SIMULTANEOUS SPECIFICATION TESTING OF MEAN AND VARIANCE STRUCTURES IN NONLINEAR TIME SERIES REGRESSION

Author(s): Song Xi Chen and Jiti Gao
Source: *Econometric Theory*, Vol. 27, No. 4 (August 2011), pp. 792-843
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/27975503
Accessed: 24-08-2017 18:48 UTC

Econometric Theory, **27**, 2011, 792–843.
doi:10.1017/S0266466610000502

SIMULTANEOUS SPECIFICATION TESTING OF MEAN AND VARIANCE STRUCTURES IN NONLINEAR TIME SERIES REGRESSION

SONG XI CHEN
Peking University

JITI GAO
The University of Adelaide

This paper proposes a nonparametric simultaneous test for parametric specification of the conditional mean and variance functions in a time series regression model. The test is based on an empirical likelihood (EL) statistic that measures the goodness of fit between the parametric estimates and the nonparametric kernel estimates of the mean and variance functions. A unique feature of the test is its ability to distribute natural weights automatically between the mean and the variance components of the goodness-of-fit measure. To reduce the dependence of the test on a single pair of smoothing bandwidths, we construct an adaptive test by maximizing a standardized version of the empirical likelihood test statistic over a set of smoothing bandwidths. The test procedure is based on a bootstrap calibration to the distribution of the empirical likelihood test statistic. We demonstrate that the empirical likelihood test is able to distinguish local alternatives that are different from the null hypothesis at an optimal rate.

1. INTRODUCTION

Let $\{(X_t, Y_t) : 1 \le t \le n\}$ be a sequence of weakly dependent stationary observations satisfying a nonparametric regression model of the form

$$Y_t = m_1(X_t) + \sigma(X_t)\, e_t, \qquad t = 1, 2, \ldots, n \tag{1.1}$$

We thank Peter C.B. Phillips, the editor, Yuichi Kitamura, the associate editor, and two referees for their constructive and insightful comments and suggestions, which have improved the presentation of the paper. We also thank Ming Li and Isabel Casas Villalba for their valuable computational assistance. Chen acknowledges the financial support from National Science Foundation grants SES-0518904 and DMS-0604533, and Gao acknowledges the financial support by Australian Research Council Discovery Grants under grant numbers DP0558602 and DP0879088.
Address correspondence to Jiti Gao, School of Economics, University of Adelaide, Adelaide SA 5005, Australia; e-mail: jiti.gao@adelaide.edu.au.

where $\{e_t\}$ is an error process with mean zero and variance one, and $m_1(x) = \mathrm{E}(Y_t|X_t = x)$ and $m_2(x) = \mathrm{var}(Y_t|X_t = x) = \sigma^2(x)$ are the unknown conditional mean and variance functions, respectively. Let

$$m(\cdot) = (m_1(x), m_2(x))^\tau \quad \text{and} \quad \left\{m_\theta(x) = (m_{1,\theta}(x), m_{2,\theta}(x))^\tau | \theta \in \Theta\right\}$$

be a family of parametric models for the conditional mean and variance of Y_t given $X_t = x$, where Θ is a parameter space and a subset of R^q.

The interest of this paper is to simultaneously test the hypotheses of the form

$$H_0 : m(x) = m_\theta(x) \quad \text{versus} \quad H_1 : m(x) = m_\theta(x) + C_n \Delta_n(x) \quad \text{for all } x \in S, \tag{1.2}$$

where S is a compact set in R^d, $C_n = (C_{1n}, C_{2n})^\tau$ is a vector of nonrandom sequences tending to zero as $n \to \infty$, and $\Delta_n(x) = (\Delta_{1n}(x), \Delta_{2n}(x))^\tau$ is a vector of bounded functions in R^{2d}.

The motivation for conducting simultaneous hypothesis testing for the conditional mean and variance functions is as follows: The time series regression model (1.1) is specified by both the conditional mean and variance while leaving the distribution of $\{e_t\}$ to be nonparametric. This is a multiple testing situation where the overall model hypothesis H_0 consists of two individual hypotheses: one on the conditional mean $H_{01} : m_1(x) = m_{1,\theta}(x)$ and the other on the conditional variance $H_{02} : m_2(x) = m_{2,\theta}(x)$ for all $x \in S$. It is known that (Simes, 1986; Benjimini and Hochberg, 1995) for testing multiple hypotheses, due to a purely random chance, a true hypothesis can be rejected, which leads to an increase of false rejection of the overall hypothesis H_0, which is rejected if either H_{01} or H_{02} is rejected. This phenomenon of increased false rejection due to multiple hypotheses is the so-called multiplicity effect.

There are two ways to correct this multiplicity effect in testing multiple hypotheses. One is to adjust the level of significance for each hypothesis via the Bonferroni procedure, which tends to be more conservative. The other is to conduct simultaneous testing as we propose in this paper by jointly testing the conditional mean and variance functions. Simultaneous testing can take into account the multiplicity effect while attaining the exact (at least asymptotically) level of signficiance. In addition, a simultaneous test is particularly useful for situations where there is no prior knowledge about whether or not the conditional mean or/and the conditional variance functions are correctly specified. For the purpose of estimating parameters involved in the conditional mean, correctly specifying the conditional mean is crucial to ensure consistency. If we ask for efficiency of the parameter estimation, however, correctly specifying the conditional variance becomes important. In the context of the diffusion process, simultaneous specification testing for both the drift and diffusion is particularly necessary when dealing with pricing options for various derivatives satisfying such a diffusion process.

A specific example that motivates our investigation is the specification testing of a continuous-time diffusion process of the form

$$dr_t = \mu(r_t)dt + \sigma(r_t)dB_t, \tag{1.3}$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are, respectively, the drift and volatility functions of the process, and B_t is the standard Brownian motion. Even though the diffusion process is a continuous–time model, the empirical observations of the process are made at discrete time points, say $\{r_{t\Delta}\}_{t=1}^n$, where Δ is the sampling interval between successive observations. And hence $\{r_{t\Delta}\}_{t=1}^n$ are discrete time series. Based on the first-order Euler approximation, the discrete time series observations satisfy

$$Y_t = \mu(X_t) + \sigma(X_t)\, e_t \tag{1.4}$$

when Δ is small, where $Y_t = \frac{r_{t\Delta} - r_{(t-1)\Delta}}{\Delta}$, $X_t = r_{(t-1)\Delta}$, and $e_t = B_{t\Delta} - B_{(t-1)\Delta}$. Then, the specification testing considered in this paper can be used to test a version of (1.2) that consists of the drift and diffusion specifications as long as Δ is small and fixed. We note that approximation is commonly used when dealing with the diffusion process, largely due to the fact that the transitional density of the process does not always have a close-form expression, as it is only implicitly defined by the Kolmogorov equations. Aït-Sahalia (1999) proposes an Edgeworth series approximation method to the transitional density function, which has since been widely used in financial econometrics. It also requires Δ being small to ensure the accuracy of the approximation. While specification testing on the diffusion process can be carried out by testing the transitional density function (Chen, Gao, and Tang, 2008; Aït-Sahalia, Fan, and Peng, 2009), rejection via testing the transitional density specification may not provide information on which part of the process, the drift or the diffusion, is misspecified. In addition, correct parametric specification of the transition density function does not necessarily imply an explicit parametric form for each of the drift and diffusion functions. It is therefore more direct and informative to specify the drift and diffusion functions simultaneously.

Nonparametric kernel estimation for the conditional mean and variance functions are well studied for both independent and dependent observations as documented in Fan and Gijbels (1996), Fan and Yao (2003), Gao (2007), and Li and Racine (2007), among many others. There is also a substantial list in the goodness-of-fit tests for a parametric conditional mean or variance model by formulating certain distance measure between the parametric model and its corresponding kernel estimator. For instance, see the works of Eubank and Spiegelman (1990), Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Fan and Li (1996), Hart (1997), Hjellvik, Yao, and Tjøstheim (1998), Li and Wang (1998), Chen and Fan (1999), Li (1999), and many others. Zhang and Dette (2004) compare the power of three kernel-based tests for the conditional mean. McKeague and Zhang (1994) consider separate testing of the conditional mean and variance specifications of a nonlinear time series regression based on some cumulative measures. Fan and Zhang (2003) propose separate tests for the conditional mean and the variance of a diffusion model. Recent studies in the field of specification testing may be found in Gao, in Li and Racine, and in Gao and Gijbels (2008). In the meantime, Escanciano (2008) proposes using a joint test for the specification of conditional mean and conditional variance function based on a generalized spectral approach. In a related paper, Escanciano and Velasco (2008) consider testing for parametric dynamic conditional quantiles. For discrete time series regression models, Chen and Fan (1999) and Li (1999) both propose testing conditional mean–variance efficiency. However, these are different from testing conditional mean and conditional variance simultaneously.

Among existing studies closely related to the topic of this paper, Tripathi and Kitamura (2003) propose an EL test for conditional moment restrictions. Fan and Zhang (2004) propose a sieve EL test for testing a general varying-coefficient regression model that extends the generalized likelihood ratio test of Fan, Zhang, and Zhang (2001). They demonstrate that the 'Wilks phenomenon' continues to hold under general assumptions on the error distribution. Both of these papers are established for independent data. For testing the conditional mean function with dependent data, Chen, Härdle, and Li (2003) develop an EL test by simulating a known Gaussian random field. The above three tests have displayed an interesting diversity in test statistic formulations via the EL. The basic idea of the EL is to maximize an objective function that is a product of probability weights allocated to observations under certain constraints that characterize the functional curve to be tested. Fan and Zhang (2004) apply kernel smoothing in both the objective function and the constraints, whereas Tripathi and Kitamura smooth only the objective function, and Chen et al. (2003) smooth only the constraints. The formulation in Tripathi and Kitamura is a one-step approach associated with a global objective function over the entire sample. Fan and Zhang (2004) and Chen et al. (2003) share a formulation of the test statistics by first constructing a local statistic at a fixed point and then integrating them to form the final test statistics; hence, both are sieve EL statistics.

A common feature among the three formulations is that the test statistics are all asymptotically pivotal. This is due to the EL's ability to internally studentize a statistic via its optimization procedure when a single smoothing bandwidth is used. This is the case for all three EL tests. Recently, Chen and Gao (2007) established an EL–based improved test over the corresponding test proposed in Chen et al. (2003). Meanwhile, Chen et al. (2008) develop a new EL test to parametrically specify the transitional distribution in a diffusion model. Such a specification testing method is an alternative to existing methods proposed in Aït–Sahalia (1996), Gao and King (2004), and Hong and Li (2005).

To the best of our knowledge, the problem of testing both the conditional mean and the conditional variance simultaneously has not been extensively discussed in the literature. This is probably because weights for any goodness-of-fit measure have to be allocated between the conditional mean and the conditional variance components, and a natural and easily computable weighting scheme that is free of secondary estimation is not readily available. In a closely related study based on a generalized spectral approach, Escanciano (2008) proposes using a specification test and then some weak convergence results for the proposed test under certain high–level technical assumptions (such as Assumptions A3–A5).

This paper proposes an EL-based test for the hypotheses in (1.2). The EL (Owen, 1988, 1990) is a technique that allows construction of nonparametric likelihood for a parameter of interest. Despite that it is intrinsically nonparametric, it posesses two important properties of a parametric likelihood: the Wilks's theorem and the Bartlett correction. Qin and Lawless (1994) establish EL for parameters defined by a set of generalized estimating equations. This is perhaps the widest framework for EL formulation. It is extended by Kitamura (1997) to weakly dependent observations. Chen and Cui (2006, 2007) show that the EL still admits Bartlett correction under this general framework. For survival data, Li, Hollander, McKeague, and Yang (1996) construct nonparametric likelihood ratio confidence bands for quantile functions, which can be used for testing purposes. Li (2003) considers a goodness-of-fit test for a parametric specification of the distribution function that is more efficient in the Bahadure sense than any weighted Kolmogorov-Smirnov test at any alternative. Einmahl and McKeague (2003) propose EL goodness-of-fit tests for a distribution function and distributional characteristics. Other closely related studies include Donald, Imbens, and Newey (2003) on empirical likelihood estimation and consistent tests of conditional moment restrictions, Kitamura, Tripathi, and Ahn (2004) on empirical likelihood–based inference in conditional moment restriction models, and Xu and Phillips (2006) on empirical likelihood estimation in diffusion models.

The EL test proposed in this paper for the joint hypothesis on the mean and variance has an attractive feature in naturally allocating weights between the conditional mean and conditional variance via the covariance matrix of the nonparametric conditional mean and variance estimators. A nicety of employing the EL is that the estimation of the weighting matrix is avoided. As a result, it avoids the task of estimating the third and fourth conditional moments. We also employ two different smoothing bandwidths, h_1 and h_2, to smooth the conditional mean and variance functions, respectively, which reflect a practical need for applying different amounts of smoothing for two different functions.

Another feature of our proposal is that the final test statistic is formulated by maximizing the EL statistics over a set of bandwidths. This is aimed at achieving the optimal rate of convergence for C_n, which defines the gap between the null and alternative hypotheses in (1.2). The existing goodness-of-fit tests for a parametric model based on a kernel estimator with a fixed bandwidth h, for instance the tests given in Härdle and Mammen (1993), require that the smallest order for C_n is $n^{-1/2}h^{-d/4}$ in order for the test to be consistent. This is much larger than $n^{-1/2}$, which is the rate achieved by tests for a finite dimensional parameter in a standard setting and by tests based on the empirical distribution function of the estimated residuals, although the latter tests assume $\Delta_n(x) \equiv \Delta(x)$

for all n. For testing parametric conditional mean models, Horowitz and Spokoiny (2001) propose an adaptive test that combines a version of the Härdle–Mammen test statistics over a set of bandwidths. The test is adaptive against the unknown smoothness of the local alternative hypothesis and reduces the order of the magnitude of C_{1n} to $n^{-1/2}\sqrt{\log\log(n)}$, which is optimal in the minimax sense of Ingster (1993a, 1993b, 1993c) and Spokoiny (1996). A similar idea is also well presented in Fan and Gijbels (1996). Meanwhile, a closely related paper by Kitamura (2001) discusses asymptotic optimality of empirical likelihood for testing moment restrictions.

In this paper we also extend the proposal of Horowitz and Spokoiny (2001) in the context of testing simultaneously for the conditional mean and variance via the EL with weakly dependent observations. Comparing with tests based on fixed bandwidths, a test based on a set of bandwidths will be less dependent on a particular choice of bandwidth/bandwidths and hence will make the test more robust. In addition, we compare our adaptive test with an adaptive version of the bivariate extension of the test proposed in Härdle and Mammen (1993) and then Horowitz and Spokoiny. To accurately approximate the distribution of the adaptive test statistic, a bootstrap procedure is used to find the critical value in each case. This combination of the EL and bootstrap is able to utilize the good features of the EL for the construction of test statistics and the effectiveness of the bootstrap in distribution approximation. There is a connection between the proposed adaptive test based on a discrete set of bandwidth values and the study of Dette and Hetzler (2007), which considers the Härdle-Mammen test as a stochastic process indexed by the coefficient c in the bandwidth $h = cn^{-1/(d+4)}$ that is at the optimal order for the curve estimation.

The paper is organized as follows: Section 2 introduces the EL test statistic based on a fixed pair of bandwidths and studies its asymptotic properties. The adaptive test that combines the EL test statistic over a set of bandwidths is proposed and analyzed in Section 3. Section 4 contains a case study that tests the goodness of fit of five diffusion models on a Federal funds rate data set. Simulation results are reported in Section 5. Mathematical assumptions and technical details are given in Appendixes A and B.

2. EMPIRICAL LIKELIHOOD TEST STATISTICS

The basic building blocks used to form the proposed test statistic are the kernel estimators of $m_1(x)$ and $m_2(x)$. Let K be a d-dimensional kernel function. We assume without loss of generality that $K(t_1,\dots,t_d)=\prod_{i=1}^d k(t_i)$, where $k(\cdot)$ is a univariate symmetric univariate probability density function. The Nadaraya–Watson (NW) estimators of $m_1(x)$ and $m_2(x)$ are, respectively,

$$\hat{m}_1(x)=\frac{\sum_{t=1}^n K_{h_1}(x-X_t)Y_t}{\sum_{t=1}^n K_{h_1}(x-X_t)} \quad \text{and} \quad \hat{m}_2(x)=\frac{\sum_{t=1}^n K_{h_2}(x-X_t)(Y_t-\hat{m}_1(X_t))^2}{\sum_{t=1}^n K_{h_2}(x-X_t)},$$

where h_1 and h_2 are the bandwidths for smoothing $m_1(x)$ and $m_2(x)$, respectively, and $K_{h_i}(u) = h_i^{-d} K(u/h_i)$. We assume that $h_1 = \beta h_2$ for a positive constant β possibly depending on n, and $h_1 \to 0$ and $nh_1^{2d}/\log^6(n) \to \infty$ as $n \to \infty$. The local polynomial estimators for $m_i(x)$ as discussed in Fan and Gijbels (1996) and Fan and Yao (1998) can be used as well without affecting the results of the paper. We choose the NW estimator for its ease of presentation.

Let $\tilde{\theta}$ be a consistent estimator of θ under H_0. Similarly to Härdle and Mammen (1993), $m_{l,\tilde{\theta}}$ are smoothed with the same kernel and bandwidths; that is, for $i = 1, 2$,

$$\tilde{m}_{i,\tilde{\theta}}(x) = \frac{\sum_{t=1}^{n} K_{h_i}(x - X_t) m_{i,\tilde{\theta}}(X_t)}{\sum_{t=1}^{n} K_{h_i}(x - X_t)}.$$

Define $\hat{m}(x) = \big(\hat{m}_1(x), \hat{m}_2(x)\big)^\tau$ and $\tilde{m}_\theta(x) = \big(\tilde{m}_{1,\theta}(x), \tilde{m}_{2,\theta}(x)\big)^\tau$. The proposed test statistic associated with a fixed bandwidth pair (h_1, h_2) is based on a weighted distance between $\hat{m}$ and $\tilde{m}_{\tilde{\theta}}$ rather than between $m_{\tilde{\theta}}$ and $\hat{m}$, in order to cancel out the bias associated with the kernel estimators so as to prevent the bias getting into the asymptotic distribution of the test statistic. Otherwise, either undersmoothing or explicit bias correction has to be carried out.

Let

$$Q_t(x) = \left[K\left(\frac{x - X_t}{h_1}\right)\left(Y_t - \tilde{m}_{1\tilde{\theta}}(x)\right),\right.$$
$$\left.\times K\left(\frac{x - X_t}{h_2}\right)\left(\{Y_t - m_{1\tilde{\theta}}(X_t)\}^2 - \tilde{m}_{2\tilde{\theta}}(x)\right)\right]^\tau.$$

There are two steps in the formulation of the EL test statistic. Let $p_t(x)$ be nonnegative values representing weights allocated to each (X_t, Y_t). In the first step, at an arbitrary $x \in S$, a compact set of R^d, the log EL ratio for $(m_1(x), m_2(x))$ evaluated at $\big(\tilde{m}_{1\tilde{\theta}}(x), \tilde{m}_{2\tilde{\theta}}(x)\big)$ is constructed as

$$\ell\{\tilde{m}_{\tilde{\theta}}(x)\} = -2 \min \sum_{t=1}^{n} \log\{np_t(x)\} \tag{2.1}$$

subject to $\sum_{t=1}^{n} p_t(x) = 1$ and $\sum_{t=1}^{n} p_t(x) Q_t(x) = 0$. A standard derivation, for instance that given in Owen (1990), shows that the optimal weights are

$$p_t(x) = \frac{1}{n}\{1 + \lambda^\tau(x) Q_t(x)\}^{-1}, \tag{2.2}$$

where $\lambda(x) = (\lambda_1(x), \lambda_2(x))^\tau$ is a pair of Lagrange multipliers defined by

$$\sum_{t=1}^{n} \frac{Q_t(x)}{1 + \lambda^\tau(x) Q_t(x)} = 0. \tag{2.3}$$

In the second step we form the test statistic

$$N_n(h) = \int \ell\{\tilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx,$$

where $h = (h_1, h_2)$ and $\pi(\cdot)$ is a nonnegative weight function supported on S satisfying $\int_{x\in S} \pi(x)dx = 1$ and $\int_{x\in S} \pi^2(x)dx < \infty$.

