), Beran and Ghosh (1998), Csörgó and Mielniczuk (1999), Gao and Anh (1999), Mielniczuk and Wu (2004), Gao (2007), and others.

By contrast, there has been little work done on parametric specification testing of $m(\cdot)$ for the case where either X_t, or e_t or both may be strictly stationary and long–range dependent time series. To the best of our knowledge, the only available work is given by Gao and Wang (2006), who consider a parametric specification of $m(\cdot)$ for the case where $\{X_t\}$ is a sequence of fixed designs while $\{e_t\}$ is a sequence of strictly stationary and long–range dependent time series errors.

This paper considers the case where the regressors X_t may exhibit some kind of long–range dependence (LRD). In the detailed discussion, we consider the case where $e_t = \sigma(X_t)\epsilon_t$, in which $\sigma(\cdot) > 0$ is an unknown function and $\{\epsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = 1$. In addition, $\{X_s : s \geq 1\}$ and $\{\epsilon_t : t \geq 1\}$ are assumed to be mutually independent. In order to clearly present both the main ideas and the key results without involving too much technicality, we assume that $\{X_t\}$ is a sequence of stationary Gaussian regressors. In Section 5 below, moreover, we point out that the case where $\{e_t\}$ is a sequence of martingale differences and $\{X_t\}$ is a sequence of strictly stationary and long–range dependent regressors may be discussed similarly.

The main interest of this paper is to consider specifying the conditional mean function while allowing the conditional variance function to be flexible. This is often the case where interest is on estimation and testing of the conditional mean function $m(x) = E[Y_t|X_t = x]$. We are thus interested in testing

$$H_0: \ m(x) = m_{\theta_0}(x) \quad \text{versus} \quad H_1: \ m(x) = m_{\theta_0}(x) + \Delta_n(x) \tag{1.2}$$

for all $x \in I\!R$, where θ_0 is a vector of unknown parameters, $m_\theta(x)$ is a known parametric function of x indexed by a vector of unknown parameters, θ, and $\{\Delta_n(x)\}$ is a sequence of unknown functions such that $\lim_{n\to\infty} \sup_{x\in I\!R} |\Delta_n(x)| = 0$. More detailed discussion and specification of $\Delta_n(x)$ is given in Section 4 below.

In some other cases, interest may be on assessing and specifying the conditional variance function $\sigma^2(\cdot)$. In such cases, parametric specification of $\sigma(\cdot)$ is an important issue in both theory and applications. Section 3 will discuss such specification issues.

The organization of this paper is as follows. Section 2 proposes a nonparametric test for (1.2) and then establishes asymptotic properties of the proposed test. Section 3 discusses some extensions. Both a bootstrap simulation procedure and its implementation in an example are given in Section 4. Section 5 concludes the paper with some remarks. The proofs of the main results are given in Appendix A.

2 Asymptotic theory

This section proposes a nonparametric test for the hypotheses (1.2) and then establishes an asymptotic distribution for the proposed test in Theorems 1 and 2 below. Their proofs, along with other proofs, are relegated to Appendix A.

Since $m(x)$ under H_1 is semiparametric, we need to establish a nonparametric or semiparametric test for (1.2). As discussed in the literature, several forms have been proposed to test (1.2) for the case where $\{(X_t, \epsilon_t) : t \geq 1\}$ is a sequence of either independent or strictly stationary short–range dependent variables.

Under H_0, the true model becomes

$$Y_t = m_{\theta_0}(X_t) + e_t \tag{2.1}$$

with $E[e_t|X_t] = 0$ under $\mathcal{H}_0$. We thus have

$$E\left[e_t E\left(e_t|X_t\right) f(X_t)\right] = E\left[\left(E^2(e_t|X_t)\right) f(X_t)\right] = 0 \tag{2.2}$$

under H_0, where $\{f(\cdot)\}$ is the marginal density function of $\{X_t\}$.

As suggested by Zheng (1996) for the independent sample case, we propose using a normalized kernel–based sample analogue of (2.2) of the form

$$\widehat{L}_n(h) = \frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n \widehat{e}_s \ a_n(X_s, X_t) \ \widehat{e}_t}{\sqrt{2 \sum_{t=1}^n \sum_{s=1,\neq t}^n \widehat{e}_s^2 \ a_n^2(X_s, X_t) \ \widehat{e}_t^2}}, \tag{2.3}$$

where $\widehat{e}_t = Y_t - m_{\widehat{\theta}}(X_t)$ and $a_n(X_s, X_t) = K\left(\frac{X_s - X_t}{h}\right)$, in which $\widehat{\theta}$ is a consistent estimator of θ_0 under H_0, $K(\cdot)$ is a probability kernel function and h is a bandwidth parameter.

It should be pointed out that several different classes of nonparametric and semiparametric tests have been proposed to deal with this kind of parametric specification testing issues. A recent literature survey in the field of model specification testing is given in Chapter 3 of Gao (2007) (see the references therein and other related references, such as Biedermann and Dette 2000).

As discussed in existing studies (such as Zheng 1996; Li and Wang 1998; Li 1999; Fan and Li 2000; Fan and Linton 2003; Arapis and Gao 2006; Gao 2007), a test statistic of the form (2.3) has a main advantage over its competitors in the situation that an indirect estimator of $\sigma^2(\cdot)$ is used to replace $\sigma^2(\cdot)$. Such feature is particularly attractive when the conditional variance function $\sigma^2(\cdot)$ as assumed in this paper is unknown nonparametrically.

It may be shown that the leading term of $\widehat{L}_n(h)$ under H_0 is given by

$$L_n(h) = \frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n e_s \ a_n(X_s, X_t) \ e_t}{\sigma_n} = \frac{\sum_{t=1}^n \sum_{s=1,\neq t}^n \epsilon_s \ K_n(X_s, X_t) \ \epsilon_t}{\sigma_n}, \tag{2.4}$$

where $K_n(X_s, X_t) = \sigma(X_s) K\left(\frac{X_s - X_t}{h}\right) \sigma(X_t)$ and $\sigma_n^2 = \frac{n^2 h}{\sqrt{\pi}} E\left[\sigma^4\left(\frac{X_1}{\sqrt{2}}\right)\right] \int_{-\infty}^{\infty} K^2(x) dx$. In order to show that $\widehat{L}_n(h)$ is an asymptotically consistent test, we need to establish an asymptotic distribution for $L_n(h)$ under the following assumptions.

Assumption 2.1. (i) $\{\epsilon_t, \, t \geq 1\}$ is a sequence of non–degenerate independent and identically distributed (i.i.d.) random errors with $E[\epsilon_1] = 0$, $E[\epsilon_1^2] = 1$ and $E[\epsilon_1^4] < \infty$. (ii) $\{X_t, \, t \geq 1\}$ is a stationary Gaussian sequence with $E[X_1] = 0$, $E[X_1^2] = 1$ and the covariance structure $\gamma(s-t) = E\left[X_s X_t\right]$ satisfying $\gamma(\tau) = |\tau|^{-\alpha} l(|\tau|) < 1$ for $\tau \geq 1$, where $0 < \alpha < 1$ and $l(x)$ is a positive function slowly varying at ∞. (iii) $\{\epsilon_t, t \geq 1\}$ is independent of $\{X_s, s \geq 1\}$.

Assumption 2.2. (i) $K(x)$ is a positive symmetric function satisfying $\int_{-\infty}^{\infty} K(x) = 1$ and $\sup_x K(x) < \infty$. (ii) $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh = \infty$.

Assumption 2.3. $\sigma(x)$ is a positive continuous function satisfying $\sigma(x) \leq C_0\left(|x|^\beta + 1\right)$ for some $\beta \geq 0$ and constant $C_0 > 0$.

The first result of this paper is given as follows; its proof is given in Appendix A.

THEOREM 1 *Under Assumptions 2.1–2.3, we have as $n \to \infty$*

$$L_n(h) \ \to_D \ N(0,1). \tag{2.5}$$

It is interesting to notice that the limit behavior in (2.5) does not depend on α involved in Assumption 2.1. As shown in Appendix A below, the asymptotic distribution of the stochastically normalized form $L_n(h)$ mainly depends on the probabilistic structure of $\{\epsilon_t\}$. In other words, the probabilistic structure of the stationary regressors $\{X_t\}$ does not affect the asymptotic distribution of $L_n(h)$. Due to the Gaussian assumption in Assumption 2.1(ii), the rest of the assumptions become probably the minimum conditions in this kind of problem. As shown in Section 5 below, some additional conditions on the joint density functions of (X_i, X_j), (X_i, X_j, X_k) and (X_i, X_j, X_k, X_l) are needed when the Gaussianity assumption is relaxed.

While the asymptotic normality in (2.5) is not unexpected, its proof cannot be derived directly using existing results for central limit theorems for quadratic forms of long–range dependent time series as discussed in Fox and Taqqu (1987), Avram (1988), Giraitis and Surgailis (1990), Giraitis and Taqqu (1997), Ho and Hsing (1996, 1997, 2003), Gao and King (2004), Hsing and Wu (2004), Gao and Wang (2006) and others. We therefore believe that the asymptotic normality result in (2.5) is a kind of extension of such existing results for the case where the random coefficient functions $K_n(\cdot,\cdot)$ reduce to a sequence of real numbers.

In addition to Assumptions 2.1–2.3, we need Assumption 2.4 below to establish an asymptotic distribution for $\widehat{L}_n(h)$.

Assumption 2.4 (i) Under the null hypothesis H_0, there is a sequence of positive real numbers η_n satisfying $\eta_n \to 0$ as $n \to \infty$ such that $||\widehat{\theta} - \theta_0|| = o_P(\eta_n)$, where $||\cdot||$

denotes the Euclidean norm. (ii) There exists some $\varepsilon_0 > 0$ such that $\frac{\partial m_\theta^2(x)}{\partial \theta^2}$ is continuous in both $x \in R$ and $\theta \in \Theta_0$, where $\Theta_0 = \{\theta : ||\theta - \theta_0|| \leq \varepsilon_0\}$. (iii)

$$\left|\left|\frac{\partial m_\theta(x)}{\partial \theta}|_{\theta=\theta_0}\right|\right| + \left|\left|\frac{\partial^2 m_\theta(x)}{\partial \theta^2}|_{\theta=\theta_0}\right|\right| \leq C(1 + |x|^{\beta_1})$$

for some constants $\beta_1 > 0$ and $C > 0$.

