Model Specification Testing in Nonparametric and Semiparametric Time Series Econometrics [1]

By Jiti Gao[2] and Maxwell King [3]

Abstract

We propose a simultaneous model specification procedure for the conditional mean and conditional variance in nonparametric and semiparametric time series econometric models. An adaptive and optimal model specification test procedure is then constructed and its asymptotic properties are investigated. The main results extend and generalize existing results for testing the mean of a fixed design nonparametric regression model to the testing of both the conditional mean and conditional variance of a class of nonparametric and semiparametric time series econometric models. In addition, we develop computer–intensive bootstrap simulation procedures for the selection of an interval of bandwidth parameters as well as the choice of asymptotic critical values. An example of implementation is given to show how to implement the proposed simultaneous model specification procedure in practice. Moreover, finite sample studies are presented to support the proposed procedure.

KEYWORDS: Continuous–time model, diffusion process, kernel estimation, nonparametric estimation, optimal test, semiparametric method, time series econometrics.

1. Introduction and Motivation

Consider a continuous–time diffusion process of the form

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t,$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are respectively the univariate drift and volatility functions of the process, and B_t is standard Brownian motion. Recently, Aït-Sahalia (1996a) developed a simple methodology for testing both the drift and the diffusion. Through using the forward Kolmogorov equation, the author derived a corresponding relationship between the marginal density of r_t and the pair (μ, σ). Then, instead of testing both the drift and the volatility

[1]The first author would like to thank Song Xi Chen, Oliver Linton and Dag Tjøstheim for some constructive discussions. The authors also acknowledge comments from seminar participants at University of Western Australia, Monash University, Catholique University de Louvain in Belgium, London School of Economics, Cornell University and Yale University, in particular, Donald Andrews, Iréne Gijbels, Yongmiao Hong, Peter Phillips, Peter Robinson, Howell Tong and Qiwei Yao. Thanks also go to the Australian Research Council for its financial support.

[2]Jiti Gao is from Department of Statistics, School of Mathematics and Statistics, The University of Western Australia, Crawley WA 6009, Australia. Email: jiti@maths.uwa.edu.au

[3]Maxwell King is with the Faculty of Business and Economics, Monash University, Melbourne, Vic. 3168, Australia. Email: max.king@buseco.monash.edu.au

simultaneously, the author considered testing whether the density function belongs to a parametric family of density functions. The approach has the advantage of using discrete data without discretizing the continuous–time model (see also Aït-Sahalia 1996b). The use of the marginal density is computationally convenient and can detect a wide range of alternatives.

For a discrete time series regression model, however, it is difficult to establish a corresponding relationship between the marginal density of the time series and the pair of the conditional mean and the conditional variance of the model. Therefore, to specify the marginal density only may not be adequate for the specification of both the conditional mean and the conditional variance of a general time series regression model. This motivates the discussion of a simultaneous model specification for both the conditional mean and the conditional variance of a class of time series econometric models of the form

$$Y_t = g(X_t) + \sigma(X_t)e_t, \ t = 1, 2, \ldots, T \tag{1.1}$$

where both $g(\cdot)$ and $\sigma(\cdot) > 0$ are unknown functions defined over R^d, the data $\{(X_t, Y_t) : t \geq 1\}$ are either independent observations or dependent time series, $\{e_t\}$ is an independent and identically distributed (i.i.d.) error with mean zero and variance one, and T is the number of observations.

In recent years, nonparametric and semiparametric techniques have been used to construct model specification tests for the mean function of model (1.1). Interest focuses on tests for a parametric form versus a nonparametric form, tests for a semiparametric (partially linear or single-index) form against a nonparametric form, and tests for the significance of a subset of the nonparametric regressors. For example, Härdle and Mammen (1993) have developed consistent tests for a parametric specification by employing the kernel regression estimation technique; Hong and White (1995) and others have applied the method of series estimation to consistent testing for a parametric regression model; Eubank and Spiegelman (1990), Eubank and Hart (1992), Wooldridge (1992), Yatchew (1992), Gozalo (1993), Samarov (1993), Whang and Andrews (1993), Horowitz and Härdle (1994), Hjellvik and Tjøstheim (1995), Fan and Li (1996), Jayasuriva (1996), Zheng (1996), Hjellvik, Yao and Tjøstheim (1998), Li and Wang (1998), Chen and Fan (1999), Li (1999), Gao and King (2001), Chen, Härdle and Li (2003), and others have developed consistent tests for a semiparametric model (partially linear or single-index) versus a nonparametric alternative for either the independent and identically distributed (i.i.d.) case or the time series case. Other related studies include Robinson (1988, 1989), Andrews (1997), Li and Hsiao (1998), Whang (2000), Aït-Sahalia, Bickel and Stoker (2001), Fan and Huang (2001), Gozalo and Linton (2001), Gao, Tong and Wolff (2002), Hong and Lee (2002), and Sperlich, Tjøstheim and Yang (2002).

Recently, Horowitz and Spokoiny (HS) (2001) have developed a new test of a parametric model of a mean function against a nonparametric alternative. The test adapts to the unknown smoothness of the alternative model and is uniformly consistent against alternatives whose distance from the parametric model converges to zero at the fastest possible rate. This rate is slower than $T^{-1/2}$, where T is the number of observations. Another feature of the HS test is that one can avoid choosing a particular bandwidth parameter for testing purposes when using kernel based test statistics. Existing studies consider using an estimation based optimal value[4] for fixing the bandwidth parameter involved. This choice may not be justified in both theory and practice, as estimation based optimal values may not be optimal for testing purposes. For a kernel based testing problem, as suggested in the HS paper, one needs to choose an optimal bandwidth parameter to ensure that the power of the resulting test can be maximized at (or near) the optimal bandwidth. The HS paper has successfully used an interval of bandwidth parameters for constructing an adaptive and optimal test for testing the mean of a fixed design nonparametric regression model.

To the best of our knowledge, however, the problem of testing both the conditional mean and the conditional variance of model (1.1) simultaneously has attracted less attention. Recently, Chen and Gao (2003) constructed an empirical likelihood (EL) based test statistic to test both the mean and the variance of a nonparametric regression model, and proposed a bootstrap simulation procedure for the implementation of the proposed test. The current paper proposes two novel classes of test statistics and constructs an adaptive and optimal test. The proposed adaptive test is consistent against some local alternatives with an optimal rate. In addition, this paper develops computer–intensive simulation procdures for the choice of kernel bandwidth parameters and asymptotic critical values.

In summary, our approach has the following features:

(i) It proposes simultaneous test procedures for testing both the conditional mean and the conditional variance of a class of nonparametric time series econometric models for both independent and strongly dependent error processes. Sound and novel theoretical properties for the simultaneous test procedures are established.

(ii) It extends and generalizes the results of Horowitz and Spokoiny (2001) for testing the mean of fixed design nonparametric regression to the simultaneous tesing of both the conditional mean and the conditional variance of a class of nonparametric and semiparametric time series econometric models.

(iii) It is applicable to a wide variety of models, which include general nonparametric regression models for both the i.i.d. case and the time series case. The test procedure is also applicable to continuous–time model specification. Both the methodology and theoretical techniques developed in this paper can be used to improve economic and financial model building and forecasting.

[4]Usually, a cross–validation selection procedure is used for choosing an optimum bandwidth parameter to ensure that the average mean square of the resulting estimator is minimized. See Härdle, Liang and Gao (2000, §2.1.3) for example.

The rest of the paper is organised as follows. Section 2 proposes two class of model specification test statistics. An adaptive test procedure is discussed in Section 3 and some asymptotic consistency results are established. Section 4 provides an application of the adaptive test procedure to a discrete nonlinear time series model. Section 5 concludes the paper with some remarks on extensions. Mathematical details are relegated to Appendices A and B.

2. Model specification tests

Throughout this section, we consider model (1.1). For convenience, let

$$m_1(x) = E(Y_t|X_t = x) = g(x) \text{ and } m_2(x) = \text{var}(Y_t|X_t = x) = \sigma^2(x)$$

for $x \in S \subset R^d$. Define $m(x) = (m_1(x), m_2(x))^\tau$ be a bivariate vector and $\{m_\theta(\cdot) = (m_{1,\theta}(\cdot), m_{2,\theta}(\cdot))^\tau | \theta \in \Theta\}$ be a parametric model that specifies parametric forms for the conditional mean and conditional variance of Y_t conditional on X_t, where $\theta \in R^q$ is an unknown parameter taking a value in the parameter space $\Theta \subset R^q$.

The interest of this paper is to test

$$\mathcal{H}_0 : m_1(x) = m_{1\theta}(x) \quad \text{and} \quad m_2(x) = m_{2\theta}(x) \tag{2.1}$$

for some $\theta \in \Theta$ against

$$\mathcal{H}_1 : m_1(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x) \quad \text{and} \quad m_2(x) = m_{2\theta}(x) + C_{2T}\Delta_{2T}(x),$$

where both $\Delta_{1T}(x)$ and $\Delta_{2T}(x)$ are continuous and bounded functions over R^d.

Note that the above hypotheses are equivalent to

$$\mathcal{H}_0 : m(x) = m_\theta(x) \text{ versus } \mathcal{H}_1 : m(x) = m_\theta(x) + C_T\Delta_T(x) \text{ for all } x \in S,$$

where $C_T = (C_{1T}, C_{2T})^\tau$ is a vector of two non–random sequences tending to zero as $T \to \infty$ and $\Delta_T(x) = (\Delta_{1T}(x), \Delta_{2T}(x))^\tau$. This contains the parametric case where $\Delta_T(\cdot) \equiv 0$. Let $\theta_0 \in \Theta$ denote the true value of θ if $\mathcal{H}_0$ is true. That is, $m(x) = m_{\theta_0}(x)$ for all $x \in S$ if $\mathcal{H}_0$ is true.

We first introduce a nonparametric kernel estimator for $m(\cdot)$. Let K be a d-dimensional bounded probability density function with a compact support on the d-dimensional cube $[-1, 1]^d$. Assume that $K(\cdot)$ satisfies the moment conditions:

$$\int uK(u)du = 0 \text{ and } \int uu^\tau K(u)du = \sigma_K^2 \mathcal{I}_d,$$

where $\mathcal{I}_d$ is the d-dimension identity matrix and σ_K^2 is a positive constant. Let h be a smoothing bandwidth satisfying $h \to 0$ and $Th^d \to \infty$ as $T \to \infty$.

Define $K_h(u) = h^{-d}K(u/h)$. The Nadaraya-Watson (NW) estimators of $m_l(x)$ for $l = 1, 2$ are defined by

$$\widehat{m}_1(x) = \frac{\sum_{t=1}^T K_h(x - X_t)Y_t}{\sum_{t=1}^T K_h(x - X_t)} \quad \text{and} \quad \widehat{m}_2(x) = \frac{\sum_{t=1}^T K_h(x - X_t)(Y_t - \widehat{m}_1(X_t))^2}{\sum_{t=1}^T K_h(x - X_t)}. \tag{2.2}$$

This paper considers using the only one smoothing parameter h. One can use two different bandwidth parameters h_1 and h_2 for $l = 1$ and $l = 2$ respectively. The representation for this case will be complicated. See Chen and Gao (2003).

Similarly, for the parametric models, one can estimate $m_{l,\theta}$ by

$$\widetilde{m}_{l,\tilde{\theta}}(x) = \frac{\sum_{t=1}^T K_h(x - X_t)m_{l,\tilde{\theta}}(X_t)}{\sum_{t=1}^T K_h(x - X_t)} \tag{2.3}$$

for $l = 1, 2$, where $\tilde{\theta}$ is a consistent estimator of θ under $\mathcal{H}_0$.

Let $\widehat{m}(x) = (\widehat{m}_1(x), \widehat{m}_2(x))^\tau$ and $\widetilde{m}_\theta(x) = (\widetilde{m}_{1,\theta}(x), \widetilde{m}_{2,\theta}(x))^\tau$. The test statistics we are going to consider are based on the difference between $\widetilde{m}_{\tilde{\theta}}(\cdot)$ and $\widehat{m}(\cdot)$, rather than directly between $m_{\tilde{\theta}}(\cdot)$ and $\widehat{m}(\cdot)$. Due to the use of (2.2) and (2.3), one can avoid the bias associated with the nonparametric estimation.

The local linear estimator can also be used to replace the NW estimator in estimating $m(\cdot)$. As we use $\widehat{m}$ and $\widetilde{m}_{\tilde{\theta}}$ to construct each test statistic, however, the possible bias associated with the NW estimator is not an issue here. In addition, the NW estimator has a simpler analytic form. Extension of our approach to the local linear estimator based test procedure can be discussed in a similar fashion, although the proof will be more technical.

We now introduce the following notation.

$$\epsilon_t = Y_t - m_1(X_t), \ \eta_t = \epsilon_t^2 - m_2(X_t),$$

$$\sigma_{ij}(x) = E\left[\epsilon_t^i \eta_t^j | X_t = x\right] \quad \text{for } i = 0, 1, 2 \quad \text{and} \quad s_0(x) = |\Sigma_0(x)|^{-1},$$

where $|A|$ is the determinant of a matrix A and

$$\Sigma_0(x) = \begin{pmatrix} \sigma_{20}(x) & \sigma_{11}(x) \\ \sigma_{11}(x) & \sigma_{02}(x) \end{pmatrix}.$$

Let $f(x)$ be the marginal density of $\{X_t\}$. We assume without loss of generality that $R(K) = \int K^2(x)dx = 1$. Let

$$\Sigma(x) = f^{-1}(x)\Sigma_0(x).$$

In this section, we then construct two different classes of model specification tests and establish their asymptotic distributions. Section 3 discusses an optimal version of one of the proposed tests. Empirical comparisons of the two tests are given in Section 4.

2.1. *Class I of Test Statistics*

To construct the first class of our test statistics, we have a look at the following null hypothesis:

$$\mathcal{H}_{01} : m_1(x) = m_{1\theta}(x) \quad \text{against} \quad \mathcal{H}_{11} : m_1(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x). \tag{2.4}$$

For testing (2.4), Härdle and Mammen (1993) suggested using the following test statistic

$$\text{HM}_T = (Th^d) \int \left(\widehat{m}_1(x) - \widetilde{m}_{1\hat{\theta}}(x)\right)^2 \pi(x)dx, \tag{2.5}$$

where $\pi(x)$ is a positive weight function satisfying $\int \pi^2(x)dx < \infty$. The authors showed that under $\mathcal{H}_{01}$

$$\overline{\text{HM}}_T = \frac{\text{HM}_T - \mu_0}{\sigma_{0h}} \rightarrow_D N(0,1), \tag{2.6}$$

where $\mu_0 = K^{(2)}(0) \int \frac{\sigma^2(x)\pi(x)}{f(x)}dx$ and $\sigma_{0h}^2 = 2h^d K^{(4)}(0) \int \left(\frac{\sigma^2(x)\pi(x)}{f(x)}\right)^2 dx$, in which $f(x)$ is the density function of X_t and $\sigma^2(x) = \text{Var}(Y_t|X_t = x)$.

For testing (2.1), equation (2.5) thus motivates the use of a test statistic of the form

$$N_{1T}(h) = (Th^d) \int \{\widehat{m}(x) - \widetilde{m}_{\tilde{\theta}}(x)\}^\tau \hat{\Sigma}^{-1}(x) \{\widehat{m}(x) - \widetilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx \tag{2.7}$$

provided that $\hat{\Sigma}^{-1}(x)$ exists, where

$$\hat{\Sigma}^{-1}(x) = \hat{f}(x)\hat{\Sigma}_0^{-1}(x),\ \hat{\Sigma}_0(x) = \begin{pmatrix} \hat{\sigma}_{20}(x) & \hat{\sigma}_{11}(x) \\ \hat{\sigma}_{11}(x) & \hat{\sigma}_{02}(x) \end{pmatrix}, \tag{2.8}$$

$\hat{f}(x) = \frac{1}{Th^d}\sum_{t=1}^T K\left(\frac{x-X_t}{h}\right)$ and for $i,j = 0,1,2$,

$$\hat{\sigma}_{ij}(x) = \frac{\sum_{t=1}^T K\left(\frac{x-X_t}{h}\right)\hat{\epsilon}_t^i \hat{\eta}_t^j}{\sum_{t=1}^T K\left(\frac{x-X_t}{h}\right)},\ \hat{\epsilon}_t = Y_t - \widehat{m}_1(X_t),\ \hat{\eta}_t = \hat{\epsilon}_t^2 - \widehat{m}_2(X_t).$$

The use of the weight function, $\pi(\cdot)$, is due to both theoretical and practical considerations. For the theoretical consideration, one does not need to assume that the support of the marginal density, $f(\cdot)$, of $\{X_t\}$ is compact. This will not exclude some important distributions such as Gaussian distributions, which is particularly important in financial modelling. For the practical consideration, when the support of $f(\cdot)$ is not compact, one can use $\pi(\cdot)$ for approximation and truncation purposes.

Before establishing the asymptotic distribution of (2.7), we give the following remark.

Remark 2.1. We should point out that (2.7) is a natural extension of (2.5) and is asymptotically equivalent to the test statistic based on the empirical likelihood method (see Chen and Gao 2003).

We now state the main result of this section and the proof is relegated to Appendix A.

Theorem 2.1. (i) *Suppose that Assumptions A.1–A.4 hold. Then under* $\mathcal{H}_0$

$$L_{1T} = L_{1T}(h) = \frac{N_{1T}(h) - 2\mu_\pi}{\sigma_h} \to_D N(0,1) \text{ as } T \to \infty, \tag{2.9}$$

where $\mu_\pi = \int \pi(x)dx$, $\sigma_h^2 = 4h^d C(K,\pi)$, $C(K,\pi) = K^{(4)}(0)R^{-2}(K)\int \pi^2(x)dx$, $K^{(j)}(\cdot)$ *denotes the* j*-times convolution product of* $K(\cdot)$, *and* $R(K) = \int K^2(u)du$.

(ii) *Assume that the conditions of (i) hold. In addition, assume that there is a random data–driven* $\hat{h}$ *such that* $\frac{\hat{h}}{h} - 1 \to_p 0$ *as* $T \to \infty$. *Then under* $\mathcal{H}_0$

$$\hat{L}_{1T} = L_{1T}(\hat{h}) = \frac{N_{1T}(\hat{h}) - 2\mu_\pi}{\sigma_{\hat{h}}} \to_D N(0,1) \tag{2.10}$$

as $T \to \infty$.

Remark 2.2. One needs to point out that either (2.9) or (2.10) is already a normalized form. It follows from (2.9) or (2.10) that L_{1T} or $\hat{L}_{1T}$ has an asymptotic normality distribution under the null hypothesis H_0. In general, H_0 should be rejected if L_{1T} or $\hat{L}_{1T}$ exceeds a critical value, L_{10}^*, of the normal distribution. As can be seen from (2.10), the test statistic, $\hat{L}_{1T}$, involves the kernel function only and is therefore applicable to real data implementation.

Remark 2.3. Theorem 2.1(ii) shows that the asymptotic normality remains unchanged when h is replaced with the random data–driven $\hat{h}$, which is known as the plug–in method. Recently, Gao and King (2001), and Lavergne (2001) suggested using the plug–in method. Apart from using the plug–in method for testing purposes, there are some other methods. For example, Horowitz and Spokoiny (2001) adopted the maximum of a test statistic over a bandwidth interval. For our case, their test statistic is similar to $\max_{h \in H_T} L_{1T}(h)$, in which H_T is an interval of bandwidths. We discuss an extension of Horowitz and Spokoiny (2001) to our case in Section 3.

Theorem 2.1 gives the asymptotic normality of the test statistics for the simultaneous testing problem. When the null hypothesis is rejected, one needs to further test

$$\mathcal{H}_{01}: \ m_1(x) = m_{1\theta}(x) \ \text{ against } \ \mathcal{H}_{11}: \ m_1(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x)$$

or

$$\mathcal{H}_{02}: \ m_2(x) = m_{2\theta}(x) \ \text{ against } \ \mathcal{H}_{12}: \ m_2(x) = m_{2\theta}(x) + C_{2T}\Delta_{2T}(x).$$

Define

$$N_{11T}(h) = (Th^d)\int \{\widehat{m}_1(x) - \widetilde{m}_{1,\tilde{\theta}}(x)\}^2 \hat{\sigma}_{20}^{-1}(x)\pi(x)dx$$

and

$$N_{12T}(h) = (Th^d)\int \{\widehat{m}_2(x) - \widetilde{m}_{2,\tilde{\theta}}(x)\}^2 \hat{\sigma}_{02}^{-1}(x)\pi(x)dx.$$

We now have the following theorem.