To appreciate the rationale of the proposed test statistic, let $\bar{U}(x) = (\bar{U}_1(x), \bar{U}_2(x))^\tau$ with

$$\bar{U}_1(x) = (nh_1^d)^{-1} \sum_{t=1}^n K\left(\frac{x - X_t}{h_1}\right)\{Y_t - \tilde{m}_{1\tilde{\theta}}(x)\} \quad \text{and}$$

$$\bar{U}_2(x) = (nh_1^d)^{-1} \sum_{t=1}^n K\left(\frac{x - X_t}{h_2}\right)\left[\{Y_t - \tilde{m}_{1\tilde{\theta}}(X_t)\}^2 - \tilde{m}_{2\tilde{\theta}}(x)\right].$$

Note that using h_1^{-d}, rather than h_2^{-d}, in $\bar{U}_2$ facilitates easy expressions. Furthermore, let $\epsilon_t = Y_t - m_1(X_t)$, $\eta_t = \epsilon_t^2 - m_2(X_t)$, $\sigma_{ij}(x) = \mathrm{E}\left[\epsilon_t^i \eta_t^j | X_t = x\right]$ for $i, j = 0, 1, 2$,

$$\Sigma_0(x) = f^{-1}(x)R(K)\begin{pmatrix} \sigma_{20}(x) & R(\beta^{-1}, K)\sigma_{11}(x) \\ R(\beta^{-1}, K)\sigma_{11}(x) & \beta^d \sigma_{02}(x) \end{pmatrix}, \tag{2.4}$$

and

$$\Sigma_1(x) = f(x)R(K)\begin{pmatrix} \sigma_{20}(x) & R(\beta, K)\sigma_{11}(x) \\ R(\beta, K)\sigma_{11}(x) & \beta^{-d} \sigma_{02}(x) \end{pmatrix}, \tag{2.5}$$

where $R(K) = \int K^2(x)dx$ and $R(\beta, K) = R^{-1}(K) \int K(x)K(\beta x)dx$.

Expansions established in (A.8) of Appendix A show that

$$\begin{aligned} \ell\{\tilde{m}_{\tilde{\theta}}(x)\} &= (nh_1^d)\bar{U}^\tau(x)\Sigma_1^{-1}(x)\bar{U}(x) + O_p\{h_1^2 \log^2(n) + (nh_1^d)^{-1/2}\log^3(n)\} \\ &= (nh_1^d)\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}^\tau \Sigma_0^{-1}(x)\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\} \\ &\quad + O_p\{h_1^2 \log^2(n) + (nh_1^d)^{-1/2}\log^3(n)\}. \end{aligned} \tag{2.6}$$

Then, the test statistic admits the following expansions:

$$\begin{aligned} N_n(h) &= (nh_1^d)\int \bar{U}^\tau(x)\Sigma_1^{-1}(x)\bar{U}(x)\pi(x)dx \\ &\quad + O_p\{h_1^2 \log^2(n) + (nh_1^d)^{-1/2}\log^3(n)\} \\ &= (nh_1^d)\int \{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}^\tau \Sigma_0^{-1}(x)\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx \\ &\quad + O_p\{h_1^2 \log^2(n) + (nh_1^d)^{-1/2}\log^3(n)\}. \end{aligned} \tag{2.7}$$

Since $(nh_1^d)^{-1}\Sigma_0(x)$ is the asymptotic covariance of $\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)$, $N_n(h)$ is asymptotically an integrated Mahalanobis distance between $\tilde{m}_{\tilde{\theta}}$ and $\hat{m}$. The covariance $\Sigma_0(x)$ naturally allocates weights between $\hat{m}_1(x) - \tilde{m}_{1\tilde{\theta}}(x)$ and $\hat{m}_2(x) - \tilde{m}_{2\tilde{\theta}}(x)$, the two components of the goodness of fit.

Before we establish the asymptotic normality of $N_n(h)$, we define $L(z) = \int K(u)K(z+u)du$ as the convolution of K, $L(z,c) = \int K(u)K(z+cu)du$,

$$\Sigma(x,y) = \begin{pmatrix} L\left(\frac{y-x}{h_1}\right)\sigma_{20}(x)f(x) & \beta^{-d}L\left(\frac{x-y}{h_1},\beta^{-1}\right)\sigma_{11}(y)f(y) \\ \beta^{-d}L\left(\frac{y-x}{h_1},\beta^{-1}\right)\sigma_{11}(x)f(x) & \beta^{-d}L\left(\frac{x-y}{h_2}\right)\sigma_{02}(y)f(y) \end{pmatrix},$$

$$\Omega(x,y) = \Sigma^{-1/2}(x,x)\Sigma(x,y)\Sigma^{-1/2}(y,y) = \left(\omega_{ij}(x,y)\right)_{2\times 2}, \qquad \textbf{(2.8)}$$

and $\sigma_n^2(h) = 2\int\int\sum_{i,j=1}^2 \omega_{ij}^2(x,y)\pi(x)\pi(y)dxdy$.

THEOREM 2.1. *Under Assumptions A.1–A.4 listed in Appendix A, for $h = (h_1,h_2)$*

$$L_n(h,\Delta_n) = \frac{N_n(h) - 2 - nh_1^d C_{1n}^2 \int \Delta_n^\tau(x)\Sigma_1^{-1}(x)\Delta_n(x)\pi(x)dx}{\sigma_n(h)} \xrightarrow{d} N(0,1)$$

as $n \to \infty$.

Remarks 2.1. It may be shown that $\sigma_n^2(h) = 2C_0h_1^d\{1+o(1)\}$ as $n\to\infty$, where C_0 is a constant not depending on n. This means that under H_0 $N_n(h) - 2 = O_p(h_1^{d/2})$, which leads to the standardization of $N_n(h)$ when constructing our adaptive test in the next section. By constructing a consistent estimator of $\sigma_n^2(h)$ and letting $\Delta_n(x) = 0$, the theorem can lead to an asymptotically normal distribution-based test for a given pair of $h = (h_1,h_2)$. However, we would not recommend it due to the facts that (i) $\sigma_h^2(h)$ has to be estimated, and (ii) the convergence to the normal distribution would be slow. A common approach is to use a bootstrap method to calibrate the distribution of the test statistic as proposed in Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Kreiss, Neumann, and Yao (1999), Franke, Kreiss, and Mammen (2002), Gao (2007), Li and Racine (2007), and others.

Remarks 2.2. As implied by the form of $\Omega(x,y)$ in (2.8), the EL statistic $N_n(h)$ is no longer an asymptotically pivotal quantity when the bandwidths h_1 and h_2 are different. When $h_1 = h_2$, however, we have $C_0 = K^{(4)}(0)$ and therefore

$$\sigma_n^2(h) = 4h_1^d K^{(4)}(0)R^{-2}(K)\int \pi^2(x)dx\ (1+o(1)),$$

where $K^{(j)}(0)$ denotes the j-times convolution of $K(\cdot)$. Thus, $N_n(h)$ is asymptotically pivotal when $h_1 = h_2$.

Remarks 2.3.

(i) We note that the choice of the weight function $\pi(\cdot)$ has little impact on the size, but can have some impact on the power of the test based on $N_n(h)$ as well as the adaptive test that we will propose in the next section. This is shown by the involvement of $\pi(\cdot)$ in the asymptotic bias and variance of the test statistics $N_n(h)$. In theory, the choice of the weight function should be made to maximize the power. A theoretical recipe for an optimal weighting function would depend on an expansion for the power function of the adaptive test, which would be quite challenging. In practice, one may first plot the kernel estimates for both the mean and variance functions, respectively, against their parametric hypothesized counterparts. One may then find the regions S_1 and S_2 that contain the regions where the two sets of curves differ, and choose $\pi(\cdot)$ to be the uniform weight over $S = S_1 \cup S_2$.

(ii) Boundary bias is an issue for the kernel estimators when the underlying curve has discontinuous boundaries. It is more of an issue when the purpose is to estimate the underlying function than model specification testing. The way we formulate the test statistic, which compares the kernel estimates with a kernel smoothed version of the parametric specifications, will cancel out the boundary bias. We also note that with a compactly supported kernel K, the boundary region is $O(h)$. This means that the contribution to the test statistic from the boundary region is negligible as compared with the contribution from the interior region.

3. AN ADAPTIVE EL TEST

Like all tests constructed via kernel smoothing with fixed bandwidths (for instance, the test of Härdle and Mammen, 1993), the test based on a pair of fixed bandwidths briefly outlined in Remark 2.1 requires both C_{1n} and C_{2n} converging to zero more slowly than $n^{-1/2}$. Indeed, it can be derived from Theorem 2.1 that the asymptotic power of the EL test based on a pair of fixed bandwidths is

$$1 - \Phi\{Z_\alpha - nh_1^d C_{1n}^2 \gamma(\Delta_n, \Sigma_1)/\sigma_n(h)\}$$

when $C_{1n} = C_{2n}$, where $\gamma(\Delta_n, \Sigma_1) = \int \Delta_n^\tau(s)\Sigma_1^{-1}(s)\Delta_n(s)\pi(s)ds$ and Z_α is the upper-α quantile of $N(0,1)$. Hence, if $C_{1n} = o(n^{-1/2}h_1^{-d/4})$, the power degenerates to α, the size of the test. This means that the test is incapable of distinguishing H_0 and H_1 if $C_{1n} = o(n^{-1/2}h_1^{-d/4})$. Horowitz and Spokoiny (2001) propose an adaptive test for testing the mean of a regression model with independent observations. The test is able to distinguish the null hypothesis from a sequence of local alternatives of varying degrees of smoothness, with C_{1n} at the optimal rate in the minimax sense of Ingster (1993a, 1993b, 1993c) and Spokoiny (1996).

Similarly to Horowitz and Spokoiny (2001), we construct in this section an adaptive EL test statistic that combines a standardized version of $N_n(h)$ over a set of bandwidths.

For $i = 1$ and 2, the set of bandwidths to smooth $m_i(\cdot)$ is

$$\mathcal{H}_{in} = \left\{h_i = h_{i\max}a_i^k : h_i \ge h_{i\min},\ k = 0, 1, 2, \ldots J_{in}\right\}, \tag{3.1}$$

where $0 < a_i < 1$, $J_{in} = \log_{1/a_i}(h_{i\max}/h_{i\min})$ is the number of bandwidths in $\mathcal{H}_{in}$, $h_{i\max} = c_{i\max}(\log\log(n))^{-1/d}$, and $h_{i\min} = c_{i\min}n^{-\gamma_i}$ for $0 < \gamma_i < \frac{1}{d}$ and some positive constants $c_{i\max}$ and $c_{i\min}$. The choice of $h_{i\max}$ is vital in reducing C_{in} to almost $n^{-1/2}$ rate. The range of γ_i allows $h_i = O\{n^{-1/(4+d)}\}$, the optimal order in the kernel estimation of $m_i(x)$. Let $\mathcal{H}_n = \mathcal{H}_{1n} \times \mathcal{H}_{2n}$ be the set of bandwidths for smoothing the bivariate function $m(x)$.

Since the omission of the known constant C_0 from $\sigma_n^2(h)$ does not affect the size and power property of $L_n(h, \Delta_n)$, we propose using an adaptive EL test statistic of the form

$$L_n = \max_{h \in \mathcal{H}_n} \frac{N_n(h) - 2}{\sqrt{h_1^d}}, \tag{3.2}$$

where 2 and $h_1^{d/2}$ are, respectively, the asymptotic mean and standarad deviation of $N_n(h)$ under H_0.

Let $l_{n\alpha}$ be the $1-\alpha$ quantile of the finite sample distribution of L_n where $\alpha \in (0, 1)$ is the significance level. We use a regression bootstrap method to approximate the quantiles of L_n by generating resamples of the innovations $\{e_t\}_{t=1}^n$, which mimics the conditional distribution of $\{e_t\}_{t=1}^n$ given all the past information.

We propose the following bootstrap procedure to approximate $l_{n\alpha}$:

1. For $t = 1, 2, \ldots, n$, let $\hat{e}_t = \frac{Y_t - m_{1\tilde{\theta}}(X_t)}{\sqrt{m_{2\tilde{\theta}}(X_t)}}$ be the estimated innovations and $Y_t^* = m_{1\tilde{\theta}}(X_t) + \sqrt{m_{2\tilde{\theta}}(X_t)}e_t^*$, where $\{e_t^*\}_{t=1}^n$ is a sample randomly generated according to the empirical distribution of $\{\hat{e}_t\}_{t=1}^n$. Let $l_{n\alpha}^*$ be the $1-\alpha$ quantile of the distribution of L_n that is induced by $\{Y_t^*\}$.
2. Let $\tilde{\theta}^*$ be the estimate of θ based on the resample $\{(X_t, Y_t^*)\}_{t=1}^n$. Compute the statistic L_n^* by replacing Y_t and $\tilde{\theta}$ with Y_t^* and $\tilde{\theta}^*$ according to (3.2).
3. Estimate $l_{n\alpha}^*$ by the $1-\alpha$ quantile of the empirical distribution of L_n^*, which can be obtained by repeating steps 1–2 many times.

The next theorem justifies the above bootstrap estimate of $l_{n\alpha}$.

THEOREM 3.1. *Suppose that Assumptions B.1–B.3 listed in Appendix B hold. Then under* H_0, $\lim_{n\to\infty} P(L_n > l_{n\alpha}^*) = \alpha$.

We now propose the adaptive test with a nominal significance level α that rejects H_0 if $L_n \geq l_{n\alpha}^*$. Theorem 3.1 guarantees that the test attains the nominal level α asymptotically. In the following we discuss the consistency of the adaptive EL test under three forms of an alternative hypothesis.

We start by evaluating the consistency of the test against a family of fixed alternatives. Let Θ be an open subset of R^q and $\mathcal{M}_{i\Theta} = \{m_{i\theta}(\cdot) : \theta \in \Theta\}$ specify a family of parametric models for the conditional mean and variance under H_0, where $i = 1, 2$. Let

$$\mathrm{E}_S[m_i(X_1) - m_{i\theta}(X_1)]^2 = \int_{x \in S} [m_i(x) - m_{i\theta}(x)]^2 f(x)dx,$$

where $f(x)$ is the density function of X_i and

$$\rho_i(m_i, \mathcal{M}_{i\Theta}) = \left[\inf_{\theta\in\Theta}\left(\mathrm{E}_S\,[m_i(X_1) - m_{i\theta}(X_1)]^2\right)\right]^{1/2} \tag{3.3}$$

is a weighted L_2 distance between $m_i(\cdot)$ and the parametric family $\mathcal{M}_{i\Theta}$. If H_0 is false, then $\min_{1\le i\le 2}\rho_i(m_i, \mathcal{M}_{i\Theta}) \ge c_\rho$ for all sufficiently large n and some $c_\rho > 0$. The following theorem establishes that the adaptive EL test is consistent against a family of fixed alternatives.

THEOREM 3.2. *Suppose that the conditions of Theorem 3.1 hold. In addition, if there is some* $c_\rho > 0$ *such that* $\min_{1\le i\le 2}\rho_i(m_i, \mathcal{M}_{i\Theta}) \ge c_\rho$, *then* $\lim_{n\to\infty} P(L_n > l^*_{n\alpha}) = 1$.

We now consider the consistency under a sequence of local alternatives of the form

$$m(x) = m_{\theta_1}(x) + C_n\Delta(x) \qquad \text{for all } x \in R^d, \tag{3.4}$$

where $C_n = (C_{1n}, C_{2n})^\tau \to 0$ as $n \to \infty$, $\theta_1 \in \Theta$ and $\Delta(x) = (\Delta_1(x), \Delta_2(x))^\tau$ is a vector of continuous functions satisfying

$$0 < D_1 \le \min_{1\le i\le 2}\int_{x\in S}\Delta_i^2(x)f(x)dx \le \max_{1\le i\le 2}\int_{x\in S}\Delta_i^2(x)f(x)dx \le D_2 < \infty \tag{3.5}$$

for some $0 < D_1 < D_2 < \infty$. In addition, suppose that $m(\cdot)$ of (3.4) satisfies, for a positive constant D_{3i},

$$\rho_i(m_i, \mathcal{M}_{i\Theta}) \ge D_{3i}\,|C_{in}| \quad \text{for } i = 1, 2. \tag{3.6}$$

Under this situation, the alternative models given in (3.4) differ from the null in a fixed direction determined by $\Delta(x)$, and the difference shrinks to zero as $n \to \infty$. The next theorem shows that the proposed test is consistent for $C_{in} \ge C_i n^{-1/2}\sqrt{\log\log(n)}$, which is a substantial improvement over the fixed bandwidth-based tests and achieves almost the conventional rate $n^{-1/2}$.

THEOREM 3.3. *Suppose that the conditions of Theorem 3.1 hold. In addition, let (3.4)–(3.6) hold with* $C_{in} \ge C_i n^{-1/2}\sqrt{\log\log(n)}$ *for some constants* $0 < C_i < \infty$ *for* $i = 1, 2$, *and let Assumption B.3(ii) hold with* $h_{1\max} = c_{\max}(\log\log(n))^{-1/d}$ *for some* $0 < c_{\max} < \infty$. *Then* $\lim_{n\to\infty} P(L_n > l^*_{n\alpha}) = 1$.

At last, we establish the consistency of the test under alternatives in a Hölder class of smooth functions with unknown degree of smoothness. Such a class is much bigger than that of alternatives considered in the previous two theorems, as it allows difference between $m(\cdot)$ and $m_\theta(\cdot)$ in any direction. In particular, we consider a general class of alternatives of the form

$$m(x) = m_{\theta_2}(x) + C_n\Delta_n(x) \quad \text{for all } x \in R^d, \tag{3.7}$$

where $\theta_2 \in \Theta$ and $\Delta_n(x) = (\Delta_{1n}(x), \Delta_{2n}(x))^\tau$ is a vector of smooth functions.

For nonnegative integers $j_1, \dots, j_d \geq 0$, let $j = (j_1, \dots, j_d)$ and

$$|j| = \sum_{k=1}^{d} j_d \quad \text{and} \quad D^j m_i(x) = \frac{\partial^{|j|} m_i(x)}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} \quad \text{for } i = 1, 2$$

whenever the derivative exists. For $i = 1, 2$, define the Hölder norm

$$||m_i||_{H;s_i} = \sup_{x \in S} \left(\sum_{|j| \leq s_i} |D^j m_i(x)| \right)$$

and a smooth class of the form $S_i(H; s_i) \equiv \{m_i : ||m_i||_{H;s_i} \leq c_{iH}\}$, which has an unknown degree of smoothness $s_i \geq \max\left(2, \frac{d}{4}\right)$ and $c_{iH} < \infty$ for $i = 1, 2$. The class of alternative models considered is

$$B_{H,n}(i) = \left\{ m_i \in S_i(H, s_i) : \rho_i(m_i, \mathcal{M}_{i\Theta}) \geq C_{im} \left(n^{-1} \sqrt{\log\log(n)}\right)^{2s_i/(4s_i+d)} \right\} \quad \textbf{(3.8)}$$

for some $0 < C_{im} < \infty$, where $\rho_i(m_i, \mathcal{M}_{i\Theta})$ is as defined in (3.3).

THEOREM 3.4. *Suppose that the conditions of Theorem 3.1 hold. If, in addition, equations (3.7) and (3.8) hold, then* $\lim_{n\to\infty} \inf_{\{m_i \in B_{H,n}(i): i=1,2\}} P(L_n > l^*_{n\alpha}) = 1$.

The conclusion of Theorem 3.4 shows that L_n is uniformly consistent over alternatives within the Hölder class of smooth functions whose distance from the parametric counterparts approaches zero at the rate of $\left(n^{-1} \sqrt{\log\log(n)}\right)^{2s_i/(4s_i+d)}$ for $i = 1, 2$, which is the fastest possible in the minmax sense of Ingster (1993a, 1993b, 1993c) and Spokoiny (1996). The most striking property of Theorem 3.4 is that it achieves the best rate of convergence for C_{in} without knowing the degree of smoothness s_i. This is the reason behind the term "adaptive and rate-optimal" used by Horowitz and Spokoiny (2001) when describing their test. We show that the same property holds for the proposed simultaneous test for the conditional mean and variance with weakly dependent observations.