THEOREM 2 *Let Assumptions 2.1–2.4 hold. If, in addition,* $n\sqrt{h}\eta_n^2 = O(1)$ *holds, where* η_n *is defined as in Assumption 2.4, then under the null hypothesis* H_0

$$\widehat{L}_n(h) \to_D N(0, 1), \qquad \textit{as } n \to \infty. \tag{2.6}$$

The proof of (2.6) is given in Appendix A. Both Theorems 1 and 2 show that asymptotic normality can still be the limiting distribution of such a test even when the process involved is long–range dependent. On the technical side, the condition that $n\sqrt{h}\eta_n^2 = O(1)$ makes a linkage between the rate of $h \to 0$ and the rate of $\widehat{\theta}$ converging to θ_0. This condition holds automatically under the conventional rate of $\eta_n = O(n^{-1/2})$, since $h \to 0$.

It is noted that Zhao and Wu (2008) have investigated the confidence bands for nonparametric estimates of $\mu(x)$ and $\sigma(x)$ in the model:

$$Y_t = \mu(X_t) + \sigma(X_t)\epsilon_t, \quad t = 1, 2, ..., n.$$

The results in Zhao and Wu (2008) might be useful in constructing a test statistic to simultaneously test whether μ and σ are of certain parametric forms, but does not provide a straightforward routine as proposed in this paper.

3 Extensions and other models

This section discusses several extensions of model (1.1) to the following cases.

3.1 Parametric specification of the conditional variance

As briefly mentioned in the introduction, it is also of interest to test

$$H_{01}: \ \sigma(x) = \sigma_{\vartheta_0}(x) \ \text{ and } \ H_{11}: \ \sigma(x) = \sigma_{\vartheta_0}(x) + \Delta_{1n}(x) \tag{3.1}$$

for all $x \in I\!R$, where $\Delta_{1n}(\cdot)$ defined similarly to $\Delta_n(\cdot)$ is chosen such that $\inf_{x\in I\!R} \sigma(x) > 0$ under H_{11}.

In this case, we may consider a transformed model of the form

$$\log(Y_t - m(X_t))^2 = \log(\sigma^2(X_t)) + \log(\epsilon_t^2) = \mu + \log(\sigma^2(X_t)) + \eta_t, \tag{3.2}$$

where $\mu = E\left[\log(\epsilon_t^2)\right]$ and $\eta_t = \log(\epsilon_t^2) - E\left[\log(\epsilon_t^2)\right]$.

We then estimate $m(\cdot)$ either nonparametrically by $\widehat{m}(\cdot)$ or parametrically by $m_{\widehat{\theta}}(\cdot)$ when H_{01} holds, the corresponding test for H_{01} may be constructed based on the following approximate model

$$Z_t = \mu + \log(\sigma^2(X_t)) + \eta_t, \tag{3.3}$$

where $Z_t = \log(Y_t - \widehat{m}(X_t))^2$ or $\log(Y_t - m_{\widehat{\theta}}(X_t))^2$.

To test H_{01}, the test $\widehat{L}_n(h)$ is still applicable with $\widehat{e}_t$ being modified as $\widehat{e}_t = Z_t - \widehat{\mu} - \log(\sigma^2_{\widehat{\vartheta}}(X_t))$, in which $\widehat{\mu}$ and $\widehat{\vartheta}$ are the respective consistent estimators of μ and ϑ_0 under H_{01}.

3.2 Additive model specification testing

In both theory and practice, we will need to consider the case of $X_t = (X_{t1}, \cdots, X_{td})^\tau$. In this case, $\{X_t\}$ involved in (1.1) is a vector of d–dimensional regressors, we may consider a hypothesis problem of the form

$$H_{02}:\ m(x) = \sum_{i=1}^{d} m_{i\theta_0}(x_i) \quad \text{versus} \quad H_{12}:\ m(x) = \sum_{i=1}^{d} m_{i\theta_0}(x_i) + \sum_{i=1}^{d} \Lambda_{in}(x_i) \tag{3.4}$$

for all $x = (x_1, \cdots, x_d)^\tau \in I\!R^d$, where each $m_{i\theta_0}(\cdot)$ is a known function indexed by θ_0, and $\{\Lambda_{in}(\cdot)\}$ is a sequence of unknown functions over $I\!R$.

In this case, the test $\widehat{L}_n(h)$ is also applicable with $\widehat{e}_t$ being modified as $\widehat{e}_t = Y_t - \sum_{i=1}^d m_{i\widehat{\theta}}(X_{ti})$. With additional conditions, the conclusion of Theorem 2 remains true.

4 Simulation and an example of implementation

This section proposes a simulation scheme to deal with the choice of both a simulated critical value and a suitable bandwidth parameter for the implementation of the test.

An example of implementation is then given to show how practically both the theory and the simulation procedure may be realized.

To study the power function of $\widehat{L}_n(h)$, we need to discuss about how to estimate $\Delta_n(x)$. Under H_1, model (1.1) becomes

$$Y_t = m(X_t) + e_t = m_{\theta_1}(X_t) + \Delta_n(X_t) + \sigma(X_t)\epsilon_t. \tag{4.1}$$

We apply a semiparametric estimation method (see, for example, Chapter 2 of Gao 2007) to estimate θ_1 by minimizing $\sum_{t=1}^n \left(Y_t - m_{\theta_1}(X_t) - \widetilde{\Delta}_n(X_t, \theta_1)\right)^2$ over θ_1, where $\widetilde{\Delta}_n(x, \theta_1) = \frac{\sum_{t=1}^n K\left(\frac{x - X_t}{\widehat{h}_{\mathrm{cv}}}\right)\left(Y_t - m_{\theta_1}(X_t)\right)}{K\left(\frac{x-X_t}{\widehat{h}_{\mathrm{cv}}}\right)}$, in which $\widehat{h}_{\mathrm{cv}}$ is chosen by a conventional cross–validation estimation method. We then estimate $\Delta_n(x)$ by $\widehat{\Delta}_n(x) = \widetilde{\Delta}_n(x, \widehat{\theta}_1)$.

Under certain conditions, it may be shown that $\lim_{n\to\infty} \sup_{x \in I\!R} \frac{\widehat{\Delta}_n(x)}{\Delta_n(x)} = 1$ and also that $\widehat{\theta}_1$ is asymptotically consistent to θ_1. Since this is a totally new topic in this kind of model specification problem, detailed discussion about suitable conditions required for the establishment of the asymptotic consistency and a rigorous proof is extremely technical. We therefore wish to leave such theoretical discussion for future research. In Example 4.1 below, we apply this estimation method for the practical implementation.

To propose the following simulation procedure, we need to introduce the following notation. Define

$$\epsilon_t^* = \overline{\epsilon}\, \eta_t, \quad \text{where } \overline{\epsilon} = \frac{1}{n}\sum_{t=1}^n \widehat{\epsilon}_t, \quad \widehat{\epsilon}_t = \frac{\left(Y_t - m_{\widehat{\theta}}(X_t)\right) I[\widehat{\sigma}_n(X_t) > 0]}{\widehat{\sigma}_n(X_t)}, \tag{4.2}$$

$$\widehat{\sigma}_n^2(x) = \frac{\sum_{t=1}^n K\left(\frac{x-X_t}{\widehat{h}_{\mathrm{cv}}}\right)\left(Y_t - m_{\widehat{\theta}}(X_t)\right)^2}{\sum_{t=1}^n K\left(\frac{x-X_t}{\widehat{h}_{\mathrm{cv}}}\right)}, \tag{4.3}$$

in which $\widehat{h}_{\mathrm{cv}}$ is chosen by a conventional cross–validation estimation method and $\{\eta_t\}$ is a sequence of i.i.d. random variables with $E[\eta_t] = 0$ and $E[\eta_t^i] = 1$ for $i = 2, 3$. In addition, we require $\epsilon_t^{*i} = \overline{\epsilon}_i\, \eta_t^i$ with $\overline{\epsilon}_i = \frac{1}{n}\sum_{t=1}^n \widehat{\epsilon}_t^i$ for $i = 2, 3$. It is noted that the choice of $\{\eta_t\}$ is not essential in the theoretical study of this paper. In practice, we choose the distribution of $\{\eta_t\}$ as follows: $P\left(\eta_1 = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $P\left(\eta_1 = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}$. Such two–point distributional structure has been used in the literature (see, for example, Li and Wang 1998).

4.1 Simulation scheme

Let l_r $(0 < r < 1)$ be the $1 - r$ quantile of the exact finite–sample distribution of $\widehat{L}_n(h)$. Because l_r may not be evaluated in practice, we suggest an approximate r–level critical value l_r^* to replace it by using the following bootstrap procedure:

- Generate $Y_t^* = m_{\widehat{\theta}}(X_t^*) + \widehat{\sigma}_n(X_t^*)\ \epsilon_t^*$ for $1 \leq t \leq n$, where $\{\epsilon_t^*\}$ is chosen as in (4.2) above, and $\{X_t^*\}$ is a sequence of stationary Gaussian regressors drawn from a stationary LRD Gaussian process with the covariance structure being given by $\gamma_{\widehat{\alpha}}(\tau) = \widehat{l}(|\tau|)\ |\tau|^{-\widehat{\alpha}}$, in which $\widehat{\alpha}$ and $\widehat{l}$ may be constructed using a spectral density estimation method (such as, Robinson 1995), and $\widehat{\sigma}_n(x)$ is give by (4.2) above.

- Use the data set $\{(X_t^*, Y_t^*) : 1 \leq t \leq n\}$ to estimate $\widehat{\theta}$ by $\widehat{\theta}^*$ and to compute $\widehat{L}_n^*(h)$, where $\widehat{L}_n^*(h)$ is the corresponding version of $\widehat{L}_n(h)$ under H_0 with $\{(X_t, Y_t) : 1 \leq t \leq n\}$ and $(\theta_0, \widehat{\theta})$ being replaced by $\{(X_t^*, Y_t^*) : 1 \leq t \leq n\}$ and $(\widehat{\theta}, \widehat{\theta}^*)$.