Theorem 2.2. (i) *Under the conditions of Theorem 2.1(i), under* $\mathcal{H}_{01}$ or $\mathcal{H}_{02}$ *we have for* $i = 1$ *or* 2,

$$L_{1iT} = \frac{N_{1iT}(h) - \mu_\pi}{\sigma_{1h}} \to_D N(0,1) \tag{2.11}$$

as $T \to \infty$, *where* $\sigma_{1h}^2 = 2h^d C(K,\pi)$.

(ii) *Under the conditions of Theorem 2.1(ii), under* $\mathcal{H}_{01}$ or $\mathcal{H}_{02}$ *we have for* $i = 1$ *or* 2,

$$L_{1iT}(\hat{h}) = \frac{N_{1iT}(\hat{h}) - \mu_\pi}{\hat{\sigma}_{1h}} \to_D N(0,1) \tag{2.12}$$

as $T \to \infty$, *where* $\hat{\sigma}_{1h}^2 = 2\hat{h}^d C(K,\pi)$.

Theorem 2.2 shows that we can test either the conditional mean or the conditional variance. The conclusion of Theorem 2.2(i) is similar to those obtained previously for kernel estimation or series estimation based test statistics. Unlike the existing test statistics, our test statistics depend only on h and K. It follows from (2.4) and (2.5) that the test statistic of Härdle and Mammen (1993) depends on $\sigma^2(x) = \mathrm{Var}(Y_t|X_t = x)$. Obviously, σ_{0h} of (2.6) needs to be estimated when using L_{0T} in practice. By contrast, σ_{1h} of (2.11) does not involve any unknown function such as $\sigma^2(x)$.

As can be seen from the construction of L_{1T}, random denominators are involved in the form. Our experience suggests that the involvement of random denominators could reduce the power of the proposed tests. This motivates the construction of the second class of our test statistics below.

2.2. *Class II of Test Statistics*

In order to explain the motivation for the construction of the second class of our test statistics, we need to have a look at some relevant test statistics for testing the null hypothesis (2.4):

$$\mathcal{H}_{01} : m_1(x) = m_{1\theta}(x) \quad \text{against} \quad \mathcal{H}_{11} : m_1(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x).$$

For testing (2.4), several authors have proposed novel test statistics. See Li and Wang (1998), and Gao and King (2001). Let $p_{st} = K((X_s - X_t)/h)$. To test (2.4), we suggest using a test statistic of the form

$$L_{21T} = L_{21T}(h) = \frac{\sum_{s=1}^T \sum_{t=1,\neq s}^T p_{st}\hat{U}_t\hat{U}_s}{S_{21T}}, \tag{2.13}$$

where $S_{21T}^2 = 2\sum_{s,t=1}^T p_{st}^2 \hat{U}_t^2 \hat{U}_s^2$ and $\hat{U}_t = Y_t - m_{1,\tilde{\theta}}(X_t)$.

Similar to T_h of Horowitz and Spokoiny (2001, pp.606), we construct a test statistic of the form

$$\widehat{L}_{21T} = \widehat{L}_{21T}(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T A_{st}\hat{U}_s\hat{U}_t}{\widehat{S}_{21T}}, \tag{2.14}$$

where $\widehat{S}_{21T}^2 = 2\sum_{t=1}^T \sum_{s=1}^T A_{st}^2 \hat{U}_t^2 \hat{U}_s^2$, $\{A_{st}\}$ is the (s,t) element of the $T \times T$ matrix $A_h = W_h^\tau W_h$, and W_h is the $T \times T$ matrix whose (s,t) element is

$$w_h(X_s, X_t) = \frac{K((X_s - X_t)/h)}{\sum_{u=1}^T K((X_s - X_u)/h)}.$$

Theoretically, $\widehat{L}_{21T}$ is much more complicated than L_{21T}, as the latter involves only a double summation while the former involves not only a triple summation, but also several random denominators.

Let $P = \{p_{st}\}$ be a $T \times T$ matrix with p_{st} as its (s,t) element and $\hat{U} = (\hat{U}_1, \ldots, \hat{U}_T)^\tau$. Then the numerator of (2.13) can be expressed as

$$\sum_{t=1}^T \sum_{s \neq t} p_{st} \hat{U}_t \hat{U}_s = \hat{U}^\tau P \hat{U} - \sum_{t=1}^T p_{tt} \hat{U}_t^2.$$

This suggests using the following form for testing the null hypothesis (2.1):

$$(\hat{U}^\tau, \hat{V}^\tau) \begin{pmatrix} P & P \\ P & P \end{pmatrix} \begin{pmatrix} \hat{U} \\ \hat{V} \end{pmatrix},$$

where $\hat{V} = (\hat{V}_1, \ldots, \hat{V}_T)^\tau$ and $\hat{V}_t = \hat{U}_t^2 - m_{2,\tilde{\theta}}(X_t)$.

A simple decomposition implies that

$$(\hat{U}^\tau, \hat{V}^\tau) \begin{pmatrix} P & P \\ P & P \end{pmatrix} \begin{pmatrix} \hat{U} \\ \hat{V} \end{pmatrix} = (\hat{U} + \hat{V})^\tau P (\hat{U} + \hat{V}). \quad (2.15)$$

Equations (2.13) and (2.15) finally motivate the use of the following test statistic for testing the null hypothesis (2.1):

$$L_{2T} = L_{2T}(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T p_{st} \hat{W}_s \hat{W}_t}{\hat{\sigma}_h} = \frac{\hat{W}^\tau P \hat{W} - \hat{\mu}_h}{\hat{\sigma}_h}, \quad (2.16)$$

where $\hat{\sigma}_h^2 = 2\sum_{s,t=1}^T p_{st}^2 \hat{W}_t^2 \hat{W}_s^2$, $\hat{W}_t = \hat{U}_t + \hat{V}_t$, $\hat{W} = (\hat{W}_1, \ldots, \hat{W}_T)^\tau$, and $\hat{\mu}_h = \sum_{t=1}^T p_{tt} \hat{W}_t^2 = K(0) \sum_{t=1}^T \hat{W}_t^2$.

Other alternatives include

$$\widehat{L}_{2T} = \hat{L}_{2T}(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T A_{st} \hat{W}_s \hat{W}_t}{\bar{\sigma}_h}, \quad (2.17)$$

where $\bar{\sigma}_h^2 = 2\sum_{s,t=1}^T A_{st}^2 \hat{W}_t^2 \hat{W}_s^2$ and $\{A_{st}\}$ is as defined in (2.14).

As can be seen from (2.16) and (2.17), there are some similarities theoretically. Empirically, our small sample studies suggest that L_{2T} is more powerful than $\widehat{L}_{2T}$. Thus, we suggest using L_{2T} of (2.16) throughout the rest of this paper.

We now conclude our construction and discussion with the following remark.

Remark 2.4. (i) Equation (2.16) extends (2.13) for the univariate case to the bivariate case. When comparing (2.13) with (2.16), one can see the similarities of the two forms. This also suggests that one can easily construct a similar form for other multiple test problems, such as testing the first four moments.

(ii) It follows from the construction of L_{2T} that the form of L_{2T} depends on the use of (2.15). Before finally using (2.15), we also considered the following alternative:

$$(\hat{U}^\tau, \hat{V}^\tau) \begin{pmatrix} P & -P \\ -P & P \end{pmatrix} \begin{pmatrix} \hat{U} \\ \hat{V} \end{pmatrix} = (\hat{U} - \hat{V})^\tau P (\hat{U} - \hat{V}).$$

Obviously, one can replace $\hat{W}_t = \hat{U}_t + \hat{V}_t$ by $\hat{W}_t = \hat{U}_t - \hat{V}_t$ in (2.16). As our asymptotic and empirical studies show that there is little difference between using the two different forms, we suggest using L_{2T} of (2.16) throughout this paper.

(iii) As can be seen from (2.7) and (2.16), the test statistic L_{1T} involves not only a triple summation, but also several random denominators. By contrast, L_{2T} involves just a double summation and no random denominator is involved in the numerator. Theoretically, the form of (2.7) looks much more complicated than that of (2.16), although the two test statistics have similar asymptotic properties. Empirically, our small sample studies in Section 4 show that L_{2T} is more powerful than L_{1T}.

We now state the main result of this section and the proof is relegated to Appendix A.

Theorem 2.3. (i) *Suppose that Assumptions A.1–A.4 hold. Then under* $\mathcal{H}_0$

$$L_{2T} = L_{2T}(h) \to_D N(0,1) \text{ as } T \to \infty.$$

(ii) *Assume that the conditions of (i) hold. In addition, assume that there is a random data–driven* $\hat{h}$ *such that* $\frac{\hat{h}}{h} - 1 \to_p 0$ *as* $T \to \infty$. *Then under* $\mathcal{H}_0$

$$L_{2T}(\hat{h}) \to_D N(0,1)$$

as $T \to \infty$.

Similar to (2.13), we can construct a test statistic for the univarate test problem H_{02} proposed above Theorem 2.2. The test statistic is given by

$$L_{22T} = L_{22T}(h) = \frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T p_{st} \hat{V}_s \hat{V}_t}{S_{22T}}, \tag{2.18}$$

where $S_{22T}^2 = 2 \sum_{t=1}^T \sum_{s=1}^T p_{st}^2 \hat{V}_s^2 \hat{V}_t^2$.

We now have the following theorem and its proof follows from that of Theorem 2.3.

Theorem 2.4. (i) *Under the conditions of Theorem 2.3(i), under* $\mathcal{H}_{01}$ *or* $\mathcal{H}_{02}$ *we have for* $i = 1$ *or* 2,

$$L_{2iT}(h) \to_D N(0,1)$$

as $T \to \infty$.

(ii) *Under the conditions of Theorem 2.3(ii), under* $\mathcal{H}_{01}$ *or* $\mathcal{H}_{02}$ *we have for* $i = 1$ *or* 2,

$$L_{2iT}(\hat{h}) \to_D N(0,1)$$

as $T \to \infty$.

Sections 2.1–2.2 mainly discuss how to establish asymptotically consistent test statistics for the null hypothesis problem of the form (2.1), in which both $m_{1\theta}(\cdot)$ and $m_{2\theta}(\cdot)$ are parametric functions. As a matter of fact, one can construct similar test statistics for two different test problems–the first one is that both $m_{1\theta}(\cdot)$ and $m_{2\theta}(\cdot)$ are of partially linear forms, and the second problem is that both $m_{1\theta}(\cdot)$ and $m_{2\theta}(\cdot)$ are of single–index forms. This extension includes some semiparametric models as alternatives to the nonparametric null models.

2.3. *Some extensions and generalizations*

Assume that there are two pairs of unknown parameters, (α, β) and (γ, δ), and a pair of unknown functions, (ϕ, ψ) such that

$$m_{1\theta}(X_t) = U_t^\tau \alpha + \phi(V_t) \quad \text{and} \quad m_{2\theta}(X_t) = Z_t^\tau \beta + \psi(W_t) \tag{2.19}$$

or

$$m_{1\theta}(X_t) = U_t^\tau \alpha + \phi(V_t^\tau \gamma) \quad \text{and} \quad m_{2\theta}(X_t) = Z_t^\tau \beta + \psi(W_t^\tau \delta), \tag{2.20}$$

where $\theta = (\alpha, \beta)$ for (2.19), $\theta = (\alpha, \gamma, \beta, \delta)$ for (2.20), and U_t, V_t, Z_t and W_t are either subsets of X_t or the entire X_t.

When $\{X_t\}$ is a sequence of i.i.d. random variables and U_t, V_t, Z_t and W_t are subsets of X_t, Härdle, Liang and Gao (2000, Chapter 2) constructed some consistent estimators for (α, β) and (ϕ, ψ) in (2.19). Similarly, one can establish consistent estimators for the parameters and functions when $\{X_t\}$ is a stationary process. See Härdle, Liang and Gao (2000, Chapter 6). Li (1999) already considered testing the conditional mean of the form of the first part of (2.19).

When $U_t = V_t = X_t$ and $\{X_t\}$ is a sequence of dependent processes in (2.20), the conditional mean becomes

$$m_{1\theta}(X_t) = X_t^\tau \alpha + \phi(X_t^\tau \gamma). \tag{2.21}$$

For model (2.21), Xia, Tong and Li (1999) established asymptotically normal estimators for the parameters and function involved. Li (1999) already constructed a consistent test statistic for testing the null hypothesis of the form of (2.21) with $\alpha \equiv 0$. Similarly, one can establish asymptotically normal estimators for the parameters and functions involved in model (2.20).

Assume that $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$, $\tilde{\delta}$, $\tilde{\phi}(\cdot)$, and $\tilde{\psi}(\cdot)$ are consistent estimators of the parameters and functions involved in (2.19) or (2.20). The detailed construction of the estimators is similar to Li (1999) and Härdle, Liang and Gao (2000, Chapter 2) for (2.19) or Li (1999), Xia, Tong and Li (1999) and Härdle, Liang and Gao (2000, Chapter 2) for (2.20). We now define

$$\widetilde{m}_{1,\tilde{\theta}}(X_t) = U_t^\tau \tilde{\alpha} + \tilde{\phi}(V_t) \text{ and } \widetilde{m}_{2,\tilde{\theta}}(X_t) = Z_t^\tau \tilde{\beta} + \tilde{\psi}(W_t)$$

for (2.19), and

$$\widetilde{m}_{1,\tilde{\theta}}(X_t) = U_t^\tau \tilde{\alpha} + \tilde{\phi}(V_t^\tau \tilde{\gamma}) \text{ and } \widetilde{m}_{2,\tilde{\theta}}(X_t) = Z_t^\tau \tilde{\beta} + \tilde{\psi}(W_t^\tau \tilde{\delta})$$

for (2.20).

Substituting the new estimator $\widetilde{m}_{\tilde{\theta}}(x) = (\widetilde{m}_{1,\tilde{\theta}}(x), \widetilde{m}_{2,\tilde{\theta}}(x))^\tau$ into (2.7), one can establish the corresponding test statistic $L_{1T}(h)$ of (2.9) for testing the null hypothesis problem of the form (2.19) or (2.20). Similarly, for the construction of the corresponding test statistic $L_{2T}(h)$ of (2.16), one needs to replace $\hat{U}_t$ and $\hat{V}_t$ there by

$$\hat{U}_t = \left[Y_t - \widetilde{m}_{1,\tilde{\theta}}(X_t)\right] \tilde{f}(V_t) \text{ and } \hat{V}_t = \left\{\left[Y_t - \widetilde{m}_{1,\tilde{\theta}}(X_t)\right]^2 - \widetilde{m}_{2,\tilde{\theta}}(X_t)\right\} \tilde{f}^2(V_t)\tilde{f}(W_t)$$

for the case of (2.19), and

$$\hat{U}_t = \left[Y_t - \widetilde{m}_{1,\tilde{\theta}}(X_t)\right] \tilde{f}(V_t^\tau \tilde{\gamma}) \text{ and } \hat{V}_t = \left\{\left[Y_t - \widetilde{m}_{1,\tilde{\theta}}(X_t)\right]^2 - \widetilde{m}_{2,\tilde{\theta}}(X_t)\right\} \tilde{f}^2(V_t^\tau \tilde{\gamma})\tilde{f}(W_t^\tau \tilde{\delta})$$

for the case of (2.20), where $\tilde{f}(\cdot)$ is the usual kernel density estimator based on the data involved.

Therefore, for the null hypothesis problem (2.19) or (2.20), we can establish the corresponding Theorems 2.1 and 2.3. The detailed conditions and the proofs of the resulting theorems are similar to those for Theorems 2.1 and 2.3. Similarily, one can consider nonparametric significance testing for both the conditional mean and conditional variance of model (1.1). To do so, one needs to extend some existing results, such as Fan and Li (1996), Lavergne and Vuong (2000), and Aït-Sahalia, Bickel and Stoker (2001) to the simultaneous setting. As they are extremely technical, we shall not provide the details, which, however, are available upon request from the first author.

We need to point out that the test statistics proposed in Sections 2.1 and 2.2 are already normalized test statistics and their asymptotic distributions are standard normal. It is expected that the rate of convergence may not be fast. Thus, Theorems 2.1–2.4 can only give some rough idea about the asymptotic behaviour of the test statistics involved when the sample size is small. Thus, in practice we need to consider using a bootstrap method when implementing the test statistics in practice. As our small sample studies suggest that $L_{2T}(h)$ is at least as powerful as $L_{1T}(h)$ for each fixed h, we need only to modify $L_{2T}(h)$ to an optimal test statistic and show that the modified test statistic is consistent against alternatives of the form (2.1) in Section 3 below.

3. An adaptive test procedure

Section 2 establishes the asymptotic normality of the test statistics for testing

$$\mathcal{H}_0: \ m(x) = m_\theta(x) \text{ versus } \mathcal{H}_1: \ m(x) = m_\theta(x) + C_T\Delta_T(x),$$

where $\Delta_T(x)$ is as defined before. The test statistics have nontrivial power only if C_T converges more slowly than $T^{-1/2}$. Define $||C_T|| = \sqrt{C_{1T}^2 + C_{2T}^2}$.

In this section, we consider that the form of the local alternative models is

$$m_T(x) = m_{\theta_1}(x) + C_T\Delta_T(x), \tag{3.1}$$

where $\theta_1 \in \Theta$.

Similar to our tests, the tests of Andrews (1997), Bierens (1982), Bierens and Ploberger (1997), and Hart (1997) are consistent against alternatives of the form (3.1) whenever C_T converges more slowly than $T^{-1/2}$. This section considers the case where the testing problem is a simultaneous one for the dependent time series case. The main results of this section correspond to Theorems 1–4 of Horowitz and Spokoiny (2001).

3.1. *Asymptotic Behaviour of the Test Statistic under the Null Hypothesis*

As discussed in Section 2, the proposed test statistics depend on the bandwidth. This section then suggests using

$$L_2^* = \max_{h\in H_T} L_{2T}(h), \tag{3.2}$$

where $H_T = \left\{h = h_{\max}a^k : \ h \geq h_{\min}, \ k = 0, 1, 2, \ldots\right\}$, in which $0 < h_{\min} < h_{\max}$, and $0 < a < 1$. Let J_T denote the number of elements of H_T. In this case, $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$.

Simulation Scheme: Throughout this section, we use the notation of $L^* = L_2^*$. We now discuss how to obtain a critical value for L^*. The exact α–level critical value, l_α^* $(0 < \alpha < 1)$ is the $1 - \alpha$ quantile of the exact finite-sample distribution of L^*. Because θ_0 is unknown, l_α^* cannot be evaluated in practice. We therefore suggest choosing a simulated α–level critical value, l_α, by using the following simulation procedure:

1. For each $t = 1, 2, \ldots, T$, generate $Y_t^* = m_{1\tilde{\theta}}(X_t) + \sqrt{m_{2\tilde{\theta}}(X_t)}e_t^*$, where $\{e_t^*\}$ is sampled randomly from a specified distribution with $E\left[e_t^*\right] = 0$ and $E\left[(e_t^*)^2\right] = 1$. In addition, assume that the third and fourth moments of $\{e_t^*\}$ exist.

2. Use the data set $\{Y_t^*, X_t : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\hat{\theta}$. Compute the statistic $\hat{L}^*$ that is obtained by replacing Y_t and $\tilde{\theta}$ with Y_t^* and $\hat{\theta}$ on the right–hand side of (3.2).

3. Repeat the above steps M times and produce M versions of $\hat{L}^*$ denoted by $\hat{L}_m^*$ for $m = 1, 2, \ldots, M$. Use the M values of $\hat{L}_m^*$ to construct their empirical bootstrap distribution function, that is, $F^*(u) = \frac{1}{M}\sum_{m=1}^{M} I(\hat{L}_m^* \leq u)$. Use the empirical bootstrap distribution function to estimate the asymptotic critical value, l_α.