To show that the conclusions of Theorems 3.1–3.4 hold unconditionally, by the dominated convergence theorem it suffices to show that the conclusions of Theorems 3.1–3.4 all hold in probability with respect to the joint distribution of $\mathcal{X} = (X_1, \dots, X_n)$. For instance, it suffices to show that

$$\lim_{n\to\infty} P_*\left(L_n > l^*_{n\alpha}\right) = \lim_{n\to\infty} P\left(L_n > l^*_{n\alpha} | \mathcal{X}\right) = \alpha \quad \text{in probability}$$

in Theorem 3.1, where $P_*\left(L_n > l^*_{n\alpha}\right) \equiv P\left(L_n > l^*_{n\alpha} | \mathcal{X}\right)$.

4. TESTING FOR DIFFUSION MODELS FOR FED FUNDS RATE DATA

We apply the proposed empirical likelihood test to a class of diffusion models that have been proposed to model the dynamics of interest rate movement in the literature. The data are the monthly Fed funds rates between January 1963 and December 1998 contained in H-15 Federal Reserve Statistical Release with $n = 432$ observed rates. Aït-Sahalia (1999) used this data set to carry out the maximum likelihood estimation of parameters based on either exact or approximate transitional density functions for the following diffusion models:

$$dr_t = \kappa(\alpha - r_t)dt + \sigma dW_t, \tag{4.1}$$

$$dr_t = \kappa(\alpha - r_t)dt + \sigma r_t^{1/2} dW_t, \tag{4.2}$$

$$dr_t = r_t\{\kappa - (\sigma^2 - \kappa\alpha)r_t\}dt + \sigma r_t^{3/2} dW_t, \tag{4.3}$$

$$dr_t = \kappa(\alpha - r_t)dt + \sigma r_t^{\rho} dW_t, \tag{4.4}$$

$$dr_t = (\alpha_{-1} r_t^{-1} + \alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2)dt + \sigma r_t^{3/2} dW_t. \tag{4.5}$$

These models are, respectively, the Ornstein-Uhlenbeck process (4.1) proposed by Vasicek (1977), the CIR model (4.2) proposed by Cox, Ingersoll, and Ross (1985), the inverse of the CIR process (4.3), the constant elasticity of volatility model (4.4) of Chan, Karolyi, Longstaff, and Sanders (1992), and the nonlinear mean reversion model (4.5) of Aït-Sahalia (1996). Note that model (4.3) is similar to but not exactly the same as the model proposed by Ahn and Gao (1999). Although both model (4.3) and the model proposed by Ahn and Gao specify a quadratic drift and a cubic function for the square of the diffusion, model (4.3) is more restrictive on the parameters, as σ^2 appears in both the drift and the diffusion. As discussed in Assumption A2′ of Aït-Sahalia (1996), equations (4.1)–(4.5) are all strictly stationary and β-mixing. A similar discussion is given in Genon-Caralot, Jeantheau, and Laredo (2000). Thus, the proposed estimation method and theory under the α-mixing assumption is directly applicable.

We consider the Euler discretization of the continuous-time diffusion models to create discrete time series models of (1.1) with $Y_t = r_{t\Delta} - r_{(t-1)\Delta}$ and $X_t = r_{(t-1)\Delta}$, where $\Delta = 20/250$ since the data were collected monthly. This will no doubt create discretization errors. The raw interest rate series and a scatter plot of $\{(X_t, Y_t)\}_{t=1}^n$ are displayed in Figure 1. The biweight kernel has been used in all the numerical works reported in this paper. Cross validation (CV) is employed to guide bandwidth selection, which gives $h_{1cv} = 0.019$ for the drift and $h_{2cv} = 0.0275$ for the diffusion.

In Figure 2 we plot the nonparametric kernel estimates of the drift $\hat{m}_1(x)$ and the diffusion $\hat{m}_2(x)$ using the bandwidths prescribed by the CV and the corresponding kernel smoothed versions of the parametric drift and diffusion functions, i.e., $\tilde{m}_{1\hat{\theta}}(x)$ and $\tilde{m}_{2\hat{\theta}}(x)$, for models (4.1) to (4.5), where $\hat{\theta}$'s are those maximum likelihood estimates given in Table VI of Aït-Sahalia (1999). The reason for not

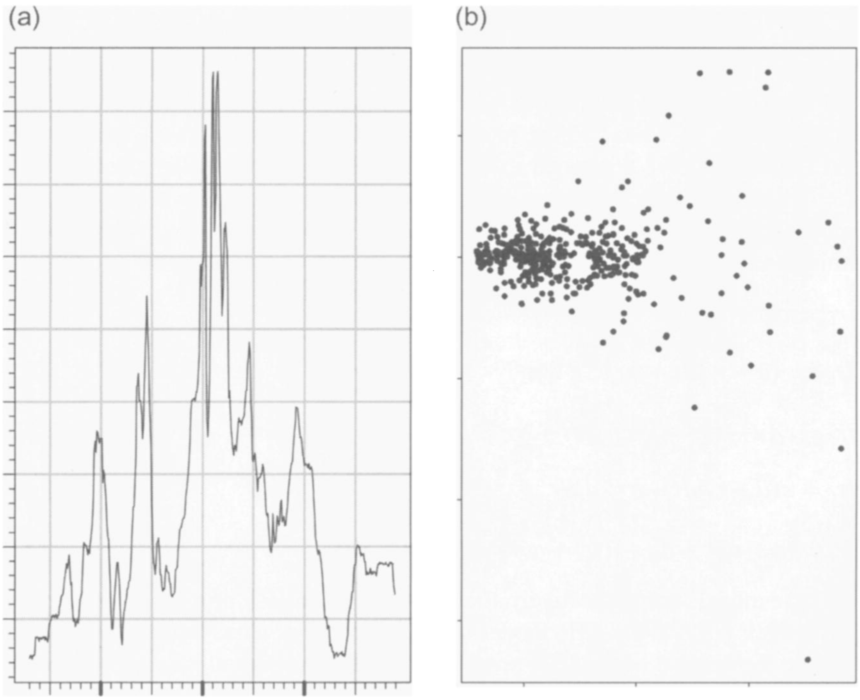

FIGURE 1. (a) The Federal funds rate series from January 1963 to December 1998, and (b) the scattered plot of X_t (the fund rates) and Y_t (the change of the fund rates).

using the bandwidths prescribed by the CV is that some initial curve plotting indicates the bandwidths given by the CV undersmooth the curves. It is clear from Figure 2(a) that the drift specified by model (4.3) is not appropriate for the data. As reported shortly, this is strongly supported by our testing results, which show very small p-values for both the simultaneous and the univariate tests for the drift. The problem with model (4.3) is probably due to the rigid relationship between its parameters, as mentioned earlier.

We observe from Figure 2 that the nonparametric kernel estimates of the drift and diffusion agree reasonably well with the parametric drift and diffusion specifications when the interest rate level is in the range of [0, 10%]. When the interest rate is at a higher level, discrepancies between the nonparametric and parametric fits start to appear in both the drift and diffusion estimation. We were alarmed when first seeing that the kernel drift estimates for the rate over 10% are positive, which is against the mechanism of mean-reverting. However, the data were extremely volatile over that range, as shown by the kernel estimate of the diffusion function in panels (c) and (d) of Figure 2. They are so volatile that a pointwise confidence band of the kernel drift estimates would cover all the parametric drifts over that range and hence the seemingly large discrepency between the kernel

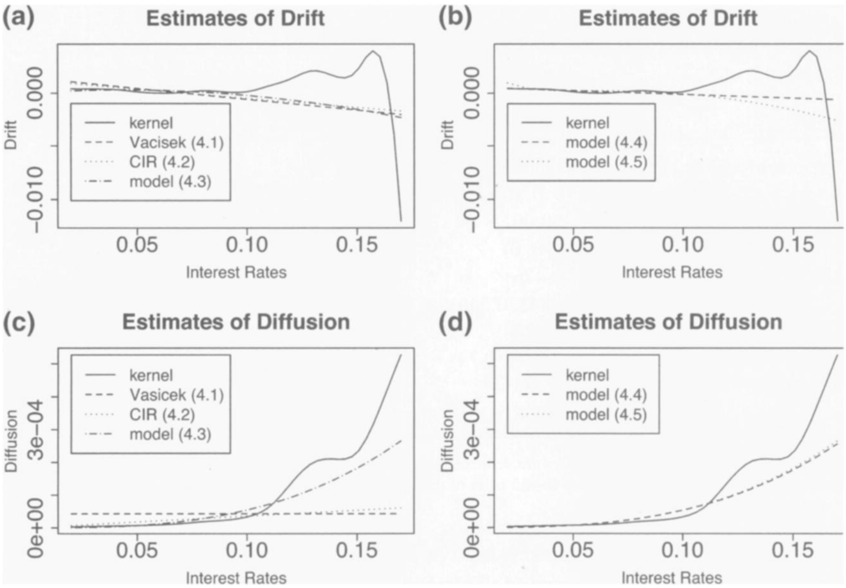

FIGURE 2. Panels (a) and (c) give the nonparametric kernel estimate (bold lines) and the parametric kernel estimates of models (4.1)–(4.3) of the drift function and the diffusion function, respectively. Panels (b) and (d) give the nonparametric kernel estimate (bold lines) and the parametric kernel estimates of models (4.4) and (4.5) of the drift function and the diffusion function, respectively.

and the parametric drift estimates may not be so significant at all after considering the large variation. It is worth pointing out that the apparent roughness in the kernel estimates of the drift and diffusion was largely due to the high volatility and data sparseness at the higher rate levels. This is not the result of using the Nadaraya-Watson kernel estimator, as the the fitted curves by the local linear kernel smoother had the same artifacts.

We first carry out the simultaneous empirical likelihood test for both the drift and diffusion for each of the models from (4.1) to (4.5) on 25 bandwidth pairs (h_1, h_2), which is the result of having five levels of h_1 ranging from $h_{1\min} = 0.01$ to $h_{1\max} = 0.025$ and five h_2s ranging from $h_{2\min} = 0.015$ and $h_{2\max} = 0.03$ according to (3.1). This range of bandwidths includes the bandwidths given by the CV and offers a wide range of smoothness. We choose the weight function $\pi(x) = \frac{1}{0.14}\mathrm{I}(0.02 < x < 0.16)$. The p-values of the adaptive EL tests are reported in the second column of Table 1. The p–values for the Vasicek model (4.1) and CIR model (4.2) a almost zero, which indicates overwhelming rejection of the these two models. Meanwhile, the p-values for the models (4.3), (4.4), and (4.5) are 0.476%, 0.628%, and 0.454%, respectively, indicating not enough statistical evidence to reject these three models.

TABLE 1. The p-values of the adaptive empirical likelihood tests for the five diffusion models

	Simultaneous test	Univariate test	
Model	Drift and diffusion	Drift only	Diffusion only
Vacicek Model (4.1)	< 0.001	0.494	< 0.001
	(0, 0, 0)	(0.144, 0.234, 0.546)	(0, 0, 0)
CIR Model (4.2)	< 0.001	0.916	< 0.001
	(0, 0, 0)	(0.612, 0.782, 0.916)	(0, 0, 0)
Model (4.3)	0.476	0.998	0.05
	(0.258, 0.41, 0.546)	(0.902, 0.976, 0.998)	(0.042, 0.104, 0.144)
Model (4.4)	0.628	1	0.05
	(0.462, 0.596, 0.73)	(0.992, 1, 1)	(0.042, 0.094, 0.146)
Model (4.5)	0.454	1	0.048
	(0.298, 0.368, 0.454)	(0.986, 1, 1)	(0.044, 0.084, 0.114)

Notes: The three numbers in parentheses are, respectively, the minimum, the medium, and the maximum p-values among the nonadaptive tests based on the 25 pairs of fixed bandwidths.

We then conducted separate adaptive EL tests for the drift and the diffusion specifications for these five models, respectively, and report the p-values in the last two columns of Table 1, which indicates that the main lack of fit came from the diffusion specification. The lack of fit of these models comes primarily from the diffusiòn specifications as shown by much smaller p-values reported in Table 1 than those for the drift. The Vasicek and CIR models are the worst two. The other three models, which allow larger volatility for higher rate values, have larger p-values than 5%. The fact that the nonparametric estimates of the diffusion are larger than the parametric estimates under models (4.4) and (4.5) indicates there is probably some extra amount of volatility that cannot be explained by these two models. This is consistent with an existing belief among financial econometricians that one-factor diffusion models cannot accommodate the full amount of volatility exhibited by real data.

At the same time, the tests for the drift specifications indicate that all the models except the Vasicek model have reasonable compatibility with the data. This may be partly due to the large volatility exhibited in the data as revealed by the kernel estimate of the diffusion function at the higher rate levels. The same can be said for the p-values of the simultaneous tests for the two nonlinear drift models, (4.3) and (4.5), and the CEV model (4.4).

5. SIMULATION STUDIES

Before we report the simulation results, we discuss the issue of computation. For the current testing of an infinite-dimensional "parameter," the computation for $N_n(h)$ involves evaluating the EL ratio $\ell\{\tilde{m}_{\hat{\theta}}(x)\}$ over a grid of x-points within the set S. And it is further increased by the adaptive test procedure and the bootstrap calibration. In spite of this, implementing the adaptive EL test based on a single time series would not be a problem, given the fact that a reliable algorithm for the EL computation is available from the authors. When carrying out the simulations, however, we need to speed up the computation, as a large number of replications are required. In the simulation studies, we use the least squares EL (LSEL) to replace the full EL when formulating $N_n(h)$. The LSEL is easier to compute, as there are closed-form solutions for the weights $p_t(x)$. Hence, some expensive nonlinear optimization can be avoided.

5.1. Least Squares Empirical Likelihood

The LSEL replaces $-\log\{np_t(x)\}$ in the objective function of the EL formulation (2.1) by $(np_t(x)-1)^2$, the quadratic Taylor expansion of $-\log(np_t(x))$ near $p_t(x)=n^{-1}$. In particular, the log LSEL ratio is

$$lsl\{\tilde{m}_{\tilde{\theta}}(x)\} = \min \sum_{t=1}^{n} \{np_t(x)-1\}^2 \tag{5.1}$$

subject to $\sum_{t=1}^{n} p_t(x) = 1$ and $\sum_{t=1}^{n} p_t(x)Q_t(x) = 0$. According to Brown and Chen (1998), the optimal LSEL weights are given by

$$p_t(x) = n^{-1} + \{n^{-1}Q(x) - Q_t(x)\}^{\tau} S^{-1}(x)Q(x),$$

where $Q(x) = \sum_{t=1}^{n} Q_t(x)$ and $S(x) = n^{-1}\sum_{t=1}^{n} Q_t(x)Q_t(x)$. Thus,

$$lsl\{\tilde{m}_{\tilde{\theta}}(x)\} = Q^{\tau}(x)S^{-1}(x)Q(x),$$

which is readily computable. The price paid for such a simple computational procedure is that the weights may be negative and the delicate second-order property of Bartlett correction is lost. However, these are entirely harmless in the current testing problem.

The LSEL counterpart to $N_n(h)$ is

$$N_n^{ls}(h) = \int lsl\{\tilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx. \tag{5.2}$$

It may be shown from Brown and Chen (1998) that $N_n^{ls}(h)$ and $N_n(h)$ are equivalent in the first order. Therefore, those first-order theoretical results established based on $N_n(h)$ in Theorems 2.1 and 3.1–3.4 are valid to the corresponding LSEL modification.

5.2. An Alternative Test

As expressed in equation (2.7) above, the leading term of the proposed test is

$$N_{ld,n}(h) = (nh_1^d)\int \{\hat{m}(x)-\tilde{m}_{\tilde{\theta}}(x)\}^{\tau}\Sigma_0^{-1}(x)\{\hat{m}(x)-\tilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx, \tag{5.3}$$

which is equivalent to the bivariate version of the corresponding test proposed in Härdle and Mammen (1993) and then Horowitz and Spokoiny (2001) of the form

$$N_{e,n}(h) = (nh_1^d) \int \{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}^\tau \widehat{\Sigma}_0^{-1}(x)\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}\pi(x)dx \tag{5.4}$$

provided that $\widehat{\Sigma}_0^{-1}(x)$ exists, where

$$\widehat{\Sigma}_0^{-1}(x) = \widehat{f}(x)\widetilde{\Sigma}_0^{-1}(x),\ \widetilde{\Sigma}_0(x) = \begin{pmatrix} \widetilde{\sigma}_{20}(x) & \widetilde{\sigma}_{11}(x) \\ \widetilde{\sigma}_{11}(x) & \widetilde{\sigma}_{02}(x) \end{pmatrix}, \tag{5.5}$$

$\widehat{f}(x) = \frac{1}{nh_1^d} \sum_{t=1}^n K\left(\frac{x-X_t}{h_1}\right)$, and for $i, j = 0, 1, 2,$

$$\widetilde{\sigma}_{20}(x) = \frac{\sum_{t=1}^n K\left(\frac{x-X_t}{h_1}\right)\widetilde{\epsilon}_t^2}{\sum_{t=1}^T K\left(\frac{x-X_t}{h_1}\right)}, \qquad \widetilde{\sigma}_{02}(x) = \frac{\sum_{t=1}^n K\left(\frac{x-X_t}{h_2}\right)\widetilde{\eta}_t^2}{\sum_{t=1}^T K\left(\frac{x-X_t}{h_2}\right)},$$

$$\widetilde{\sigma}_{11}(x) = \frac{\sum_{t=1}^n K\left(\frac{x-X_t}{h_1}\right)\widetilde{\epsilon}_t\widetilde{\eta}_t}{\sum_{t=1}^T K\left(\frac{x-X_t}{h_1}\right)},$$

in which $\widetilde{\epsilon}_t = Y_t - \tilde{m}_{1\tilde{\theta}}(X_t)$ and $\widetilde{\eta}_t = \widetilde{\epsilon}_t^2 - \tilde{m}_{2\tilde{\theta}}(X_t)$.

We then compare L_n of (3.2) with its national competitor of the form

$$L_{e,n} = \max_{h \in \mathcal{H}_n} \frac{N_{e,n}(h) - 2}{\sqrt{h_1^d}} \tag{5.6}$$

in the finite-sample analysis in Section 5.3 below. Note that the conclusions of Theorems 3.1–3.4 hold for $L_{e,n}$, since $N_{e,n}(h)$ is the leading term of $N_h(h)$.

5.3. Simulation Results

We report in this section results of some simulation studies designed to evaluate the empirical performance of the proposed adaptive EL test. The model considered was an ARCH(1) model of the form

$$Y_t = \alpha + \beta Y_{t-1} + C_{1n}\cos(8Y_{t-1}) + \left\{\sigma\sqrt{Y_{t-1}^2 + 1} + C_{2n}\cos(8Y_{t-1})\right\} e_t, \tag{5.7}$$

where $\{e_t\}_{t=1}^n$ were independent and identically distributed innovations independent of Y_{t-1}. The assigned parameter values were $\alpha = 0.25$, $\beta = 0.5$, and $\sigma = 0.25$. We considered two distributions for e_t: $e_t \sim N(0,1)$ and $e_i \sim \frac{\chi_{15}^2 - 15}{\sqrt{30}}$. The choice of the chi-square distributed innovation was to assess the performance of the test in the presence of skewness, which is $\frac{8}{15}$, whereas the kurtosis is $3 + \frac{4}{5}$.

We also chose the cosine function as the local shift functions $\Delta_i(x)$ to make the models under H_1 fairly close to those under H_0 and hence make it more difficult to distinguish between H_0 and H_1. In evaluating the power of the test we chose $C_{1n} = C_{2n} = 0.04$ and 0.06, respectively. The sample sizes considered in the simulation were $n = 300$ and $n = 500$, whereas the number of simulations was 500, with the number of bootstrap resamples being 300.

The vector of parameters $\theta = (\alpha, \beta, \sigma^2)$ was estimated using the pseudo-maximum likelihood, which is commonly used in the estimation of autoregressive conditional hetoroskedasticity (ARCH) models. From information collected from the simulations, the parameters were estimated with good precision even under H_1 for both types of innovations. The maximum averaged mean square errors in estimating α, β, and σ were, respectively, 0.00092, 0.00359, and 0.0129 for $n = 300$, and 0.00054, 0.00203, and 0.0102 for $n = 500$. As the ARCH model is only asymptotically stationary, in each simulation the model was prerun 200 times.

The biweight kernel $k(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1)$ was used throughout this section for kernel estimation. The weight function $\pi(x)$ was chosen to be the uniform density ranging from the 5% to 95% quantiles of the asymptotic stationary distribution of Y_t obtained by a pilot simulation. In each simulation, the likelihood ratio was evaluated over 50 equally spaced grid points within the support of $\pi(\cdot)$.