- Repeat the above step M times and produce M versions of $\widehat{L}_n^*(h)$ denoted by $\widehat{L}_{n,m}^*(h)$ for $m = 1, 2, \ldots, M$. Use the M values of $\widehat{L}_{n,m}^*(h)$ to construct their empirical distribution function. The bootstrap distribution of $\widehat{L}_n^*(h)$ given $\mathcal{W}_n = (X_1, \cdots, X_n; Y_1, \cdots, Y_n)$ is defined by $P^*\left(\widehat{L}_n^*(h) \leq x\right) = P\left(\widehat{L}_n^*(h) \leq x | \mathcal{W}_n\right)$. Then let l_r^* $(0 < r < 1)$ satisfy $P^*\left(\widehat{L}_n^*(h) \geq l_r^*\right) = r$ and estimate l_r by l_r^*.

It is pointed out that $l_r^* = l_r^*(h)$ is a function of h. A critical problem raised in the implementation of the proposed test is the choice of a suitable bandwidth h. To solve this problem, define the size and power functions of $\widehat{L}_n^*(h)$ as

$$\gamma_n(h) = P^*\left(\widehat{L}_n^*(h) > l_r^* | H_0 \text{ true}\right) \text{ and } \beta_n(h) = P^*\left(\widehat{L}_n^*(h) > l_r^* | H_0 \text{ false}\right). \tag{4.4}$$

Let $H_n = \{h : \ r - \varepsilon_0 < \gamma_n(h) < r + \varepsilon_0\}$ and define $\widehat{h}_{\text{gwy}}$ such that $\beta_n(\widehat{h}_{\text{gwy}}) = \max_{h \in H_n} \beta_n(h)$, where $0 < \varepsilon_0 < r$ is some small constant.

In general, the issue of how to find $\widehat{h}_{\text{gwy}}$ theoretically has not been addressed in this kind of long–range dependent time series case. Since the regressors $\{X_t\}$ are still stationary, the theory and methodology developed in Gao and Gijbels (2008) for the stationary time series case may still be applicable (see also Chapter 3 of Gao 2007). In the following example, we therefore propose using the leading term of an asymptotically approximated version of $\widehat{h}_{\mathrm{gwy}}$ of the form

$$\widehat{h}_{\mathrm{gwy}} = \widehat{a}(K)^{-\frac{1}{2}} \widehat{t_n}^{-\frac{3}{2}}, \tag{4.5}$$

where $\widehat{t}_n = n\ \widehat{C}_n^2$, $\widehat{C}_n^2 = \frac{\frac{1}{n}\sum_{t=1}^n \widehat{\Delta}_n^2(X_t)\widehat{f}(X_t)}{\widehat{\sigma}_0^2\sqrt{2\widehat{\nu}_0\int K^2(v)dv}}$, $\widehat{a}(K) = \frac{\sqrt{2}K^{(3)}(0)}{3\left(\sqrt{\int K^2(u)du}\right)^3}\ \widehat{c}(f)$ and $\widehat{c}(f) = \frac{1}{n}\sum_{t=1}^n \widehat{f}^2(X_t)\widehat{\sigma}_n^6(X_t)\cdot\left(\frac{1}{n}\sum_{t=1}^n \widehat{f}(X_t)\widehat{\sigma}_n^4(X_t)\right)^{-\frac{3}{2}}$, in which $K^{(3)}(\cdot)$ is the three–time convolution of $K(\cdot)$ with itself, $\widehat{\sigma}_0^2 = \frac{1}{n}\sum_{t=1}^n \widehat{\sigma}_n^2(X_t)$, $\widehat{\nu}_0 = \frac{1}{n}\sum_{t=1}^n \widehat{f}^2(X_t)$ and $\widehat{f}(x) = \frac{1}{n\widehat{h}_{\mathrm{cv}}}\sum_{t=1}^n K\left(\frac{x-X_t}{\widehat{h}_{\mathrm{cv}}}\right)$ is the conventional nonparametric kernel density estimate.

4.2 An example of implementation

Example 4.1 Consider a linear model of the form

$$Y_t = \theta_0 + \theta_1 X_t + \sigma_0\ \sqrt{1+X_t^2}\ \epsilon_t, \tag{4.6}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. observations sampled from either $N(0,1)$ or a normalized χ^2 distribution of the form $\frac{\chi_2^2-2}{2}$, and $\{X_t\}$ is a sequence of stationary Gaussian regressors with $E[X_1] = E[X_1^2] = 1$ and $E[X_s X_t] = \gamma(s-t)$ for $s \neq t$ with $\gamma(k) = \eta|k|^{-\alpha}$ for $k = \pm 1, \pm 2, \cdots$. The true values involved in (4.6) are $\theta_0 = \theta_1 = 1$, $\sigma_0 = 1$ and $\alpha = \eta = 0.5$.

To compute the sizes of the test, generate $\{Y_t\}$ from

$$H_0:\ Y_t = 1 + X_t + \sqrt{1+X_t^2}\ \epsilon_t. \tag{4.7}$$

To generate the data under H_1, we consider the case of $\Delta_n(x) = c_n\ \Delta(x)$ in (1.2) and generate $\{Y_t\}$ from

$$H_1:\ Y_t = 1 + X_t + c_n\ X_t^2 + \sqrt{1+X_t^2}\ \epsilon_t, \tag{4.8}$$

with $\Delta(x) = x^2$ and $c_n = d_{1n} = n^{-\frac{1}{2}}\sqrt{\log\log(n)}$ or $c_n = d_{2n} = n^{-\frac{13}{30}}$.

The reasoning for the choice of c_n is as follows. The rate of $d_{1n} = n^{-\frac{1}{2}}\sqrt{\mathrm{loglog}(n)}$ is the optimal rate of testing in this kind of nonparametric kernel testing problem as discussed in Horowitz and Spokoiny (2001). The rate of $d_{2n} = n^{-\frac{13}{30}}$ implies that the optimal bandwidth $\widehat{h}_{\mathrm{gwy}}$ is proportional to $n^{-\frac{1}{5}}$.

We choose the standard Normal kernel for $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ in the implementation. Let z_r be the $1-r$ quantile of the standard Normal distribution. Note that $z_{0.01} = 2.33$ at the 1% level, $z_{0.05} = 1.645$ at the 5% level and $z_{0.10} = 1.28$ at the 10% level.

For $i, j = 1, 2$, let $\widehat{h}_{\rm igwy}(j)$ denote $\widehat{h}_{\rm gwy}(j)$ corresponding to d_{in} for either the case of $\epsilon_t \sim N(0,1)$ (with $j=1$) or the case of $\epsilon_t \sim \frac{\chi_2^2 - 2}{2}$ (with $j=2$), $L_{\rm igwy}(j) = \widehat{L}_n(\widehat{h}_{\rm igwy}(j))$, $l_{ir}^*(j) = l_{ir}^*(\widehat{h}_{\rm igwy}(j))$, $L_{\rm cv}(j) = \widehat{L}_n(\widehat{h}_{\rm cv}(j))$ under H_0, and $L_{i{\rm cv}}(j) = \widehat{L}_n(\widehat{h}_{\rm cv}(j))$ corresponding to d_{in} for $i, j = 1, 2$ under H_1, where $\widehat{h}_{cv}$ is chosen such that

$$\widehat{h}_{cv} = \arg \min_{h \in H_{\rm cv}} \frac{1}{n} \sum_{t=1}^{n} \left(Y_t - \widehat{m}_{-t}(X_t, h)\right)^2, \tag{4.9}$$

in which $\widehat{m}_{-t}(X_t, h) = \frac{\sum_{s=1, \neq t}^{n} K\left(\frac{X_s - X_t}{h}\right) Y_s}{\sum_{u=1, \neq t}^{n} K\left(\frac{X_u - X_t}{h}\right)}$ and $H_{\rm cv} = \left[n^{-1}, n^{-(1-\delta_0)}\right]$ for $0 < \delta_0 < 1$.

In Tables 4.1–4.6 below, we use $N = 250$ as the number of the bootstrap resamples and $M = 500$ as the number of replications. For $i, j = 1, 2$, let $f_{\rm igwy}(j)$ denote the frequency of $L_{\rm igwy}(j) > l_{ir}^*(j)$, $f_{\rm cv}(j)$ be the frequency of $L_{\rm cv}(j) > z_r$ under H_0, and $f_{i{\rm cv}}(j)$ be the frequency of $L_{i{\rm cv}}(j) > z_r$ under H_1 for $r = 1\%$, 5% or 10%.

Table 4.1. Rejection Rates for Testing the Conditional Mean at the 1% level

Observation	Three Versions of the Test							
	Null Hypothesis Is True							
n	$f_{\rm cv}(1)$		$f_{\rm cv}(2)$		$f_{\rm 1gwy}(1)$	$f_{\rm 1gwy}(2)$	$f_{\rm 2gwy}(1)$	$f_{\rm 2gwy}(2)$
250	0.002		0.008		0.014	0.010	0.022	0.010
500	0.014		0.008		0.020	0.014	0.014	0.012
750	0.016		0.002		0.016	0.006	0.016	0.006
	Null Hypothesis Is False							
n	$f_{\rm 1cv}(1)$	$f_{\rm 1cv}(2)$	$f_{\rm 2cv}(1)$	$f_{\rm 2cv}(2)$	$f_{\rm 1gwy}(1)$	$f_{\rm 1gwy}(2)$	$f_{\rm 2gwy}(1)$	$f_{\rm 2gwy}(2)$
250	0.080	0.090	0.088	0.088	0.182	0.162	0.120	0.126
500	0.086	0.094	0.080	0.094	0.198	0.194	0.138	0.132
750	0.102	0.092	0.102	0.090	0.212	0.256	0.170	0.180

Table 4.2. Rejection Rates for Testing the Conditional Mean at the 5% level

Observation	Three Versions of the Test							
Null Hypothesis Is True								
n	$f_{\mathrm{cv}}(1)$		$f_{\mathrm{cv}}(2)$		$f_{1\mathrm{gwy}}(1)$	$f_{1\mathrm{gwy}}(2)$	$f_{2\mathrm{gwy}}(1)$	$f_{2\mathrm{gwy}}(2)$
250	0.028		0.026		0.072	0.066	0.078	0.052
500	0.016		0.022		0.054	0.046	0.054	0.058
750	0.018		0.032		0.046	0.058	0.056	0.056
Null Hypothesis Is False								
n	$f_{1\mathrm{cv}}(1)$	$f_{1\mathrm{cv}}(2)$	$f_{2\mathrm{cv}}(1)$	$f_{2\mathrm{cv}}(2)$	$f_{1\mathrm{gwy}}(1)$	$f_{1\mathrm{gwy}}(2)$	$f_{2\mathrm{gwy}}(1)$	$f_{2\mathrm{gwy}}(2)$
250	0.128	0.154	0.128	0.142	0.276	0.304	0.214	0.224
500	0.124	0.112	0.116	0.126	0.284	0.320	0.216	0.248
750	0.152	0.152	0.162	0.162	0.382	0.352	0.304	0.286