We now state the following result and its proof is relegated to Appendix B.

Theorem 3.1. *Assume that Assumptions A.1–A.2 and B.1–B.3 hold. Then under* $\mathcal{H}_0$

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

The main result on the behavior of the test statistic L^* under $\mathcal{H}_0$ is that l_α is an asymptotically correct α–level critical value under any model in $\mathcal{H}_0$.

3.2. *Consistency Against a Fixed Alternative*

We now show that L^* is consistent against a fixed alternative model. Assume that model (1.1) holds. Let the parameter set Θ be an open subset of R^q. Let $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$ satisfy Assumption B.1 listed in Appendix B. For $i = 1, 2$, let

$$M_i(\theta) = (m_{i\theta}(X_1), \ldots, m_{i\theta}(X_T))^\tau, \quad \overline{m}_i = (m_i(X_1), \ldots, m_i(X_T))^\tau,$$

$$M(\theta) = (M_1(\theta)^\tau, M_2(\theta)^\tau)^\tau \text{ and } \overline{m} = (\overline{m}_1^\tau, \overline{m}_2^\tau)^\tau.$$

Measure the distance between m and $\mathcal{M}$ by the normalized l_2 distance

$$\rho(m, \mathcal{M}) = \left[\inf_{\theta\in\Theta}\left(\frac{1}{2T}||\overline{m} - M(\theta)||^2\right)\right]^{1/2} \tag{3.3}$$

$$= \left[\inf_{\theta\in\Theta}\left(\frac{1}{2T}||\overline{m}_1 - M_1(\theta)||^2 + \frac{1}{2T}||\overline{m}_2 - M_2(\theta)||^2\right)\right]^{1/2}.$$

If $\mathcal{H}_0$ is false, then $\rho(m, \mathcal{M}) \geq c_\rho$ for all sufficiently large T and some $c_\rho > 0$. A consistent test will reject a false $\mathcal{H}_0$ with probability approaching one as $T \to \infty$.

The following theorem establishes the consistency.

Theorem 3.2. *Assume that the conditions of Theorem 3.1 hold. In addition, if there is some* $C_\rho > 0$ *such that* $\lim_{T\to\infty} P(\rho(m, \mathcal{M}) \geq C_\rho) = 1$ *holds, then*

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

The proof of Theorem 3.2 is relegated to Appendix B.

3.3. *Consistency Against a Sequence of Local Alternatives*

In this section, we consider the consistency of L^* under local alternatives of the form

$$m_T(x) = m_{\theta_1}(x) + C_T \Delta_T(x)$$

with $||C_T|| \geq C_0 T^{-1/2} h_{\max}^{-d/4} \left(\mathrm{loglogT}\right)^{1/4}$ for some constant $C_0 > 0$ and $\theta_1 \in \Theta$, where

$$m_T(x) = (m_{1T}(x), m_{2T}(x))^\tau,$$

$$m_{1T}(x) = m_{1\theta}(x) + C_{1T}\Delta_{1T}(x) \quad \text{and } m_{2T}(x) = m_{2\theta}(x) + C_{2T}\Delta_{2T}(x).$$

Throughout this section, for $i = 1, 2$ let

$$\overline{m}_{iT} = (m_{iT}(X_1), \ldots, m_{iT}(X_T))^\tau, \ \overline{\Delta}_{iT} = (\Delta_i(X_1), \ldots, \Delta_i(X_T))^\tau,$$

$$\overline{m}_T = (\overline{m}_{1T}^\tau, \overline{m}_{2T}^\tau)^\tau, \ \overline{\Delta}_T = (\overline{\Delta}_{1T}^\tau, \overline{\Delta}_{2T}^\tau)^\tau,$$

For $k = 1, 2$, let $\bigtriangledown_\theta M_k(\theta)$ be the $T \times q$ matrix whose (i, j) element is $\frac{\partial m_{k\theta}(X_i)}{\partial \theta_j}$ and $\bigtriangledown_\theta M(\theta) = ((\bigtriangledown_\theta M_1(\theta))^\tau, (\bigtriangledown_\theta M_2(\theta))^\tau)^\tau$.

We assume that $\Delta_T(x)$ is a continuous function that is normalized so that

$$\frac{1}{2T}||\overline{\Delta}_T||^2 = \frac{1}{2T}\left(\sum_{t=1}^{T} |\Delta_{1T}(X_t)|^2 + \sum_{t=1}^{T} |\Delta_{2T}(X_t)|^2\right) \geq 1. \tag{3.4}$$

We also suppose that $\overline{\Delta}_T$ is not an element of the space spanned by the columns of $\Delta_\theta M(\theta)$. That is,

$$|| \bigtriangledown_\theta M(\theta) - \Pi_1 \bigtriangledown_\theta M(\theta)|| \geq \delta || \bigtriangledown_\theta M(\theta)|| \tag{3.5}$$

for some $\delta > 0$, where

$$\Pi_1 = \bigtriangledown_\theta M(\theta_1) \left(\bigtriangledown_\theta M(\bigtriangledown_1)^\tau \bigtriangledown_\theta M(\theta_1)\right)^{-1} \bigtriangledown_\theta M(\theta_1)^\tau$$

is the projection operator into the column space of $\bigtriangledown_\theta M(\theta_1)$.

Conditions (3.4) and (3.5) exclude functions $\Delta_T(\cdot)$ for which $||\overline{m}_T - M(\theta_{T,0})|| = o(||C_T||)$ for some nonstochastic sequence $\{\theta_{T,0}\} \in \Theta$. Thus, (3.4) and (3.5) ensure that the rate of convergence of m_T to the parametric model $M(\theta_1)$ is the same as the rate of convergence of C_T to zero. In particular, when (3.4) and (3.5) hold in probability,

$$\left[\inf_{\theta \in \Theta}\left(\frac{1}{2T}||\overline{m}_T - M(\theta)||^2\right)\right]^{1/2} \geq \delta ||C_T||(1 - o(1)) \tag{3.6}$$

holds in probability.

We now state the following consistency result and its proof is relegated to Appendix B.

Theorem 3.3. *Assume that Assumptions A.1–A.2 and B.1–B.3 hold. Let $\tilde{\theta}$ be a $\sqrt{T}$–consistent estimator of θ. Let m_T satisfy (3.1) with $||C_T|| \geq C T^{-1/2} h_{\max}^{-d/4} \left(\mathrm{loglogT}\right)^{1/4}$ for some constant $C > 0$. In addition, let conditions (3.4) and (3.5) hold in probability. Then*

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

The result shows that the power of the adaptive, rate-optimal test approaches one as $T \to \infty$ for any function $\Delta_T(\cdot)$ and sequence $\{C_T\}$ that satisfy the conditions of Theorem 3.3.

3.4. *Consistency Against a Sequence of Smooth Alternatives*

This section discusses that L^* is consistent uniformly over alternatives in a Hölder smoothness class whose distance from the parametric model approaches zero at the fastest possible rate. The results can be extended to Sobolev and Besov classes under more technical conditions.

Before specifying our smoothness classes, we introduce the following notation. Let $j = (j_1, \ldots, j_d)$, where $j_1, \ldots, j_d \geq 0$ are integers, be a multi-index. For $i = 1, 2$, define

$$|j| = \sum_{i=1}^{d} j_i \quad \text{and} \quad D^j m_i(x) = \frac{\partial^{|j|} m_i(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$$

whenever the derivative exists. Define the Hölder norm

$$||m||_{H,s} = \sup_{x \in S} \sum_{|j| \leq s} \left(|D^j m_1(x)| + |D^j m_2(x)| \right).$$

The smoothness classes that we consider consist of functions $m \in S(H, s) \equiv \{m : ||m||_{H,s} \leq c_H\}$ for some (unknown) $s \geq \max(2, d/4)$ and $c_H < \infty$.

For some $s \geq \max(2, d/4)$ and all sufficiently large $c_m < \infty$, define

$$B_{H,T} = \left\{ m \in S(H, s) : \lim_{T \to \infty} P \left(\rho(m, \mathcal{M}) \geq c_m \left(T^{-1} \sqrt{\text{loglogT}} \right)^{2s/(4s+d)} \right) = 1 \right\}, \qquad (3.7)$$

where $\rho(m, \mathcal{M})$ is as defined in (3.3).

We now state the following consistency result and its proof is relegated to Appendix B.

Theorem 3.4. *Assume that Assumptions A.1–A.2 and B.1–B.3 hold. Then for* $0 < \alpha < 1$ *and* $B_{H,T}$ *as defined in (3.7)*

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

Remark 3.1. Theorems 3.1–3.4 extend Theorems 1–4 of Horowitz and Spokoiny (2001) from testing the mean of a fixed design regression model to the testing of both the conditional mean and the conditional variance of nonparametric α–mixing time series. Moreover, we consider the simultaneous test case where both the mean and variance functions can be simultaneously tested. Due to the property, we do not need to estimate the conditional variance directly for the simulation procedure proposed at the beginning of Section 3.

Remark 3.2. As can be seen from the above, the implementation of the adaptive test requires an intensive computing process. In particular, one needs to select both the interval of bandwidth parameters, H_T, and the asymptotic critical value, l_α. In particular, it is quite difficult to select a bandwidth parameter, h, for implementing the test statistic, L_{1T}, as existing theory provides no theoretical criteria on how this kind of choice should be done. It should be pointed out that existing selection criteria for h for estimation purposes may not be applicable and suitable, as estimation based optimal h values are not necessarily optimal for testing purposes. Our experience suggests that the choice of h should be based on the assessment of the power of the test involved. In Section 4 below, we provide two detailed simulation procedures for the choice of both H_T and the asymptotic critical value.

4. An example of implementation

This section then illustrates the proposed adaptive tests by a simulated example. In this example, we use simulated data to compare some small sample properties of $L_{1T}(h)$ and the adaptive test statistic L_2^* of (3.2).

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

$$Y_t = \alpha + \beta X_t + \sigma \cdot \sqrt{1 + 0.5X_t^2} \cdot e_t,$$

in which

$$X_t = 0.5X_{t-1} + \epsilon_t, \ t = 1, 2, \ldots, T, \tag{4.1}$$

where α, β and $\sigma > 0$ are unknown parameters to be estimated, both $\{\epsilon_t : t \geq 1\}$ and $\{e_t : t \geq 1\}$ are mutually independent and identically distributed, and independent of X_0, $\epsilon_t \sim U(-0.5, 0.5)$, $X_0 \sim U(-1, 1)$, and $\{e_t\}$ is either the standard $N(0, 1)$ or the normalized exponential $\text{Exp}(1) - 1$ error, which has mean zero and variance one.

Define the true forms of the conditional mean and conditional variance by

$$g_\theta(X_t) = \alpha + \beta X_t \ \text{ and } \ \sigma_\theta(X_t) = \sigma\sqrt{1 + 0.5X_t^2}.$$

We now consider a sequence of alternative models of the form

$$Y_t = g_T(X_t) + \sigma_T(X_t)e_t, \tag{4.2}$$

where

$$g_T(x) = g_\theta(x) + C_T\phi(x/D_T) \quad \text{and} \quad \sigma_T(x) = \sigma_\theta(x) + C_T\phi(x/D_T), \tag{4.3}$$

in which $D_T = \left(T^{-1}\sqrt{\text{loglogT}}\right)^{1/9}$, $C_T = D_T^4$ and $\phi(\cdot)$ is the probability density function of the standard normal distribution. The choice of (4.2) and (4.3) ensures that (3.7) holds with $s = 2$ and $d = 1$. This implies that the adaptive test is consistent against the sequence with an optimal rate.

In the following detailed simulation, we consider using a class of alternatives of the form

$$Y_t = \alpha + \beta X_t + \frac{1}{\psi}\phi(X_t/\psi) + \left(\sigma \cdot \sqrt{1 + 0.5X_t^2} + \frac{1}{\psi}\phi(X_t/\psi)\right) e_t, \tag{4.4}$$

where $\psi \neq 0$ is defined as the truncation parameter to be chosen, and the others are as defined in (4.1). In Table 4.1 below, we calculate both the size and the power of our adaptive test for various cases.

The vector of unknown parameters, $\theta = (\alpha, \beta, \sigma)$, involved in (4.1) was then estimated using the pseudo–maximum likelihood method, which is quite common in the estimation of parametric ARCH models. Due to the structure of (4.1), we choose the following weight function and the kernel function given by

$$\pi(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

and

$$K(x) = \begin{cases} \frac{15}{16}(1-x^2)^2 & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

Let $x_i = \frac{i}{n}$ and $n = [T^{1/5}]$ ($[x] \leq x$ denotes the largest integer part of x). Define

$$\hat{N}_{1T}(h) = \frac{1}{n}\sum_{i=1}^{n}(Th)\{\widehat{m}(x_i) - \widetilde{m}_{\tilde{\theta}}(x_i)\}^{\tau}\hat{\Sigma}^{-1}(x_i)\{\widehat{m}(x_i) - \widetilde{m}_{\tilde{\theta}}(x_i)\}, \tag{4.7}$$

where $\widehat{m}(x) = (\widehat{m}_1(x), \widehat{m}_2(x))^{\tau}$, $\widetilde{m}_{\theta}(x) = (\widetilde{m}_{1,\theta}(x), \widetilde{m}_{2,\theta}(x))^{\tau}$, $\tilde{\theta}$ is an estimator of θ,

$$\widehat{m}_1(x) = \frac{\sum_{t=1}^{T} K((x - X_t)/h)Y_t}{\sum_{t=1}^{T} K((x - X_t)/h)}, \quad \widehat{m}_2(x) = \frac{\sum_{t=1}^{T} K((x - X_t)/h)(Y_t - \hat{m}_1(X_t))^2}{\sum_{t=1}^{T} K((x - X_t)/h)},$$

$$\widetilde{m}_{l,\theta}(x) = \frac{\sum_{t=1}^{T} K((x - X_t)/h)m_{l,\theta}(X_t)}{\sum_{t=1}^{T} K((x - X_t)/h)}$$

for $l = 1, 2$, $m_{1,\theta}(X_t) = \alpha + \beta X_t$, $m_{2,\theta}(X_t) = \sigma^2\left[1 + 0.5X_t^2\right]$,

$$\hat{\Sigma}^{-1}(x) = \hat{f}(x)\hat{\Sigma}_0^{-1}(x), \ \hat{\Sigma}_0(x) = \begin{pmatrix} \hat{\sigma}_{20}(x) & \hat{\sigma}_{11}(x) \\ \hat{\sigma}_{11}(x) & \hat{\sigma}_{02}(x) \end{pmatrix},$$

$\hat{f}(x) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{x - X_t}{h}\right)$ and for $i, j = 0, 1, 2$,

$$\hat{\sigma}_{ij}(x) = \frac{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)\hat{\epsilon}_t^i\hat{\eta}_t^j}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}, \ \hat{\epsilon}_t = Y_t - \hat{m}_1(X_t), \ \hat{\eta}_t = \hat{\epsilon}_t^2 - \hat{m}_2(X_t),$$

and $K(\cdot)$ is as defined in (4.6). Alternatively, one can generate x_i from the density $\pi(\cdot)$ as many as $Q = 1000$ times and then define

$$\widehat{N}_{1T}(h) = \frac{Th}{Q}\sum_{\text{Q replications}}\left\{\frac{1}{n}\sum_{i=1}^{n}\{\widehat{m}(x_i) - \widetilde{m}_{\tilde{\theta}}(x_i)\}^{\tau}\hat{\Sigma}^{-1}(x_i)\{\widehat{m}(x_i) - \widetilde{m}_{\tilde{\theta}}(x_i)\}\right\}. \tag{4.8}$$

With the choice of $\pi(\cdot)$ and $K(\cdot)$ in (4.5) and (4.6), the constant $C(K, \pi)$ involved in L_{1T} is $\frac{93}{10}$. In order to calculate L_2^* of (3.2), one needs to find H_T, which is chosen by the following simulation procedure:

- For the simulation, we start with some initial values for θ_0 and X_0.
- For each $t = 1, 2, \ldots, T$, generate the data (X_t, Y_t) from (4.2) and (4.3).
- Use the data set $\{(Y_t, X_t) : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\tilde{\theta}$. For each fixed h, compute the resulting function of h given by

$$\widehat{L}_1(h) = \widehat{L}_{1T}(h) = \frac{\widehat{N}_{1T}(h) - 2}{\sqrt{\frac{186h}{5}}}.$$

- Repeat the above steps $M = 1000$ times and produce M versions of $\hat{L}_1(h)$ denoted by $\hat{L}_{1m}(h)$ for $m = 1, 2, \ldots, M$. Use the M functions of h, $\hat{L}_{1m}(h)$ for $m = 1, 2, \ldots, M$, to construct their empirical bootstrap distribution function, that is,

$$F_{1h}(u) = \frac{1}{M} \sum_{m=1}^{M} I(\hat{L}_{1m}(h) \leq u),$$

  where $I(U \leq u)$ is the usual indicator function.

- For the given empirical value $l_{0.05} = 1.65$, one can calculate the following power function

$$\phi_1(h) = 1 - F_{1h}(l_{0.05}).$$

- Find approximately at which h value the power function $\phi_1(h)$ is maximized. Denote the maximizer by h^*. Similarly, one can find the maximizer, h_*, of the corresponding power function $\phi_2(h)$ for

$$\widehat{L}_2(h) = \frac{\sum_{t=1}^{T} \left(\sum_{s=1, \neq t}^{T} p_{st} \hat{W}_s \right) \hat{W}_t}{\hat{\sigma}_h},$$

  where $\hat{\sigma}_h^2 = 2 \sum_{t=1}^{T} \sum_{s=1}^{T} p_{st}^2 \hat{W}_t^2 \hat{W}_s^2$, $\hat{W}_t = \hat{U}_t + \hat{V}_t$, $\hat{U}_t = Y_t - m_{1,\tilde{\theta}}(X_t)$, $\hat{V}_t = \hat{U}_t^2 - m_{2,\tilde{\theta}}(X_t)$, $p_{ts} = K((X_t - X_s)/h)$, and $K(\cdot)$ and $\tilde{\theta}$ are as defined before.

- Using h_*, construct H_T.

We now can calculate the following test statistic

$$L_1^* = \widehat{L}_1(h^*) = \frac{\widehat{N}_{1T}(h^*) - 2}{\sqrt{\frac{186h^*}{5}}}. \tag{4.9}$$

For the chosen H_T, we can compute L_2^* of (3.2) given by

$$L_2^* = \max_{h \in H_T} \left(\frac{\sum_{t=1}^{T} \left(\sum_{s=1, \neq t}^{T} p_{st} \hat{W}_s \right) \hat{W}_t}{\hat{\sigma}_h} \right). \tag{4.10}$$

In order to compute the rejection rates of the test statistics, one needs to find the corresponding simulated critical values.

We suggest choosing two simulated 5%–level critical values, $l_{1,0.05}$ and $l_{2,0.05}$, by using the following simulation procedure:

- For the simulation, we start with some initial values θ_0 and X_0.
- For each $t = 1, 2, \ldots, T$, generate the data (X_t, Y_t) from model (4.1).
- Use the data set $\{(Y_t, X_t) : t = 1, 2, \ldots, T\}$ to estimate θ. Denote the resulting estimate by $\tilde{\theta}$. For the chosen H_T, compute the statistics L_1^* and L_2^* given by (4.9) and (4.10).
- Repeat steps 2–3 $M = 1000$ times and produce M versions of L_1^* and L_2^* denoted by L_{1m}^* and L_{2m}^* for $m = 1, 2, \ldots, M$. Use the M values of L_{1m}^* and L_{2m}^* to construct their empirical bootstrap distribution functions, that is, $F_i^*(u) = \frac{1}{M} \sum_{m=1}^{M} I(L_{im}^* \leq u)$ for $i = 1, 2$. Use the empirical bootstrap distribution functions to calculate the two bootstrap simulated critical values, $l_{1,0.05}$ and $l_{2,0.05}$.