We need to choose $\mathcal{H}_{1n}$ and $\mathcal{H}_{2n}$ in order to form the adaptive test. The CV was employed to select h_1 by prerunning the simulations reported above. In the case of the normal innovation, the averaged bandwidth (standard error) prescribed by the CV was 0.25 (0.1) for $n = 300$ and 0.23 (0.08) for $n = 500$, respectively. The figures for the chi-square innovations were similar. Note that $(\log\log n)^{-1} = 0.5743397$ and 0.547374 for $n = 300$ and $n = 500$, respectively. In view of these and Assumption B.3(iii), we chose $\mathcal{H}_{1n} = \{0.3, 0.332, 0.367, 0.407, 0.45\}$ with $a_1 = 0.903$ for $n = 300$ and $\mathcal{H}_{1n} = \{0.25, 0.281, 0.316, 0.356, 0.4\}$ with $a_1 = 0.889$ for $n = 500$. Here we chose $h_{1\min}$ to be slightly smaller than the averaged CV and a scaled-down value of $(\log\log(n))^{-1}$ as $h_{1\max}$.

The selection of h_2 depends on the choice of h_1 used to smooth $m_1(\cdot)$ in order to obtain the estimated residuals. After substituting in the h_1 values prescribed by the CV, the averaged CV-based h_2 values were found to be smaller than the corresponding h_1 values. Considering the variation in the CV h_1-values and the observation that h_2 was in general smaller than h_1, we simply chose $\mathcal{H}_{2n} = \{0.25, 0.281, 0.316, 0.356, 0.4\}$ with $a_2 = 0.889$ for $n = 300$ and $\mathcal{H}_{2n} = \{0.2, 0.2300, 0.2646, 0.3043, 0.35\}$ with $a_2 = 0.869$ for $n = 500$ for both types of innovations. These gave 25 combinations of (h_1, h_2) when formulating the adaptive test statistic L_n in (3.1).

We first carried out the simultaneous test at the nominal 5% significance level for H_0 against H_1 where the mean and variance were shifted by the same amount, i.e., $C_{1n} = C_{2n} = 0.0, 0.04$, and 0.06, respectively. The results are reported in Table 2. We observe that the empirical sizes of the test are quite close to 5% and improve as the sample size increases from 300 to 500. This indicates that the bootstrap approximation to the distribution of the test statistic is of good quality.

TABLE 2. Size and power of the adaptive least square empirical likelihood test at 5% nominal significance level for the normally distributed and the chi-square distributed innovations

	$n = 300$		$n = 500$	
$C_{n1} = C_{n2}$	Normal	Chi-square	Normal	Chi-square
0 (Size)	0.064	0.07	0.054	0.062
	(0.06, 0.066, 0.074)	(0.068,0.072,0.08)	(0.034, 0.046,0.056)	(0.052, 0.062, 0.08)
0.03 (Power)	0.218	0.132	0.418	0.286
	(0.13, 0.166, 0.22)	(0.104, 0.138,0.17)	(0.3, 0.376, 0.426)	(0.166, 0.244, 0.304)
0.04 (Power)	0.410	0.194	0.718	0.486
	(0.226, 0.316, 0.418)	(0.14, 0.184,0.246)	(0.462, 0.638, 0.722)	(0.288, 0.432, 0.516)

Notes: The three numbers in parentheses are, respectively, the minimum, the medium, and the maximum size/power among the nonadaptive tests based on the fixed bandwidths.

The power of the test is quite respectable considering that H_0 and H_1 were made deliberately close. As expected, when each C_{in} was increased, the power of the test increased; and for a fixed level of C_{in}, the power increased when n increased. The latter was because the distance between H_0 and H_1 became smaller when n became larger despite the fact that C_{in} was kept the same.

We then compared the power of the simultaneous test with two univariate adaptive tests on the univariate hypotheses $H_{01}: m_1(x) = m_{1\theta}(x)$ versus $H_{11}: m_1(x) = m_{1\theta}(x) + C_{1n}\Delta_{1n}(x)$ on the conditional mean and $H_{02}: m_2(x) = m_{2\theta}(x)$ versus $H_{12}: m_2(x) = m_{2\theta}(x) + C_{2n}\Delta_{2n}(x)$ on the conditional variance. Univariate test statistics can be formulated as follows; Let

$$Q_{1t}(x) = K\left(\frac{x - X_t}{h}\right)\{Y_t - \tilde{m}_{1\tilde{\theta}}(x)\} \quad \text{and}$$

$$Q_{2t}(x) = K\left(\frac{x - X_t}{h}\right)\left[\{Y_t - \tilde{m}_{1\tilde{\theta}}(X_t)\}^2 - \tilde{m}_{2\tilde{\theta}}(x)\right].$$

The LSEL ratio for $m_j(x)$ evaluated at $\tilde{m}_{j\tilde{\theta}}(x)$ is $\ell\{\tilde{m}_{j\tilde{\theta}}(x)\} = \min \sum_{t=1}^n \{np_{jt}(x) - 1\}^2$ subject to $\sum_{t=1}^n p_{jt}(x) = 1$ and $\sum_{t=1}^n p_{jt}(x) Q_{jt}(x) = 0$. At a given h_j, $j = 1$ and 2, $N_{jn}(h_j) = \int \ell\{\tilde{m}_{j\tilde{\theta}}(x)\}\pi(x)dx$, which then leads to the adaptive test statistic L_{jn} like the formation of L_n. The same bootstrap procedure outlined in Section 3 can also be used to profile the distribution of L_{jn} and to formulate the test procedures.

We were interested to see if there was a significant reduction of power for the simultaneous test while H_1 was different from H_0 in only one component. Both the sizes and the power values of the simultaneous and the corresponding univariate tests are presented in Table 3 for the normal innovation and in Table 4 for the chi-square innovation. The univariate tests had reasonable sizes as well. As expected, there was reduction in the power of the simultaneous test. The reduction was relatively small for $C_{1n} = 0.03$ and $n = 300$. It is encouraging to see there was only very small power reduction from the univariate variance test at all the levels considered. We observe that the power values of both the simultaneous and the univariate tests were higher for the case of $C_{1n} = 0$ than for the case of $C_{2n} = 0$. This was probably due to different amounts of variability in $\hat{m}_1(x)$ and $\hat{m}_2(x)$.

The adaptive simultaneous (univariate) test statistics are constructed over 25 (5) pairs of bandwidths. To understand more about the adaptive tests, we also carried out both the simultaneous and univariate tests based on certain sets of fixed bandwidths. In Tables 2–4 we report in parentheses, beneath the size/power of the adaptive tests, the smallest, medium, and largest size/power of the 25 (5) fixed bandwidth tests. It was found that (i) the sizes of the fixed bandwidth tests were generally clustered tight range around the 5% significance level despite the range of the bandwidths being quite wide, considering that all the design variable values were confined in [0, 1]; (ii) more importantly, the power of the adaptive test was larger than the average and often was close to the maximum power of

TABLE 3. Size and power of the adaptive simultaneous test versus the univariate tests for the conditional mean and the conditional variance for normal innovation

	$n = 300$		$n = 500$	
	Simultaneous	Univariate	Simultaneous	Univariate
(a) Comparing with the univariate test for conditional mean, i.e., $C_{n2} = 0$				
C_{n1}				
0 (Size)	0.064	0.04	0.054	0.056
	(0.06, 0.066, 0.074)	(0.038, 0.04, 0.044)	(0.034, 0.046,0.056)	(0.056, 0.058, 0.062)
0.03 (Power)	0.094	0.14	0.168	0.23
	(0.066,0.072, 0.09)	(0.126,0.132,0.14)	(0.1,0.14,0.172)	(0.18,0.204, 0.236)
0.04 (Power)	0.124	0.22	0.23	0.38
	(0.066, 0.106, 0.144)	(0.18, 0.199,0.22)	(0.136, 0.2,0.248)	(0.322.0.352,0.388)
(b) Comparing with the univariate test for conditional variance, i.e., $C_{n1} = 0$				
C_{n2}				
0 (Size)	0.064	0.072	0.054	0.052
	(0.06, 0.066, 0.074)	(0.058,0.064,0.076)	(0.034, 0.046,0.056)	(0.044, 0.046, 0.054)
0.03 (Power)	0.15	0.174	0.31	0.33
	(0.1120.134,0.176)	(0.114,0.14,0.176)	(0.192,0.29,0.342)	(0.188,0.258,0.336)
0.04 (Power)	0.294	0.328	0.516	0.588
	(0.174,0.248,0.326)	(0.156,0.232,0.33)	(0.338, 0.494,0.578)	(0.326, 0.502, 0.592)

Notes: The three numbers in parentheses are, respectively, the minimum, the medium, and the maximum size/power among the nonadaptive tests based on the fixed bandwidths.

TABLE 4. Size and power of the adaptive simultaneous test versus the univariate tests for the conditional mean and the conditional variance for the chi-square innovation

	$n = 300$		$n = 500$	
	Simultaneous	Univariate	Simultaneous	Univariate
(a) Comparing with the univariate test for conditional mean, i.e., $C_{n2} = 0$				
C_{n1}				
0 (Size)	0.07	0.04	0.062	0.056
	(0.068,0.072,0.08)	(0.06, 0.064,0.07)	(0.052,0.062,0.08)	(0.038, 0.042, 0.05)
0.03 (Power)	0.108	0.128	0.158	0.196
	(0.086, 0.1,0.11)	(0.12,0.122,0.132)	(0.102,0.134,0.162)	(0.0.198,0.208,0.218)
0.04 (Power)	0.142	0.224	0.206	0.352
	(0.098, 0.132,0.148)	(0.218, 0.22,0.226)	(0.106,0.172,0.226)	(0.3,0.344,0.39)
(b) Comparing with the univariate test is for conditional variance, i.e., $C_{n1} = 0$				
C_{n2}				
0 (Size)	0.07	0.082	0.062	0.068
	(0.068,0.072,0.08)	(0.064,0.071,0.082)	(0.052,0.062,0.08)	(0.056,0.066,0.076)
0.03 (Power)	0.134	0.156	0.274	0.276
	(0.092, 0.108, 0.148)	(0.088,0.13,0.156)	(0.16,0.238,0.296)	(0.118,0.223,0.278)
0.04 (Power)	0.206	0.258	0.454	0.476
	(0.13,0.192,0.224)	(0.14,0.191,0.256)	(0.26,0.43,0.508)	(0.226,0.362,0.478)

Notes: The three numbers in parentheses are, respectively, the minimum, the medium, and the maximum size/power among the nonadaptive tests based on the fixed bandwidths.

TABLE 5. Size and power for L_n and $L_{e,n}$ at the 5% level with $C_{n1} = C_{n2}$

	Normal error distribution			Chi-square error distribution	
	Observation	Null hypothesis is true		Null hypothesis is true	
	n	L_n	$L_{e,n}$	L_n	$L_{e,n}$
	300	0.064	0.060	0.070	0.056
	500	0.054	0.056	0.062	0.047
Observation	Departure	Null hypothesis is false		Null hypothesis is false	
n	C_{n1}	L_n	$L_{e,n}$	L_n	$L_{e,n}$
300	0.03	0.218	0.187	0.132	0.094
500	0.03	0.418	0.344	0.286	0.229
300	0.04	0.410	0.372	0.194	0.137
500	0.04	0.718	0.653	0.486	0.429

the 25 fixed bandwidth tests. This indicates that the adaptive tests do enhance the power as revealed theoretically in Section 3.

In addition, we compare both the size and power performance of L_n with its natural competitor: $L_{e,n}$ in Table 5 below. As $L_{e,n}$ is constructed based on the leading term of L_n under H_0, there are just minor differences between the sizes. Similarly to the corresponding results for the univariate case discussed in Chen and Gao (2007), however, there is some substantial power reduction when just using $L_{e,n}$, the bivariate version of the test proposed in Horowitz and Spokoiny (2001). This further demonstrates the advantage of using the EL–based adaptive test-L_n over its natural competitor-$L_{e,n}$.

6. CONCLUSIONS

This paper has proposed an EL–based simultaneous test for parametric specification of both the conditional mean and conditional variance functions in a nonlinear time series regression model. The proposed simultaneous test is particularly useful to deal with the case where there is no knowledge about whether the conditional mean and/or variance functions are correctly specified. Both an asymptotic distribution of the proposed simultaneous test and asymptotic consistency results of an adaptive version of the proposed test have been established and proved.

The proposed simultaneous test has been implemented using both simulated and real data examples. As shown in Section 5 above, the proposed test performs well numerically even when one of the conditional mean and conditional variance functions is already correctly specified. In this case, there is only a small power reduction in each individual situation when using the simultaneous test while a univariate test should be used instead.

Future discussion includes the following two issues. The first issue is whether one could extend the proposed test to accommodate the case where the dimensionality of $\{X_t\}$ is sufficiently large. This could involve using a nonparametric additive form to approximate each of the conditional mean and variance functions. The second issue is whether one could allow for the inclusion of discrete components in $\{X_t\}$. To be able to deal with this, one would need to extend recent work on univariate specification testing (see, for example, Li and Racine, 2007, Ch. 12) to a simultaneous setting.

REFERENCES

Ahn, D.H. & B. Gao (1999) A parametric nonlinear model of term structure dynamics. *Review of Financial Studies* 12, 721–762.

Aït-Sahalia, Y. (1996) Testing continuous–time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.

Aït-Sahalia, Y., J. Fan, & H. Peng (2009) Nonparametric transition–based tests for jump diffusions. *Journal of the American Statistical Association* 487, 1102–1116.

Benjamini, Y. & Y. Hochberg (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society Series B* 57, 289–300.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes* (*Lecture Note in Statistics*, 110). Springer.

Brown, B. & S.X. Chen (1998) Combined and least squares empirical likelihood. *Annals of the Institute of Statistical Mathematics* 50, 697–714.

Chan, K.C., D.A. Karolyi, F.A. Longstaff, & A.B Sanders (1992) An empirical comparison of alternative models of the short–term interest rate. *Journal of Finance* 47, 1209–1227.

Chen, S.X. & H. Cui (2006) On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* 93, 215–220.

Chen, S.X. & H. Cui (2007) On the second order properties of empirical likelihood with moment restrictions. *Journal of Econometrics* 141, 492–516.

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

Chen, S.X., J. Gao, & C. Tang (2008) A test for model specification of diffusion processes. *Annals of Statistics* 36, 167–198.

Chen, S.X., W. Härdle, & M. Li (2003) An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society, Series B* 65, 663–678.

Chen, X. & Y. Fan (1999) Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of term structure of interest rates. *Econometrica* 53, 385-407.

Dette, H. & B. Hetzler (2007) Specification tests indexed by bandwidths. *Sankhy Series A* 69, 28–54.

Donald, S., G. Imbens, & W.K. Newey (2003) Empirical likelihood estimation and tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.

Einmahl, J.H.J. & I.W. McKeague (2003) Empirical likelihood based hypothesis testing. *Bernoulli* 9, 267–290.

Escanciano, J.C. (2008) Joint diagnostic tests for conditional mean and variance specifications. *Journal of Econometrics* 143, 74–87.

Escanciano, J.C. & C. Velasco (2008) Specification tests of parametric dynamic conditional quantiles. CAEPR Working paper #2008–021. Indiana University Bloomington.

Eubank, R.L. & 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. & I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*. Chapman & Hall.

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

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

Fan, J. & C. Zhang (2003) A re-examination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association* 98, 118–134.

Fan, J., C. Zhang, & J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Fan, J. & J. Zhang (2004) Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics* 32, 1858–1907.

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

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

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

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

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

Gao, J., H. Tong & R.C.L. Wolff (2002) Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis* 83, 324–359.

Genon-Caralot, V., T. Jeantheau, & C. Laredo (2000) Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6, 1051–1079.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hart, J. (1997) *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer.

Hjellvik, V. & D. Tjøstheim (1995) Nonparametric tests of linearity for time series. *Biometrika* 82, 351–368.

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

Hong, Y. & H. Li (2005) Nonparametric specification testing for continuous-time models with application to spot interest rates. *Review of Financial Studies* 18, 37–84.

Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–632.

Ingster, Y.I. (1993a) Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Mathematical Methods of Statistics* 2, 85–114.

Ingster, Y.I. (1993b) Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Mathematical Methods of Statistics* 2, 171–189.

Ingster, Y.I. (1993c) Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Mathematical Methods of Statistics* 2, 249–268.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.

Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69, 1661–1672.

Kitamura, Y., G. Tripathi, & H. Ahn, (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667–1714.

Kreiss, J.P., M. Neumann, & Q. Yao (1999) Bootstrap tests for simple structures in nonparametric time series regression. Private communication.
Li, G. (2003) Nonparametric likelihood ratio goodness-of-fit tests for survival data. *Journal of Multivariate Analysis* 86, 166–182.
Li, G., M. Hollander, I. W. McKeague, & J. Yang (1996) Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Annals of Statistics* 24, 628–640.
Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
Li, Q. & J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics* 87, 145–165.
Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.
McKeague, I.W. & M.J. Zhang (1994) Identification of nonlinear time series from first order cumulative characteristics. *Annals of Statistics* 22, 495–514.
Owen, A. (1988) Empirical likelihood confidence intervals for a single functional. *Biometrika* 75, 237–249.
Owen, A. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.
Owen, A. (2001) *Empirical Likelihood*. Chapman & Hall.
Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating functions. *Annals of Statistics* 22, 300–325.
Simes, R.J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73, 751–754.
Spokoiny, V.G. (1996) Adaptive hypothesis testing using wavelets. *Annals of Statistics* 24, 2477–2498.
Tripathi, G. & Y. Kitamura (2003) Testing conditional moment restrictions. *Annals of Statistics* 31, 2059–2095.
Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
Xu, K.L. & P.C.B. Phillips (2006) Empirical likelihood re-weighted functional estimation of diffusion models. Mimeo, Yale University.
Zhang, C. & H. Dette (2004) A power comparison between nonparametric regression tests. *Statistics and Probability Letters* 66, 289–301.

APPENDIX A: Proof of Theorem 2.1

The proof of Theorem 2.1 requires several assumptions.

Assumptions A.1.

(i) Let the process (X_t, Y_t) be strictly stationary and α-mixing with the mixing coefficient $\alpha(t) \le C_\alpha \alpha^t$ 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 $s, t \ge 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants, and Ω_i^j denotes the σ-field generated by $\{(X_k, Y_k) : i \le k \le j\}$.

(ii) Let $\{e_i\}$ defined in (1.1) satisfy for all $i \ge 1$

$$P\left(\mathrm{E}[e_i|\Omega_{i-1}] = 0\right) = P\left(\mathrm{E}[e_i^2|\Omega_{i-1}] = 1\right) = 1,$$

where $\Omega_i = \sigma\{(X_{j+1}, Y_j) : 1 \le j \le i\}$ is a sequence of σ-fields generated by $\{(X_{j+1}, Y_j) : 1 \le j \le i\}$.

(iii) Let ζ_i be either $\epsilon_i = Y_i - m_1(X_i) = \sqrt{m_2(X_i)}e_i$ or $\eta_i = \epsilon_i^2 - m_2(X_i) = m_2(X_i)$ $[e_i^2 - 1]$. Assume that there is some positive constant $r > 4$ such that $\mathrm{E}\left[|\zeta_i|^r\right] < \infty$.

Assumptions A.2.

(i) Assume $\inf_{x\in S} m_2(x) \geq C_m > 0$ for some constant C_m. For $k,l = 0,1,2$, let $\sigma_{kl}(x) = \mathrm{E}[\epsilon_i^k \eta_i^l | X_i = x]$. Assume that $m_1(x)$ and $\sigma_{kl}(x)$ for $(k,l) = (2,0)$, $(0,2)$, and $(1,1)$ all have continuous derivatives of up to the second order and satisfy the Lipschitz condition

$$|m_1(u) - m_1(v)| \leq C_0||u-v|| \quad \text{and} \quad \left|\sigma_{jk}(u) - \sigma_{jk}(v)\right| \leq C_{kl}||u-v||$$

with $u, v \in S$ and some constants $0 < C_0, C_{kl} < \infty$ for $(k,l) = (2,0)$, $(1,1)$, and $(0,2)$. In addition, suppose that there are two constants $0 < c_\sigma < C_\sigma < \infty$ such that $0 < c_\sigma \leq \sigma_{20}(x)\sigma_{02}(x) - \sigma_{11}^2(x) \leq C_\sigma < \infty$ uniformly in $x \in S$.