Table 4.3. Rejection Rates for Testing the Conditional Mean at the 10% level

Observation	Three Versions of the Test							
Null Hypothesis Is True								
n	$f_{\mathrm{cv}}(1)$		$f_{\mathrm{cv}}(2)$		$f_{1\mathrm{gwy}}(1)$	$f_{1\mathrm{gwy}}(2)$	$f_{2\mathrm{gwy}}(1)$	$f_{2\mathrm{gwy}}(2)$
250	0.038		0.024		0.080	0.116	0.080	0.114
500	0.046		0.034		0.104	0.078	0.102	0.084
750	0.030		0.038		0.106	0.098	0.106	0.090
Null Hypothesis Is False								
n	$f_{1\mathrm{cv}}(1)$	$f_{1\mathrm{cv}}(2)$	$f_{2\mathrm{cv}}(1)$	$f_{2\mathrm{cv}}(2)$	$f_{1\mathrm{gwy}}(1)$	$f_{1\mathrm{gwy}}(2)$	$f_{2\mathrm{gwy}}(1)$	$f_{2\mathrm{gwy}}(2)$
250	0.164	0.154	0.180	0.152	0.380	0.368	0.328	0.284
500	0.162	0.134	0.160	0.148	0.430	0.354	0.354	0.284
750	0.176	0.166	0.162	0.172	0.408	0.452	0.332	0.364

Tables 4.1–4.3 show that there is some size distortion when using $\widehat{h}_{\mathrm{cv}}$ and z_r in practice. The size performance may be significantly improved when using the simulated

critical value $l_r^*(\widehat{h}_{\rm gwy})$ associated with the power–based $\widehat{h}_{\rm gwy}$. As expected from the theory, the test associated with $\widehat{h}_{\rm gwy}$ is more powerful than that based on $\widehat{h}_{\rm cv}$. In addition, $\widehat{L}_n(\widehat{h}_{\rm 1gwy})$ corresponding to d_{1n} is more powerful than that of $\widehat{L}_n(\widehat{h}_{\rm 2gwy})$ corresponding to d_{2n} while their sizes are comparable. This is not surprising, because d_{1n} has been shown to be the optimum rate for this kind of nonparametric testing (see, for example, Horowitz and Spokoiny 2001; Chapter 3 of Gao 2007).

In addition, Tables 4.1–4.3 also show that the choice of the distribution of $\{\epsilon_t\}$ has little impact on the simulated sizes and power values. This may show that robustness of the proposed bootstrap simulation procedure.

In summary, our small and medium–sample studies in the simulated example have shown that the use of an asymptotically normal test associated with a cross–validation estimation–based bandwidth may not make such a test practically applicable due to poor size and power properties. However, the performance of such a test can be significantly improved when it is coupled with a power–based optimal bandwidth as well as a bootstrap simulated critical value.

5 Conclusions and discussion

This paper has considered a class of nonlinear time series models with possible LRD in the regressors. A simple kernel test has been proposed and then studied both theoretically and practically. The small and medium–sample studies have shown that both the theory and the simulation procedure work well.

As briefly mentioned in the introductory section, the assumptions on X_t and e_t may be relaxed. For the error part, it is possible to show that Theorems 1 and 2 remain true when $\{e_t\}$ is a sequence of martingale differences.

For the regressor case, we may allow $\{X_t\}$ to be a sequence of strictly stationary and long–range dependent regressors. In this case, we need to introduce the following additional assumption.

Assumption 5.1. (i) $\{X_t, t \geq 1\}$ is a sequence of strictly stationary and long–range dependent regressors with $E[X_1] = 0$, $E[X_1^2] = 1$ and the covariance structure

$\gamma(s-t) = E[X_s X_t]$ satisfying $\gamma(\tau) = |\tau|^{-\alpha} l(|\tau|) < 1$ for $\tau \geq 1$, where $0 < \alpha < 1$ and $l(x)$ is a positive function slowly varying at ∞.

(ii) Let $f_{i,j,k,l}(\cdot)$ be the joint probability density of (X_i, X_j, X_k, X_l). Assume that all $f_{i,j,k,l}(\cdot,\cdot,\cdot,\cdot)$ are uniformly continuous.

(iii) In addition, for $m = 2$ or 4

$$\limsup_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \left\{ \int \sigma^{2m}(x) f_{(i,j)}(x,x) dx \right\} < \infty,$$

$$\limsup_{n\to\infty} \frac{1}{n^3} \sum_{i,j,k=1}^{n} \left\{ \int \sigma^{2m}(x) f_{(i,j,k)}(x,x,x) dx \right\} < \infty,$$

$$\limsup_{n\to\infty} \frac{1}{n^4} \sum_{i,j,k,l=1}^{n} \left\{ \int \sigma^{8}(x) f_{(i,j,k,l)}(x,x,x,x) dx \right\} < \infty,$$

$$\limsup_{n\to\infty} \frac{1}{n^4} \sum_{i,j,k,l=1}^{n} \left\{ \iint \sigma^{4}(x)\sigma^4(y) f_{(i,j,k,l)}(x,x,y,y) dxdy \right\} < \infty.$$

It is expected that the conclusion of Theorem 1 remains true when Assumptions 2.1(i)(iii)–2.3 and 5.1 hold, and that the conclusion of Theorem 2 also remains valid when Assumptions 2.1(i)(iii)–2.4 and 5.1 hold. Further rigorous proofs of such conclusions are left for future research.

When $\{X_t\}$ involved in model (1.1) is allowed to be a linear process, model (1.1) will have more practical applications. One of the special cases is a nonparametric autoregressive model when $X_t = Y_{t-1}$. In this case, we expect that such a model may be applicable to check whether the conditional mean function of a long–range dependent time series, such as the S&P 500 Index, may be parametrically specified. Such issues are also left for future research.

REFERENCES

Arapis, M. and Gao, J. (2006). Empirical comparisons in short–term interest rate models using nonparametric methods. *Journal of Financial Econometrics* **4** 310–345.

Avram, F. (1988). On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probability Theory and Related Fields* **79** 37–45.

Beran, J. (1994). *Statistics for Long–Memory Processes*. Chapman & Hall, London.

Beran, J. and Ghosh, S. (1998). Root–n–consistent estimation in partial linear models with long–memory errors. *Scandinavian Journal of Statistics* **25** 345–357.

Biedermann, S. and Dette, H. (2000). Testing linearity of regression models with dependent errors by kernel based methods. *Test* **9** 417–438.

Cheng, B. and Robinson, P. M. (1994). Semiparametric estimation from time series with long–range dependence. *Journal of Econometrics* **64** 335–353.

Csörgö, S. and Mielniczuk, J. (1999). Random-design regression under long–range dependence. *Bernoulli* **5** 209–224.

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

Fan, Y. and Li, Q. (2000). Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16** 1016–1041.

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

Fox, R. and Taqqu, M. S. (1987). Central limit theorems for quadratic forms in random variables having long–range dependence. *Probability Theory and Related Fields* **74** 213–240.

Gao, J. (2007). *Nonlinear Time Series*: semiparametric methods. Chapman & Hall/CRC.

Gao, J. and Anh, V. (1999). Semiparametric regression with long-range dependent error processes. *Journal of Statistical Planning & Inference* **80** 37–57.

Gao, J. and Gijbels, I. (2008). Bandwidth parameter selection in nonparametric kernel testing. Forthcoming in the *Journal of the American Statistical Association* (December issue).

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

Gao, J. and Wang, Q. (2006). Long–range dependent time series specification. Working paper available from the first author.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields* **86** 87–104.

Giraitis, L. and Taqqu, M. S. (1997). Limit theorems for bivariate Appell polynomials. I. Central limit theorems. *Probability Theory and Related Fields* **107** 359–381.

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

Hidalgo, F. J. (1997). Nonparametric estimation with strongly dependent multivariate time series. *Journal of Time Series Analysis* **8** 95–122.

Ho, H. C. and Hising, T. (1996). On the asymptotic expansion of the empirical processes of long–memory moving averages. *Annals of Statistics* **24** 992–1024.

Ho, H. C. and Hising, T. (1997). Limit theorems for functional of moving averages. *Annals of Probability* **25** 1636–1669.

Ho, H. C. and Hising, T. (2003). A decomposition for generalized U-statistics of long-memory linear processes. *Theory and Applications of Long–Range Dependence* 143–155, Birkhuser Boston, Boston, MA.

Horowitz, J. and Spokoiny, V. (2001). An adaptive rate–optimal test of a parametric mean–regression model against a nonparametric alternative. *Econometrica* **69** 599–631.

Hsing, T. and Wu, W. B. (2004). On weighted U–statistics for stationary processes. *Annals of Probability* **32** 1600–1631.

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

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

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

Mielniczuk, J. and Wu, W. B. (2004). On random–design model with dependent errors. *Statistica Sinica* **14** 1105–1126.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long–range dependence. *Annals of Statistics* **23** 1630–1661.

Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics* **25** 2054–2083.

Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *Annals of Statistics* **36** 1854–1878.

6 Appendix A

This appendix provides the full proofs of Theorems 1 and 2 as well as the necessary lemmas and their proofs. For notational simplicity, we denote all different i, j, k, l ($i \neq j,\ i \neq k,\ i \neq l,\ j \neq k,\ j \neq l,\ k \neq l$) simply by $i \neq j \neq k \neq l$, and constants by $C, C_1, ...$, which may have different values at each appearance. Also, $\{X_k, k \geq 1\}$ and $\{\epsilon_k, k \geq 1\}$ are assumed to satisfy Assumption 2.1. $h \to 0$ and $nh \to \infty$, as $n \to \infty$, as in Assumption 2.2(ii).