For each case where both ψ and T are chosen, we can compute the rejection rates. For calculating the rejection rates when $\mathcal{H}_0$ is true, one needs to use the data $\{(X_t, Y_t)\}$ where each (X_t, Y_t) is generated from (4.1). For calculating the rejection rates when $\mathcal{H}_1$ is true, one needs to use the data $\{(X_t, Y_t)\}$ where each (X_t, Y_t) is generated from (4.2). The number of simulations in producing Table 4.1 below was 1000. The detailed results are given in Table 4.1 below.

Table 4.1 near here

Remark 4.1. (i) First, one needs to point out that before modifying $L_{2T}(h)$ of (2.16) to be adaptive, we conducted some small sample studies for both $L_{1T}(h)$ and $L_{2T}(h)$. Our studies showed that $L_{2T}(h)$ was more powerful than $L_{1T}(h)$ uniformly in h. Moreover, Table 4.1 shows that L_2^* of (4.10) is more powerful than L_1^* of (4.9) for all the cases under consideration. We were also trying to compare the power of L_2^* of (3.2) with that of the proposed CGL test given in (3.1) of Chen and Gao (2003). Because the detailed comparison requires some very intensive and extremely lengthy computation as well as the implementation of both the proposed simulation scheme given in §3.1 and the so–called empirical likelihood based bootstrap simulation procedure proposed in Chen and Gao (2003), we have not been able to finish the detailed comparison for Example 4.1.

(ii) As can be seen from the first part of Table 4.1, for the standard Normal error the power can be close to one when $T = 500$ and the value of ψ^{-1} is between 4% and 10%. This may show that L_2^* is not only asymptotically optimal but also practically applicable to both the small and medium sample cases, since the differences between $\mathcal{H}_0$ and $\mathcal{H}_1$ were made deliberately close. We also computed the power of the tests for the case where $\psi = 1$ or 0.25, our small sample results showed that the power of L_2^* was already 100% even when $T = 250$. In the second part of Table 4.1, we have provided some small sample results for the case where the error is the normalized exponential random variable. The results show that the power of L_2^* is uniformly higher than that for the standard $N(0, 1)$ case. This may show that L_2^* is capable of capturing the skewness and kurtosis due to the flexible structure of $\{e_t^*\}$ allowed in the Simulation Scheme.

As pointed out in Section 2.2, for some cases one may need only to test either the conditional mean or the conditional variance. For the one–sided test case, it would be interesting to know whether there would be any significant reduction of the power when using L_2^* while $\mathcal{H}_1$ was different from $\mathcal{H}_0$ only in either the conditional mean or the conditional variance. In other words, we would be interested to know whether L_2^* would be much more powerful than either $L_{21}^* = \max_{h \in H_{1T}} L_{21T}(h)$ or $L_{22}^* = \max_{h \in H_{2T}} L_{22T}(h)$ when testing an one–sided problem, where $L_{21T}(h)$ and $L_{22T}(h)$ are as defined in (2.13) and (2.18) respectively, and the choice of H_{1T} and H_{2T} is similar to that for H_T. We have conducted some small sample studies for L_{21}^*, L_{22}^* and L_2^* for the one–sided test case. The number of simulations in producing Table 4.1 below was 1000. The detailed results are given in Tables 4.2 and 4.3 below.

Table 4.2 near here

Table 4.3 near here

Remark 4.2. (i) Tables 4.2 and 4.3 provide some detailed values for the power of the simultaneous test and the power of the two one–sided tests when $C_{2T} \equiv 0$ or $C_{1T} \equiv 0$. Our small sample results show that the power of the simultaneous test was just slightly less powerful than the corresponding one–sided test for both the cases even when the simultaneous test was used for testing either the conditional mean or the conditional variance. This may suggest that one can consider testing both the conditional mean and the conditional variance simultaneously when it is difficult to determine which component (the conditional mean or the conditional variance) may cause a model specification problem. We observed that the reduction of the power of the simultaneous test for the case of $C_{2T} \equiv 0$ was smaller than that for the case of $C_{1T} \equiv 0$. We also observed that both the simultaneous and the one–sided tests for the case of $C_{1T} \equiv 0$ were less powerful than the corresponding tests for the case of $C_{2T} \equiv 0$. We have not been able to explain these phenomena, although we think that this may be due to the increase in variability when testing the conditional variance only. It is also observed that the sizes of the three tests were all quite close to 5%.

(ii) When comparing the individual values for the power of the simultaneous test with those for the power of the one–sided tests for the Normal error distribution and the normalized exponential error distribution, we found some kind of superiority of the tests for the normalized exponential error distribution over those for the Normal error distribution, although the superiority may not be significant. This finding is similar to that drawn from Table I of Horowitz and Spokoiny (2001).

5. Conclusion

In this paper, we considered the general nonparametric time series regression model (1.1) and then proposed several model specification test statistics for testing the mean and the variance under the α–mixing condition. Furthermore, we established the adaptive test. Several consistency results about the test power of the test statistics were then developed. The consistency results extend the main results of Theorems 1–4 of Horowitz and Spokoiny (2001) from the fixed design case to the α–mixing time series case. The proposed optimal tests were illustrated through using a simulated example in Section 4.

The results given in this paper can be extended in a number of directions. First, the results for the short-range dependent time series case can be extended to the long-range dependent time series case, which is also relevant to some economic and financial data problems. Second, one can relax the strict stationarity and the mixing condition, as the recent work by Karlsen and Tjøstheim (2001) indicates that it is possible to do such work without the stationarity and the mixing condition.[5] This part is particularly important for the two reasons: (i) for the long-range dependent case one needs to avoid assuming both the long-range dependence and the mixing condition, as they contradict each other; and (ii) some important models are nonstationary and long–range dependent. See for example, Robinson (1995, 1997), and Gao (2002). Some of these issues are left for possible future research.

Appendix A

This appendix lists the necessary assumptions for the establishment and the proof of the main results given in Section 2.

A.1. *Assumptions*

Assumption A.1. (i) *Assume that the process (X_t, Y_t) is strictly stationary and α-mixing with the mixing coefficient $\alpha(t) = C_\alpha \alpha^t$ defined by*

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

for all $s, t \geq 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants, and Ω_i^j denotes the σ-field generated by $\{(X_t, Y_t) : i \leq t \leq j\}$.

(ii) *Assume that $P\left(0 < \min_{t\geq 1} \sigma(X_t) \leq \max_{t \geq 1} \sigma(X_t) < \infty\right) = 1$ and that for all $t \geq 1$ and $1 \leq i \leq 4$,*

$$P\left(E[e_t^i|\Omega_{t-1}] = \mu_i\right) = 1,$$

where $\mu_1 = 0$, $\mu_2 = 1$, μ_3 and μ_4 are real constants, and $\Omega_t = \sigma\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$ is a sequence of σ-fields generated by $\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$.

(iii) *Let $\zeta_t = \epsilon_t$ or η_t. In addition,*

$$E[|\zeta_t^{4+\alpha}|] < \infty \text{ and } E\left[\left|\zeta_{t_1}^{i_1}\zeta_{t_2}^{i_2} \cdot \zeta_{t_l}^{i_l}\right|^{1+\beta}\right] < \infty$$

[5] One also needs to point out that for the continuous–time case, Aït-Sahalia (1999) is applicable to the nonstationarity case.

for some small constants $\alpha > 0$ *and* $\beta > 0$*, where* $2 \leq l \leq 4$ *is an integer,* $0 \leq i_j \leq 4$ *and* $\sum_{j=1}^{l} i_j \leq 8$.

ASSUMPTION A.2. (i) *Let* $\zeta_t = \epsilon_t$ *or* η_t *and* $\mu_i(x) = E[\zeta_t^i | X_t = x]$ *for* $1 \leq i \leq 4$*. Assume that the following Lipschitz condition is satisfied:*

$$\max_{1 \leq i \leq 4} |\mu_i(u+v) - \mu_i(u)| \leq D(u)||v||$$

with $v \in S$ *(any compact set of* R^q*) and* $E\left[|D(X_t)|^{2+\gamma}\right] < \infty$ *for some small constant* $\gamma > 0$*, where* $||\cdot||$ *denotes the Euclidean norm.*

(ii) *Let* S_π *be a compact subset of* R^d*. Assume that* $\pi(\cdot)$ *is a positive weight function supported on* S_π *and satisfies* $0 < \int \pi^2(x)dx \leq C$ *for some constant* C*. Let* $S_f = \{x \in R^d : f(x) > 0\}$ *and* S_X *be the projection of* S_π *in* S_f*.*[6] *Assume that the marginal density function,* $f(x)$*, of* X_t*, and that all the first two derivatives of* $f(x)$ *and* $m_i(x)$*,* $i = 1, 2$*, are continuous on* R^d*,* $\inf_{x \in S_X} m_2(x) \geq C_m > 0$ *for some constant* C_m*, and on* S_X *the density function* $f(x)$ *is bounded below by* C_f *and above by* C_f^{-1} *for some* $C_f > 0$*, where* $m_1(x) = E[Y_t | X_t = x]$ *and* $m_2(x) = \mathrm{var}[Y_t | X_t = x]$.

(iii) *Let* $f_{\tau_1, \tau_2, \cdots, \tau_l}(\cdot)$ *be the joint probability density of* $(X_{1+\tau_1}, \ldots, X_{1+\tau_l})$ $(1 \leq l \leq 4)$*. Assume that* $f_{\tau_1, \tau_2, \cdots, \tau_l}(\cdot)$ *exists and satisfies the following Lipschitz condition:*

$$|f_{\tau_1, \tau_2, \cdots, \tau_l}(x_1 + v_1, \cdots, x_l + v_l) - f_{\tau_1, \tau_2, \cdots, \tau_l}(x_1, \cdots, x_l)| \leq D_{\tau_1, \cdots, \tau_l}(x_1, \cdots, x_l)||v||$$

for $v \in S$*, where* S *is any compact subset of* R^d *and* $D_{\tau_1, \cdots, \tau_l}(x_1, \cdots, x_l)$ *is integrable and satisfies the following conditions*

$$\int D_{\tau_1, \cdots, \tau_l}(x_1, \cdots, x_l)||x||^{2\theta} dx < M_1 < \infty,$$

$$\int D_{\tau_1, \cdots, \tau_l}(x_1, \cdots, x_l) f_{\tau_1, \tau_2, \cdots, \tau_l}(x_1, \cdots, x_l) dx < M_2 < \infty$$

for some $\theta > 1$ *and constants* $M_1 > 0$ *and* $M_2 > 0$.

ASSUMPTION A.3. (i) *Assume that the univariate kernel function* $k(\cdot)$ *is nonnegative, symmetric, and supported on* $[-1, 1]$*. In addition,* $k(x)$ *is continuous on* $[-1, 1]$*. This paper considers using*

$$K(x_1, \cdots, x_d) = \prod_{i=1}^{d} k(x_i).$$

(ii) *The bandwidth parameter* h *satisfies that*

$$\lim_{T \to \infty} h = 0, \quad \lim_{T \to \infty} Th^d = \infty \quad \text{and} \quad \limsup_{T \to \infty} Th^{5d} < \infty.$$

ASSUMPTION A.4. *Assume that for any parametric estimator,* $\overline{\theta}$*, of* θ

$$\max_{1 \leq i \leq 2} \max_{1 \leq t \leq T} |m_{i\overline{\theta}}(X_t) - m_{i\theta}(X_t)| = O_p(T^{-1/2}).$$

[6]In other words, $S_X = S_\pi \cap S_f$.

REMARK A.1. Assumptions A.1(i)(ii), A.2(ii) and A.3 and A.4 are novel conditions. Assumptions A.1(iii) and A.2(i)(iii) are similar to some parts of Condition (A1) of Li (1999, p.107). All the conditions are quite natural in this kind of problem. Note that we have not assumed the independence between $\{X_t\}$ and $\{e_t\}$. When $\{X_t\}$ and $\{e_t\}$ are independent, Assumption A.1(ii) holds naturally. For this case, model (1.1) becomes a nonparametric ARCH model when $X_t = (Y_{t-1}, \cdots, Y_{t-d})$ and $\{e_t\}$ is a sequence of i.i.d. random errors. We also have not assumed that the marginal density of X_t has a compact support. Instead, we impose some restrictions on the support of the weight function $\pi(\cdot)$. Assumption A.2 ensures that $0 < \inf_{x \in S_X} \mu_2(x) \leq \sup_{x \in S_X} \mu_2(x) < \infty$ and $0 < \inf_{x \in S_X} \mu_4(x) \leq \sup_{x \in S_X} \mu_2(x) < \infty$. These two conditions are required to ensure that $\Sigma^{-1}(x)$ exists and that the smallest eigenvalue of $\Sigma^{-1}(x)$ is positive uniformly in x. Assumption A.4(i) that requires the $\sqrt{T}$–rate of convergence for the parametric case is a standard condition. It holds when each $m_{i\theta}(\cdot)$ is differentiable in θ and $\overline{\theta}$ is an $\sqrt{T}$–consistent estimator of θ.

A.2. *Technical Lemmas*

The following lemmas are necessary for the proof of the main results stated in Section 2. Throughout the rest of this paper, we use $f(x_{i_1}, \ldots, x_{i_d})$ to represent the joint density function of $(X_{i_1}, \ldots, X_{i_d})$ for $1 \leq i_1 < \ldots < i_d \leq d$.

LEMMA A.1. *Suppose that M_m^n are the σ-fields generated by a stationary α-mixing process ξ_i with the mixing coefficient $\alpha(i)$. For some positive integers m let $\eta_i \in M_{s_i}^{t_i}$ where $s_1 < t_1 < s_2 < t_2 < \cdots < t_m$ and suppose $t_i - s_i > \tau$ for all i. Assume further that $||\eta_i||_{p_i}^{p_i} = E|\eta_i|^{p_i} < \infty$ for some $p_i > 1$ for which $Q = \sum_{i=1}^{l} \frac{1}{p_i} < 1$. Then*

$$\left| E\left[\prod_{i=1}^{l} \eta_i\right] - \prod_{i=1}^{l} E[\eta_i] \right| \leq 10(l-1)\alpha(\tau)^{(1-Q)} \prod_{i=1}^{l} ||\eta_i||_{p_i}.$$

PROOF: See Roussas and Ionnides (1987).

LEMMA A.2. *Let ξ_t be a r-dimensional strictly stationary and strong mixing (α–mixing) stochastic process. Let $\phi(\cdot, \cdot)$ be a symmetric Borel function defined on $R^r \times R^r$. Assume that for any fixed $x \in R^r$, $E[\phi(\xi_1, x)] = 0$ and $E[\phi(\xi_i, \xi_j)|\Omega_0^{j-1}] = 0$ for any $i < j$, where Ω_i^j denotes the σ–field generated by $\{\xi_s : i \leq s \leq j\}$. Let $\phi_{st} = \phi(\xi_s, \xi_t)$, $\sigma_{st}^2 = \mathrm{var}(\phi_{st})$ and $\sigma_T^2 = \sum_{1 \leq s < t \leq T} \sigma_{st}^2$. For some small constant $0 < \delta < 1$, let*

$$M_{T1} = \max_{1 \leq i < j < k \leq T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{1+\delta}, \int |\phi_{ik}\phi_{jk}|^{1+\delta} dP(\xi_i) dP(\xi_j, \xi_k) \right\},$$

$$M_{T21} = \max_{1 \leq i < j < k \leq T} \max \left\{ E|\phi_{ik}\phi_{jk}|^{2(1+\delta)}, \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j, \xi_k) \right\},$$

$$M_{T22} = \max_{1 \leq i < j < k \leq T} \max \left\{ \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i, \xi_j) dP(\xi_k), \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j) dP(\xi_k) \right\},$$

$$M_{T3} = \max_{1 \leq i < j < k \leq T} E|\phi_{ik}\phi_{jk}|^2, \ M_{T4} = \max_{\substack{1 < i,j,k \leq 2T \\ i,j,k \text{ different}}} \left\{ \max_P \int |\phi_{1i}\phi_{jk}|^{2(1+\delta)} dP \right\},$$

where the maximization over P in the equation for M_{T4} is taken over the four probability measures $P(\xi_1, \xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_{i_1})P(\xi_{i_2}, \xi_{i_3})$, and $P(\xi_1)P(\xi_i)P(\xi_j)P(\xi_k)$, where (i_1, i_2, i_3) is the permutation of (i, j, k) in ascending order;

$$M_{T51} = \max_{1 \le i < j < k \le T} \max \left\{ E \left| \int \phi_{ik}\phi_{jk}\phi_{ik}\phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} \right\},$$

$$M_{T52} = \max_{1 \le i < j < k \le T} \max \left\{ \int \left| \int \phi_{ik}\phi_{jk}\phi_{ik}\phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} dP(\xi_j) dP(\xi_k) \right\},$$

$$M_{T6} = \max_{1 \le i < j < k \le T} E \left| \int \phi_{ik}\phi_{jk} dP(\xi_i) \right|^2 .$$

Assume that all the $M_T's$ are finite. Let

$$M_T = \max \left\{ T^2 M_{T1}^{\frac{1}{1+\delta}}, T^2 M_{T51}^{\frac{1}{2(1+\delta)}}, T^2 M_{T52}^{\frac{1}{2(1+\delta)}}, T^2 M_{T6}^{\frac{1}{2}} \right\}$$

and

$$N_T = \max \left\{ T^{\frac{3}{2}} M_{T21}^{\frac{1}{2(1+\delta)}}, T^{\frac{3}{2}} M_{T22}^{\frac{1}{2(1+\delta)}}, T^{\frac{3}{2}} M_{T3}^{\frac{1}{2}}, T^{\frac{3}{2}} M_{T4}^{\frac{1}{2(1+\delta)}} \right\}.$$

If $\lim_{T\to\infty} \frac{\max\{M_T, N_T\}}{\sigma_T^2} = 0$, then as $T \to \infty$

$$\frac{1}{\sigma_T} \sum_{1 \le s < t \le T} \phi(\xi_s, \xi_t) \to_D N(0,1). \qquad (A.1)$$

Remark A.2. Lemma A.2 establishes central limit theorems for degenerate U–statistics of strongly dependent processes. The lemma extends and complements some existing results for the β–mixing case. See for example, Lemma 3.2 of Hjellvik, Yao and Tjøstheim (1998) and Theorem 2.1 of Fan and Li (1999).

Proof: See the proof of Lemma B.1 of Gao and King (2001).