(ii) The weight function π is supported on the compact set S and $0 < \pi \leq C$ for some constant C; the marginal density function, $f(x)$, of X_t has continuous first two derivatives on R^d and $0 < c_f \leq f(x) \leq C_f < \infty$ for all $x \in S$ for two positive constants c_f and C_f.

(iii) Let $f_{\tau_1,\tau_2,\ldots,\tau_l}(\cdot)$ 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:

$$\left|f_{\tau_1,\tau_2,\ldots,\tau_l}(u_1,\ldots,u_l) - f_{\tau_1,\tau_2,\ldots,\tau_l}(v_1,\ldots,v_l)\right| \leq D_f||u-v||$$

for $u = (u_1, \ldots, u_l)$ and $v = (v_1, \ldots, v_l) \in S$, $1 \leq l \leq 4$, and $0 < D_f < \infty$ is a constant.

Assumptions A.3.

(i) The kernel K is a product kernel defined by $K(x_1, \ldots, x_d) = \prod_{i=1}^d k(x_i)$, where $k(\cdot)$ is a univariate symmetric probability density function and Lipschitz continuous in its support; that is, $|k(t_1) - k(t_2)| \leq C|t_1 - t_2|$ for some positive constant C.

(ii) The smoothing bandwidths h_1 and h_2 satisfy $\lim_{n\to\infty} h_1 = 0$ and $\lim_{n\to\infty} \frac{nh_1^{2d}}{\log^6(n)} = \infty$. There is a constant $0 < \beta_0 < \infty$ such that $\lim_{n\to\infty} \frac{h_1}{h_2} = \beta_0$. Furthermore, we restrict $1 \leq d \leq 3$.

Assumptions A.4.

(i) Suppose that for any parametric estimator, $\overline{\theta}$, of θ

$$\max_{1\leq i\leq 2} \sup_{x\in S} |m_{i\overline{\theta}}(x) - m_{i\theta}(x)| = O_p\left(n^{-1/2}\right).$$

(ii) Assume that both $\Delta_{1n}(x)$ and $\Delta_{2n}(x)$ defined in (1.2) are continuous and uniformly bounded with respect to $x \in R^d$, and $C_{in} = n^{-1/2}h_i^{-d/4}$ for $i = 1, 2$.

Remarks A.1. The geometric strong mixing (GSM) assumed in Assumption A.1(i) can be weakened to $\alpha(t) \sim t^{-\eta(d)}$ for some sufficiently large $\eta(d)$ that depends on d. The GSM has been established for ARCH models by Masry and Tjøstheim (1995). Assumption A.1(ii) holds naturally. For example, when $\{X_i\}$ and $\{e_i\}$ are independent, Assumption A.1(ii) requires only that $\{e_i\}$ is a sequence of martingales satisfying $\mathrm{E}[e_i|e_{i-1}, \ldots, e_1] = 0$

and $E[e_i^2|e_{i-1},\dots,e_1]=1$. For this case, model (1.1) becomes a nonparametric ARCH model when $X_i=(Y_{i-1},\dots,Y_{i-d})$ and $\{e_i\}$ is a sequence of martingale differences. Assumption A.2 is a set of standard regularity conditions imposed on m_i, f, the joint conditional moment fuctions σ_{jk}, and the joint probability density functions. We have also not assumed that the marginal density of X_t has compact support. Instead, we impose some restrictions on the support of the weight function $\pi(\cdot)$. Assumption A.3(ii) on bandwidths h_l $(l=1,2)$ includes $h_l=O\left(n^{-1/d+4}\right)$; which is the optimal bandwidth that minimizes the mean integrated square errors of the curve estimates $\hat{m}_1(x)$ and $\hat{m}_2(x)$ and is also the optimal order selected by either the cross–validation or the plug-in method. The requirement on $d\le 3$ is to ensure that $E\{S(x)\}-\Sigma_1(x)$, which is $O(h_1^2)$, is of an order smaller than $h_1^{d/2}$. It is known that the kernel method will encounter the curse of dimensionality when $d\ge 4$. To allow for $d\ge 4$, a κth $(\kappa>2)$ order kernel may need to be employed such that $E\{S(x)\}-\Sigma_1(x)$ is reduced to h_1^κ. This then permits d to be extended to $d<2\kappa$. Assumption A.4(i), which requires the $\sqrt{n}$–rate of convergence for the parametric case, is a standard condition. It holds when $\overline{\theta}$ is a $\sqrt{n}$–consistent estimator of θ. Assumption A.4(ii) imposes a reasonable restriction on C_{in} for this kind of kernel test.

To evaluate the asymptotic properties of log-EL ratio $\ell\{\tilde{m}_{\tilde{\theta}}(x)\}$, the uniform order of the Lagrange multiplier $\lambda(x)$ defined in (2.3) is studied in the next lemma.

LEMMA A.1. Under Assumptions *A.1–A.3,* we have

$$\sup_{x\in S}||\lambda(x)||=o_p\{(nh_1^d)^{-1/2}\log(n)\}.$$

Proof. Let $S_n(x)=n^{-1}\sum_{i=1}^n Q_i(x)Q_i^\tau(x)$ and recall that

$$\bar{U}_1(x)=(nh_1^d)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{h_1}\right)\{Y_i-\tilde{m}_{1\tilde{\theta}}(x)\}\quad\text{and}\tag{A.1}$$

$$\bar{U}_2(x)=(nh_1^d)^{-1}\sum_{i=1}^n K\left(\frac{x-X_i}{h_2}\right)\left[\{Y_i-\tilde{m}_{1\tilde{\theta}}(X_i)\}^2-\tilde{m}_{2\tilde{\theta}}(x)\right].\tag{A.2}$$

Following Owen (1990), to prove the lemma we need to show that, for any $\eta\in R^2$ and $||\eta||=1$,

$$\sup_{x\in S}|\bar{U}_j(x)|=o_p\{(nh_1^d)^{-1/2}\log(n)\},\quad\text{for } j=1,2\tag{A.3}$$

$$P\{\inf_{x\in S}\eta^\tau S_n(x)\eta h_1^{-d}\ge d_0\}=1\quad\text{for a positive } d_0>0,\quad\text{and}\tag{A.4}$$

$$\max_{1\le i\le n}\sup_{x\in S}||Q_i(x)||=o_p\{(nh_1^d)^{1/2}\log^{-1}(n)\},\tag{A.5}$$

where $Q_i(x)=\left[K\left(\frac{x-X_i}{h_1}\right)\{Y_i-\tilde{m}_{1\tilde{\theta}}(x)\},K\left(\frac{x-X_i}{h_2}\right)\left[\{Y_i-\tilde{m}_{1\tilde{\theta}}(x)\}^2-\tilde{m}_{2\tilde{\theta}}(x)\right]\right]^\tau$.

The proof of (A.3) for $j=1$ has been given in Chen et al. (2003), and for $j=2$ it is almost the same. To prove (A.4), we note that, following standard techniques to establish uniform convergence for α-mixing sequences, for instance those given in Bosq (1998),

$$h_1^{-d}S_n(x)=\Sigma_1(x)+\tilde{O}_p\{(nh^d)^{-1/2}\log n+h_1^2\},\tag{A.6}$$

where $\Sigma_1(x)$ is defined in (2.5) and $\tilde{O}_p(\delta_n)$ denotes the term that is $O_p(\delta_n)$ after taking suprema over $x \in S$ for a nonrandom sequence δ_n. Similar understanding should be given for $\tilde{o}_p(\delta_n)$.

It can be shown by applying the Cauchy-Schwarz inequality that $R(\beta^{-1})R(\beta) \leq 1$. This, along with Assumption A.2(ii), implies that $\Sigma_1(x)$ is positive definite at each x and that the smallest eigenvalue of $\Sigma_1(x)$ is uniformly bounded away from zero. This then implies (A.4).

Let $w_i = \sup_{x \in S} ||Q_i(x)||$. As K, m, and Δ_n are all bounded in S, then $w_i \leq C_1(|\epsilon_i| + |\eta_i|) + C_2$. From the Chebyshev inequality and Assumption A.1(iii),

$$P\left(w_i > (nh_1^d)^{1/2}\{\log(n)\}^{-1}\right)$$
$$\leq P\left(|\epsilon_i| \geq C_3(nh_1^d)^{1/2}\{\log(n)\}^{-1}\right) + P\left(|\eta_i| \geq C_3(nh_1^d)^{1/2}\{\log(n)\}^{-1}\right)$$
$$\leq C_4(nh_1^d)^{-r/2}\log^r(n)\}$$

for $r > 4$. This yields

$$\sum_{n=1}^{\infty} P\left(w_i > (nh_1^d)^{1/2}\{\log(n)\}^{-1}\right) \leq C_5 \sum_{n=1}^{\infty} (nh_1^d)^{-r/2}\log^r(n)$$
$$\leq C_5 \sum_{n=1}^{\infty} n^{-r/4}\log^r(n) < \infty$$

using $\lim_{n\to\infty} nh_1^{2d} = \infty$.

The Borel-Cantelli lemma implies that $w_i > (nh_1^d)^{1/2}\{\log(n)\}^{-1}$ finitely often with probability one. This means that $Z_n = \max_{1\leq i\leq n} w_i > (nh_1^d)^{1/2}\{\log(n)\}^{-1}$ finitely often. Equation (A.5) is therefore proved. ■

Derivation of (2.6). From Lemma A.1, a Taylor expansion from (2.3) yields

$$\sum_{i=1}^{n} Q_i(x)[1 - Q_i^{\tau}(x)\lambda(x)] + \tilde{O}_p\{(nh_1^d)^{-1/2}\log(n)\} = 0.$$

Inverting the above expansion,

$$\lambda(x) = S^{-1}(x)\bar{U}(x) + \tilde{o}_p\{(nh_1^d)^{-1}\log^2(n)\}$$
$$= \Sigma_1^{-1}(x)\bar{U}(x) + \tilde{O}_p\{(nh_1^d)^{-1}\log^2(n) + h_1^2(nh_1^d)^{-1/2}\log(n)\}. \qquad \textbf{(A.7)}$$

An expansion to the log-empirical likelihood ratio is

$$\ell\{\tilde{m}_{\tilde{\theta}}(x)\} = 2\sum_{i=1}^{n} \log\{1 + \lambda^{\tau}(x)Q_i(x)\}$$
$$= 2\lambda^{\tau}(x)\sum Q_i(x) - \lambda^{\tau}(x)\sum Q_i(x)Q_i(x)\lambda(x) + \tilde{o}_p\{(nh_1^d)^{-1/2}\log^3(n)\}$$

$$= (nh_1^d)\bar{U}^\tau(x)\Sigma_1^{-1}(x)\bar{U}(x) + \tilde{o}_p\{(nh_1^d)^{-1/2}\log^3(n) + h_1^2\log^2(n)\}$$

$$= nh_1^d\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}^\tau \Sigma_0^{-1}(x)\{\hat{m}(x) - \tilde{m}_{\tilde{\theta}}(x)\}$$

$$+ \tilde{o}_p\{(nh_1^d)^{-1/2}\log^3(n) + h_1^2\log^2(n)\}, \qquad \textbf{(A.8)}$$

which establishes (2.6). ■

Proof of Theorem 2.1. Recall that $\epsilon_i = Y_i - m_1(X_i) = \sqrt{m_2(X_i)}e_i$ and $\eta_i = m_2(X_i)(e_i^2 - 1)$. Hence,

$$\bar{U}_1(x) = n^{-1}\sum K_{h_1}(x - X_i)\{\epsilon_i + C_{1n}\Delta_{n1}(X_i)\} + O_p(n^{-1/2}), \qquad \textbf{(A.9)}$$

$$\bar{U}_2(x) = n^{-1}\beta^d\sum K_{h_2}(x - X_i)\left[\{\epsilon_i + C_{1n}\Delta_{n1}(X_i)\}^2 - \tilde{m}_{2\tilde{\theta}}(x)\right]$$

$$= n^{-1}\beta^d\sum K_{h_2}(x - X_i)\{\eta_i + C_{2n}\Delta_{n2}(X_i)\}$$

$$+ O_p\{n^{-1/2} + n^{-1}h_1^{-3d/4}\log n\} \qquad \textbf{(A.10)}$$

as $\hat{\theta}$ is an $\sqrt{n}$–consistent estimator of θ and $\sup_x |m_\theta(x) - m_{\tilde{\theta}}(x)| = O_p(n^{-1/2})$.

From (2.5),

$$\Sigma_1^{-1}(x) = \frac{f^{-1}(x)R^{-1}(K)}{\beta^d\sigma_{02}(x)\sigma_{20}(x) - R^2(\beta, K)\sigma_{11}^2(x)}\begin{pmatrix} \beta^d\sigma_{02}(x) & -R(\beta, K)\sigma_{11}(x) \\ -R(\beta, K)\sigma_{11}(x) & \sigma_{20}(x)\end{pmatrix}$$

$$= \begin{pmatrix} \nu_{11}(x) & \nu_{12}(x) \\ \nu_{21}(x) & \nu_{22}(x)\end{pmatrix}.$$

For $i = 1, 2$, let $W_{ti}(x) = n^{-1}K_{h_i}(x - X_t)$, and for $s, t = 1, 2, \ldots, n$,

$$a_{st} = nh_1^d\int W_{s1}(x)W_{t1}(x)\nu_{11}(x)\pi(x)dx,$$

$$b_{st} = nh_1^d\int W_{s1}(x)W_{t2}(x)\nu_{12}(x)\pi(x)dx \quad \text{and}$$

$$c_{st} = nh_1^d\int W_{s2}(x)W_{t2}(x)\nu_{22}(x)\pi(x)dx.$$

Then (2.7), (A.9), and (A.10) lead to $N_n(h) = S_{n1}(h) + S_{n2}(h) + S_{n3}(h) + o_p(h_1^{d/2})$, where

$$S_{1n}(h) = \sum_{1\le s\ne t\le n}\{a_{st}\epsilon_s\epsilon_t - 2b_{st}\epsilon_s\eta_t + c_{st}\eta_s\eta_t\},$$

$$S_{2n}(h) = \sum_{s=1}^{n} \{a_{ss}\epsilon_s^2 - 2b_{ss}\epsilon_s\eta_s + c_{ss}\eta_s^2\},$$

$$S_{3n}(h) = nh_1^d \left[C_{1n}^2 \int \left(\sum_{s=1}^{n} W_{s1}(x)\Delta_{n1}(X_s)\right)^2 \nu_{11}(x)\pi(x)dx\right]$$
$$+ nh_1^d \left[C_{2n}^2\beta^{2d} \int \left(\sum_{s=1}^{n} W_{s2}(x)\Delta_{n2}(X_s)\right)^2 \nu_{22}(x)\pi(x)dx\right]$$
$$- 2nh_1^d \left[C_{1n}C_{2n}\beta^d \int \left(\sum_{s=1}^{n} W_{s1}(x)\Delta_{n1}(X_s)\right)\right.$$
$$\left.\times \left(\sum_{t=1}^{n} W_{t2}(x)\Delta_{n2}(X_t)\right)\nu_{12}(x)\pi(x)dx\right].$$

Standard derivations show that $\mathrm{E}\{S_{n2}(h)\} = 2 + O(h_1^2)$ and $\mathrm{Var}\{S_{n2}(h)\} = o(h_1^d)$. Hence,

$$S_{n2}(h) = 2 + o_p(h_1^{d/2}). \tag{A.11}$$

Since $\sum_{i=1}^{n} W_{sl}(x)\Delta_{nl}(X_s) = \Delta_{nl}(x)f(x) + \tilde{o}_p\{(nh_1^d)^{-1/2}\log n + h_1^2\}$ for $l = 1$ and 2, and the fact that $C_{ni} = n^{-1/2}h_i^{-d/4}$,

$$S_{n3}(h) = h_1^{d/2} \int f^2(x)\Delta_n^\tau(x)\Sigma_1^{-1}(x)\Delta_n(x)\pi(x)dx + o_p(h_1^{d/2}). \tag{A.12}$$

Therefore, both $S_{1n}(h)$ and $S_{2n}(h)$ contribute only to the mean of $N_n(h)$. It remains to establish the variance and then the asymptotic normality of $S_{n1}(h)$. Let $\phi_{st} = \phi_{st}(h) = a_{st}\epsilon_s\epsilon_t - 2b_{st}\epsilon_s\eta_t + c_{st}\eta_s\eta_t$. Then, $S_{1n}(h) = \sum_{1\le s\ne t\le n}\phi_{st}$ is a degenerate U-statistic. Let $\sigma_{st}^2 = \mathrm{Var}(\phi_{st})$ and $\sigma_n^2(h) = \sum_{1\le i\ne j\le n}\sigma_{ij}^2$.

Let us derive $\sigma_n^2(h)$ and then prove the asymptotic normality of $S_{n1}(h)$. Observe that

$$\sigma_n^2(h) = \sum_{1\le s\ne t\le n} \mathrm{E}\left[\phi_{st}^2\right] = \sum_{1\le s\ne t\le n} \mathrm{E}\left[a_{st}^2\epsilon_s^2\epsilon_t^2 + 4b_{st}^2\epsilon_s^2\eta_t^2 + c_{st}^2\eta_s^2\eta_t^2\right]$$
$$+2\sum_{1\le s\ne t\le n} \mathrm{E}\left[a_{st}c_{st}\epsilon_s\epsilon_t\eta_s\eta_t - 2a_{st}b_{st}\epsilon_s^2\epsilon_t\eta_t - 2b_{st}c_{st}\epsilon_s\eta_s\eta_t^2\right]$$
$$\equiv \sum_{l=1}^{6}\sigma_{ln}^2(h). \tag{A.13}$$

It can be shown according to the definition of a_{st} that

$$a_{st}^2 = \int\int \frac{1}{(nh_1^d)^2} K\left(\frac{x-X_s}{h_1}\right) K\left(\frac{y-X_s}{h_1}\right) K\left(\frac{x-X_t}{h_1}\right) K\left(\frac{y-X_t}{h_1}\right)$$
$$\times \nu_{11}(x)\pi(x)\nu_{11}(y)\pi(y)dxdy.$$

Thus,

$$\mathrm{E}\left[a_{st}^2\epsilon_s^2\epsilon_t^2\right] = \mathrm{E}\left\{a_{st}^2\mathrm{E}\left[\epsilon_s^2\epsilon_t^2|(X_s,X_t)\right]\right\} = \mathrm{E}\left[a_{st}^2\sigma_{20}(X_s)\sigma_{20}(X_t)\right]$$
$$= \frac{1}{(nh_1^d)^2}\int\int \nu_{11}(x)\pi(x)\nu_{11}(y)\pi(y)\mathrm{E}\left[K\left(\frac{x-X_s}{h_1}\right)K\left(\frac{y-X_s}{h_1}\right)\right.$$
$$\left.\times K\left(\frac{x-X_t}{h_1}\right)K\left(\frac{y-X_t}{h_1}\right)\sigma_{20}(X_s)\sigma_{20}(X_t)\right]dxdy.$$

Using Assumptions A.2 and A.3, we have as $n \to \infty$

$$\mathrm{E}\left[K\left(\frac{x-X_s}{h_1}\right)K\left(\frac{y-X_s}{h_1}\right)K\left(\frac{x-X_t}{h_1}\right)K\left(\frac{y-X_t}{h_1}\right)\sigma_{20}(X_s)\sigma_{20}(X_t)\right]$$
$$= h_1^{2d}\int\int K\left(s+\frac{x-y}{h}\right)K(s)K(t)K\left(t-\frac{x-y}{h_1}\right)\sigma_{20}(x-th_1)\sigma_{20}(y-sh_1)$$
$$\times f(x-th_1,y-sh_1)dsdt$$
$$= h_1^{2d}L^2\left(\frac{x-y}{h_1}\right)\sigma_{20}(x)\sigma_{20}(y)f(x,y)(1+o(1)),$$

where $L(x) = \int K(x+y)K(y)dy$ and $f(x,y)$ is the joint density of (X_s, X_t). Therefore, as $n \to \infty$,

$$\sigma_{1n}^2 \equiv \sum_{1\le s\neq t\le n} \mathrm{E}\left[a_{st}^2\epsilon_s^2\epsilon_t^2\right]$$
$$= \int\int \nu_{11}(x)\pi(x)\nu_{11}(y)\pi(y)\sigma_{20}(x)\sigma_{20}(y)L^2\left(\frac{x-y}{h_1}\right)f(x,y)$$
$$\times dxdy(1+o(1)). \tag{A.14}$$

The other terms can be derived in a similar fashion. It may be shown by matching the terms with the definition of Σ that

$$\sigma_n^2(h) = 2\int\int \mathrm{tr}\{\Sigma(x,y)\Sigma^{-1}(y,y)\Sigma^{\tau}(x,y)\Sigma^{-1}(x,x)\}\pi(x)\pi(y)dxdy$$
$$= 2\int\int\sum_{i,j=1}^{2}\omega_{ij}^2(x,y)\pi(x)\pi(y)dxdy, \tag{A.15}$$

where $\{\omega_{ij}(x,y)\}$ is the (i,j) element of $\Omega(x,y)$ and $\mathrm{tr}(M)$ denotes the trace of M, in which $\Omega(x,y)$ is the same as in (2.8).