6.1 Technical lemmas

In addition to the notation in Section 2, define $\tau_m = \int_{-\infty}^{\infty} K^m(x)dx$, $m \geq 1$, $\gamma_{ij} = \gamma(i-j)$,

$$\Sigma = \begin{pmatrix} 1 & \gamma_{ij} & \gamma_{ik} & \gamma_{il} \\ \gamma_{ji} & 1 & \gamma_{jk} & \gamma_{jl} \\ \gamma_{ki} & \gamma_{kj} & 1 & \gamma_{kl} \\ \gamma_{li} & \gamma_{lj} & \gamma_{lk} & 1 \end{pmatrix},$$

and $\Lambda_n = \{(i, j, k, l) : |s - t| \geq \delta_n, \text{where } s, t = i, j, k \text{ or } l\}$ with some $\delta_n \to \infty$.

LEMMA 1 *(i) If $\tau_m < \infty$, then for all $i \neq j$,*

$$E\left[K_n^m(X_i, X_j)\right] = \frac{\tau_m\, h}{2\sqrt{\pi}[1-\gamma_{ij}]^{1/2}}\, E\left[\sigma^{2m}\left(\sqrt{[1+\gamma_{ij}]/2}X_1\right)\right]\{1+o(1)\}. \quad \text{(A.1)}$$

(ii) Assume $(i, j, k, l) \in \Lambda_n$. If $\tau_m < \infty$, then

$$E\left[K_n^m(X_i, X_j)\right] = \frac{\tau_m\, h}{2\sqrt{\pi}}\, E\left[\sigma^{2m}(X_1/\sqrt{2})\right]\{1+o(1)\}; \quad \text{(A.2)}$$

if $\tau_2 < \infty$, then

$$E\{K_n^2(X_i, X_j)\, K_n^2(X_k, X_l)\} = \frac{\tau_2^2\, h^2}{4\pi}\left(E\left[\sigma^4(X_1/\sqrt{2})\right]\right)^2\{1+o(1)\}. \quad \text{(A.3)}$$

(iii). Under Assumption 2.2, for all $i \neq j \neq k \neq l$ and all $m_1, m_2, m_3, m_4 \geq 0$, we have

$$E\Big\{(1+|X_i|^{m_1})(1+|X_j|^{m_2})K\left(\frac{X_i - X_j}{h}\right)\Big\} \;\leq\; C\,h, \tag{A.4}$$

$$E\Big\{(1+|X_i|^{m_1})(1+|X_j|^{m_2})(1+|X_k|^{m_3}) \times K\left(\frac{X_i - X_j}{h}\right)K\left(\frac{X_j - X_k}{h}\right)\Big\} \;\leq\; C\,h^2, \tag{A.5}$$

$$E\Big\{(1+|X_i|^{m_1})(1+|X_j|^{m_2})(1+|X_k|^{m_3})(1+|X_l|^{m_4}) K\left(\frac{X_i - X_k}{h}\right)K\left(\frac{X_j - X_k}{h}\right)K\left(\frac{X_i - X_l}{h}\right)K\left(\frac{X_j - X_l}{h}\right)\Big\} \;\leq\; C\,h^3, \tag{A.6}$$

where C is a constant depending only on $\max \gamma_{ij}$ and m_j.

Remark A.1. More detailed calculation shows that, for all $i \neq j \neq k \neq l$, we have

$$E\big\{K_n^2(X_i, X_j)\,K_n^2(X_k, X_l)\big\} \;\sim\; \frac{\tau_2^2\,h^2}{(2\pi)^2\,(\det\Sigma)^{1/2}} \iint \sigma^4(x)\sigma^4(y)e^{-\mu^\tau\Sigma^{-1}\mu/2}dxdy,$$

where $\mu^\tau = (x, x, y, y)$. We omit the details as (A.3) is sufficient for this paper.

Proof. We first prove (A.1). Write $\rho = \gamma_{ij}$. It is readily seen that, as $h \to 0$,

$$\begin{aligned}
E\left[K_n^m(X_i, X_j)\right] &= \frac{1}{2\pi(1-\rho^2)^{1/2}} \iint \sigma^m(x)\sigma^m(y)K^m\big[(x-y)/h\big]e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)} \\
&\times\; dxdy \text{ (letting } x = y_1 + hx_1, y = y_1 \text{ and simple reorganization)} \\
&= \frac{h}{2\pi(1-\rho^2)^{1/2}} \iint \sigma^m(y_1+hx_1)\sigma^m(y_1)K^m(x_1) \\
&\times\; e^{-\frac{1}{2(1-\rho^2)}\left[2(1-\rho)y_1^2+2h(1-\rho)x_1y_1+h^2x_1^2\right]}dx_1dy_1 \\
&= \frac{h}{2\pi(1-\rho^2)^{1/2}} \iint \sigma^m(y+hx)\sigma^m(y)K^m(x)e^{-\frac{(y+hx/2)^2}{1+\rho}-\frac{(3-\rho)h^2x^2}{4(1-\rho^2)}}dxdy \\
&= \frac{h}{2\pi(1-\rho^2)^{1/2}} \iint \sigma^m(y+3hx/2)\sigma^m(y+hx/2)K^m(x)e^{-\frac{y^2}{1+\rho}-\frac{(3-\rho)h^2x^2}{4(1-\rho^2)}} \\
&\times\; dxdy \\
&= \frac{h}{2\pi(1-\rho^2)^{1/2}} \int_{-\infty}^{\infty}\sigma^{2m}(y)e^{-\frac{y^2}{1+\rho}}dy \int_{-\infty}^{\infty}K^m(x)dx\,\{1+o(1)\} \\
&= \frac{\tau_m\,h}{2\sqrt{\pi}(1-\gamma_{ij})^{1/2}}E\left[\sigma^{2m}\left(X_1\sqrt{(1+\gamma_{ij})/2}\right)\right]\big\{1+o(1)\big\},
\end{aligned} \tag{A.7}$$

where, in the second step from below, we have used the dominate convergence theorem and the continuity of $\sigma(x)$. This proves (A.1).

By recalling $\gamma(k) = |k|^{-\alpha}l(|k|)$ and noting that, for $(i, j, k, l) \in \Lambda_n$

$$\gamma_{st} = |s-t|^{-\alpha}l(|s-t|) = o(\delta_n^{-\alpha/2}), \quad \text{where } s, t = i, j, k \text{ or } l, \tag{A.8}$$

(A.2) is obvious as $\sigma(x)$ is continuous and $\sigma(x) \le C_0\,(|x|^\beta + 1)$, for some $\beta > 0$.

We next prove (A.3). First note that, similarly to the proof of (A.7), as $h \to 0$,

$$\begin{aligned}
\frac{1}{\sqrt{\pi}} \iint K_n^2(x,y)\, e^{-(x^2+y^2)/2} dxdy &= \tau_2\, h\, E\left[\sigma^4(X_1/\sqrt{2})\right] \big\{1+o(1)\big\}, \\
\frac{1}{\sqrt{2\pi}} \iint K_n^2(x,y)\, e^{-(x^2+y^2)/4} dxdy &= \tau_2\, h\, E\left[\sigma^4(X_1)\right] \big\{1+o(1)\big\}.
\end{aligned}$$

These facts yield that, with $\mu^\tau = (x,y,s,t)$,

$$\begin{aligned}
I_n &:= \int\cdots\int K_n^2(x,y)\, K_n^2(s,t) e^{-\mu^\tau \mu/2} dxdydsdt \\
&= \pi\, \left(\tau_2\, h\, E\left[\sigma^4(X_1/\sqrt{2})\right]\right)^2 \big\{1+o(1)\big\}, \qquad (A.9) \\
I_{1n} &:= \int\cdots\int K_n^2(x,y)\, K_n^2(s,t) e^{-\mu^\tau \mu/4} dxdydsdt \\
&= 2\,\pi\, \left(\tau_2\, h\, E\left[\sigma^4(X_1)\right]\right)^2 \big\{1+o(1)\big\}. \qquad (A.10)
\end{aligned}$$

We are now ready to prove (A.3). By virtue of (A.8), we may rewrite Σ as $\Sigma = I + \delta_n^{-\alpha/2} D$, where I is an identity matrix of order 4 and maximum element of D is bounded by an absolute constant. This implies that $\det \Sigma \sim 1$ and there exists a matrix D_1 whose element may depend on h such that maximum element of D_1 is bounded by an absolute constant and as n large enough,

$$\Sigma^{-1} = I + \delta_n^{-\alpha/2} D_1. \qquad (A.11)$$

Recall $h \to 0$ as $n \to \infty$. It follows easily from (A.11) that, as n large enough,

$$\left|e^{-\mu^\tau \Sigma^{-1}\mu/2} - e^{-\mu^\tau \mu/2}\right| \le e^{-\mu^\tau \mu/2} \left|e^{-\delta_n^{-\alpha/2} \mu^\tau D \mu/2} - 1\right| \le \delta_n^{-\alpha/4}\, e^{-\mu^\tau \mu/4}, \qquad (A.12)$$

where $\mu^\tau = (x,y,s,t)$. This, together with (A.9) and (A.10), yields that

$$\begin{aligned}
& E\big\{K_n^2(X_i, X_j)\, K_n^2(X_k, X_l)\big\} \\
&= \frac{1}{(2\pi)^2\, (\det \Sigma)^{1/2}} \int\cdots\int K_n^2(x,y)\, K_n^2(s,t) e^{-\mu' \Sigma^{-1} \mu/2} dxdydsdt \\
&= \frac{1}{(2\pi)^2\, (\det \Sigma)^{1/2}} \left[I_n + O(\delta_n^{-\alpha/4})\, I_{1n}\right] \\
&= \frac{\tau_2^2\, h^2}{4\pi} \left(E\left[\sigma^4\left(X_1/\sqrt{2}\right)\right]\right)^2 \big\{1+o(1)\big\},
\end{aligned}$$

which implies (A.3).