Before stating the next lemma, we define and recall the following notation.

$$W_t(x) = \frac{1}{Th^d} K\left(\frac{x - X_t}{h}\right), \ \epsilon_t = Y_t - m_1(X_t), \ \eta_t = \epsilon_t^2 - m_2(X_t),$$

$$\sigma_{ij}(x) = E\left[\epsilon_t^i \eta_t^j | X_t = x\right] \quad \text{for } i = 0, 1, 2 \quad \text{and} \quad s_0(x) = |\Sigma_0(x)|^{-1}$$

where $|A|$ is the determinant of a matrix A and

$$\Sigma_0(x) = \begin{pmatrix} \sigma_{20}(x) & \sigma_{11}(x) \\ \sigma_{11}(x) & \sigma_{02}(x) \end{pmatrix}.$$

For $s, t = 1, 2, \ldots$, let

$$a_{st} = Th^d \int W_s(x) W_t(x) \sigma_{02}(x) s_0(x) f^{-1}(x) \pi(x) dx,$$

$$b_{st} = Th^d \int W_s(x) W_t(x) \sigma_{11}(x) s_0(x) f^{-1}(x) \pi(x) dx,$$

$$c_{st} = Th^d \int W_s(x) W_t(x) \sigma_{20}(x) s_0(x) f^{-1}(x) \pi(x) dx,$$

$$\phi_{st} = a_{st}\epsilon_s\epsilon_t - 2b_{st}\epsilon_s\eta_t + c_{st}\eta_s\eta_t,$$

$$N_{0T} = N_{0T}(h) = \sum_{s=1}^{T}\sum_{t=1}^{T} \phi_{st}. \quad (A.2)$$

Without loss of generality, we assume throughout the rest of this paper that

$$\int k(x)dx = \int k^2(x)dx = R(k) \equiv 1 \text{ and } \int \pi(x)dx = \int \pi^2(x)dx \equiv 1.$$

LEMMA A.3. *Under Assumptions A.1–A.3, we have as* $T \to \infty$

$$E\left[N_{0T}(h)\right] = 2 \quad \text{and} \quad \text{var}\left[N_{0T}(h)\right] = 4h^d K^{(4)}(0)(1+o(1)).$$

PROOF: It follows from Assumptions A.2–A.3 that as $T \to \infty$

$$\begin{aligned} a_{tt} &= Th^d \int W_t^2(x)\sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)dx \\ &= \int \frac{1}{Th^d} K^2\left(\frac{x-X_t}{h}\right)\sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)dx \\ &= \frac{1}{T}\left(\int K^2(u)du\right)\sigma_{02}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)(1+o(1)). \end{aligned} \quad (A.3)$$

Thus, as $T \to \infty$

$$\begin{aligned} \sum_{t=1}^{T} E\left[a_{tt}\epsilon_t^2\right] &= E\left[\sigma_{02}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\epsilon_t^2\right](1+o(1)) \\ &= E\left[\sigma_{02}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\sigma_{20}(X_t)\right](1+o(1)) \\ &= \int \sigma_{02}(x)s_0(x)\pi(x)\sigma_{20}(x)dx(1+o(1)). \end{aligned} \quad (A.4)$$

Similarly, we can obtain that as $T \to \infty$

$$\begin{aligned} \sum_{t=1}^{T} E\left[c_{tt}\eta_t^2\right] &= E\left[\sigma_{20}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\eta_t^2\right](1+o(h^d)) \\ &= E\left[\sigma_{20}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\sigma_{02}(X_t)\right](1+o(h^d)) \\ &= \int \sigma_{20}(x)s_0(x)\pi(x)\sigma_{02}(x)dx(1+o(h^d)) \end{aligned} \quad (A.5)$$

and

$$\begin{aligned} -2\sum_{t=1}^{T} E\left[b_{tt}\eta_t^2\right] &= -2E\left[\sigma_{11}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\epsilon_t\eta_t\right](1+o(h^d)) \\ &= -2E\left[\sigma_{11}(X_t)s_0(X_t)f^{-1}(X_t)\pi(X_t)\sigma_{11}(X_t)\right](1+o(h^d)) \\ &= -2\int \sigma_{11}^2(x)s_0(x)\pi(x)dx(1+o(h^d)). \end{aligned} \quad (A.6)$$

In view of (A.4)–(A.6), we have

$$E\left[N_{0T}(h)\right] = \sum_{t=1}^{T} E\left[\phi_{tt}\right]$$

$$= 2\int \left[\sigma_{20}(x)\sigma_{02}(x) - \sigma_{11}^2(x)\right] s_0(x)\pi(x)dx = 2\int \pi(x)dx = 2.$$

This finishes the proof of the first part of Lemma A.3. For the proof of the second part of Lemma A.3, let

$$\sigma_{st}^2 = \mathrm{E}[\phi_{st}^2] \ \text{ and } \ \sigma_T^2 = 2\sum_{1\le s,t\le T} \sigma_{st}^2.$$

Then

$$\sigma_T^2 = 2\sum_{1\le s,t\le T}\sigma_{st}^2 = 2\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\phi_{st}^2\right] = 2\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[a_{st}\epsilon_s\epsilon_t - 2b_{st}\epsilon_s\eta_t + c_{st}\eta_s\eta_t\right]^2$$

$$= 2\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[a_{st}^2\epsilon_s^2\epsilon_t^2 + 4b_{st}^2\epsilon_s^2\eta_t^2 + c_{st}^2\eta_s^2\eta_t^2 + 2a_{st}c_{st}\epsilon_s\epsilon_t\eta_s\eta_t - 4a_{st}b_{st}\epsilon_s^2\epsilon_t\eta_t - 4b_{st}c_{st}\epsilon_s\eta_s\eta_t^2\right].$$

We first look at the main part of σ_T^2. Similar to (A.3), we can have

$$a_{st}^2 = \int\int \frac{1}{(Th^d)^2} K\left(\frac{x-X_s}{h}\right) K\left(\frac{y-X_s}{h}\right) K\left(\frac{x-X_t}{h}\right) K\left(\frac{y-X_t}{h}\right) \times$$

$$\sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{02}(y)s_0(y)f^{-1}(y)\pi(y)dxdy.$$

Thus,

$$E\left[a_{st}^2\epsilon_s^2\epsilon_t^2\right] = E\left\{a_{st}^2 E\left[\epsilon_s^2\epsilon_t^2|(X_s,X_t)\right]\right\} = E\left[a_{st}^2\sigma_{20}(X_s)\sigma_{20}(X_t)\right]$$

$$= \frac{1}{(Th^d)^2}\int\int \sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{02}(y)s_0(y)f^{-1}(y)\pi(y)\times$$

$$E\left[K\left(\frac{x-X_s}{h}\right) K\left(\frac{y-X_s}{h}\right) K\left(\frac{x-X_t}{h}\right) K\left(\frac{y-X_t}{h}\right)\sigma_{20}(X_s)\sigma_{20}(X_t)\right] dxdy.$$

We now have a look at the following component. Using Assumptions A.2 and A.3, we have as $T\to\infty$

$$E\left[K\left(\frac{x-X_s}{h}\right) K\left(\frac{y-X_s}{h}\right) K\left(\frac{x-X_t}{h}\right) K\left(\frac{y-X_t}{h}\right)\sigma_{20}(X_s)\sigma_{20}(X_t)\right]$$

$$= \int\int K\left(\frac{x-u}{h}\right) K\left(\frac{y-u}{h}\right) K\left(\frac{x-v}{h}\right) K\left(\frac{y-v}{h}\right)\sigma_{20}(u)\sigma_{20}(v)f(u,v)dudv$$

$$= \int\int K\left(\frac{x-y}{h} + \frac{y-u}{h}\right) K\left(\frac{y-u}{h}\right) K\left(\frac{x-v}{h}\right) K\left(\frac{x-v}{h} - \frac{x-y}{h}\right)\sigma_{20}(u)\sigma_{20}(v)f(u,v)dudv$$

$$= h^{2d}\int\int K\left(s + \frac{x-y}{h}\right) K(s)K(t)K\left(t - \frac{x-y}{h}\right)\sigma_{20}(x-th)\sigma_{20}(y-sh)f(x-th,y-sh)dsdt$$

$$= h^{2d} L\left(\frac{x-y}{h}\right) L\left(\frac{y-x}{h}\right)\sigma_{20}(x)\sigma_{20}(y)f(x,y)(1+o(1)),$$

where $L(x) = \int K(x+y)K(y)dy$.

Therefore, as $T \to \infty$

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[a_{st}^2\epsilon_s^2\epsilon_t^2\right] = \int\int \sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{02}(y)s_0(y)f^{-1}(y)\pi(y)\sigma_{20}(x)\sigma_{20}(y)\times$$

$$L\left(\frac{x-y}{h}\right)L\left(\frac{y-x}{h}\right)f(x,y)dxdy(1+o(1)). \qquad (A.7)$$

Similarly,

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[c_{st}^2\eta_s^2\eta_t^2\right] = \int\int \sigma_{02}(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{02}(y)s_0(y)f^{-1}(y)\pi(y)\sigma_{20}(x)\sigma_{20}(y)\times$$

$$L\left(\frac{x-y}{h}\right)L\left(\frac{y-x}{h}\right)f(x,y)dxdy(1+o(1)),$$

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[b_{st}^2\epsilon_s^2\eta_t^2\right] = \int\int \sigma_{11}^2(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{11}^2(y)s_0(y)f^{-1}(y)\pi(y)\times$$

$$L\left(\frac{x-y}{h}\right)L\left(\frac{y-x}{h}\right)f(x,y)dxdy(1+o(1)), \qquad (A.8)$$

and

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[a_{st}c_{st}\epsilon_s\epsilon_t\eta_s\eta_t - 2a_{st}b_{st}\epsilon_s^2\epsilon_t\eta_t - 2b_{st}c_{st}\epsilon_s\eta_s\eta_t^2\right]$$

$$= -2\int\int \sigma_{20}(x)s_0(x)f^{-1}(x)\pi(x)\sigma_{02}(y)s_0(y)f^{-1}(y)\pi(y)\sigma_{11}(x)\sigma_{11}(y)\times$$

$$L\left(\frac{x-y}{h}\right)L\left(\frac{y-x}{h}\right)f(x,y)dxdy(1+o(1)). \qquad (A.9)$$

In view of (A.7)–(A.9), as as $T \to \infty$

$$\sigma_T^2 = \int\int \left[\sigma_{02}(x)\sigma_{02}(y)\sigma_{20}(x)\sigma_{20}(y) + \sigma_{11}^2(x)\sigma_{11}^2(y) - 2\sigma_{20}(x)\sigma_{02}(y)\sigma_{11}(x)\sigma_{11}(y)\right]\times$$

$$s_0(x)\frac{\pi(x)\pi(y)}{f(x)f(y)}s_0(y)L\left(\frac{x-y}{h}\right)L\left(\frac{y-x}{h}\right)f(x,y)dxdy(1+o(1))$$

$$= 4h^d K^{(4)}(0)\left(\int \pi^2(x)dx\right)(1+o(1)) \qquad (A.10)$$

using $s_0(x) = \left(\sigma_{02}(x)\sigma_{20}(x) - \sigma_{11}^2(x)\right)^{-1}$, where $K^{(4)}$ denotes the 4-times convolution product of K.

By Lemma A.1 (with $\eta_1 = \phi_{ik}$, $\eta_2 = \phi_{jk}$, $l = 2$, $p_i = 2(1+\delta)$ and $Q = \frac{1}{1+\delta}$),

$$E\left|\phi_{ik}\phi_{jk}\right| \le 10 M_{T1}^{\frac{1}{1+\delta}}\alpha^{\frac{\delta}{1+\delta}}(j-i),$$

where M_{T1} is as defined in Lemma A.2.

Therefore, using the fact that $\sum_{i=1}^{\infty}\alpha^{\frac{\delta}{1+\delta}}(i) < \infty$,

$$\sum_{1\le i<j<k\le T} E\left|\phi_{ik}\phi_{jk}\right| \le 10T^2 M_{T1}^{\frac{1}{1+\delta}}\sum_{i=1}^{T}\left(1-\frac{i}{T}\right)\alpha^{\frac{\delta}{1+\delta}}(i) \le CT^2 M_{T1}^{\frac{1}{1+\delta}} = o(\sigma_T^2), \qquad (A.11)$$

which follows from (A.13)–(A.15) below.

Equations (A.10) and (A.11) imply

$$\text{VAR}\left[N_{0T}(h)\right] = 4\left[\sum_{1\le s<t\le T}\text{var}(\phi_{st}) + 2\sum_{1\le i<j<k\le T} E\left(\phi_{ik}\phi_{jk}\right)\right]$$

$$= 4h^d K^{(4)}(0)\left(\int \pi^2(x)dx\right)(1+o(1)).$$

This finishes the proof of the second part of Lemma A.3.

A.3. *Proof of Theorem 2.1*

Proof of Theorem 2.1(i): To prove Theorem 2.1(i), we first show that as $T\to\infty$

$$\frac{N_{0T}(h)-2}{\sigma_h} \to N(0,1).$$

To apply Lemma A.2, let $\xi_t = (\epsilon_t, \eta_t, X_t^\tau)$ and $\phi(\xi_s,\xi_t) = \phi_{st}$ defined in (A.2). Let M_T and N_T be defined as in Lemma A.2. We now verify only the following condition listed in Lemma A.2,

$$\frac{\max\{M_T, N_T\}}{\sigma_h^2} \to 0 \;\text{ as } T\to\infty \tag{A.12}$$

for M_{T1}, M_{T21}, M_{T3}, M_{T51}, M_{T52} and M_{T6}. The others follow similarly.

For the M_T part, one justifies only

$$\frac{T^2 M_{T1}^{\frac{1}{1+\delta}}}{\sigma_h^2} \to 0 \;\text{ as } T\to\infty.$$

The others follow similarly.

Let $\psi_{st} = a_{st}\epsilon_s\epsilon_t$. It follows that for some $0<\delta<1$ and $1\le i<j<k\le T$

$$E\left[|\psi_{ik}\psi_{jk}|^{1+\delta}\right] = E\left[|\epsilon_i\epsilon_j\epsilon_k^2 a_{ik}a_{jk}|^{1+\delta}\right]$$

$$\le \left\{E\left[|\epsilon_i\epsilon_j\epsilon_k^2|^{2(1+\delta)(1+\delta_2)}\right]\right\}^{\frac{1}{2(1+\delta_2)}}\left\{E\left[|a_{ij}a_{ik}|^{(1+\delta)(1+\delta_1)}\right]\right\}^{\frac{1}{(1+\delta_1)}}$$

$$\le C_\epsilon\left\{E\left[|a_{ij}a_{ik}|^{(1+\delta)(1+\delta_1)}\right]\right\}^{\frac{1}{(1+\delta_1)}}, \tag{A.13}$$

using Assumption A.1(iii), where C_ϵ is a constant.

Since $0<\delta_1<1$ and $0<\delta_2<1$ satisfy $\frac{1}{1+\delta_1}+\frac{1}{2(1+\delta_2)}=1$ and $\frac{1+\delta}{3-\delta}<\delta_1<\frac{1-\delta}{1+\delta}$, we have that

$$1<\zeta_1=(1+\delta)(1+\delta_2)<2 \;\text{ and }\; 1<\zeta_2=(1+\delta)(1+\delta_1)<2.$$

Let $p(x)=\sigma_{02}(x)s_0(x)\pi(x)f^{-1}(x)$. Similar to (A.3), we have

$$a_{ik}a_{jk} = (Th^d)^{-2}\int\int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_k}{h}\right)K\left(\frac{y-X_j}{h}\right)K\left(\frac{y-X_k}{h}\right)p(x)p(y)dxdy$$

$$= T^{-2}\int\int K(u)K\left(u+\frac{X_i-X_k}{h}\right)K(v)K\left(v+\frac{X_j-X_k}{h}\right)p(X_i+uh)p(X_j+vh)dudv$$

$$= T^{-2}p(X_i)p(X_j)L\left(\frac{X_i-X_k}{h}\right)L\left(\frac{X_j-X_k}{h}\right)(1+o(1)).$$

For convenience, we use $\zeta = \zeta_2$ and ignore the small order $o(1)$ throughout the rest of the proof of Theorem 2.1(i). For the given $1 < \zeta < 2$ and T sufficiently large, we obtain

$$\begin{aligned} M_{T11} &= E\left|a_{ik}a_{jk}\right|^{\zeta} \\ &= T^{-2\zeta}\int\int\int |p(u)p(v)|^{\zeta}\left|L\left(\frac{u-w}{h}\right)\right|^{\zeta}\left|L\left(\frac{v-w}{h}\right)\right|^{\zeta} f(u,v,w)dudvdw \\ &= T^{-2\zeta}h^{2d}\int\int\int |p(z+xh)p(z+yh)|^{\zeta}\left|L(x)L(y)\right|^{\zeta} f(z+xh,z+yh,z)dxdydz \\ &= C_p T^{-2\zeta}h^{2d}, \end{aligned} \tag{A.14}$$

using Assumptions A.2 and A.3, where C_p is a constant.

Thus, as $T \to \infty$

$$\frac{T^2 M_{T11}^{\frac{1}{1+\delta}}}{\sigma_h^2} = C\frac{T^2\left(T^{-2\zeta}h^{2d}\right)^{1/\zeta}}{h^d} = h^{\frac{(2-\zeta)d}{\zeta}} \to 0. \tag{A.15}$$

Hence, (A.13)–(A.15) show that (A.12) holds for the first part of M_{T1}. The proof for the second part of M_{T1} follows similarly.

Similar to (A.8) and (A.14), we have that as $T \to \infty$

$$\begin{aligned} M_{T3} &= E\left|\psi_{ik}\psi_{jk}\right|^2 = E\left[a_{ik}^2 a_{jk}^2 \epsilon_i^2\epsilon_j^2\epsilon_k^4\right] \\ &= (Th^d)^{-4}h^{4d}E\left[p^2(X_i)p^2(X_j)L^2\left(\frac{X_i-X_k}{h}\right)L^2\left(\frac{X_j-X_k}{h}\right)\sigma_{20}(X_i)\sigma_{20}(X_j)\mu_4(X_k)\right] \\ &= T^{-4}\int\int\int p^2(x)p^2(y)L^2\left(\frac{x-z}{h}\right)L^2\left(\frac{y-z}{h}\right)\sigma_{20}(x)\sigma_{20}(y)\mu_4(z)f(x,y,z)dxdydz \\ &= T^{-4}h^{2d}\int\int\int p^2(uh+w)p^2(vh+w)L^2(u)L^2(v)\sigma_{20}(uh+w)\sigma_{20}(vh+w)\mu_4(w)dudvdw \\ &= CT^{-4}h^{2d}, \end{aligned} \tag{A.16}$$

using Assumptions A.2–A.3, where $\mu_4(x) = E[\epsilon_k^4|X_k]$.

This implies that as $T \to \infty$

$$\frac{T^{3/2} M_{T3}^{\frac{1}{2}}}{\sigma_h^2} = C\frac{T^{3/2}T^{-2}h^d}{h^d} = CT^{-1/2} \to 0. \tag{A.17}$$

Thus, (A.16) and (A.17) now show that (A.12) holds for M_{T3}. It follows from the structure of $\{\psi_{ij}\}$ that (A.12) holds automatically for M_{T51}, M_{T52} and M_{T6}, since $E[\epsilon_i|X_i] = 0$.

We now start to prove that (A.12) holds for M_{T21}.

Similar to (A.13), it follows that for some $0 < \delta < 1$ and $1 \leq i < j < k \leq T$

$$\begin{aligned} M_{T21} &= E\left[\left|\psi_{ik}\psi_{jk}\right|^{2(1+\delta)}\right] = E\left[\left|\epsilon_i\epsilon_j\epsilon_k^2 a_{ik}a_{jk}\right|^{2(1+\delta)}\right] \\ &\leq \left\{E\left[\left|\epsilon_i\epsilon_j\epsilon_k^2\right|^{2(1+\delta)(1+\delta_3)}\right]\right\}^{\frac{1}{1+\delta_3}}\left\{E\left[\left|a_{ij}a_{ik}\right|^{2(1+\delta)(1+\delta_4)}\right]\right\}^{\frac{1}{(1+\delta_4)}}, \end{aligned}$$

where $0 < \delta_3 < 1$ and $0 < \delta_4 < 1$ satisfy $\frac{1}{1+\delta_3} + \frac{1}{1+\delta_4} = 1$,

$$1 < \zeta_3 = (1+\delta)(1+\delta_3) < 2 \quad \text{and} \quad 1 < \zeta_4 = (1+\delta)(1+\delta_4) < 2.$$

Similar to (A.14) and (A.15), we obtain that as $T \to \infty$

$$\frac{T^{3/2} M_{T21}^{\frac{1}{2(1+\delta)}}}{\sigma_h^2} = C \frac{T^{3/2} T^{-2} \left(h^{2d}\right)^{1/(2\zeta_4)}}{h^d} = C \frac{1}{T^{1/2} h^{(1-\zeta_4^{-1})d}} \to 0$$

using the fact that $\lim_{T\to\infty} Th^d = \infty$ and $(1 - \zeta_4^{-1}) < \frac{1}{2}$.

This finally completes the proof of (A.12) for M_{T21} and thus (A.12) holds for the first part of $\{\phi_{st}\}$. Similarly, one can show that (A.12) holds for the other parts of $\{\phi_{st}\}$. Thus, we have shown that under $\mathcal{H}_0$

$$\frac{N_{0T}(h) - 2}{\sigma_h} \to N(0, 1) \text{ as } T \to \infty. \qquad (A.18)$$

The proof of Theorem 2.1(i) therefore follows from (3.6) and Assumptions A.3(ii) and A.4.