We now need to establish the asymptotic normality of $S_{n1}(h)$. Let $\gamma_i = (X_i, \epsilon_i, \eta_i)$, $P(\gamma_i)$, $P(\gamma_i,\gamma_j)$, $P(\gamma_i,\gamma_j,\gamma_k)$, and $P(\gamma_i,\gamma_j,\gamma_k,\gamma_l)$ be the probability measures of γ_i, (γ_i,γ_j), $(\gamma_i,\gamma_j,\gamma_k)$, and $(\gamma_i,\gamma_j,\gamma_k,\gamma_l)$ for different $i,j,k,l \in \{1,\ldots,n\}$, respectively. Define for some constant $\delta > 0$,

$$M_{n1} = \max_{1<i<j\le n}\max\left\{\mathrm{E}\left[|\phi_{1j}\phi_{ij}|^{1+\delta}\right], \int|\phi_{1j}\phi_{ij}|^{1+\delta}dP(\gamma_1)dP(\gamma_i,\gamma_j)\right\},$$

$$M_{n21} = \max_{1<i<j\le n} \max\left\{ \mathrm{E}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right], \int |\phi_{1j}\phi_{ij}|^{2(1+\delta)} dP(\gamma_1) dP(\gamma_i,\gamma_j) \right\},$$

$$M_{n22} = \max_{1<i<j\le n} \max\left\{ \int |\phi_{1j}\phi_{ij}|^{2(1+\delta)} dP(\gamma_1,\gamma_i) dP(\gamma_j), \right.$$

$$\left. \times \int |\phi_{1j}\phi_{ij}|^{2(1+\delta)} dP(\gamma_1) dP(\gamma_i) dP(\gamma_j) \right\},$$

$$M_{n2} = \max\{M_{n21}, M_{n22}\},$$

$$M_{n3} = \max_{1<i<j\le n} \mathrm{E}|\phi_{1j}\phi_{ij}|^2, \quad M_{n4} = \max_{1<i,j,k\le n} \left\{ \max_P \int |\phi_{1j}\phi_{jk}|^{2(1+\delta)} dP \right\},$$

$$M_{n5} = \max_{1<i<j} \max\left\{ \mathrm{E}\left|\int \phi_{1i}\phi_{1j} dP(\gamma_1)\right|^{2(1+\delta)}, \right.$$

$$\left. \times \int \left|\int \phi_{1i}\phi_{1j} dP(\gamma_1)\right|^{2(1+\delta)} dP(\gamma_i) dP(\gamma_j) \right\},$$

$$M_{n6} = \max_{1<i<j} \mathrm{E}\left|\int \phi_{1i}\phi_{1j} dP(\gamma_1)\right|^2,$$

where the maximization over P is taken over $P(\gamma_1,\gamma_i,\gamma_j,\gamma_k)$, $P(\gamma_1)P(\gamma_i,\gamma_j,\gamma_k)$, $P(\gamma_1)$ $P(\gamma_i)P(\gamma_j,\gamma_k)$, and $P(\gamma_1)P(\gamma_i)P(\gamma_j)P(\gamma_k)$ for mutually different i, j, k.

According to Theorem A.1 of Gao (2007), $\{\sigma_n(h)\}^{-1} S_{n1}(h)$ is asymptotically normal if for some $\delta > 0$, as $n \to \infty$,

$$\max \sigma_n^{-2} \left\{ n^2 \left(M_{n1}^{1/1+\delta} + M_{n5}^{1/2(1+\delta)} + M_{n6}^{1/2} \right), \right.$$

$$\left. \times n^{3/2} \left(M_{n2}^{1/2(1+\delta)} + M_{n3}^{1/2} + M_{n2}^{1/2(1+\delta)} \right) \right\} \to 0. \quad \textbf{(A.16)}$$

To verify (A.16), we evaluate only the order of magnitude of M_{n2}, as the other terms can be investigated in a similar fashion. We first work on M_{n2}. Notice that for $l = 1$ and 2.

$$\left| \int K_{h_l}(x - X_i) K_{h_l}(x - X_j) v_{ll}(x) \pi(x) dx \right| \le C L_{h_l}(X_i - X_j) \quad \text{and} \quad \textbf{(A.17)}$$

$$\left| \int K_{h_1}(x - X_i) K_{h_2}(x - X_j) v_{ll}(x) \pi(x) dx \right| \le C L_{h_2}(X_i - X_j, \beta), \quad \textbf{(A.18)}$$

where $L_{h_l}(t) = h_l^{-d} L(t/h_l) = h_l^{-d} K^{(2)}(t/h_l)$ and $L_{h_l}(t,\beta) = h_l^{-d} L(t/h_l,\beta)$. Assumption A.1(iii) implies that $\mathrm{E}\left[|\zeta_1\zeta_i\zeta_j^2|^{2(1+\delta)p}\right]$ are all bounded where ζ_l is either ϵ_l or η_l for

$l = 1, i$ and j, and $p > 1$. Applying Hölder's inequality, for $1 < i < j \le n$ and some $r > 1$ such that $p^{-1} + r^{-1} = 1$,

$$\begin{aligned}
\mathrm{E}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right] \le C\Bigg\{ & \sum_{l=1}^{2} \left[\mathrm{E}\{L_{h_1}(X_1 - X_j)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r} \\
& + \left[\mathrm{E}\{L_{h_1}(X_1 - X_j)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r} \\
& + \sum_{l=1}^{2} \left[\mathrm{E}\{L_{h_2}(X_1 - X_j)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r} \\
& + \left[\mathrm{E}\{L_{h_2}(X_1 - X_j)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r} \\
& + \sum_{l=1}^{2} \left[\mathrm{E}\{L_{h_2}(X_1 - X_j, \beta)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r} \\
& + \left[\mathrm{E}\{L_{h_2}(X_1 - X_j, \beta)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r} \Bigg\}. \qquad \textbf{(A.19)}
\end{aligned}$$

Note that $L_h(t, \beta_1) = L_h(t)$ if $\beta_1 = 1$. Let f_{1ij} be the joint density of (X_1, X_i, X_j). Then for $\beta_1 = 1$ or β and $l, m = 1$ or 2,

$$\begin{aligned}
& \left[\mathrm{E}\{L_{h_l}(X_1 - X_j, \beta_1)L_{h_m}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r} \qquad \textbf{(A.20)} \\
& = (h_l h_m)^{-2d(1+\delta)+2d/r} \\
& \quad \times \left[\int\int\int \{L(u, \beta_1)L(v, \beta_2)\}^{2(1+\delta)r} f_{1ij}(z - h_l u, z - h_m v, z)\,du\,dv\,dz\right]^{1/r} \\
& \le C h_1^{-4d(1+\delta)+2d/r}.
\end{aligned}$$

Therefore, if we choose r such that $1 < r < \frac{2}{1+\delta}$, then

$$\left\{\mathrm{E}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right]\right\}^{1/2(1+\delta)} \le C h^{-2d+d/r(1+\delta)} = o(n^{1/2}h_1^{-d}). \qquad \textbf{(A.21)}$$

Now let us consider the second term in M_{n2}. Let E_i and E_{ij} be expectations with respect to γ_i and (γ_i, γ_j), respectively. From (A.17) and (A.18), and applying the same argument as in (A.19), we have

$$\begin{aligned}
& \mathrm{E}_1\mathrm{E}_{ij}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right] \\
& \le C\mathrm{E}_1\Bigg\{ |\epsilon_1|^{2(1+\delta)} \sum_{l=1}^{2} \left[\mathrm{E}_{ij}\{L_{h_1}(X_1 - X_j)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r} \\
& \quad + |\epsilon_1|^{2(1+\delta)} \left[\mathrm{E}_{ij}\{L_{h_1}(X_1 - X_j)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r} \\
& \quad + |\eta_1|^{2(1+\delta)} \sum_{l=1}^{2} \left[\mathrm{E}_{ij}\{L_{h_2}(X_1 - X_j)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r}
\end{aligned}$$

$$+ |\eta_1|^{2(1+\delta)} \left[\mathrm{E}_{ij}\{L_{h_2}(X_1 - X_j)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r}$$
$$+ (|\epsilon_1| + |\eta_1|)^{2(1+\delta)} \sum_{l=1}^{2} \left[\mathrm{E}_{ij}\{L_{h_2}(X_1 - X_j, \beta)L_{h_l}(X_i - X_j)\}^{2(1+\delta)r}\right]^{1/r}$$
$$+ (|\epsilon_1| + |\eta_1|)^{2(1+\delta)} \left[\mathrm{E}_{ij}\{L_{h_2}(X_1 - X_j, \beta)L_{h_2}(X_i - X_j, \beta)\}^{2(1+\delta)r}\right]^{1/r}\Bigg\}.$$

Note that (A.20) is still true if we replace E there by $\mathrm{E}_1\mathrm{E}_{ij}$. Therefore, if we choose q such that $1 < r < \frac{2}{1+\delta}$,

$$\left\{\mathrm{E}_1\mathrm{E}_{ij}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right]\right\}^{1/2(1+\delta)} \le Ch_1^{-2d+d/r(1+\delta)} = o(n^{1/2}h_1^{-d}). \qquad \textbf{(A.22)}$$

The third term in M_{n2} is almost the same as the second term we have just evaluated, and hence

$$\left\{\mathrm{E}_j\mathrm{E}_{1i}\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right]\right\}^{1/2(1+\delta)} \le Ch_1^{-2d+d/r(1+\delta)} = o(n^{1/2}h_1^{-d}). \qquad \textbf{(A.23)}$$

The last term in M_{n2} is in fact $\left\{\mathrm{E}_1\mathrm{E}_i\mathrm{E}_j\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right]\right\}^{1/2(1+\delta)}$, and it may be shown by applying the same method as above that

$$\left\{\mathrm{E}_1\mathrm{E}_i\mathrm{E}_j\left[|\phi_{1j}\phi_{ij}|^{2(1+\delta)}\right]\right\}^{1/2(1+\delta)} \le Ch_1^{-2d+d/r(1+\delta)} = o(n^{1/2}h_1^{-d}). \qquad \textbf{(A.24)}$$

Combining (A.21), (A.22), (A.23), and (A.24), we have $n^{3/2}M_{n2}^{1/2(1+\delta)} = o(\sigma_n^2(h))$. Thus, the verification of (A.16) is completed.

Therefore, in the light of (A.11), (A.12), (A.15), and the asymptotic normality of S_{n1}, we have, as $n \to \infty$,

$$\sigma_n(h)^{-1}\left(N_n(h) - 2 - h_1^{d/2}\int f^2(x)\Delta_n^\tau(x)\Sigma_1^{-1}(x)\Delta_n(x)\pi(x)dx\right) \xrightarrow{d} N(0,1). \qquad \textbf{(A.25)}$$

This completes the proof of Theorem 2.1. ∎

APPENDIX B: Proofs of Theorems 3.1–3.4

To avoid repeating the conditioning argument (given $\mathcal{X} = (X_1, \ldots, X_n)$) for each case in the following proofs of Lemmas B.1–B.10, we use P_* and E_* to represent the respective conditional probability and conditional expectation given $\mathcal{X}$. Unless otherwise stated, the corresponding conditioning arguments are all understood to be held in probability with respect to the joint distribution of $\mathcal{X} = (X_1, \ldots, X_n)$.

B.1. Assumptions. Let $\mathcal{M} = \{m_\theta(\cdot) : \theta \in \Theta\}$. For $i = 1, 2$, define $\nabla_\theta m_{i\theta}(x) = \frac{\partial m_{i\theta}(x)}{\partial\theta}$ and $\nabla_\theta^2 m_{i\theta}(x) = \frac{\partial^2 m_{i\theta}(x)}{\partial\theta\partial\theta'}$, 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 $v = (v_1, \ldots, v_q)^\tau$.

Assumptions 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 $x \in S$ and $i = 1, 2$, $m_{i\theta}(x)$ is twice differentiable almost surely with respect to $\theta \in \Theta$. In addition, suppose that there exist constants $0 < G_{ji} < \infty$ for $1 \leq i \leq 2$ and $1 \leq j \leq 3$ such that

$$\sup_{\theta\in\Theta} \int_{x\in S} |m_{i\theta}(x)|^2 f(x)dx \leq G_{1i} < \infty,$$

$$\sup_{\theta\in\Theta} \int_{x\in S} ||\nabla_\theta m_{i\theta}(x)||^2 f(x)dx \leq G_{2i} < \infty,$$

$$\sup_{\theta\in\Theta} \int_{x\in S} \left|\left|\nabla_\theta^2 m_{i\theta}(x)\right|\right|_m^2 f(x)dx \leq G_{3i},$$

where $||B||_m^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$ for $B = (b_{ij})_{1\leq i,j\leq q}$ and $f(x)$ denotes the marginal density of X_i.

(ii) For each $i = 1, 2$ and $\theta \in \Theta$, $m_{i\theta}(x)$ is twice differentiable almost surely with respect to $x \in R^d$.

(iii) Assume that there is a finite $C_I > 0$ such that for every $\varepsilon > 0$ and $i = 1, 2$

$$\inf_{\theta,\theta'\in\Theta:||\theta-\theta'||\geq\varepsilon} \int_{x\in S} \left[m_{i\theta}(x) - m_{i\theta'}(x)\right]^2 f(x)dx \geq C_I.$$

Assumptions B.2.

(i) Let H_0 be true. Then $\theta_0 \in \Theta$ and $\lim_{n\to\infty} P_*\left(\sqrt{n}||\tilde{\theta} - \theta_0|| > C_L\right) < \varepsilon$ for any $\varepsilon > 0$ and all sufficiently large C_L.

(ii) Under H_1 there is a $\tilde{\theta}_1 \in \Theta$ such that $\lim_{n\to\infty} P_*\left(\sqrt{n}||\tilde{\theta} - \tilde{\theta}_1|| > C_L\right) < \varepsilon$ for any $\varepsilon > 0$ and all sufficiently large C_L.

(iii) Assume that given Ω_1^n, there is a sequence of independent random errors $\{e_t^*\}$ such that for $1 \leq j \leq 4$,

$$P\left(\mathrm{E}\left[e_t^{*j}|\Omega_1^n\right] = \mathrm{E}[e_t^j|\Omega_{t-1}]\right) = 1, \tag{B.1}$$

where Ω_1^n and Ω_{i-1} are as defined in Assumption A.1. In addition, there is some constant $\delta_\zeta > 0$ such that $\mathrm{E}_*\left(\left|\zeta_{t_1}^{i_1}\zeta_{t_2}^{i_2}\cdots\zeta_{t_l}^{i_l}\right|^{1+\delta_\zeta}\right) < \infty$ holds in probability, where $\zeta_t = \sqrt{m_2(X_i)}e_{t*}$ or $m_2(X_i)\left[e_{t*}^2 - 1\right]$ with $e_{t*} = e_t$ or e_t^*, $1 \leq l \leq 4$ and $\sum_{j=1}^l i_j \leq 8$.

(iv) Let $\{\theta_{n,0} : n = 1, 2, \ldots\}$ be a sequence in Θ whose limit points, if any, are all in Θ. Define $Y_t^* = m_{1\theta_{n,0}}(X_t) + \sqrt{m_{2\theta_{n,0}}(X_t)}e_t^*$, where $\{e_t^*\}$ is sampled randomly from a distribution with zero mean and unit variance. Let $\hat{\theta}_n^*$ be the estimator of $\theta_{n,0}$ that is obtained from $\{Y_t^*, X_t : t = 1, 2, \ldots, n\}$. Then

$$\lim_{n\to\infty} P_*\left(\sqrt{n}||\hat{\theta}_n^* - \theta_{n,0}|| > C_L\right) < \varepsilon$$

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

Assumptions B.3.

(i) Assume that Assumptions A.1–A.3 hold.

(ii) Assume that $h_{i\max} = c_{i\max}(\mathrm{loglog}n)^{-1/d} > h_{i\min} = c_{i\min}n^{-\gamma_i}$ for $i = 1, 2$, where γ_i, $c_{i\min}$ and $c_{i\max}$ are some constants satisfying $0 < \gamma_i < \frac{1}{2d}$ and $0 < c_{i\min}, c_{i\max} < \infty$.

Assumptions B.1, B.2, and B.3(i) are quite standard in this kind of problem, and mirror Assumptions 1, 2, and 4 of Horowitz and Spokoiny (2001) for the fixed-design case. Assumption B.3(ii) allows the estimation-based optimal rate of $n^{-1/4+d}$ to be included in the range of $h_{i\min}$ and hence the bandwidth prescribed by either the cross-validation or the plug-in bandwidth selector can be used to guide the bandwidth set selection. Conditions similar to (B.1) are assumed in Franke et al. (2002). In addition, equation (B.1) is made deliberately general such that the theorems of Section 3 are valid for a wider range of situations. For example, equation (B.1) follows if there is a sequence of independent random errors $\{\tilde{\delta}_{tj}\}$ with $\mathrm{E}[\tilde{\delta}_{tj}|\Omega_1^n] = 0$ such that, for $j = 1, \ldots, 4$,

$$e_t^{*j} = \mathrm{E}[e_t^j|\Omega_{t-1}] + \tilde{\delta}_{tj}. \tag{B.2}$$

For practical implementations, more details on the innovation process are needed to facilitate the bootstrap generation of $\{e_t^*\}_{t=1}^n$. Under the assumption that

$$P\left(\mathrm{E}[e_t^j|\Omega_{t-1}] = \mathrm{E}[e_i^j]\right) = 1 \quad \text{for } 1 \le j \le 4, \tag{B.3}$$

we use the standard bootstrap for random design regression as outlined in Section 3. The justification is the following: Let $\hat{e}_t = \frac{Y_t - m_{1\tilde{\theta}}(X_t)}{\sqrt{m_{2\tilde{\theta}}(X_t)}}$, $\hat{m}_j = \frac{1}{n}\sum_{t=1}^n \hat{e}_t^j$ for $j = 1, 2, 3, 4$, and F_n be the empirical distribution of $\left\{\hat{e}_i : 1 \le i \le n\right\}$. Given $\{(X_t, Y_t) : 1 \le t \le n\}$, we draw a sequence of independent and identically distributed bootstrap resamples $\{e_t^* : t \ge 1\}$ from F_n.