We finally prove (A.6). The proofs of (A.4) and (A.5) are similar but simpler. Let $\mu^\tau = (x,y,s,t)$ as before. By virtue of $\max \gamma_{ij} < 1$, we have $\det \Sigma > 0$. It follows from this fact that

$\mu^\tau \Sigma^{-1}\mu \geq \lambda_0 \mu^\tau \mu$, where $\lambda_0 = \min\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} > 0$ and $\lambda_j, j = 1, ..., 4$ are the eigenvalues of Σ. Now it is readily seen that if we denote the value of left hand in (A.6) by I_{2n}, then

$$
\begin{aligned}
I_{2n} &= \frac{1}{(2\pi)^2 \, (\det \Sigma)^{1/2}} \int \cdots \int (1+|x|^{m_1})(1+|y|^{m_2})(1+|s|^{m_3})(1+|t|^{m_4}) \\
&\times \; K\big(\frac{x-s}{h}\big) \, K\big(\frac{y-s}{h}\big) \, K\big(\frac{x-t}{h}\big) K\big(\frac{y-t}{h}\big) e^{-\mu^\tau \Sigma^{-1}\mu/2} \, dxdydsdt \\
&\leq \; C \int \cdots \int K\big(\frac{x-s}{h}\big) \, K\big(\frac{y-s}{h}\big) \, K\big(\frac{x-t}{h}\big) \, K\big(\frac{y-t}{h}\big) e^{-\lambda_0 \mu^\tau \mu/4} \, dxdydsdt \\
&= \; C \, h^3 \iiint K(s) \, K(y-s) \, K(t) \, K(y-t) \; \big(\int_{-\infty}^{\infty} e^{-\lambda_0 \mu_1^\tau \mu_1/4} \, dx \big) dydsdt \\
&\leq \; C_1 \, h^3 \iiint K(s) \, K(y-s) \, K(t) \, dydsdt \;\leq\; C_2 h^3, \qquad \text{(A.13)}
\end{aligned}
$$

where $\mu_1^\tau = (x, x-hy, x-hs, x-ht)$ and we have used the fact:

$$
\int_{-\infty}^{\infty} e^{-\lambda_0 \mu_1^\tau \mu_1/4} \, dx \leq \int_{-\infty}^{\infty} e^{-\lambda_0 x^2} \, dx < \infty,
$$

since

$$
\begin{aligned}
\mu_1^\tau \mu_1 &= 4x^2 - 2hx(y+s+t) + h^2(y^2+s^2+t^2) \\
&= 4\big[x - h(y+s+t)/4\big]^2 + h^2\big[y^2+s^2+t^2 - (y+s+t)^2/4\big] \\
&\geq 4\big[x - h(y+s+t)/4\big]^2.
\end{aligned}
$$

This proves (A.6) and also completes the proof of Lemma 1.

LEMMA 2 *If $\tau_4 < \infty$, then*

$$
\frac{1}{n^2 h} \sum_{1 \leq i \neq j \leq n} \big(K_n^2(X_i, X_j) - E\left[K_n^2(X_i, X_j)\right] \big) \;=\; o_P(1), \qquad \text{(A.14)}
$$

and hence

$$
\frac{1}{n^2 h} \sum_{1 \leq i \neq j \leq n} K_n^2(X_i, X_j) \;\to_P\; \frac{\tau_2}{2\sqrt{\pi}} \, E\left[\sigma^4\big(X_1/\sqrt{2}\big)\right]. \qquad \text{(A.15)}
$$

Moreover, we also have

$$
\frac{1}{n^2 h} \sum_{1 \leq i \neq j \leq n} \epsilon_i^2 K_n^2(X_i, X_j) \epsilon_j^2 \;\to_P\; \frac{\tau_2}{2\sqrt{\pi}} \, E\left[\sigma^4\big(X_1/\sqrt{2}\big)\right]. \qquad \text{(A.16)}
$$

Proof. Write $\widetilde{K}_n(X_i, X_j) = K_n^2(X_i, X_j) - E\left[K_n^2(X_i, X_j)\right]$ and define Λ_n as before with $\delta_n = (nh)^{1/2}$. We have

$$\begin{aligned} E\Big(\sum_{1\le i\ne j\le n} \widetilde{K}_n(X_i, X_j)\Big)^2 &= \sum_{(i,j,k,l)\in\Lambda_n} E\big\{\widetilde{K}_n(X_i, X_j)\,\widetilde{K}_n(X_k, X_l)\big\} \\ &+ \sum_{(i,j,k,l)\notin\Lambda_n} E\big\{\widetilde{K}_n(X_i, X_j)\,\widetilde{K}_n(X_k, X_l)\big\} \\ &= \Delta_{n1} + \Delta_{n2}, \qquad \text{say.} \end{aligned} \tag{A.17}$$

Recall $\max \gamma_{ij} < 1$. It follows easily from (A.1) with $m = 4$ that

$$E\left[\widetilde{K}_n^2(X_i, X_j)\right] \le 4E\left[K_n^4(X_i, X_j)\right] \le 4C\tau_4\, h$$

and

$$E\left[\widetilde{K}_n(X_i, X_j)\,\widetilde{K}_n(X_k, X_l)\right] \le \Big(E\left[\widetilde{K}_n^2(X_i, X_j)\right]\Big)^{1/2} \Big(E\left[\widetilde{K}_n^2(X_k, X_l)\right]\Big)^{1/2} \le 4C\tau_4\, h.$$

Therefore, whenever $\tau_4 < \infty$ and $nh \to \infty$,

$$\Delta_{n2} \;\le\; 4C\tau_4\,\delta_n\, n^3\, h \;=\; o(n^4h^2). \tag{A.18}$$

As for Δ_{n1}, by noting that, uniformly for $(i, j, k, l) \in \Lambda_n$,

$$\begin{aligned} E\big\{\widetilde{K}_n(X_i, X_j)\,\widetilde{K}_n(X_k, X_l)\big\} &= E\big\{K_n^2(X_i, X_j)\,K_n^2(X_k, X_l)\big\} \\ &- E\left[K_n^2(X_i, X_j)\right]\, E\left[K_n^2(X_k, X_l)\right] \;=\; o(h^2), \end{aligned}$$

by (A.2) with $m = 2$ and (A.3), it is readily seen that

$$\Delta_{n1} \;=\; \sum_{(i,j,k,l)\in\Lambda_n} E\big\{\widetilde{K}_n(X_i, X_j)\,\widetilde{K}_n(X_k, X_l)\big\} = o(n^4h^2). \tag{A.19}$$

Now (A.14) follows from (A.17)-(A.19) and Markov's inequality.

Similarly, it follows easily from (A.2) and $E[K_n^2(X_i, X_j)] \le C\,h$ [by (A.1)] that

$$\begin{aligned} &\frac{1}{n^2h} \sum_{1\le i\ne j\le n} E\left[K_n^2(X_i, X_j)\right] \\ =\;& \frac{1}{n^2h}\left(\sum_{|i-j|\ge(nh)^{1/2}} E\left[K_n^2(X_i, X_j)\right] + \sum_{|i-j|<(nh)^{1/2}} E\left[K_n^2(X_i, X_j)\right]\right) \\ =\;& \frac{\tau_2}{2\sqrt{\pi}}\, E\left[\sigma^4\big(X_1/\sqrt{2}\big)\right]\, \frac{1}{n^2}\#\{(i,j) : |i-j| \ge (nh)^{1/2}, 1 \le i, j \le n\} \,+\, o(1) \\ =\;& \frac{\tau_2}{2\sqrt{\pi}}\, E\left[\sigma^4\big(X_1/\sqrt{2}\big)\right] \,+\, o(1), \end{aligned}$$

where $\#(A)$ denotes the number of elements in A. This, together with (A.14), yields (A.15).

By recalling $\{\epsilon_k\}$ is a sequence of i.i.d. random errors with $E[\epsilon_1^2] = 1$ and independent of X_k, the proof of (A.16) is the same as that of (A.15). We omit the details. The proof of Lemma 2 is now finished.

The following lemma is needed in the proof of Theorems 1 and 2. The lemma is also useful in itself.

LEMMA 3 *Let $\{\eta_k, k \geq 1\}$ be a sequence of i.i.d. random variables. Let a_{nij} be a sequence of constants with $a_{nij} = a_{nji}$ for all $n \geq 1$. Let $\varphi_n(x, y)$ be symmetric Borel-measurable functions such that, for all $n \geq 1$,*

$$E\left[\varphi_n^2(\eta_1, \eta_2)\right] > 0 \quad \text{and} \quad E\big(\varphi_n(\eta_1, \eta_2) \mid \eta_1\big) = 0. \tag{A.20}$$

Then there exists an absolute constant $A > 0$ such that

$$\sup_x \left|P\big(B_n^{-1} Q_{1n} \leq x\big) - \Phi(x)\right| \leq \; A\, B_n^{-4/5}\big(A_{1n} E\varphi_n^4(\eta_1, \eta_2) + A_{2n}\, \mathcal{L}_n\big)^{1/5}, \tag{A.21}$$

where $Q_{1n} = \sum_{1 \leq i < j \leq n} a_{nij}\, \varphi_n(\eta_i, \eta_j)$, $B_n^2 = \sum_{1 \leq i < j \leq n} a_{nij}^2 E\left[\varphi_n^2(\eta_1, \eta_2)\right]$,

$$\begin{aligned} A_{1n} &= \sum_{i=2}^{n} \Big(\sum_{j=1}^{i-1} a_{nij}^2\Big)^2, \quad A_{2n} = \sum_{i=2}^{n-1} \sum_{j=i+1}^{n} \Big(\sum_{k=1}^{i-1} a_{nik}\, a_{njk}\Big)^2, \\ \mathcal{L}_n &= E\big[\varphi_n(\eta_1, \eta_3)\, \varphi_n(\eta_1, \eta_4)\, \varphi_n(\eta_2, \eta_3)\, \varphi_n(\eta_2, \eta_4)\big]. \end{aligned}$$

PROOF. In the proof of Lemma 3, we omit the subscripts n in a_{nij} and φ_n for convenience. Set, for $i = 2, \cdots, n$, $Z_i = \sum_{k=1}^{i-1} a_{ik} \varphi(\eta_i, \eta_k)$ and $\mathcal{F}_i = \sigma(\eta_1, ..., \eta_i)$. It is readily seen that $Q_{1n} = \sum_{i=2}^{n} Z_i$ with $E(Z_i \mid \mathcal{F}_{i-1}) = 0$, $i = 2, \cdots, n$, by (A.20). This implies that $\{Q_{1j}, \mathcal{F}_j, 2 \leq j \leq n\}$ forms a martingale sequence. Hence it follows from Theorem 3.9 with $\delta = 1$ in Hall and Heyde (1980) that there exists an absolute constant $A > 0$ such that

$$\sup_x \left|P\big(B_n^{-1} Q_{1n} \leq x\big) - \Phi(x)\right| \leq A B_n^{-4/5}\, M_n^{1/5}, \tag{A.22}$$

where $U_n^2 = \sum_{i=2}^{n} Z_i^2$ and $M_n = \sum_{i=2}^{n} E[Z_i^4] + E(U_n^2 - B_n^2)^2$.