Proof of Theorem 2.1(ii): Note that as $T \to \infty$

$$L_{1T}(\hat{h}) = \frac{N_{1T}(\hat{h}) - 2}{\sigma_{\hat{h}}} = \frac{\frac{N_{1T}(\hat{h})}{N_{1T}(h)} [N_{1T}(h) - 2] + 2\left[\frac{N_{1T}(\hat{h})}{N_{1T}(h)} - 1\right]}{\left[\frac{\sigma_{\hat{h}}}{\sigma_h} - 1\right]\sigma_h + \sigma_h}$$

$$= \frac{\frac{N_{1T}(\hat{h})}{N_{1T}(h)} [N_{1T}(h) - 2] + 2\left[\frac{N_{1T}(\hat{h})}{N_{1T}(h)} - 1\right]}{\left[\frac{\hat{h}^{d/2}}{h^{d/2}} - 1\right]\sigma_h + \sigma_h} = \frac{N_{1T}(h) - 2}{\sigma_h}(1 + o_p(1))$$

using the continuity of $N_{1T}(h)$ in h. This completes the proof of Theorem 2.1(ii).

Proof of Theorems 2.2: The proof follows immediately from that of Theorem 2.1.

Proof of Theorems 2.3: As expected, the proof of Theorem 2.3 is much less complicated than that of Theorem 2.1. To prove Theorem 2.3, it suffices to show that as $T \to \infty$

$$\frac{\sum_{t=1}^T \sum_{s=1,\neq t}^T p_{st}\xi_s\xi_t}{\sqrt{2\sum_{s,t=1}^T p_{st}^2 \xi_s^2 \xi_t^2}} \to_D N(0, 1)$$

under H_0, where $\xi_t = \epsilon_t + \eta_t$.

The main technique is still Lemma A.2. The detailed proof is very similar to that of Theorem 3.1 of Gao and King (2001) for the univariate case. Thus, we shall not provide the detailed proof. However, it is available upon request.

Appendix B

This appendix lists the necessary assumptions for the establishment and the proof of the main results given in Section 3.

B.1. *Assumptions*

Let the parameter set Θ be an open subset of R^q. Let $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$. For $i = 1, 2$, define $\bigtriangledown_\theta m_{i\theta}(x) = \frac{\partial m_{i\theta}(x)}{\partial\theta}$, $\bigtriangledown_\theta^2 m_{i\theta}(x) = \frac{\partial^2 m_{i\theta}(x)}{\partial\theta\partial\theta'}$, and $\bigtriangledown_\theta^3 m_{i\theta}(x) = \frac{\partial^3 m_{i\theta}(x)}{\partial\theta\partial\theta'\partial\theta''}$ whenever these derivatives exist. For any $q \times q$ matrix D, define

$$||D||_\infty = \sup_{v \in R^q} \frac{||Dv||}{||v||},$$

where $||v||^2 = \sum_{i=1}^q v_i^2$ for $v = (v_1, \ldots, v_q)^\tau$.

Assumption B.1. *The parameter set* Θ *is an open subset of* R^q *for some* $q \geq 1$. *The parametric family* $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$ *satisfies*:

(i) *For each* $x \in R^d$ *and* $i = 1, 2$, $m_{i\theta}(x)$ *is three times differentiable almost surely with respect to* $\theta \in \Theta$. *Assume that* $\{G(x)\}$ *is a positive and integrable function with* $E[G(X_t)] < \infty$ *uniformly in* $t \geq 1$ *such that* $\max_{1\leq i\leq 2} \sup_{\theta\in\Theta} |m_{i\theta}(X_t)|^2 \leq G(X_t)$ *and* $\max_{1\leq i\leq 2} \sup_{\theta\in\Theta} ||\bigtriangledown_\theta^j m_{i\theta}(X_t)||^2 \leq G(X_t)$ *for* $j = 1, 2, 3$, *where for* $B = \{b_{ij}\}_{1\leq i,j\leq q}$, $||B||^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$. [7]

(ii) *For each* $i = 1, 2$ *and* $\theta \in \Theta$, $m_{i\theta}(x)$ *is continuous with respect to* $x \in R^d$. [8]

(iii) *Assume that there is a finite* $C_I > 0$ *such that for every* $\varepsilon > 0$

$$\inf_{\theta,\theta'\in\Theta:||\theta-\theta'||\geq\varepsilon} \min_{1\leq i\leq 2} [m_{i\theta}(X_1) - m_{i\theta'}(X_1)]^2 \geq C_I \varepsilon^2$$

holds with probability one (almost surely). [9]

Assumption B.2. (i) *Let* $\mathcal{H}_0$ *be true. Then* $\theta_0 \in \Theta$ *and*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\tilde{\theta} - \theta_0|| > C_L\right) < \varepsilon$$

for any $\varepsilon > 0$ *and all sufficiently large* C_L.

(ii) *Let* $\mathcal{H}_0$ *be false. Then there is a* $\theta^* \in \Theta$ *such that*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\tilde{\theta} - \theta^*|| > C_L\right) < \varepsilon$$

for any $\varepsilon > 0$ *and all sufficiently large* C_L.

(iii) *Let* $\{\theta_{T,0} : T = 1, 2, \ldots\}$ *be a sequence in* Θ *whose limit points, if any, are all in* Θ. *Define* $Y_t^* = m_{1\theta_{T,0}}(X_t) + \sqrt{m_{2\theta_{T,0}}(X_t)} e_t^*$, *where* $\{e_t^*\}$ *is sampled randomly from the specified distribution defined in the Simulation Scheme of Section 3.1. Let* $\hat{\theta}_T$ *be the estimator of* $\theta_{T,0}$ *that is obtained from the data set* $\{Y_t^*, X_t : t = 1, 2, \ldots, T\}$. *Then*

$$\lim_{T\to\infty} P\left(\sqrt{T}||\hat{\theta}_T - \theta_{T,0}|| > C_L\right) < \varepsilon$$

for any $\varepsilon > 0$ *and all sufficiently large* C_L.

Assumption B.3. (i) *Assume that Assumption A.3(i) holds.*

(ii) *Assume that the set* H_T *has the structure of (3.2) with* $h_{\max} > h_{\min} \geq T^{-\gamma}$ *for some constant* γ *such that* $0 < \gamma < 1/(d+2)$, *and* $J_T^2 h_{\max}^d \to 0$ *as* $T \to \infty$.

Remark B.1. Assumptions B.1–B.3 are quite standard in this kind of problem. Assumptions B.1 and B.2 extend Assumptions 1–2 and 4 of Horowitz and Spokoiny (2001) to the time series

[7]Note that Condition (i) may not be the weakest set of conditions imposed on $\{m_{i\theta}(x)\}$. For example, one can modify the corresponding restrictions on $\{m_{i\theta}(x)\}$ to: $\max_{1\leq i\leq 2} \sup_{\theta\in\Theta} E\left[m_{i\theta}(X_t)^2\right] < \infty$ and $\max_{1\leq i\leq 2} \sup_{\theta\in\Theta} E\left[||\bigtriangledown_\theta^j m_{i\theta}(X_t)||^2\right] < \infty$ for $1 \leq j \leq 3$. For this case, the proof of (B.15) and (B.18) below will become more tedious.

[8]Note that in Assumption A.2(ii), some smoothness conditions on $\{m_i(x)\}$ have already been imposed. We therefore do not need to impose similar conditions on $\{m_{i\theta}(x)\}$, as $m_i(x) = m_{i\theta}(x)$ holds for some θ when H_0 holds and $m_i(x) = m_{i\theta}(x) + C_{iT}\Delta_{iT}(x)$ when H_1 holds.

[9]This condition is to ensure that $\{m_{i\theta}(x)\}$ is identifiable with respect to θ.

case. Assumption B.3 is silghtly different from Assumption 6 of Horowitz and Spokoiny (2001). Actually, we have been unable to verify whether Assumption 6 is necessary for the proof of Lemma 10 in particular. Assumption B.3 holds in many cases. For example, it allows the use of the estimation based optimal value $h_{\text{optimal}} = CT^{-\frac{1}{2s+d}}$ in case it can also be optimal for testing purposes. Assumption B.3(ii) holds when $h_{\max} = (\log(T))^{-\frac{2}{d}-\epsilon}$ for some small $\epsilon > 0$.

B.2. *Technical Lemmas*

Before stating the necessary lemmas for the proof of the results given in Section 3, we introduce the following notation.

For $j = 1, 2$, let $\epsilon_t = Y_t - m_1(X_t)$, $\eta_t = \epsilon_t^2 - m_2(X_t)$, $\xi_t = \epsilon_t + \eta_t$,

$$\lambda_{jt}(\theta) = \lambda_j(X_t, \theta) = m_j(X_t) - m_{j\theta}(X_t) = m_{j\theta_0}(X_t) - m_{j\theta}(X_t),$$

$$\pi_t(\theta) = \lambda_{1t}^2(\theta) + 2\epsilon_t\lambda_{1t}(\theta), \quad \pi(\theta) = (\pi_1(\theta), \ldots, \pi_T(\theta))^\tau,$$

$$\lambda_t(\theta) = \lambda(X_t, \theta) = \lambda_{1t}(\theta) + \lambda_{2T}(\theta), \quad \lambda(\theta) = (\lambda_1(\theta), \ldots, \lambda_T(\theta))^\tau,$$

$$Q_T(\theta) = \lambda(\theta)^\tau P \lambda(\theta) = \sum_{s=1}^T \sum_{t=1}^T p_{st}\lambda_s(\theta)\lambda_t(\theta),$$

$$\Pi_T(\theta) = \pi(\theta)^\tau P \pi(\theta) = \sum_{s=1}^T \sum_{t=1}^T p_{st}\pi_s(\theta)\pi_t(\theta), \qquad (B.1)$$

where $p_{st} = K((X_s - X_t)/h)$.

Lemma B.1. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold.*

(i) *For every* $\delta > 0$

$$\max_{h \in H_T} \sup_{||\theta - \theta_0|| \leq \delta} \frac{Q_T(\theta)}{T^2 h^d} \leq C\delta^2$$

in probability, where $C > 0$ *is a constant.*

(ii) *For each* $\theta \in \Theta$,

$$\lim_{T \to \infty} P\left(\frac{Q_T(\theta)}{\lambda(\theta)^\tau \lambda(\theta)} \geq Th^{d+1}\right) = 1.$$

Proof: (i) It follows from the definition of $Q_T(\theta)$ that $Q_T(\theta) \leq ||P||_\infty ||\lambda(\theta)||^2$.

In order to prove Lemma B.1(i), one first needs to show that

$$||P||_\infty \leq CTh^d \qquad (B.2)$$

in probability for some constant $C > 0$.

It follows from the uniform convergence of $\hat{f}(X_t)$ (see Lemmas A.1 and A.3 of Härdle, Liang and Gao 2000) that $\max_{1 \leq t \leq T} \left|\hat{f}(X_t) - f(X_t)\right| = o_p(1)$ as $T \to \infty$.

This implies that as $T \to \infty$

$$\max_{1 \leq t \leq T} \frac{1}{Th^d} \sum_{s=1}^T p_{ts} = \max_{1 \leq t \leq T} \frac{1}{Th^d} \sum_{s=1}^T K\left(\frac{X_t - X_s}{h}\right) = \max_{1 \leq t \leq T} \hat{f}(X_t) \leq C$$

in probability. This finally finishes the proof of (B.2).

In view of (B.2), in order to prove Lemma B.1(i), it suffices to show that

$$\sup_{||\theta-\theta_0||\leq\delta} ||\lambda(\theta)||^2 \leq CT\delta^2 \tag{B.3}$$

in probability.

A Taylor series expansion to $m_{i\theta}(X_t) - m_{i\theta_0}(X_t)$ and an application of Assumption B.1(i) imply (B.3). This finishes the proof of Lemma B.1(i).

(ii). To prove Lemma B.1(ii), it suffices to show that as $T \to \infty$

$$P\left(\lambda(\theta)^\tau\lambda(\theta) > T^{-1}h^{-d-1}\lambda(\theta)^\tau P\lambda(\theta)\right) \to 0.$$

In view of the definition of $Q_T(\theta)$, one needs only to show that for any given small $\epsilon > 0$

$$P\left(\frac{\lambda(\theta)^\tau\lambda(\theta)}{E\left[\lambda(\theta)^\tau P\lambda(\theta)\right]} > T^{-1}h^{-d-1}\right) < \epsilon.$$

Similar to the proof of Lemma A.3, we can easily calculate that as $T \to \infty$

$$E\left[\lambda(\theta)^\tau P\lambda(\theta)\right] = T^2\int\int K^2\left(\frac{u-v}{h}\right)\lambda(u,\theta)\lambda(v,\theta)f(u,v)dudv = CT^2h^d(1+o(1)) \tag{B.4}$$

and

$$E\left[\lambda(\theta)^\tau\lambda(\theta)\right] = T\int\lambda^2(x,\theta)f(x)dx. \tag{B.5}$$

Thus, equations (B.4) and (B.5) imply

$$P\left(\frac{\lambda(\theta)^\tau\lambda(\theta)}{E\left[\lambda(\theta)^\tau P\lambda(\theta)\right]} > T^{-1}h^{-d-1}\right) \leq Th^{d+1}\frac{E\left[\lambda(\theta)^\tau\lambda(\theta)\right]}{E\left[\lambda(\theta)^\tau P\lambda(\theta)\right]} = C\frac{T^2h^{d+1}}{T^2h^d} \leq Ch \to 0$$

as $T \to \infty$. This finally finishes the proof of Lemma B.1(ii).

For simplicity, in the following lemmas and their proofs, we let $q = 1$. For $1 \leq i \leq 2$ and $1 \leq j \leq 3$, define

$$\psi_{ij}(X_t,\theta) = m_{i\theta}^{(j)}(X_t) = \frac{d^j m_{i\theta}(X_t)}{d\theta^j}.$$

Lemma B.2. *Under Assumptions A.1–A.2 and B.1–B.3, we have for any given $\theta \in \Theta$ and $i, j = 1, 2$*

$$T^{-1}J_T^{-1/2}\max_{h\in H_T} h^{-d/2}\left|\sum_{s=1}^T\sum_{t=1}^T p_{st}\xi_s\psi_{ij}(X_t,\theta)\right| = O_p(1). \tag{B.6}$$

Proof: It suffices to show that for any large constant $C_0 > 0$

$$P\left[T^{-1}J_T^{-1/2}\max_{h\in H_T} h^{-d/2}\left|\sum_{s=1}^T\sum_{t=1}^T p_{st}\xi_s\psi_{ij}(X_t,\theta)\right| > C_0\right]$$

$$\leq \sum_{h\in H_T} P\left[\left|\sum_{s=1}^T\sum_{t=1}^T p_{st}\xi_s\psi_{ij}(X_t,\theta)\right| > C_0TJ_T^{1/2}h^{d/2}\right] \leq \sum_{h\in H_T}\frac{1}{C_0^2T^2J_Th^d}E\left[\sum_{s=1}^T\sum_{t=1}^T p_{st}\xi_s\psi_{ij}(X_t,\theta)\right]^2$$

$$\leq \sum_{h\in H_T}\frac{1}{C_0^2T^2J_Th^d}\left\{\sum_{s=1}^T\sum_{t=1}^T E\left[p_{st}\xi_s\psi_{ij}(X_t,\theta)\right]^2 + \Lambda_{ijT}(\theta)\right\}, \tag{B.7}$$

where

$$\Lambda_{ijT}(\theta) = E\left[\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s\psi_{ij}(X_t,\theta)\right]^2 - \sum_{s=1}^{T}\sum_{t=1}^{T} E\left[p_{st}\xi_s\psi_{ij}(X_t,\theta)\right]^2.$$

Similar to (A.7) and (A.8), one can show that as $T \to \infty$

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[p_{st}\xi_s\psi_{ij}(X_t,\theta)\right]^2 = T^2 \int\int K\left(\frac{x-y}{h}\right)\sigma_\xi^2(x)\psi_{i1}^2(y,\theta)f(x,y)dxdy$$

$$= C(\theta)T^2h^d(1+o(1)) \quad (B.8)$$

for some function $C(\theta)$, where $\sigma_\xi^2(x) = E[\xi_t^2|X_t = x]$.

Similar to (A.11), one can show that as $T \to \infty$

$$\Lambda_{ijT}(\theta) = o(T^2h^d). \quad (B.9)$$

Thus, equations (B.7)–(B.9) complete the proof.

LEMMA B.3. *Under Assumptions A.1–A.2 and B.1–B.3, we have as $T \to \infty$*

$$T^{-1}J_T^{-1/2}\max_{h\in H_T} h^{-d/2}\max_{1\le s\le T}\left|\sum_{t=1}^{T} p_{st}\xi_t\right| = O_p(1). \quad (B.10)$$

PROOF: Similar to (B.7), we have for large constant $C_0 > 0$

$$P\left[T^{-1}J_T^{-1/2}\max_{h\in H_T} h^{-d/2}\max_{1\le s\le T}\left|\sum_{t=1}^{T} p_{st}\xi_t\right| > C_0\right]$$

$$\le \sum_{h\in H_T}\sum_{s=1}^{T} P\left[T^{-1}J_T^{-1/2}h^{-d/2}\left|\sum_{t=1}^{T} p_{st}\xi_t\right| > C_0\right] \le \frac{1}{C_0^2J_TT^2}\sum_{h\in H_T} h^{-d}\sum_{s=1}^{T} E\left[\sum_{t=1}^{T} p_{st}\xi_t\right]^2$$

$$= \frac{1}{C_0^2J_TT^2}\sum_{h\in H_T} h^{-d}\left\{\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[p_{st}^2\xi_t^2\right] + \sum_{s=1}^{T}\sum_{t_1\neq t_2} E\left[p_{st_1}p_{st_2}\xi_{t_1}\xi_{t_2}\right]\right\}. \quad (B.11)$$

Similar to (B.8), we can have as $T \to \infty$

$$\sum_{s=1}^{T}\sum_{t=1}^{T} E\left[p_{st}^2\xi_t^2\right] = T^2\int\int K^2\left(\frac{x-y}{h}\right)\sigma_\xi^2(x)f(x,y)dxdy = CT^2h^d(1+o(1)). \quad (B.12)$$

Analogous to (B.9), one can show that as $T \to \infty$

$$\sum_{s=1}^{T}\sum_{t_1\neq t_2} E\left[p_{st_1}p_{st_2}\xi_{t_1}\xi_{t_2}\right] = o(T^2h^d). \quad (B.13)$$

Thus, equations (B.11)–(B.13) complete the proof of (B.10).

LEMMA B.4. *Under Assumptions A.1–A.2 and B.1–B.3, we have for each $u > 0$ and $i, j = 1, 2$,*

$$\max_{h\in H_T}\sup_{|\theta-\theta_0|\le T^{-1/2}u} h^{-d/2}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s\lambda_{it}^j(\theta)\right| = O_p\left(J_T^{1/2}T^{1/2}\right) \quad (B.14)$$

under $\mathcal{H}_0$.

PROOF: We prove (B.14) for $i = j = 1$ only. Using a Taylor series expansion to $m_{1\theta}(X_t) - m_{1\theta_0}(X_t)$ and Assumption B.1, we have for θ' between θ and θ_0

$$\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_{1t}(\theta)\right| = \left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \left[m_{1\theta}(X_t) - m_{1\theta_0}(X_t)\right]\right|$$

$$\leq \left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \psi_{11}(X_t,\theta_0)\right| + \frac{1}{2}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \psi_{12}(X_t,\theta_0)\right| |\theta-\theta_0|^2$$

$$+\frac{1}{6}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \psi_{13}(X_t,\theta')\right| |\theta-\theta_0|^3 \leq \left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \psi_{11}(X_t,\theta_0)\right| |\theta-\theta_0|$$

$$+\frac{1}{2}T|\theta-\theta_0|^2\left|\sum_{s=1}^{T} p_{st}\xi_s \psi_{12}(X_t,\theta_0)\right| + \frac{1}{6}T|\theta-\theta_0|^3 \max_{1\leq t\leq T}\left|\sum_{s=1}^{T} p_{st}\xi_s\right| \cdot \max_{1\leq t\leq T}\left|\psi_{13}(X_t,\theta')\right|. \quad (B.15)$$

Hence, (B.6), (B.10), (B.15) and Assumption B.1(i) imply

$$\max_{h\in H_T}\sup_{||\theta-\theta_0||\leq T^{-1/2}u} h^{-d/2}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_{1t}(\theta)\right| \leq O_p\left(J_T^{1/2}T^{1/2}\right). \quad (B.16)$$

The proof of (B.14) follows from (B.15) and (B.16).