B.2. Technical Lemmas. Before stating the necessary lemmas for the proofs of Theorems 3.1–3.4 given in Section 3, we introduce the following notation:

For $i = 1, 2$, let $W_{ti}(x) = n^{-1}K_{h_i}(x - X_t)$,

$$\lambda_{it}(\theta) = \lambda_i(X_t, \theta) = m_i(X_t) - m_{i\theta}(X_t) = m_{i\theta_0}(X_t) - m_{i\theta}(X_t), \tag{B.4}$$

$$\lambda_i(\theta) = (\lambda_{i1}(\theta), \ldots, \lambda_{in}(\theta))^\tau, \qquad \lambda(\theta) = (\lambda_1(\theta)^\tau, \lambda_2(\theta)^\tau)^\tau,$$

$$\Sigma_1^{-1}(x) = \begin{pmatrix} \nu_{11}(x) & \nu_{12}(x) \\ \nu_{21}(x) & \nu_{22}(x) \end{pmatrix},$$

$$W_i(x) = (W_{1i}(x), \ldots, W_{ni}(x))^\tau, \qquad W(x) = (W_1(x), W_2(x))^\tau,$$

$$A_{ij}(x) = W(x)\nu_{ij}(x)W(x)^\tau, \qquad A_{ij} = \int A_{ij}(x)\pi(x)dx,$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{B.5}$$

Recall from Appendix A that

$$a_{st} = nh_1^d \int W_{s1}(x)W_{t1}(x)\nu_{11}(x)\pi(x)dx,$$

$$b_{st} = nh_1^d \int W_{s1}(x)W_{t2}(x)\nu_{12}(x)\pi(x)dx,$$

$$c_{st} = nh_1^d \int W_{s2}(x)W_{t2}(x)\nu_{22}(x)\pi(x)dx,$$

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

Let

$$N_{0n}(h) = \sum_{s=1}^{n}\sum_{t=1}^{n}\phi_{st}, \qquad N_{1n}(h,\tilde{\theta}) = (nh_1^d)\int \tilde{U}(x)^\tau \Sigma_1^{-1}(x)\tilde{U}(x)\pi(x)dx \quad \text{and}$$

$$Q_n(\theta) = \lambda(\theta)^\tau A\lambda(\theta) = \sum_{s=1}^{n}\sum_{t=1}^{n}\left[a_{st}\lambda_{1s}(\theta)\lambda_{1t}(\theta) - 2b_{st}\lambda_{1s}(\theta)\lambda_{2t}(\theta) + c_{st}\lambda_{2s}(\theta)\lambda_{2s}(\theta)\right].$$

From (2.7),

$$N_n(h) = N_{1n}(h,\tilde{\theta}) + o_p(h_1^{d/2}) \quad \text{and} \quad N_{1n}(h,\tilde{\theta}) = N_{0n}(h) + Q_n(\tilde{\theta}) + \Pi_n(\tilde{\theta}), \tag{B.6}$$

where $\Pi_n(\theta) = N_{1n}(h,\tilde{\theta}) - N_{0n}(h) - Q_n(\tilde{\theta})$.

LEMMA B.1. *Let Assumptions B.1–B.3 hold.*

(i) *For every $\delta > 0$ and sufficiently large n, $\sup_{||\theta-\theta_0||\le\delta} Q_n(\theta) \le Cnh_1^d\delta^2$, in probability uniformly in $h \in \mathcal{H}_n$, where $C > 0$ is a constant.*

(ii) *For each $\theta \in \Theta$ and sufficiently large n, $C_1 h_1^d \cdot \lambda(\theta)^\tau\lambda(\theta) \le Q_n(\theta) \le C_2 h_1^d \cdot \lambda(\theta)^\tau\lambda(\theta)$, in probability for some $0 < C_1 \le C_2 < \infty$.*

Proof.

(i) It follows from the definition of $Q_n(\theta)$ that

$$Q_n(\theta) \le ||A||_\infty ||\lambda(\theta)||^2.$$

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

$$||A||_\infty \le Ch_1^d \tag{B.7}$$

in probability for some constant $C > 0$.

Note that $||A||_\infty \le \max_{1\le i,j\le 2}\{||A_{ij}||_\infty\}$. Thus, we just evaluate $||A_{11}||_\infty$ as the other three terms can be done similarly. Let $q(x) = \nu_{11}(x)\pi(x)$ and $\tilde{f}(x) = \frac{1}{nh_1^d}\sum_{i=1}^n K\left(\frac{x-X_t}{h_1}\right)$ be the kernel density estimator of f. Since $A_{11} = \{a_{st}\}_{n\times n}$, we have

$$||A_{11}||_\infty \le \max_{1\le t\le n}\sum_{s=1}^n |a_{st}| \le \max_{1\le t\le n}\int \tilde{f}(x)K\left(\frac{x-X_t}{h_1}\right)|q(x)|dx$$

$$= C(1+o_p(1))h_1^d \max_{1\le t\le n}\left(\int K(u)f(X_t+uh_1)|q(X_t+uh_1)|du\right) \le Ch_1^d.$$

Similarly, one can show that (B.7) is true for the other parts of $Q_n(\theta)$. In view of (B.7), in order to prove Lemma B.1(i), it suffices to show that in probability

$$\sup_{||\theta-\theta_0||\leq\delta} ||\lambda(\theta)||^2 \leq Cn\delta^2. \tag{B.8}$$

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) Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A, respectively. In view of

$$\lambda_{\min}(A)\cdot||\lambda(\theta)||^2 \leq Q_n(\theta) \leq \lambda_{\max}(A)\cdot||\lambda(\theta)||^2,$$

in order to prove Lemma B.1(ii), it suffices to show that, for n large enough,

$$\lambda_{\min}(A) \geq Ch_1^d(1+o_P(1)) \quad \text{in probability} \tag{B.9}$$

for some $C > 0$. Analogously to the proof of Lemma A.2 of Gao, Tong, and Wolff (2002), one can finish the proof of (B.9).

For simplicity, in the following lemmas and their proofs, we let $q = 1$. For $1 \leq i$, $j \leq 2$, 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 B.1–B.3, we have for any given $\theta \in \Theta$ and $i = 1, 2$*

$$J_{1n}^{-1/2} \max_{h_1\in\mathcal{H}_{1n}} h_1^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\epsilon_s\psi_{i1}(X_t,\theta)\right| = O_P(1), \tag{B.10}$$

$$J_{12n}^{-1/2} \max_{h\in\mathcal{H}_{1n}} h_1^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n b_{st}\epsilon_s\psi_{i1}(X_t,\theta)\right| = O_P(1), \tag{B.11}$$

$$J_{2n}^{-1/2} \max_{h_2\in\mathcal{H}_{2n}} h_2^{-d/2}\left|\sum_{s=1}^T\sum_{t=1}^T c_{st}\eta_s\psi_{i1}(X_t,\theta)\right| = O_P(1), \tag{B.12}$$

where $J_{12n} = J_{1n} \times J_{2n}$.

Proof. We prove (B.10) only; the others follow similarly. It suffices to show that for any large constant $C_0 > 0$

$$P_*\left[J_{1n}^{-1/2} \max_{h_1\in\mathcal{H}_{1n}} h_1^{-d/2}\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\epsilon_s\psi_{i1}(X_t,\theta)\right| > C_0\right]$$

$$\leq \sum_{h_1\in\mathcal{H}_{1n}} P_*\left[\left|\sum_{s=1}^n\sum_{t=1}^n a_{st}\epsilon_s\psi_{i1}(X_t,\theta)\right| > C_0 J_{1n}^{1/2}h_1^{d/2}\right]$$

$$\leq \sum_{h\in\mathcal{H}_{1n}} \frac{1}{C_0^2 J_{1n} h_1^d} \mathrm{E}_* \left[\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \psi_{i1}(X_t,\theta)\right]^2$$

$$\leq \sum_{h_1\in\mathcal{H}_{1n}} \frac{1}{C_0^2 J_{1n} h_1^d} \left\{\sum_{s=1}^n \sum_{t=1}^n \mathrm{E}_* \left[a_{st}\epsilon_s \psi_{i1}(X_t,\theta)\right]^2 + \Pi_{1n}(\theta)\right\}, \tag{B.13}$$

where $\Pi_{1n}(\theta) = \mathrm{E}_* \left[\sum_{(s,t)\neq(s',t')} a_{st}\epsilon_s \psi_{i1}(X_t,\theta) a_{s't'}\epsilon_{s'} \psi_{i1}(X_{t'},\theta)\right]$.

Similarly to (A.13) and (A.14), one can show that as $n\to\infty$

$$\sum_{s=1}^n \sum_{t=1}^n \mathrm{E}_* \left[a_{st}\epsilon_s \psi_{i1}(X_t,\theta)\right]^2 = C(\theta) h_1^d (1+o(1)) \tag{B.14}$$

for some function $C(\theta)$.

Analogously to Theorem A.1 of Gao (2007), we can show that as $n\to\infty$

$$\Pi_{1n}(\theta) = o(h_1^d). \tag{B.15}$$

Thus, equations (B.13)–(B.15) complete the proof. ■

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

$$J_n^{-1/2} \max_{h\in\mathcal{H}_n} h_1^{-d/2} \max_{1\leq s\leq n} \left|\sum_{t=1}^n d_{st}\zeta_t\right| = O_P(1), \tag{B.16}$$

where $\zeta_t = \epsilon_t$ or η_t, $d_{st} = a_{st}, b_{st}$, or c_{st}, and $J_n = J_{1n}$, J_{12n}, or J_{2n}.

Proof. The proof is similar to that of Lemma B.2 and therefore is omitted. ■

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

$$\max_{h\in\mathcal{H}_n} \sup_{|\theta-\theta_0|\leq n^{-1/2}u} h_1^{-d/2} \left|\sum_{s=1}^n \sum_{t=1}^n d_{st}\zeta_s \lambda_{it}(\theta)\right| = O_P\left(J_{12n}^{1/2} n^{-1/2}\right) = o_P(1) \tag{B.17}$$

under H_0, and for each $u>0$ and some $h\in\mathcal{H}_n$

$$\sup_{|\theta-\theta_0|\leq n^{-1/2}u} h_1^{-d/2} \left|\sum_{s=1}^n \sum_{t=1}^n d_{st}\zeta_s \lambda_{it}(\theta)\right| = o_P(1) \tag{B.18}$$

under H_1, where $d_{st} = a_{st}, b_{st}$, or c_{st} is as defined before.

Proof. We prove (B.17) for $d_{st} = a_{st}$, $\zeta_t = \epsilon_t$, and $i=1$ only. Using a Taylor series expansion for $m_{1\theta}(X_t) - m_{1\theta_0}(X_t)$ and Assumption B.1, we have for θ' between θ and θ_0

$$\left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \lambda_{1t}(\theta)\right| = \left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \left[m_{1\theta}(X_t) - m_{1\theta_0}(X_t)\right]\right| \tag{B.19}$$

$$\leq \left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \psi_{11}(X_t,\theta_0)\right| |\theta-\theta_0|$$

$$+\frac{1}{2}\left|\sum_{s=1}^n \sum_{t=1}^n a_{st}\epsilon_s \psi_{21}(X_t,\theta')\right| |\theta-\theta_0|^2.$$

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

$$\max_{h_1 \in \mathcal{H}_{1n}} \sup_{|\theta-\theta_0| \le n^{-1/2}u} h_1^{-d/2} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st} \zeta_s \lambda_{1t}(\theta) \right| \le O_P\left(J_{1n}^{1/2} n^{-1/2}\right)$$

as required. The proof of (B.18) follows similarly. ■

LEMMA B.5. *Let Assumptions B.1–B.3 hold. Then for every $u > 0$, some $h \in \mathcal{H}_n$, $i = 1, 2$, and any sequence $\{D_n\} \to \infty$ as $n \to \infty$*

$$\sup_{|\theta-\theta_0| \le n^{-1/2}u} \left| \sum_{s=1}^{n} \sum_{t=1}^{n} d_{st} \zeta_s \lambda_i(X_t, \theta) \right| = o_p(h_1^{d/2} D_n). \tag{B.20}$$

Proof. The proof is similar to that of Lemma B.4 and therefore omitted. ■

In view of the definition of $L_n(h) = \frac{N_n(h)-2}{h_1^{d/2}}$ and (B.6), let us define

$$L_{0n}(h) = \frac{N_{0n}(h) - 2}{h_1^{d/2}},$$

$$L_{1n}(h) = \frac{N_{1n}(h, \tilde{\theta}) - 2}{h_1^{d/2}}, \quad \text{and}$$

$$L_{2n}(h) = \frac{N_{1n}(h, \tilde{\theta}_1) - 2}{h_1^{d/2}}, \tag{B.21}$$

where $\tilde{\theta}_1 = \theta_0$ when H_0 is true and $\tilde{\theta}_1$ is as defined in Assumption B.2(ii) when H_0 is false.

Let $\tilde{\epsilon}_t^* = \sqrt{m_{2\tilde{\theta}}(X_t)} e_t^*$, $\tilde{\eta}_t^* = m_{2\tilde{\theta}}(X_t)[e_t^{*2} - 1]$, and $L_{0n}^*(h)$ be the version of $L_{0n}(h)$ with ϵ_t and η_t replaced by $\tilde{\epsilon}_t^*$ and $\tilde{\eta}_t^*$, respectively. For each $t = 1, 2, \ldots, n$, generate $Y_t^* = m_{1\tilde{\theta}}(X_t) + \sqrt{m_{2\tilde{\theta}}(X_t)} e_t^*$. Use the data set $\{Y_t^*, X_t : 1 \le t \le n\}$ to re-estimate θ and denote the resulting estimate by $\hat{\theta}^*$. Let $L_{1n}^*(h)$ be the version of $L_{1n}(h)$ with $\tilde{\theta}$, X_t, Y_t replaced with $\hat{\theta}^*$, X_t, and Y_t^*, respectively. In addition, let $\tilde{\phi}_{st}^* = a_{st} \tilde{\epsilon}_s^* \tilde{\epsilon}_t^* - 2b_{st} \tilde{\epsilon}_s^* \tilde{\eta}_t^* + c_{st} \tilde{\eta}_s^* \tilde{\eta}_t^*$.

LEMMA B.6.

(i) *Let Assumptions B.1–B.3 hold. Then, as $n \to \infty$,*

$$L_n(h) = L_{1n}(h) + o_P(1) = L_{2n}(h) + o_P(1) \quad \text{and} \quad L_{1n}^*(h) = L_{0n}^*(h) + o_P(1) \tag{B.22}$$

uniformly over $h \in \mathcal{H}_n$. Under H_0, uniformly over $h \in \mathcal{H}_n$,

$$L_{1n}(h) = L_{0n}(h) + o_P(1). \tag{B.23}$$

(ii) *Let Assumptions B.1–B.3 hold. Then, as $n \to \infty$,*

$$\max_{h \in \mathcal{H}_n} h_1^{-d/2} \left(\sum_{s=1}^{n} \phi_{ss} - 2 \right) = o_P(1) \quad \text{and} \quad \max_{h \in \mathcal{H}_n} h_1^{-d/2} \left(\sum_{s=1}^{n} \tilde{\phi}_{ss}^* - 2 \right) = o_P(1). \tag{B.24}$$

Proof. In view of the definitions given in (B.6) and (B.21), in order to prove the first part of (B.22), it suffices to show that

$$Q_n(\tilde{\theta}) - Q_n(\tilde{\theta}_1) = o_P\left(h_1^{d/2}\right) \quad \text{and} \quad \Pi_n(\tilde{\theta}) - \Pi_n(\tilde{\theta}_1) = o_P\left(h_1^{d/2}\right) \qquad \textbf{(B.25)}$$

uniformly over $h \in \mathcal{H}_n$.

The first part of (B.25) follows from Lemma B.1(i) using Assumption B.2. To prove the second part of (B.22), in view of (B.6), we have

$$\Pi_n(\theta) = 2\sum_{s=1}^{n}\sum_{t=1}^{n}\left[a_{st}\lambda_{1t}(\theta)\epsilon_s + b_{st}\lambda_{2t}(\theta)\epsilon_s + b_{st}\lambda_{1t}(\theta)\eta_s + c_{st}\lambda_{2t}(\theta)\eta_s\right].$$

Thus, the proof of the second part of (B.25) follows from Lemma B.4. Analogously, the proof of (B.23) follows using Assumptions B.1–B.3. The proof of the second part of (B.22) follows similarly using Assumptions B.1–B.3.

In view of the definitions of both ϕ_{st} and $\tilde{\phi}^*_{st}$, to prove (B.24) it suffices to show that as $n \to \infty$,

$$\max_{h\in\mathcal{H}_n} h_1^{-d/2}\sum_{s=1}^{n}\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right) = o_P(1) \quad \text{and}$$

$$\max_{h\in\mathcal{H}_n} h_1^{-d/2}\sum_{s=1}^{n}\left(\tilde{\phi}^*_{ss} - \mathrm{E}_*\left[\tilde{\phi}^*_{ss}\right]\right) = o_P(1). \qquad \textbf{(B.26)}$$

As the proof of the second part of (B.24) is the same as that of the first part, we prove only the first part. For any given small constant $\delta > 0$, similar to the proof of Lemma B.2 we have, as $n \to \infty$,

$$P_*\left(\max_{h\in\mathcal{H}_n} h_1^{-d/2}\left|\sum_{s=1}^{n}\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)\right| > \delta\right)$$

$$\le \sum_{h\in\mathcal{H}_n} h_1^{-d}\delta^{-2}\mathrm{E}_*\left[\sum_{s=1}^{n}\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)\right]^2$$

$$= \sum_{h\in\mathcal{H}_n} h_1^{-d}\delta^{-2}\left\{\sum_{s=1}^{n}\mathrm{E}_*\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)^2 + \sum_{s=1}^{n}\sum_{t=1,\neq s}^{n}\mathrm{E}_*\left[\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)\left(\phi_{tt} - \mathrm{E}_*\left[\phi_{tt}\right]\right)\right]\right\}$$

$$\le C\sum_{h\in\mathcal{H}_n}\frac{1}{nh_1^{2d}} \le C\frac{J_{12n}}{nh_{1\min}^{2d}} \to 0$$

using $\frac{J_{12n}}{nh_{1\min}^{2d}} = c_{\min}^{-1}\frac{J_{12n}}{n^{1-2d\gamma}} \to 0$ implied from Assumption B.3(ii), and

$$\sum_{s=1}^{n}\mathrm{E}_*\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)^2 = \frac{C}{n}, \qquad \textbf{(B.27)}$$

$$\sum_{s=1}^{n}\sum_{t=1,\neq s}^{n}\mathrm{E}_*\left[\left(\phi_{ss} - \mathrm{E}_*\left[\phi_{ss}\right]\right)\left(\phi_{tt} - \mathrm{E}_*\left[\phi_{tt}\right]\right)\right] = o\left(\frac{C}{n}\right),$$

which may be proved as in (B.14) and (B.15). Therefore, we have finished the proof of Lemma B.6. ■

LEMMA B.7. *Let Assumptions B.1–B.3 hold. Then,* $\max_{h\in\mathcal{H}_n} L_{1n}(h)$ *and* $\max_{h\in\mathcal{H}_n} L_{1n}^*(h)$ *have the same asymptotic distribution under* H_0.

Proof. In view of Lemma B.6, to prove Lemma B.7 it suffices to show that the distributions of $\max_{h\in\mathcal{H}_n}\sum_{s\neq t=1}^n \phi_{st}$ and $\max_{h\in\mathcal{H}_n}\sum_{s\neq t=1}^n \phi_{st}^*$ are asymptotically the same.

In order to prove that the result holds under H_0, in view of Assumption B.2(i), it suffices to show that the result holds under the notation

$$\epsilon_i = \sqrt{m_{2\theta_0}(X_i)}e_i, \qquad \eta_i = m_{2\theta_0}(X_i)(e_i^2 - 1),$$

$$\epsilon_i^* = \sqrt{m_{2\theta_0}(X_i)}e_i^*, \qquad \eta_i^* = m_{2\theta_0}(X_i)(e_i^{*2} - 1).$$

For $1 \le s \neq t \le n$, let $\xi_s = (\epsilon_s, \eta_s)^\tau$ and $\phi_{st}^* = \xi_s^{*\tau}\tilde{A}_{st}\eta_t^*$ with

$$\xi_s^* = (\epsilon_s^*, \eta_s^*)^\tau \quad \text{and} \quad \tilde{A}_{st} = (2h_1^{d/2})^{-1}\begin{pmatrix} a_{st} & -b_{st} \\ -b_{st} & c_{st}\end{pmatrix}.$$

We now need to show that the distributions of $\max_{h\in\mathcal{H}_n}\sum_{s\neq t=1}^n \phi_{st}$ and $\max_{h\in\mathcal{H}_n}\sum_{s\neq t=1}^n \phi_{st}^*$. We are going to take the approach of Horowitz and Spokoiny (2001) with modifications, as the stochastic quantities being considered are different. For $h \in \mathcal{H}_n$, let $u_t = \xi_t$ or ξ_t^* define

$$B_{hn}(u_1, \ldots, u_n) = \sum_{s\neq t} u_s^\tau \tilde{A}_{st} u_t. \tag{B.28}$$

Let $B_n(u_1, \ldots, u_n)$ be the sequence that is obtained by stacking $B_{hn}(u_1, \ldots, u_n)$ over $h \in \mathcal{H}_n$. Let $G(\cdot) = G_n(\cdot)$ be a three-times continuously differentiable function over $R^{J_{12n}}$. Define

$$C_n(G) = \sup_{x\in R^{J_{12n}}} \max_{i,j,k=1,2,\ldots,J_{12n}} \left|\frac{\partial^3 G(v)}{\partial v_i \partial v_j \partial v_k}\right|.$$

Like Horowitz and Spokoiny (2001), there are two steps in the proof of Lemma B.7. First, we want to show that

$$\left|E_*\left[G(B_n(\xi_1, \ldots, \xi_n))\right] - E_*\left[G(B_n(\xi_1^*, \ldots, \xi_n^*))\right]\right| \le C_0 C_n \left(\frac{J_{12n}}{nh_{1\min}^d}\right)^{3/2} \tag{B.29}$$

for any three-times differentiable $G(\cdot)$, some finite constant C_0, and all sufficiently large n. Then in the second step, (B.29) is used to show that $B_n(\xi_1, \ldots, \xi_n)$ and $B_n(\xi_1^*, \ldots, \xi_n^*)$ have the same asymptotic distribution.