Next we will show that

$$M_n \leq 10\, A_{1n}\, E\left[\varphi_n^4(\eta_1, \eta_2)\right] + 4\, A_{2n}\, \mathcal{L}_n, \tag{A.23}$$

and then (A.21) follows immediately. In fact, by noting $B_n^2 = E[U_n^2]$, we have

$$M_n = \sum_{i=2}^{n} E[Z_i^4] + EU_n^4 - B_n^4 = 2\sum_{i=2}^{n} E[Z_i^4] + 2\sum_{2 \leq i < j \leq n} E\left[Z_i^2 Z_j^2\right] - B_n^4. \tag{A.24}$$

Using the second part of (A.20), we obtain that, for all $i < j$,

$$\begin{aligned}
E\left[Z_i^2 Z_j^2\right] &= \sum_{k,k_1=1}^{i-1} \sum_{l,l_1=1}^{j-1} a_{ik} a_{ik_1} a_{jl} a_{jl_1} E\big[\varphi(\eta_i,\eta_k)\varphi(\eta_i,\eta_{k_1})\varphi(\eta_j,\eta_l)\varphi(\eta_j,\eta_{l_1})\big] \\
&= \sum_{k=1}^{i-1}\sum_{l=1}^{j-1} a_{ik}^2 a_{jl}^2 E\big[\varphi^2(\eta_i,\eta_k)\,\varphi^2(\eta_j,\eta_l)\big] \\
&+ 2\sum_{\substack{k,l=1\\ k\neq l}}^{i-1} a_{ik}a_{il}a_{jk}a_{jl}\, E\big[\varphi(\eta_i,\eta_k)\,\varphi(\eta_i,\eta_l)\,\varphi(\eta_j,\eta_k)\,\varphi(\eta_j,\eta_l)\big] \\
&+ 2\sum_{k=1}^{i-1} a_{ik}^2 a_{ji} a_{jk}\, E\big[\varphi^2(\eta_i,\eta_k)\,\varphi(\eta_j,\eta_i)\,\varphi(\eta_j,\eta_k)\big] \\
&= R_{1ij} + R_{2ij} + R_{3ij}, \qquad\qquad (A.25)
\end{aligned}$$

where

$$\begin{aligned}
R_{1ij} &= \sum_{k=1}^{i-1}\sum_{\substack{l=1\\ l\neq k}}^{j-1} a_{ik}^2 a_{jl}^2 \left(E\left[\varphi^2(\eta_1,\eta_2)\right]\right)^2 \\
R_{2ij} &= 2\sum_{k,l=1}^{i-1} a_{ik}a_{il}a_{jk}a_{jl}\, E\big[\varphi(\eta_1,\eta_3)\varphi(\eta_1,\eta_4)\varphi(\eta_2,\eta_3)\varphi(\eta_2,\eta_4)\big] \\
|R_{3ij}| &\le \sum_{k=1}^{i-1} a_{ik}^2\big(3a_{jk}^2 + 2|a_{ji}a_{jk}|\big) E\left[\varphi^4(\eta_1,\eta_2)\right] \\
&\le 4\sum_{k=1}^{i-1} a_{ik}^2\big(a_{jk}^2 + a_{ji}^2\big) E\left[\varphi^4(\eta_1,\eta_2)\right].
\end{aligned}$$

Similarly, for all $2 \le i \le n$,

$$\begin{aligned}
E[Z_i^4] &= \sum_{j=1}^{i-1} a_{ij}^4 E[\varphi^4(\eta_1,\eta_2)] + \sum_{1\le j\neq k\le i-1} a_{ij}^2 a_{ik}^2 E\big[\varphi^2(\eta_1,\eta_2)\varphi^2(\eta_1,\eta_3)\big] \\
&\le \Big(\sum_{j=1}^{i-1} a_{ij}^2\Big)^2 E\left[\varphi^4(\eta_1,\eta_2)\right]. \qquad\qquad (A.26)
\end{aligned}$$

By virtue of (A.25) and (A.26), it is readily seen that $\sum_{i=2}^n E\left[Z_i^4\right] \le A_{1n} E\left[\varphi^4(\eta_1,\eta_2)\right]$ and

$$2\sum_{2\le i<j\le n} E\left[Z_i^2 Z_j^2\right] \le B_n^4 + 4A_{2n}\mathcal{L}_n + 8\,A_{1n}\,E\left[\varphi^4(\eta_1,\eta_2)\right].$$

Substituting these upper bounds back into (A.24), we obtain the inequality (A.23). The proof of Lemma 3 is now completed.

6.2 Proofs of Theorems

Proof of Theorem 1. Let $\widetilde{B}_n^2 = \sum_{1\le i<j\le n} K_n^2(X_i, X_j)$ and $\widetilde{M}_n = \sum_{1\le i<j\le n} \epsilon_i\epsilon_j K_n(X_i, X_j)$. By virtue of (A.15) and symmetry of $K_n(x, y)$, in order to prove Theorem 1, it suffices to show that

$$\widetilde{M}_n/\widetilde{B}_n \quad \to_D \quad N(0,1). \tag{A.27}$$

We now apply Lemma 3 to prove (A.27). Write

$$\begin{aligned}
\tilde{A}_{1n} &= \sum_{1\le i\ne j\le n} K_n^4(X_i, X_j) + \sum_{1\le i\ne j\ne k\le n} K_n^2(X_i, X_j) K_n^2(X_i, X_k), \\
\tilde{A}_{2n} &= \sum_{1\le i\ne j\ne k\ne l\le n} K_n\big(X_i, X_k\big)\, K_n\big(X_j, X_k\big)\, K_n\big(X_i, X_l\big)\, K_n\big(X_j, X_l\big) \\
&\quad + \sum_{1\le i\ne j\ne k\le n} K_n^2(X_i, X_k) K_n^2(X_j, X_k).
\end{aligned}$$

By recalling $K_n(X_i, X_j) \le C_0^2(1+|X_i|^\beta)(1+|X_j|^\beta)K(\frac{X_i-X_j}{h})$, it follows easily from Lemma 1 (iii) that if $h\to 0$ and $nh\to\infty$, then $E\tilde{A}_{1n} = o(n^4h^2)$ and $E\tilde{A}_{2n} = O(n^4h^3) = o(n^4h^2)$. This, together with (A.15) and the fact that $\tilde{A}_{1n} \le \widetilde{B}_n^4$ and $\tilde{A}_{2n} \le 2\,\widetilde{B}_n^4$, yields that

$$\begin{aligned}
E\big\{(\tilde{A}_{1n}+\tilde{A}_{2n})/\widetilde{B}_n^4\big\} &= E\big\{(\tilde{A}_{1n}+\tilde{A}_{2n})\big[I_{(\widetilde{B}_n^2\le A_0^2n^2h/4)} + I_{(\widetilde{B}_n^2> A_0^2n^2h/4)}\big]/\widetilde{B}_n^4\big\} \\
&\le 3P\big(\widetilde{B}_n^2 \le A_0^2n^2h/4\big) + 16A_0^{-4}n^{-4}h^{-2}E(\tilde{A}_{1n}+\tilde{A}_{2n}) \\
&= o(1),
\end{aligned} \tag{A.28}$$

where $A_0^2 = \frac{1}{\sqrt{\pi}}\, E\left[\sigma^4(X_1/\sqrt{2})\right] \int_{-\infty}^{\infty} K^2(x)dx$. Therefore, by recalling $E\left[\epsilon_0^4\right] < \infty$, it follows easily from (A.28) and Lemma 3 with $\varphi_n(x, y) = xy$ and $a_{nij} = K_n(X_i, X_j)$ that

$$\begin{aligned}
\sup_x \Big|P\big(\widetilde{M}_n/\widetilde{B}_n \le x\big) - \Phi(x)\Big| &\le E\sup_x \Big|P\big(\widetilde{M}_n/\widetilde{B}_n \le x \mid X_1, \cdots, X_n\big) - \Phi(x)\Big| \\
&\le A\left[(E\eta_1^4)^{2/5} + (E\eta_2^2)^{4/5}\right]\cdot E\big\{(\tilde{A}_{1n}+\tilde{A}_{2n})/\widetilde{B}_n^4\big\} = o(1)
\end{aligned}$$

as $n\to\infty$. This proves (A.27) and also complete the proof of Theorem 1.

Proof of Theorem 2. Set $\Lambda(x) = m_{\theta_0}(x) - m_{\widehat{\theta}}(x)$. Under H_0, we may write

$$\begin{aligned}
\widehat{M}_n(h) &:= \sum_{t=1}^{n}\sum_{s=1,\ne t}^{n} \widehat{e}_s\ a_n(X_s, X_t)\ \widehat{e}_t \\
&= \sum_{t=1}^{n}\sum_{s=1,\ne t}^{n} \Big\{\sigma(X_s)\epsilon_s + \Lambda(X_s)\Big\}\, K\Big(\frac{X_s-X_t}{h}\Big)\, \Big\{\sigma(X_t)\epsilon_t + \Lambda(X_t)\Big\} \\
&= M_n(h) + 2R_{1n} + R_{2n},
\end{aligned} \tag{A.29}$$

where $M_n(h) = \sum_{t=1}^n \sum_{s=1,\neq t}^n \epsilon_s\, \epsilon_t\, K_n(X_s, X_t)$,

$$R_{1n} = \sum_{t=1}^n \epsilon_t \sum_{s=1,\neq t}^n K\Big(\frac{X_s - X_t}{h}\Big)\, \Lambda(X_s)\, \sigma(X_t),$$
$$R_{2n} = \sum_{t=1}^n \sum_{s=1,\neq t}^n K\Big(\frac{X_s - X_t}{h}\Big)\, \Lambda(X_s)\, \Lambda(X_t).$$