LEMMA B.5. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Then for every* $u > 0$, $i, j = 1, 2$, *some* $h \in H_T$ *and as* $T \to \infty$

$$\sup_{|\theta-\theta^*|\leq T^{-1/2}u}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_i^j(X_t,\theta)\right| = o_p(Q_T(\theta^*)). \quad (B.17)$$

under $\mathcal{H}_1$.

PROOF: Similar to the proof of Lemma B.1(ii), in order to prove (B.17), it suffices to show that

$$\sup_{|\theta-\theta^*|\leq T^{-1/2}u}\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_i^j(X_t,\theta)\right| = o_p(q_T), \quad (B.18)$$

where $q_T = E\left[Q_T(\theta^*)\right]$.

We consider the case of $i = j = 1$ only, as the others follow similarly. Note that

$$\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_1(X_t,\theta)\right| \leq \left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \lambda_1(X_t,\theta^*)\right|$$

$$+\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s \psi_{11}(X_t,\theta^*)\right| |\theta-\theta^*| + \frac{1}{2}T|\theta-\theta^*|^2\left|\sum_{s=1}^{T} p_{st}\xi_s \psi_{12}(X_t,\theta^*)\right|$$

$$+\frac{1}{6}T|\theta-\theta^*|^3 \max_{1\leq t\leq T}\left|\sum_{s=1}^{T} p_{st}\xi_s\right| \cdot \max_{1\leq t\leq T}\left|\psi_{13}(X_t,\theta')\right|, \quad (B.19)$$

where θ' lies between θ and θ^*.

In view of (B.6), (B.10), (B.19), Assumptions B.1(i) and B.3(ii), in order to prove (B.17), it suffices to show that for any $\delta > 0$ and as $T \to \infty$

$$P\left[\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s\lambda_1(X_t,\theta^*)\right| > \delta q_T\right] \to 0.$$

Similar to (B.8) and (B.9), one can show that as $T \to \infty$

$$E\left[\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s\lambda_1(X_t,\theta^*)\right]^2 = CT^2h^d(1+o(1)). \qquad (B.20)$$

Thus, equations (B.19) and (B.20) imply that as $T \to \infty$

$$P\left[\left|\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}\xi_s\lambda_1(X_t,\theta^*)\right| > \delta q_T\right] \leq \frac{1}{\delta^2 q_T^2} E\left[\sum_{s=1}^{T}\sum_{t=1}^{T} a_{st}\xi_s\lambda_1(X_t,\theta^*)\right]^2 = \frac{CT^2h^d(1+o(1))}{q_T^2} \to 0$$

using $q_T = CT^2h^d(1+o(1))$ given in (B.4) above, where C is a constant independent of T. Lemma B.5 is therefore proved.

Before establishing some other lemmas, we introduce the following notation. Let $\xi_t = \epsilon_t + \eta_t$, $\xi = (\xi_1, \ldots, \xi_T)^\tau$,

$$N_{20}(h) = \xi^\tau P\xi, \quad \text{and} \quad N_{2T}(h) = N_{2T}(h,\tilde{\theta}) = \hat{W}^\tau P\hat{W}.$$

Let $\mu_h = E\left[\hat{\mu}_h\right]$ and $\sigma_h^2 = E\left[\hat{\sigma}_h^2\right]$. It can easily be calculated that

$$\mu_h = T(1+o(1))K(0)\sigma_\xi^2(x)f(x)dx,$$

$$\sigma_h^2 = E\left[\hat{\sigma}_h^2\right] = 2\sum_{s,t=1}^{T} E\left[p_{st}^2\hat{W}_s^2\hat{W}_t^2\right]$$

$$= 2T^2h^d(1+o(1))\int K^2(x)dx \cdot \int \sigma_\xi^4(y)f^2(y)dy \equiv C_\sigma^2T^2h^d(1+o(1)),$$

where $\sigma_\xi^2(x) = E[\xi_t^2|X_t = x]$, $f(x)$ is the marginal density of $\{X_t\}$, and $C_\sigma^2 > 0$ is a constant.

It follows that

$$N_{2T}(h,\theta) = N_{20}(h) + Q_T(\theta) + \Pi_T(\theta) + R_T(\theta), \qquad (B.21)$$

where $Q_T(\theta)$ and $\Pi_T(\theta)$ are as defined in (B.1) above, and $R_T(\theta)$ is the remainder term given by

$$R_T(\theta) = N_{2T}(h,\theta) - N_{20}(h) - Q_T(\theta) - \Pi_T(\theta).$$

Define

$$L_{20}(h) = \frac{N_{20}(h)-\mu_h}{\sigma_h},\ L_{2T}(h) = \frac{N_{2T}(h,\tilde{\theta})-\hat{\mu}_h}{\hat{\sigma}_h} \ \text{and}\ \tilde{L}_{2T}(h) = \frac{N_{2T}(h,\theta^*)-\mu_h}{\sigma_h}. \qquad (B.22)$$

Note that $L_{20}(h) = \tilde{L}_{2T}(h)$ when $\theta^* = \theta_0$.

Assume that $\{e_t^*\}$ is as defined in the Simulation Scheme of Section 3.1. Let

$$\tilde{\epsilon}_t = \sigma(X_t)e_t^* \ \text{and}\ \tilde{\eta}_t = \sigma^2(X_t)[(e_t^*)^2 - 1].$$

Let $\tilde{L}_{20}(h)$ be the version of $L_{20}(h)$ with $\epsilon_t = \sigma(X_t)e_t$ and $\eta_t = \sigma^2(X_t)[e_t^2 - 1]$ be replaced with $\tilde{\epsilon}_t$ and $\tilde{\eta}_t$ respectively.

For each $t = 1, 2, \ldots, T$, generate $Y_t^* = m_{1\tilde{\theta}}(X_t) + \sqrt{m_{2\tilde{\theta}}(X_t)}e_t^*$. Use the data set $\{Y_t^*, X_t : 1 \leq t \leq T\}$ to re-estimate θ. Denote the resulting estimate by $\hat{\theta}$. Let $\hat{L}_{2T}(h)$ be the version of $L_{2T}(h)$ of (B.22) with $\tilde{\theta}$, ϵ_t and η_t replaced with $\hat{\theta}$, $\sqrt{m_{2\hat{\theta}}(X_t)}e_t^*$ and $m_{2\hat{\theta}}(X_t)[(e_t^*)^2 - 1]$ respectively.

LEMMA B.6. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Then as $T \to \infty$*

$$L_{2T}(h) = \tilde{L}_{2T}(h) + o_p(1) \quad \text{and} \quad \hat{L}_{2T}(h) = \tilde{L}_{20}(h) + o_p(1) \tag{B.23}$$

uniformly over $h \in H_T$.

PROOF: The proof of (B.23) follows from Lemmas B.1–B.5.

LEMMA B.7. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Then $\max_{h \in H_T} \tilde{L}_{2T}(h)$ and $\max_{h \in H_T} \tilde{L}_{20}(h)$ have identical asymptotic distributions under $\mathcal{H}_0$.*

PROOF: In view of Lemmas B.1–B.6, in order to prove Lemma B.7, it suffices to show that $\max_{h \in H_T} L_{20}(h)$ and $\max_{h \in H_T} \tilde{L}_{20}(h)$ are asymptotically the same. For $h \in H_T$, let $u_t = \xi_t$ or $\tilde{\xi}_t = \tilde{\epsilon}_t + \tilde{\eta}_t$, define

$$B_{hT}(u_1, \ldots, u_T) = (C_\sigma T h^{d/2})^{-1} \left[\sum_{s \neq t} p_{st} u_s u_t\right]. \tag{B.24}$$

Let $B_T(u_1, \ldots, u_T)$ be the sequence obtained by stacking the corresponding $B_{hT}(u_1, \ldots, u_T)$. Let $G(\cdot) = G_T(\cdot)$ be a 3–times continuously differentiable function over R^{J_T}. Define

$$C_T(G) = \sum_{x \in R^{J_T}} \max_{i,j,k=1,2,\ldots,J_T} \left|\frac{\partial^3 G(v)}{\partial v_i \partial v_j \partial v_k}\right|.$$

The proof of Lemma B.7 is divided into two steps. The first step is to show that

$$\left|E\left[G(B_T(\xi_1, \ldots, \xi_T))\right] - E\left[G(B_T(\tilde{\xi}_1, \ldots, \tilde{\xi}_T))\right]\right| \leq C_0 C_T(G) J_T^2 h_{\max}^d \tag{B.25}$$

for any 3–times differentiable $G(\cdot)$, some finite constant C_0, and all sufficiently large T.

The second step is to use (B.25) to show that $B_T(\xi_1, \ldots, \xi_T)$ and $B_T(\tilde{\xi}_1, \ldots, \tilde{\xi}_T)$ have the same asymptotic distribution.

Throughout the rest of the proof of Lemma B.7, we assume without loss of generality that $\sigma(X_t) = C_\sigma = 1$ and replace p_{st} in (B.24) with $\tilde{p}_{st}(h) = (Th^{d/2})^{-1} p_{st}$.

We can easily show that

$$\left|E\left[G(B_T(\xi_1, \ldots, \xi_T))\right] - E\left[G(B_T(\tilde{\xi}_1, \ldots, \tilde{\xi}_T))\right]\right|$$

$$\leq \sum_{t=1}^{T} \left|E\left[G(B_T(\xi_1, \ldots, \xi_t, \tilde{\xi}_{t+1}, \cdots, \tilde{\xi}_T))\right] - E\left[G(B_T(\xi_1, \ldots, \xi_{t-1}, \tilde{\xi}_t, \ldots, \tilde{\xi}_T))\right]\right|, \tag{B.26}$$

where $B_T(\xi_1, \ldots, \xi_T, \tilde{\xi}_{T+1}) = B_T(\xi_1, \ldots, \xi_T)$ and $B_T(\xi_0, \tilde{\xi}_1, \ldots, \tilde{\xi}_T) = B_T(\tilde{\xi}_1, \ldots, \tilde{\xi}_T)$.

We now derive an upper bound on the last term of the sum on the right–hand side of (B.26).

Let U_{T-1}, Λ_T and $\tilde{\Lambda}_T$, respectively, denote the vectors that are obtained by stacking

$$U_{h,T} = \sum_{s=1}^{T-1} \sum_{t=1,\neq s}^{T-1} \tilde{p}_{st}(h)\xi_s\xi_t, \quad \Lambda_{h,T} = 2\xi_T \sum_{s=1}^{T-1} \tilde{p}_{sT}(h)\xi_s, \quad \tilde{\Lambda}_{h,T} = 2\tilde{\xi}_T \sum_{s=1}^{T-1} \tilde{p}_{sT}(h)\xi_s.$$

Then a Taylor expansion of the last term of the sum on the right–hand side of (B.26) about $\xi_T = \tilde{\xi}_T = 0$ yields

$$\left| E\left[G(B_T(\xi_1, \ldots, \xi_T))\right] - E\left[G(B_T(\xi_1, \ldots, \xi_{T-1}, \tilde{\xi}_T))\right] \right| \leq \left| E\left[G'(U_{T-1})(\Lambda_T - \tilde{\Lambda}_T)\right] \right|$$

$$+\frac{C_{1T}(G)}{2} \left\{ E\left[||\Lambda_T||^2\right] + E\left[||\tilde{\Lambda}_T||^2\right] \right\} + \frac{C_{2T}(G)}{6} \left\{ E\left[||\Lambda_T||^3\right] + E\left[||\tilde{\Lambda}_T||^3\right] \right\},$$

where G' and G'' denote the gradient and matrix of second derivatives of G, and $C_{1T}(G)$ and $C_{2T}(G)$ are positive constants possibly depending on T.

Since Assumption A.1(ii) implies

$$E[\xi_T|\xi_1, \ldots, \xi_{T-1}] = E[\tilde{\xi}_T|\xi_1, \ldots, \xi_{T-1}] = 0,$$

we have

$$E\left[\Lambda_T - \tilde{\Lambda}_T|\xi_1, \ldots, \xi_{T-1}\right] = 0.$$

Therefore

$$\left| E\left[G(B_T(\xi_1, \ldots, \xi_T))\right] - E\left[G(B_T(\xi_1, \ldots, \xi_{T-1}, \tilde{\xi}_T))\right] \right|$$

$$\leq \frac{C_{1T}(G)}{2} \left\{ E\left[||\Lambda_T||^2\right] + E\left[||\tilde{\Lambda}_T||^2\right] \right\} + \frac{C_{2T}(G)}{6} \left\{ E\left[||\Lambda_T||^3\right] + E\left[||\tilde{\Lambda}_T||^3\right] \right\}. \qquad (B.27)$$

To find an upper bound for the right–hand side of (B.27), let $\tilde{P}_{sT}$ be the vector that is obtained by stacking $\tilde{p}_{sT}(h)$ $(h \in H_T)$. Let $C_1 = 4E\left[\xi_T^2\right]$ and $C_2 = 8E\left[|\xi_T|^3\right]$. We then have as $T \to \infty$

$$E\left[||\Lambda_T||^2\right] = 4E\left[|\xi_T|^2\right] E\left[\left|\left|\sum_{s=1}^{T-1} \tilde{P}_{sT}\xi_s\right|\right|^2\right] = 4E\left[|\xi_T|^2\right] \cdot E\left[\sum_{h\in H_T}\left(\sum_{s=1}^{T-1} \tilde{p}_{sT}(h)\xi_s\right)\right]^2$$

$$= C_1 \sum_{k\in H_T} \sum_{h\in H_T} E\left[\sum_{s=1}^{T-1} \tilde{p}_{sT}(h)\tilde{p}_{sT}(k)\xi_s^2 + \sum_{s=1}^{T-1} \sum_{t=1,t\neq s}^{T-1} \tilde{p}_{sT}(h)\tilde{p}_{tT}(k)\xi_s\xi_t\right]$$

$$= C_1 \sum_{k\in H_T} \sum_{h\in H_T} (hk)^{-d/2} T^{-2} \left(\sum_{s=1}^{T-1} E\left[p_{sT}(h)p_{sT}(k)\xi_s^2\right] + \sum_{s=1}^{T-1} \sum_{t=1,t\neq s}^{T-1} E\left[p_{sT}(h)p_{tT}(k)\xi_s\xi_t\right]\right)$$

$$= \frac{C_1}{T} \sum_{k\in H_T} \sum_{h\in H_T} (hk)^{-d/2} \int\int\int K\left(\frac{u-w}{h}\right) K\left(\frac{u-w}{k}\right) x^2 f(u, w, x) du dw dx$$

$$+C_1 \sum_{k\in H_T} \sum_{h\in H_T} (hk)^{-d/2} \int \cdots \int K\left(\frac{u-w}{h}\right) K\left(\frac{v-w}{k}\right) xy f(u, v, w, x, y) du dv dw dx dy$$

$$= \frac{C_1}{T} \sum_{k\in H_T} \sum_{h\in H_T} h^{d/2} k^{-d/2} \int\int\int K(x_1) K\left(\frac{h}{k} x_1\right) x_3^2 f(x_2 + x_1 h, x_2, x_3) dx_1 dx_2 dx_3$$

$$+C_1 \sum_{k\in H_T} \sum_{h\in H_T} (hk)^{d/2} \int \cdots \int K(x_1)K(x_2)x_4x_5 f(x_3 + hx_1, x_3 + kx_2, x_3, x_4, x_5) dx_1 \cdots dx_5,$$

$$\leq CJ_T^2 h_{\max}^d,$$

where $f(u,w,x)$ denotes the joint density function of (X_s, X_T, ξ_s) and $f(u,v,w,x,y)$ denotes the joint density function of $(X_s, X_t, X_T, \xi_s, \xi_t)$.

Similarly, we obtain

$$\sum_{h,k\in H_T} (hk)^{-d} E\left[\sum_{s,t,u,v=1}^{T-1} \tilde{p}_{sT}(h)\xi_s \tilde{p}_{tT}(h)\xi_t \tilde{p}_{uT}(k)\xi_u \tilde{p}_{vT}(k)\xi_v\right]$$

$$\begin{aligned}
&= \sum_{h,k\in H_T} (hk)^{-d} E\left(\sum_{s=1}^{T-1}\left[p_{sT}^2(h)p_{sT}^2(k)\xi_s^4\right] + \sum_{s\neq t}^{T-1}\left[p_{sT}^2(h)p_{tT}^2(k) + 2p_{sT}(h)p_{sT}(k)p_{tT}(h)p_{tT}(k)\right]\xi_s^2\xi_t^2\right)\\
&+ 2\sum_{h,k\in H_T} (hk)^{-d} E\left(\sum_{s\neq t}^{T-1} p_{sT}^2(h)p_{sT}(k)p_{tT}(k)\xi_s^3\xi_t + \sum_{s\neq t\neq u}^{T-1} p_{sT}^2(h)p_{tT}(k)p_{uT}(k)\xi_s^2\xi_t\xi_u\right)\\
&+ 2\sum_{h,k\in H_T} (hk)^{-d} E\left(\sum_{s\neq t\neq u\neq v}^{T-1} p_{sT}(h)p_{tT}(k)p_{uT}(k)p_{vT}(k)\xi_s\xi_t\xi_u\xi_v\right)\\
&= \sum_{h,k\in H_T} \frac{(hk)^{-d}}{T^3}\int\int\int K^2\left(\frac{u-w}{h}\right)K^2\left(\frac{u-w}{k}\right)x^4 f(u,w,x)dudwdx\\
&+ \sum_{h,k\in H_T} \frac{(hk)^{-d}}{T^2}\int\cdots\int K^2\left(\frac{u-w}{h}\right)K^2\left(\frac{v-w}{k}\right)x^2y^2 f(u,v,w,x,y)dudvdwdxdy\\
&+ \sum_{h,k\in H_T} \frac{2}{T^2h^dk^d}\int\cdots\int K\left(\frac{u-w}{h}\right)K\left(\frac{u-w}{k}\right)K\left(\frac{v-w}{h}\right)K\left(\frac{v-w}{k}\right)x^2y^2 f(u,v,w,x,y)dudvdwdxdy\\
&+ \sum_{h,k\in H_T} \frac{4}{T^2h^dk^d}\int\cdots\int K^2\left(\frac{u-w}{h}\right)K\left(\frac{u-w}{k}\right)K\left(\frac{v-w}{k}\right)x^3y f(u,v,w,x,y)dudvdwdxdy\\
&+ \sum_{h,k\in H_T} \frac{4}{Th^dk^d}\int\cdots\int K^2\left(\frac{u-w}{h}\right)K\left(\frac{v-w}{k}\right)K\left(\frac{s-w}{k}\right)x^2yz\\
&\times\ f(u,v,w,s,x,y,z)dudvdwdsdxdydz\\
&+ \sum_{h,k\in H_T} \frac{1}{h^dk^d}\int\cdots\int K\left(\frac{u-w}{h}\right)K\left(\frac{v-w}{h}\right)K\left(\frac{s-w}{k}\right)K\left(\frac{t-w}{k}\right)x_1x_2x_3x_4\\
&\times\ f(u,v,w,s,t,x_1,x_2,x_3,x_4)dudvdwdsdtdx_1dx_2dx_3dx_4\\
&\leq\ CJ_T^2h_{\max}^{2d}
\end{aligned}$$

using Assumptions A.1(iii), A.2(ii) and B.3, where $C > 0$ is a constant independent of T.