Similarly, we also have

$$\begin{aligned}
\widehat{\sigma}_{1n}^2(h) &:= \sum_{t=1}^n \sum_{s=1,\neq t}^n \widehat{e}_s^2\ a_n^2(X_s, X_t)\ \widehat{e}_t^2 \\
&= \sum_{t=1}^n \sum_{s=1,\neq t}^n \Big\{\sigma(X_s)\epsilon_s + \Lambda(X_s)\Big\}^2 K^2\Big(\frac{X_s - X_t}{h}\Big)\, \Big\{\sigma(X_t)\epsilon_t + \Lambda(X_t)\Big\}^2 \\
&= \widetilde{\sigma}_{1n}^2(h) + R_{3n},
\end{aligned}$$

where $\widetilde{\sigma}_{1n}^2(h) = \sum_{t=1}^n \sum_{s=1,\neq t}^n \epsilon_s^2\, \epsilon_t^2\, K_n^2(X_s, X_t)$ and

$$\begin{aligned}
R_{3n} &= \sum_{t=1}^n \sum_{s=1,\neq t}^n \Big\{2\epsilon_t\, \Lambda(X_t)\, \sigma(X_t) + \Lambda^2(X_t)\Big\}\, K^2\Big(\frac{X_s - X_t}{h}\Big) \\
&\times\ \Big\{\big[\sigma(X_s)\epsilon_s + \Lambda(X_s)\big]^2 + \sigma^2(X_s)\, \epsilon_s^2\Big\}.
\end{aligned}$$

Recall that $M_n(h)/(A_0 n\sqrt{h}) \to_D N(0,1)$ and $\frac{2}{n^2 h}\tilde{\sigma}_{1n}^2(h) \to_P A_0^2$, where

$$A_0^2 = \frac{1}{\sqrt{\pi}}\, E\left[\sigma^4(X_1/\sqrt{2})\right] \int_{-\infty}^{\infty} K^2(x)dx,$$

by Theorem 1 and (A.16). Theorem 2 will follow if we prove

$$R_{1n} = o_p(n\sqrt{h}), \quad R_{2n} = o_p(n\sqrt{h}) \quad \text{and} \quad R_{3n} = o_p(n^2 h). \tag{A.30}$$

To prove (A.30), for $\forall\delta > 0$, write $\Omega_n = \{\widehat{\theta} : ||\widehat{\theta} - \theta_0|| \le \delta\, \eta_n\}$.

First deal with R_{2n} and R_{3n}. Note that, by virtue of (ii) and (iii) in Assumption 2.4,

$$|\Delta(X_s)| \le C\, \big|\big|\widehat{\theta} - \theta_0\big|\big|\, \Big|\Big|\frac{\partial m_\theta(X_s)}{\partial\theta}|_{\theta=\theta_0}\Big|\Big| \le C_1\, \delta\, \eta_n\, (1 + |X_s|^{\beta_1}), \tag{A.31}$$

for n sufficiently large such that $\Omega_n \subseteq \Theta_0$. It follows from (A.31) that

$$|R_{2n}| I(\widehat{\theta} \in \Omega_n) \le C\, \delta^2\, \eta_n^2 \sum_{t=1}^n \sum_{s=1,\neq t}^n (1 + |X_s|^{\beta_1})\, (1 + |X_t|^{\beta_1})\, K\Big(\frac{X_s - X_t}{h}\Big)$$

and

$$|R_{3n}| I(\widehat{\theta} \in \Omega_n) \leq C\,\delta\,\eta_n \sum_{t=1}^{n} \sum_{s=1, \neq t}^{n} (1 + |X_s|^{2(\beta+\beta_1)})\,(1 + |X_t|^{2(\beta+\beta_1)}) K\Big(\frac{X_s - X_t}{h}\Big)(1 + |\epsilon_t|)(1 + \epsilon_s^2),$$

for n sufficiently large, and hence by (A.4) and (A.5), for $\forall \delta > 0$,

$$\begin{aligned} P(|R_{2n}| \geq \delta^{1/2} n \sqrt{h}) &\leq P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) \\ &\quad + (\delta^{1/2} n \sqrt{h})^{-1}\,E\left[|R_{2n}| I(\widehat{\theta} \in \Omega_n)\right] \\ &\leq P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) + C\sqrt{\delta}\,n\,h^{1/2}\,\eta_n^2 \end{aligned}$$

and

$$\begin{aligned} P(|R_{3n}| \geq \delta^{1/2} n^2 h) &\leq P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) \\ &\quad + (\delta^{1/2} n^2 h)^{-1}\,E\left[|R_{3n}| I(\widehat{\theta} \in \Omega_n)\right] \\ &\leq P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) + C\sqrt{\delta}\,\eta_n. \end{aligned}$$

This proves $|R_{2n}| = o_P(n\sqrt{h})$ and $|R_{3n}| = o_P(n^2 h)$, where we have used the facts that $n\,h^{1/2}\,\eta_n^2 = O(1)$, $\eta_n \to 0$ and $\widehat{\theta} - \theta_0 = o_P(\eta_n)$.

We next prove $|R_{1n}| = o_P(n\sqrt{h})$. Let

$$\begin{aligned} J_1(s,t) &= K\left(\frac{X_s - X_t}{h}\right) \frac{\partial m_\theta(X_s)}{\partial \theta}|_{\theta=\theta_0}\,\sigma(X_t), \\ J_2(s,t) &= K\left(\frac{X_s - X_t}{h}\right) \left\{\Lambda(X_s) + \frac{\partial m_\theta(X_s)}{\partial \theta}|_{\theta=\theta_0}\left(\theta_0 - \widehat{\theta}\right)\right\} \sigma(X_t). \end{aligned}$$

Under these notation, we have

$$R_{1n} = \left(\theta_0 - \widehat{\theta}\right) \sum_{t=1}^{n} \epsilon_t \sum_{s=1, \neq t}^{n} J_1(s,t) + \sum_{t=1}^{n} \epsilon_t \sum_{s=1, \neq t}^{n} J_2(s,t). \tag{A.32}$$

Recall Assumption 2.3 and Assumption 2.4(iii). It follows from the results (A.4) and (A.5) that, for all $s \neq s_1 \neq t$ and for n sufficiently large,

$$\begin{aligned} E\left[J_1^2(s,t)\right] &\leq C\,E\left\{(1 + |X_s|^{2\beta_1})(1 + |X_t|^{2\beta})\,K^2\Big(\frac{X_s - X_t}{h}\Big)\right\} \\ &\leq C_1\,h, \\ E\left[|J_1(s,t) J_1(s_1,t)|\right] &\leq C\,E\Big[(1 + |X_{s_1}|^{\beta_1})\,(1 + |X_{s_2}|^{\beta_1})\,(1 + |X_t|^{2\beta}) \\ &\qquad K\Big(\frac{X_{s_1} - X_t}{h}\Big) K\Big(\frac{X_{s_2} - X_t}{h}\Big)\Big] \\ &\leq C_1\,h^2. \end{aligned}$$

These facts imply that for any $1 \le t \le n$,

$$E\Big[\sum_{s=1,\neq t}^{n} J_1(s,t)\Big]^2 \le C\,(nh + n^2h^2) \,\le 2C\,n^2h^2,$$

since $nh \to \infty$. Hence, by the iid properties of ϵ_t with $E[\epsilon_1] = 0$ and the independence between ϵ_t and X_s, we obtain that

$$E\Big[\sum_{t=1}^{n}\epsilon_t \sum_{s=1,\neq t}^{n} J_1(s,t)\Big]^2 \;\; \le \;\; 2C\,n^2h^2 \sum_{t=1}^{n} E[\epsilon_t^2] \le C_1\,n^3h^2. \tag{A.33}$$

On the other hand, it follows easily from Taylor's expansion of $m_\theta(x)$ (respect to θ) that, under H_0, for all $s \neq s_1 \neq t$ and for n large enough such that $\Omega_n \subseteq \Theta_0$,

$$\begin{aligned}
& E\left[|\epsilon_t||J_2(s,t)|I(\widehat{\theta} \in \Omega_n)\right] \\
& \le \; C\,\delta^2\,\eta_n^2\,E|\epsilon_t|\,E\left[K\Big(\frac{X_s - X_t}{h}\Big)\Big|\Big|\frac{\partial^2 m_\theta(X_s)}{\partial\theta^2}|_{\theta=\theta_0}\Big|\Big|\,|\sigma(X_t)|\right] \\
& \le \; C_1\,\delta^2\,\eta_n^2\;\;E\left[K\Big(\frac{X_s - X_t}{h}\Big)\,(1 + |X_s|^{\beta_1})\,(1 + |X_t|^{\beta})\right] \\
& \le \; C_1\,\delta^2\,\eta_n^2\,h.
\end{aligned} \tag{A.34}$$

It follows from (A.32)–(A.34) that

$$\begin{aligned}
E\left[|R_{1n}|I(\widehat{\theta} \in \Omega_n)\right] \;\; & \le \;\; \delta\,\eta_n\,E\Big|\sum_{t=1}^{n}\epsilon_t \sum_{s=1,\neq t}^{n} J_1(s,t)\Big| \\
& \quad\quad + \sum_{t=1}^{n}\sum_{s=1,\neq t}^{n} E\left[|\epsilon_t|\,|J_2(s,t)I(\widehat{\theta} \in \Omega_n)\right] \\
& \le \;\; C\,(\delta\,n^{3/2}h\eta_n + \delta^2\,n^2\eta_n^2\,h).
\end{aligned} \tag{A.35}$$

This, together with Markov's inequality, yields that, for $\forall\delta > 0$,

$$\begin{aligned}
P\left(|R_{1n}| \ge \delta^{1/2}n\sqrt{h}\right) \;\; & \le \;\; P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) \\
& \quad\quad + C\delta^{-1/2}(n^2h)^{-1/2}E\left[\big|\,R_{1n}\big|I(\widehat{\theta} \in \Omega_n)\right] \\
& \le \;\; P\left(||\widehat{\theta} - \theta_0|| > \delta\,\eta_n\right) + C\delta^{1/2}\,nh^{1/2}\eta_n^2,
\end{aligned} \tag{A.36}$$

since $h \to 0$. This yields $R_{1n} = o_P(n\sqrt{h})$, by recalling $nh^{1/2}\eta_n^2 = O(1)$ and $\widehat{\theta} - \theta_0 = o_P(\eta_n)$. The proof of Theorem 2 is now complete.