This implies that as $T \to \infty$

$$E\left[||\Lambda_T||^3\right] = C_2 E\left[\left|\left|\sum_{s=1}^{T-1}\tilde{P}_{sT}\xi_s\right|\right|^3\right] \leq C_2\left\{E\left[\sum_{h\in H_T}\left(\sum_{s=1}^{T-1}\tilde{p}_{sT}(h)\xi_s\right)^2\right]^2\right\}^{3/4}$$

$$= C_2\left\{\sum_{h\in H_T}\sum_{k\in H_T} E\left[\sum_{s,t,u,v=1}^{T-1}\tilde{p}_{sT}(h)\xi_s\tilde{p}_{tT}(h)\xi_t\tilde{p}_{uT}(k)\xi_u\tilde{p}_{vT}(k)\xi_v\right]\right\}^{3/4}$$

$$\leq C\left(J_Th_{\max}^d\right)^{3/2}.$$

A similar result holds for $E\left[||\tilde{\Lambda}_T||^3\right]$. Thus

$$E\left[||\Lambda_T||^3\right] + E\left[||\tilde{\Lambda}_T||^3\right] \leq 2C\left(J_T h_{\max}^d\right)^{3/2}. \qquad (B.28)$$

Equations (B.27)–(B.28) therefore imply

$$\left|E\left[G(B_T(\xi_1, \ldots, \xi_T))\right] - E\left[G(B_T(\tilde{\xi}_1, \ldots, \tilde{\xi}_T))\right]\right|$$

$$\leq C_0 C_T(G)\left(J_T^2 h_{\max}^d \left[\frac{1}{T h_{\min}^d} + 1\right] + \left(J_T h_{\max}^d\right)^{3/2}\right) \leq C_0 C_T(G) J_T^2 h_{\max}^d.$$

This finishes the first step.

Step 2: It suffices to show that for any real x

$$\lim_{T\to\infty}\left\{P\left[\max_{h\in H_T} B_{hT}(\xi_1, \ldots, \xi_T) \leq x\right] - P\left[\max_{h\in H_T} B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) \leq x\right]\right\} = 0$$

or, equivalently, that

$$\lim_{T\to\infty}\left|E\left(\prod_{h\in H_T} I\left[B_{hT}(\xi_1, \ldots, \xi_T) \leq x\right]\right) - E\left(\prod_{h\in H_T} I\left[B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) \leq x\right]\right)\right| = 0.$$

Assume that $G(\cdot)$ is a nondecreasing and three times continuously differentiable function on the real line and that it satisfies $G(v) = 0$ if $v \leq -1$ and $G(v) = 1$ if $v \geq 0$. Let $\delta_T = J_T^{-2}$. A simple calculation shows that

$$\left|E\left(\prod_{h\in H_T} I\left[B_{hT}(\xi_1, \ldots, \xi_T) \leq x\right]\right) - E\left(\prod_{h\in H_T} I\left[B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) \leq x\right]\right)\right|$$

$$\leq \left|E\left(\prod_{h\in H_T} G\left[\frac{B_{hT}(\xi_1, \ldots, \xi_T) - x}{\delta_T}\right]\right) - E\left(\prod_{h\in H_T} G\left[\frac{B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) - x}{\delta_T}\right]\right)\right|$$

$$+ \sum_{h\in H_T} E\left|G\left[\frac{B_{hT}(\xi_1, \ldots, \xi_T) - x}{\delta_T}\right] - I\left[B_{hT}(\xi_1, \ldots, \xi_T) \leq x\right]\right|$$

$$+ \sum_{h\in H_T} E\left|G\left[\frac{B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) - x}{\delta_T}\right] - I\left[B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) \leq x\right]\right|. \qquad (B.29)$$

Each term of the summands of the second two sums on the right–hand side of (B.29) is bounded from above by $J_T \delta_T = J_T^{-1}$. Thus, using (B.26) to bound the first term on the right–hand side of (B.29) yields

$$\left|P\left[\max_{h\in H_T} B_{hT}(\xi_1, \ldots, \xi_T) \leq x\right] - P\left[\max_{h\in H_T} B_{hT}(\tilde{\xi}_1, \ldots, \tilde{\xi}_T) \leq x\right]\right|$$

$$\leq C_0 C_T(G) J_T^2 h_{\max}^d + 2J_T^{-1} \to 0 \qquad (B.30)$$

as $T \to \infty$. This finally completes the proof of Lemma B.7.

Remark B.2. As a result of Lemma B.7, we can obtain that both $\max_{h\in H_T} \tilde{L}_{20}(h)$ and $\max_{h\in H_T} L_{20}(h)$ have identical asymptotic distributions. This result will be used in the proof of Lemma B.10 below.

Lemma B.8. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Then for any* $x \geq 0$, $h \in H_T$ *and all sufficiently large* T

$$P\left(\tilde{L}_{20}(h) > x\right) \leq \exp\left(-\frac{x^2}{4}\right).$$

Proof: Similar to the proof of (A.18), we obtain that for any small $\delta > 0$ there exists a large integer $T_0 \geq 1$ such that for $T \geq T_0$

$$\left|P(\tilde{L}_{20}(h) \leq x) - \Phi(x)\right| < \delta,$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$.

This implies for any $T \geq T_0$ and $x \geq 0$

$$\begin{aligned} & P(\tilde{L}_{20}(h) > x) \leq 1 - \Phi(x) + \delta \\ &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du + \delta = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{4}} e^{-\frac{u^2}{4}} du + \delta \\ &\leq e^{-\frac{x^2}{4}} \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{4}} du + \delta \leq e^{-\frac{x^2}{4}} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{u^2}{4}} du + \delta \\ &= e^{-\frac{x^2}{4}} \frac{\sqrt{2}}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{v^2}{2}} dv + \delta = \frac{\sqrt{2}}{2} e^{-\frac{x^2}{4}} + \delta \end{aligned}$$

using $\frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{v^2}{2}} dv = \frac{1}{2}$.

The proof follows by letting $0 < \delta \leq \left(1 - \frac{\sqrt{2}}{2}\right) e^{-\frac{x^2}{4}}$ for any $x \geq 0$.

For $0 < \alpha < 1$, define $\tilde{l}_\alpha$ to be the $1-\alpha$ quantile of $\max_{h \in H_T} \tilde{L}_{20}(h)$.

Lemma B.9. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Then for large enough* T

$$\tilde{l}_\alpha \leq 2\sqrt{\log(J_T) - \log(\alpha)}.$$

Proof: The proof is trivial and similar to that of Lemma 12 of Horowitz and Spokoiny (2001).

Lemma B.10. *Suppose that Assumptions A.1–A.2 and B.1–B.3 hold. Suppose that*

$$\lim_{T \to \infty} P\left(\frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right) = 1 \tag{B.31}$$

for some $h \in H_T$, *where*

$$\tilde{l}_\alpha^* = \max\left(\tilde{l}_\alpha, \sqrt{2\log(J_T) + \sqrt{2\log(J_T)}}\right).$$

Then

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

Proof: By Lemma B.6, L^* can be replaced by $\max_{h \in H_T} \tilde{L}_{2T}(h)$. By Lemmas B.6 and B.7, l_α can be replaced by $\tilde{l}_\alpha$. Thus, it suffices to show that

$$\lim_{T \to \infty} P(\max_{h \in H_T} \tilde{L}_{2T}(h) > \tilde{l}_\alpha) = 1,$$

which holds if $\lim_{T\to\infty} P(\tilde{L}_{2T}(h) > \tilde{l}_\alpha) = 1$ for some $h \in H_T$. For any $h \in H_T$, using Remark B.2 and then Lemma B.5 we have [10]

$$\tilde{L}_{2T}(h) = L_{20}(h) + \frac{Q_T(\theta^*) + \Pi_T(\theta^*)}{\sigma_h} = \tilde{L}_{20}(h) + \frac{Q_T(\theta^*) + \Pi_T(\theta^*)}{\sigma_h} + o_p(1)$$
$$= \tilde{L}_{20}(h) + \frac{Q_T(\theta^*)(1 + o_p(1))}{\sigma_h} + o_p(1). \qquad (B.32)$$

On the other hand, condition (B.31) implies that as $T \to \infty$

$$P\left(\frac{Q_T(\theta^*)}{\sigma_h} < 2\tilde{l}_\alpha^*\right) \to 0. \qquad (B.33)$$

Observe that

$$P(\tilde{L}_{2T}(h) > \tilde{l}_\alpha) = P\left(\tilde{L}_{2T}(h) > \tilde{l}_\alpha, \frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right) + P\left(\tilde{L}_{2T}(h) > \tilde{l}_\alpha, \frac{Q_T(\theta^*)}{\sigma_h} < 2\tilde{l}_\alpha^*\right)$$
$$\equiv I_{1T} + I_{2T}.$$

Thus, it follows from (B.31)–(B.32) that as $T \to \infty$

$$I_{1T} = P\left(\tilde{L}_{20}(h) + \frac{Q_T(\theta^*) + \Pi_T(\theta^*)}{\sigma_h} > \tilde{l}_\alpha | \frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right) P\left(\frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right)$$
$$\geq P\left(\tilde{L}_{20}(h) > \tilde{l}_\alpha - 2\tilde{l}_\alpha^* | \frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right) P\left(\frac{Q_T(\theta^*)}{\sigma_h} \geq 2\tilde{l}_\alpha^*\right) \to 1 \qquad (B.34)$$

because $\tilde{L}_{20}(h)$ is asymptotically normal and therefore bounded in probability and $\tilde{l}_\alpha - 2\tilde{l}_\alpha^* \to -\infty$ as $T \to \infty$.

Because of (B.33), as $T \to \infty$

$$I_{2T} \leq P\left(\frac{Q_T(\theta^*)}{\sigma_h} < 2\tilde{l}_\alpha^*\right) \to 0. \qquad (B.35)$$

Equations (B.34) and (B.35) complete the proof. B.3. *Proofs of Theorems 3.1–3.4*

Proof of Theorem 3.1: By Lemma B.6, $\max_{h\in H_T} \tilde{L}_{2T}(h) = \max_{h\in H_T} L_{2T}(h) + o_p(1)$. By Lemma B.7, $\max_{h\in H_T} \hat{L}_{2T}(h) - \max_{h\in H_T} \tilde{L}_{20}(h) \to 0$ in distribution as $T \to \infty$. Furthermore, Lemma B.6 implies $\max_{h\in H_T} L_{2T}(h) - \max_{h\in H_T} \hat{L}_{2T}(h) \to 0$ in distribution as $T \to \infty$ when $\mathcal{H}_0$ holds. This finishes the proof.

Proof of Theorem 3.4: For the proof of Theorem 3.4, one needs to use the conditions of Theorem 3.4 to finish the proof. For our case, we don't need Lemma 14 of Horowitz and Spokoiny (2001), although it holds in probability in our case. In our proof, we mainly use Lemma B.1(ii). It follows from Lemma B.1(ii) that for every $\theta \in \Theta$

$$\lim_{T\to\infty} P\left(Q_T(\theta) \geq h^{d+1}\lambda(\theta)^\tau\lambda(\theta)\right) = 1. \qquad (B.36)$$

[10]Note that the derivation of $T_{h0} = \tilde{T}_{h0} + \frac{||b_h(\theta^*)||^2 + 2b_h^\tau(\theta^*)W_h\epsilon}{V_h}$ in the proof of Lemma 13 of Horowitz and Spokoiny (2001) should be $T_{h0} = T_h^0 + \frac{||b_h(\theta^*)||^2 + 2b_h^\tau(\theta^*)W_h\epsilon}{V_h}$, where $T_h^0 = \frac{||W_h\epsilon||^2 - N_h}{V_h}$. Thus, in the proof of Lemma 13 of Horowitz and Spokoiny (2001), one needs to use their Lemma 10 again.

In view of (B.36) and the definition of $\tilde{l}_\alpha^*$, in order to verify (B.31), it suffices to show that

$$\lim_{T\to\infty} P\left(h^{d+\eta}\lambda(\theta)^\tau\lambda(\theta) \geq 4\tilde{l}_\alpha^* h^{d/2}\right) = 1,$$

which follows from the condition of Theorem 3.4 that

$$\lim_{T\to\infty} P\left(\rho(m,\mathcal{M}) \geq C_m \left(T^{-1}\sqrt{\mathrm{loglogT}}\right)^{2s/(4s+d)}\right) = 1$$

and the fact that for an absolute constant $C_0 > 0$,

$$\liminf_{T\to\infty} \frac{Th^{\frac{(d+2)(4s+d)}{2(2s+d)}}}{\sqrt{\mathrm{loglogT}}} \geq \liminf_{T\to\infty} \frac{Th_{\min}^{\frac{(d+2)(4s+d)}{2(2s+d)}}}{\sqrt{\mathrm{loglogT}}} \geq C_0$$

using Assumption B.3. This completes the proof of Theorem 3.4.

Proofs of Theorem 3.2–3.3: One can follow the corresponding proofs of Theorems 2–3 of Horowitz and Spokoiny (2001). For the proofs of Theorems 3.2 and 3.3, one needs only to modify the proofs of their Theorems 2 and 3 slightly by using the fact that conditions (3.3)–(3.5) now hold in probability. Alternatively, similar to the proof of Theorem 3.4 above one can use (B.34) and the fact that conditions (3.3)–(3.5) now hold in probability to complete the proofs.

REFERENCES

Aït-Sahalia, Y. (1996a): "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Aït-Sahalia, Y. (1996b): "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

Aït-Sahalia, Y., P. Bickel, and T. Stoker (2001): "Goodness-of-fit Tests for Regression using Kernel Methods," *Journal of Econometrics*, 105, 363–412

Andrews, D. W. K. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1028.

Bierens, H. J. (1982) "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Chen, S., and J. Gao (2003): "Simultaneous Specification Tests for the Mean and Variance Structures of Time Series Regression," working paper, Department of Statistics and Applied Probability, National University of Singapore.

Chen, S. X., W. Härdle, and M. Li (2003): "An Empirical Likelihood Goodness–of–Fit Test for Time Series," *Journal of the Royal Statistical Society, Ser. B.*, 65, 663–678.

Chen, X., and Y. Fan (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373–401.

Eubank, R. L., and J. D. Hart (1992): "Testing Goodness-of-Fit in Regression via Order Selection," *The Annals of Statistics*, 20, 1412–1425.

Eubank, R. L., and C. H. Spiegelman (1990): "Testing the Goodness of Fit of a Linear Model via

Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387–392.

FAN, J., AND L. S. HUANG (2001): "Goodness-of-fit Tests for Parametric Regression Models," *Journal of the American Statistical Association*, 453, 640–652.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865–890.

FAN, Y., AND Q. LI (1998): "Central Limit Theorem for Degenerate U–Statistics of Absolutely Regular Processes with Applications to Model Specification Testing," *Journal of Nonparametric Statistics*, 10, 245–271.

GAO, J. (2002): "Estimation of Continuous–Time Financial Models with Long–Range Dependence," working paper, School of Mathematics and Statistics, The University of Western Australia, Australia. Available at www.maths.uwa.edu.au/˜jiti/jap.pdf.

GAO, J., AND M. KING (2001): "Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression Models," working paper, School of Mathematics and Statistics, The University of Western Australia, Australia. Available at www.maths.uwa.edu.au/˜jiti/jems.pdf.

GAO, J., H. TONG, AND R. WOLFF (2002): "Model Specification Tests in Nonparametric Stochastic Regression Models," *Journal of Multivariate Analysis*, 83, 324–359.

GOZALO, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9, 451–477.

GOZALO, P. L., AND O. LINTON (2001): "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," *Journal of Econometrics*, 104, 1–48.

HÄRDLE, W., H. LIANG, H., AND J. GAO (2000): *Partially Linear Models.* New York: Physica–Verlag.

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

HART, J. (1997): *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

HJELLVIK, V., AND D. TJØSTHEIM (1995): "Nonparametric Tests of Linearity for Time Series," *Biometrika*, 82, 351–368.

HJELLVIK, V., Q. YAO, Q., AND D. TJØSTHEIM (1998): "Linearity Testing using Local Polynomial Approximation," *Journal of Statistical Planning and Inference*, 68, 295-321.

HONG, Y., AND Y. J. LEE (2002): "Generalized Spectral Tests for Conditional Mean Specification in Time Series with Conditional Heteroscedasticity of Unknown Form," working paper, Department of Economics and Department of Statistical Science, Cornell University.

HONG, Y., AND H. LI (2002): "Nonparametric Specification Testing for Continuous–Time Models with Application to Spot Interest Rates," working paper, Department of Economics and Department of Statistical Science, Cornell University.

HONG, Y., AND H. WHITE (1995): "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica*, 63, 1133–1159.

HOROWITZ, J. L., AND W. HÄRDLE (1994): "Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory*, 10, 821–848.

HOROWITZ, J. L., AND V. G. SPOKOINY (2001): An Adaptive, Rate–Optimal Test of a Parametric Mean–Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–632.

JAYASURIYA, B. (1996): "Testing for Polynomial Regression Using Nonparametric Regression Techniques.

Journal of the American Statistical Association, 91, 1626–1631.

Karlsen, H., and D. Tjøstheim (2001): "Nonparametric Estimation in Null Recurrent Time Series," *The Annals of Statistics*, 29, 372–416.

Li, Q. (1999): "Consistent Model Specification Tests for Time Series Econometric Models," *Journal of Econometrics*, 92, 101–147.

Li, Q., and C. Hsiao (1998): "Testing Serial Correlation in Semiparametric Panel Data Models," *Journal of Econometrics*, 87, 207–237.

Li, Q., and S. Wang (1998): "A Simple Consistent Bootstrap Tests for a Parametric Regression Functional Form," *Journal of Econometrics*, 87, 145–165.

Robinson, P. (1988): "Root–N–Consistent Semiparametric Regression," *Econometrica*, 56, 931–964.

Robinson, P. (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies*, 56, 511–534.

Robinson, P. (1995): "Log-periodogram Regression of Time Series with Long-range Dependence," *Annals of Statistics*, 23, 1048–1072.

Robinson, P. (1997): "Large-sample Inference for Nonparametric Regression with Dependent Errors," *Annals of Statistics*, 25, 2054–2083.

Roussas, G., and D, Ioannides (1987): "Moment Inequalities for Mixing Sequences of Random Variables," *Stochastic Analysis and Applications*, 5, 61–120.

Samarov, A. M. (1993) "Exploring Regression Structure Using Nonparametric Functional Estimation," *Journal of the American Statistical Association*, 423, 836–847.

Sperlich, S., D. Tjøstheim, and L. Yang (2002): "Nonparametric Estimation and Testing of Interaction in Additive Models," *Econometric Theory*, 18, 197–251.

Whang, Y. J. (2000): "Consistent Bootstrap Tests of Parametric Regression Functions," *Journal of Econometrics*, 98, 27–46.

Whang, Y. J., and D. W. K. Andrews (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.

Wooldridge, J. (1992): "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452–475.

Xia, Y., H. Tong, and W. K. Li (1999): "On Extended Partially Linear Single-Index Models," *Biometrika*, 86, 831–842.

Yatchew, A. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.

Zheng, J. (1996): "A consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.

Table 4.1. Rejection Rates for the Simultaneous Tests at the 5% level

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_1^*	L_2^*
	250	0.054	0.060
	500	0.063	0.056
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_1^*	L_2^*
10	250	0.551	0.723
10	500	0.776	1.000
25	250	0.357	0.533
25	500	0.691	0.866
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_1^*	L_2^*
	250	0.049	0.062
	500	0.053	0.058
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_1^*	L_2^*
10	250	0.679	0.887
10	500	0.847	1.000
25	250	0.462	0.667
25	500	0.717	0.933

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

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{21}^*
	250	0.052	0.059
	500	0.047	0.054
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{21}^*
40	250	0.198	0.267
40	500	0.401	0.478
25	250	0.602	0.667
25	500	0.827	0.866
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{21}^*
	250	0.047	0.053
	500	0.057	0.049
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{21}^*
40	250	0.362	0.404
40	500	0.617	0.733
25	250	0.643	0.679
25	500	0.933	1.000

Table 4.3. Rejection Rates for Testing the Conditional Variance at the 5% level

Normal Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{22}^*
	250	0.052	0.046
	500	0.061	0.058
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{22}^*
40	250	0.193	0.264
40	500	0.467	0.591
25	250	0.278	0.376
25	500	0.593	0.632
Normalized Exponential Error Distribution			
Truncation	Observation	Null Hypothesis Is True	
	T	L_2^*	L_{22}^*
	250	0.051	0.055
	500	0.047	0.059
Truncation	Observation	Null Hypothesis Is False	
ψ	T	L_2^*	L_{22}^*
40	250	0.267	0.404
40	500	0.523	0.732
25	250	0.309	0.443
25	500	0.764	0.898