Note that

$$\begin{aligned} &\left|E_*\left[G(B_n(\xi_1, \ldots, \xi_n))\right] - E_*\left[G(B_n(\xi_1^*, \ldots, \xi_n^*))\right]\right| \qquad \text{(B.30)} \\ &\quad \le \sum_{t=1}^n \left|E_*\left[G(B_n(\xi_1, \ldots, \xi_t, \xi_{t+1}^*, \cdots, \xi_n^*))\right] - E_*\left[G(B_n(\xi_1, \ldots, \xi_{t-1}, \xi_t^*, \ldots, \xi_n^*))\right]\right|, \end{aligned}$$

where $B_n(\xi_1, \ldots, \xi_n, \xi_{n+1}^*) = B_n(\xi_1, \ldots, \xi_n)$ and $B_n(\xi_0, \xi_1^*, \ldots, \xi_n^*) = B_n(\xi_1^*, \ldots, \xi_n^*)$.

We now derive an upper bound on the last term of the sum on the right-hand side of (B.30). Similar bounds can be derived for the other terms. Let U_{n-1}, Λ_n, and Λ_n^* denote the respective vectors that are obtained by stacking $U_{h,n}$, $\Lambda_{h,n}$, and $\Lambda_{h,n}^*$ over $h \in \mathcal{H}_n$, where

$$U_{h,n} = \sum_{s=1}^{n-1} \sum_{t=1,\neq s}^{n-1} \xi_s^\tau \tilde{A}_{st} \xi_t, \qquad \Lambda_{h,n} = 2 \sum_{s=1}^{n-1} \xi_s^\tau \tilde{A}_{sn} \xi_n, \qquad \Lambda_{h,n}^* = 2 \sum_{s=1}^{n-1} \xi_s^\tau \tilde{A}_{sn} \xi_n^*.$$

Using a Taylor expansion to the last term of the sum on the right-hand side of (B.30) about $\xi_n = \xi_n^* = 0$ gives

$$\begin{aligned} &\left|\mathrm{E}_*\left[G(B_n(\xi_1,\ldots,\xi_n))\right] - \mathrm{E}_*\left[G(B_n(\xi_1,\ldots,\xi_{n-1},\xi_n^*))\right]\right| \\ &\quad \le \left|\mathrm{E}_*\left[G'(U_{n-1})\left(\Lambda_n - \Lambda_n^*\right)\right]\right| + \frac{1}{2}\left|\mathrm{E}_*\left[\Lambda_n^\tau G''(U_{n-1})\Lambda_n - \Lambda_n^{*\tau} G''(U_{n-1})\Lambda_n^*\right]\right| \\ &\qquad + \frac{C_n(G)}{6}\left\{\mathrm{E}_*\left[||\Lambda_n||^3\right] + \mathrm{E}_*\left[||\Lambda_n^*||^3\right]\right\}, \end{aligned}$$

where G' and G'' denote the gradient and matrix of second derivatives of G, respectively, and $C_n(G)$ is a positive and finite constant.

Using (B.1), we have

$$\mathrm{E}_*\left[G'(U_{n-1})\left(\Lambda_n - \Lambda_n^*\right)\right] = 0, \tag{B.31}$$

$$\mathrm{E}_*\left[\Lambda_n^\tau G''(U_{n-1})\Lambda_n - \Lambda_n^{*\tau} G''(U_{n-1})\Lambda_n^*\right] = 0.$$

This therefore implies that

$$\begin{aligned} &\left|\mathrm{E}_*\left[G(B_n(\xi_1,\ldots,\xi_n))\right] - \mathrm{E}_*\left[G(B_n(\xi_1,\ldots,\xi_{n-1},\xi_n^*))\right]\right| \\ &\quad \le C_0 \frac{C_n(G)}{6}\left(\mathrm{E}_*\left[||\Lambda_n||^3\right] + \mathrm{E}_*\left[||\Lambda_n^*||^3\right]\right) \end{aligned} \tag{B.32}$$

for some constant $0 < C_0 < \infty$.

To complete the proof, we need to introduce the notation

$$V_{1n}(h) = 2\sum_{s=1}^{n-1} \tilde{a}_{sn}\epsilon_s\epsilon_n, \qquad V_{2n}(h) = -2\sum_{s=1}^{n-1} \tilde{b}_{sn}\eta_s\epsilon_n, \tag{B.33}$$

$$V_{3n}(h) = -2\sum_{s=1}^{n-1} \tilde{b}_{sn}\epsilon_s\eta_n, \qquad V_{4n}(h) = 2\sum_{s=1}^{n-1} \tilde{c}_{sn}\eta_s\eta_n,$$

where $\tilde{a}_{st} = (2h_1^{d/2})^{-1}a_{st}$, $\tilde{b}_{st} = (2h_1^{d/2})^{-1}b_{st}$, and $\tilde{c}_{st} = (2h_1^{d/2})^{-1}c_{st}$. Let $\{V_{in}\}$ be the respective sequence that is obtained by stacking $\{V_{in}(h)\}$ over $h \in \mathcal{H}_n$ for $1 \le i \le 4$. It is obvious that for $i = 1, 2$

$$||\Lambda_n||^3 \le C_3 \sum_{j=1}^{4} ||V_{jn}||^3,$$

where C_3 is a constant.

To estimate the upper bound of (B.32), we now calculate only the bound for V_{4n}. The others follow similarly. Observe that

$$c_{st} = nh_1^d \int W_{s2}(x)W_{t2}(x)\upsilon_{22}(x)\pi(x)dx \tag{B.34}$$
$$= \frac{h_1^d}{nh_2^{2d}} \int K\left(\frac{x-X_s}{h_2}\right) K\left(\frac{x-X_t}{h_2}\right) p(x)dx$$
$$= \frac{h_1^d}{nh_2^d} \int K(u)K\left(u+\frac{X_s-X_t}{h_2}\right) p(X_s+uh_2)du$$
$$= \frac{h_1^d}{nh_2^d} L_2\left(\frac{X_s-X_t}{h_2}, X_s\right) = \frac{\beta_n^d}{n} L_2\left(\frac{X_s-X_t}{h_2}, X_s\right),$$

where $p(x) = \upsilon_{22}(x)\pi(x)$ and $L_2(x,y) = \int K(u)K(u+x)p(y+uh_2)du$.

Using (B.34), we have as $n \to \infty$

$$\sum_{h\in\mathcal{H}_n}\sum_{k\in\mathcal{H}_n} \mathrm{E}_*\left[\sum_{s=1}^{n-1}\sum_{t=1,\neq s}^{n-1} \tilde{c}_{sn}^2(h)\tilde{c}_{tn}^2(k)\eta_s^2\eta_t^2\eta_n^4\right] \tag{B.35}$$
$$\leq \sum_{h\in\mathcal{H}_n}\sum_{k\in\mathcal{H}_n} \frac{\beta_n^{2d}\gamma_n^{2d}}{n^2h_1^dk_1^d}\mathrm{E}_*\left[L_2^2\left(\frac{X_s-X_n}{h_2}, X_s\right) L_2^2\left(\frac{X_t-X_n}{k_2}, X_t\right)\eta_s^2\eta_t^2\eta_n^4\right]$$
$$\leq C\sum_{h\in\mathcal{H}_n}\sum_{k\in\mathcal{H}_n} \frac{\beta_n^{2d}\gamma_n^{2d}}{n^2h_1^dk_1^d}\left(\mathrm{E}_*\left[\left|\eta_s^2\eta_t^2\eta_n^4\right|^{1+\delta_\zeta}\right]\right)^{1/1+\delta_\zeta}$$
$$\leq C\cdot\left(\frac{J_{12n}}{nh_{1\min}}\right)^2 \beta_n^d\gamma_n^d \leq C\cdot\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^2 \beta_0^{2d}(1+o(1))$$

using Assumptions B.1 and B.2, where $k=(k_1,k_2)$ with $k_i \in \mathcal{H}_{in}$, $\gamma_n = \frac{k_1}{k_2}$, $f(x,y,z,u,v,w)$ is the joint density function of $(X_s, X_t, X_n, \eta_s, \eta_t, \eta_n)$, and $0 < C < \infty$ is a constant.

Similarly to (B.15), we can show that as $n \to \infty$

$$\sum_{h,k\in\mathcal{H}_n} D_{hk}\mathrm{E}_*\left[\sum_{1\leq s\neq t\leq n-1} c_{sn}^2(h)c_{sn}(k)c_{tn}(k)\eta_s^3\eta_t\eta_n^4\right] = o\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^2, \tag{B.36}$$

$$\sum_{h,k\in\mathcal{H}_n} D_{hk}\mathrm{E}_*\left(\sum_{1\leq s\neq t, s\neq u, t\neq u\leq n-1} c_{sn}^2(h)c_{tn}(k)c_{un}(k)\eta_s^2\eta_t\eta_u\eta_n^4\right) = o\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^2,$$

$$\sum_{h,k\in\mathcal{H}_n} D_{hk}\mathrm{E}_*\left(\sum_{\substack{1\leq s\neq t, s\neq u, s\neq v,\\ t\neq u, t\neq v, u\neq v\leq n-1}} c_{sn}(h)c_{tn}(h)c_{un}(k)c_{vn}(k)\eta_s\eta_t\eta_u\eta_v\eta_n^4\right)$$
$$= o\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^2,$$

using the fact that for every given x,

$$\mathrm{E}\left[\sigma^2(X_t)L_2\left(\frac{X_t-x}{h_2},X_t\right)(e_t^2-1)\right]=\mathrm{E}\left[\sigma^2(X_t)L_2\left(\frac{X_t-x}{h_2},X_t\right)\mathrm{E}\left[(e_t^2-1)|\Omega_{t-1}\right]\right]$$
$$=0$$

implied from Assumption B.1, where $D_{hk}=\frac{1}{h_1^d k_1^d}$. The detailed verification of (B.36) is similar to that of Theorem A.1 of Gao (2007) using their Lemmas A.1 and A.2.

Equations (B.35) and (B.36) imply that as $n\to\infty$

$$\sum_{h\in\mathcal{H}_n}\sum_{k\in\mathcal{H}_n}\mathrm{E}_*\left[\sum_{s,t,u,v=1}^{n-1}\tilde{c}_{sn}(h)\eta_s\tilde{c}_{tn}(h)\eta_t\tilde{c}_{un}(k)\eta_u\tilde{c}_{vn}(k)\eta_v\eta_n^4\right] \tag{B.37}$$
$$=\sum_{h,k\in\mathcal{H}_n}(h_1k_1)^{-d}\mathrm{E}_*\left(\sum_{s=1}^{n-1}\left[c_{sn}^2(h)c_{sn}^2(k)\eta_s^4\eta_n^4\right]\right)$$
$$+\sum_{h,k\in\mathcal{H}_n}(h_1k_1)^{-d}\mathrm{E}_*\left(\sum_{1\le s\ne t\le n-1}\left[c_{sn}^2(h)c_{tn}^2(k)+2c_{sn}(h)c_{sn}(k)c_{tn}(h)c_{tn}(k)\right]\times\eta_s^2\eta_t^2\eta_n^4\right)$$
$$+4\sum_{h,k\in\mathcal{H}_n}(h_1k_1)^{-d}\mathrm{E}_*\left(\sum_{1\le s\ne t\le n-1}c_{sn}^2(h)c_{sn}(k)c_{tn}(k)\eta_s^3\eta_t\eta_n^4\right)$$
$$+4\sum_{h,k\in\mathcal{H}_n}(h_1k_1)^{-d}\mathrm{E}_*\left(\sum_{1\le s\ne t,s\ne u,t\ne u\le n-1}c_{sn}^2(h)c_{tn}(k)c_{un}(k)\eta_s^2\eta_t\eta_u\eta_n^4\right)$$
$$+\sum_{h,k\in\mathcal{H}_n}(h_1k_1)^{-d}\mathrm{E}_*\left(\sum_{1\le s\ne t,s\ne u,s\ne v,t\ne u,t\ne v,u\ne v\le n-1}\times c_{sn}(h)c_{tn}(h)c_{un}(k)c_{vn}(k)\eta_s\eta_t\eta_u\eta_v\eta_n^4\right)$$
$$\le C\cdot\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^2,$$

where $C>0$ is a constant independent of n.

Let $\tilde{C}_{sn}$ be the vector that is obtained by stacking $\tilde{c}_{sn}(h)$ over $h \in \mathcal{H}_n$. Then, equation (B.37) implies that as $n \to \infty$

$$\begin{aligned} E_*\left[\|V_{4n}\|^3\right] &= 8E_*\left[\left\|\sum_{s=1}^{n-1} \tilde{C}_{sn}\eta_s\eta_n\right\|^3\right] \le 8\left\{E_*\left[\left\|\sum_{s=1}^{n-1} \tilde{C}_{sn}\eta_s\eta_n\right\|^4\right]\right\}^{3/4} \\ &= 8\left\{\sum_{h\in\mathcal{H}_n}\sum_{k\in\mathcal{H}_n} E_*\left[\sum_{s,t,u,v=1}^{n-1} \tilde{c}_{sn}(h)\eta_s\tilde{c}_{tn}(h)\eta_t\tilde{c}_{un}(k)\eta_u\tilde{c}_{vn}(k)\eta_v\eta_n^4\right]\right\}^{3/4} \\ &\le C\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^{3/2}. \end{aligned} \qquad \textbf{(B.38)}$$

Thus, we can show that

$$E_*\left[\|\Lambda_n\|^3\right] \le C\sum_{j=1}^{4} E_*\left[\|V_{jn}\|^3\right] \le C\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^{3/2}. \qquad \textbf{(B.39)}$$

A similar result holds for $E_*\left[\|\Lambda_n^*\|^3\right]$. Thus

$$E_*\left[\|\Lambda_n\|^3\right] + E_*\left[\|\Lambda_n^*\|^3\right] \le 2C\left(\frac{J_{12n}}{nh_{1\min}^d}\right)^{3/2}. \qquad \textbf{(B.40)}$$

Equations (B.30), (B.32), and (B.40) therefore imply (B.29).

As demonstrated in Horowitz and Spokoiny (2001),

$$\lim_{n\to\infty}\left\{P_*\left[\max_{h\in\mathcal{H}_n} B_{hn}(\xi_1,\ldots,\xi_n) \le x\right] - P_*\left[\max_{h\in\mathcal{H}_n} B_{hn}(\xi_1^*,\ldots,\xi_n^*) \le x\right]\right\} = 0$$

for any real x is equivalent to

$$\lim_{n\to\infty}\left|E_*\left(\prod_{h\in\mathcal{H}_n} I\left[B_{hn}(\xi_1,\ldots,\xi_n) \le x\right]\right) - E_*\left(\prod_{h\in\mathcal{H}_n} I\left[B_{hn}(\xi_1^*,\ldots,\xi_n^*) \le x\right]\right)\right| = 0.$$

Following the lines of Horowitz and Spokoiny (2001) and utilizing the above established bound in (B.40), it can then be shown that as $n \to \infty$

$$\left|P_*\left[\max_{h\in\mathcal{H}_n} B_{hn}(\xi_1,\ldots,\xi_n) \le x\right] - P_*\left[\max_{h\in\mathcal{H}_n} B_{hn}(\xi_1^*,\ldots,\xi_n^*) \le x\right]\right| \to 0. \qquad \textbf{(B.41)}$$

This finally completes the proof of Lemma B.7. ■

It follows from Lemma B.7 that $\max_{h\in\mathcal{H}_n} L_{0n}^*(h)$ and $\max_{h\in\mathcal{H}_n} L_{0n}(h)$ have identical asymptotic distributions. This result is used in the proof of Lemma B.10, below.

LEMMA B.8. *Let Assumptions B.1–B.3 hold. Then for any $x \ge 0$, $h \in \mathcal{H}_n$, and sufficiently large n*

$$P_*\left(L_{0n}^*(h) > x\right) \le \exp\left(-\frac{x^2}{4}\right).$$

Proof. Since the distribution of e_t^* is not necessarily normal, the argument used in the proof of Lemma 11 of Horowitz and Spokoiny (2001) may not be true. We therefore adopt a different proof. It follows from the beginning of the proof of Theorem 2.1 that for any small $\delta > 0$ there exists a large integer $n_0 \geq 1$ such that, for $n \geq n_0$ and any $x \geq 0$,

$$\left| P_*(L_{0n}^*(h) \leq x) - \Phi(x) \right| < \delta,$$

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

This implies that, for any $n \geq n_0$ and $x \geq 0$,

$$\begin{aligned} P_*(L_{0n}^*(h) > x) &\leq 1 - \Phi(x) + \delta \\ &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/2} du + \delta = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/4} e^{-u^2/4} du + \delta \\ &\leq e^{-x^2/4} \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-u^2/4} du + \delta \leq e^{-x^2/4} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u^2/4} du + \delta \\ &= e^{-x^2/4} \frac{\sqrt{2}}{\sqrt{2\pi}} \int_0^{\infty} e^{-v^2/2} dv + \delta = \frac{\sqrt{2}}{2} e^{-x^2/4} + \delta \end{aligned}$$

using $\frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-v^2/2} dv = \frac{1}{2}$. The proof follows by letting $0 < \delta \leq \left(1 - \frac{\sqrt{2}}{2}\right) e^{-x^2/4}$ for any $x \geq 0$.

Before we present the next lemma, let us define, for $0 < \alpha < 1$, $l_{n\alpha}^{0*}$ to be the $1 - \alpha$ quantile of $\max_{h \in \mathcal{H}_n} L_{0n}^*(h)$. ■

LEMMA B.9. *Let Assumptions B.1–B.3 hold. Then for large enough n*

$$l_{n\alpha}^{0*} \leq 2\sqrt{\log(J_{12n}) - \log(\alpha)}.$$

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

LEMMA B.10. *Let Assumptions B.1–B.3 hold. In addition,*

$$\lim_{n\to\infty} P_* \left(\frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^* \right) = 1 \tag{B.42}$$

for some $h \in \mathcal{H}_n$, where $\tilde{l}_{n\alpha}^* = \max\left(l_{n\alpha}^{0*}, \sqrt{2\log(J_{12n}) + \sqrt{2\log(J_{12n})}}\right)$. Then the adaptive test statistic L_n defined in (3.2) satisfies

$$\lim_{n\to\infty} P(L_n > l_{n\alpha}) = 1.$$

Proof. By Lemma B.6, L_n can be replaced with $\max_{h \in \mathcal{H}_n} L_{2n}(h)$. By Lemma B.7, $l_{n\alpha}$ therefore can be replaced by $l_{n\alpha}^{0*}$. Thus, we need only to show that

$$\lim_{n\to\infty} P_*(\max_{h \in \mathcal{H}_n} L_{2n}(h) > l_{n\alpha}^{0*}) = 1,$$

which holds if $\lim_{n\to\infty} P_*(L_{2n}(h) > l_{n\alpha}^{0*}) = 1$ for some $h \in \mathcal{H}_n$. For any $h \in \mathcal{H}_n$, using the conclusion from Lemma B.7 that $\max_{h\in\mathcal{H}_n} L_{0n}^*(h)$ and $\max_{h\in\mathcal{H}_n} L_{0n}(h)$ have identical asymptotic distributions, we have

$$L_{2n}(h) = L_{0n}(h) + \frac{Q_n(\tilde{\theta}_1) + \Pi_n(\tilde{\theta}_1)}{h_1^{d/2}} = L_{0n}^*(h) + \frac{Q_n(\tilde{\theta}_1) + \Pi_n(\tilde{\theta}_1)}{h_1^{d/2}} + o_p(1). \quad \textbf{(B.43)}$$

Applying Lemma B.5 with $D_n = \tilde{l}_{n\alpha}^*$,

$$\frac{\Pi_n(\tilde{\theta}_1)}{h_1^{d/2}} = o_p(\tilde{l}_{n\alpha}^*). \quad \textbf{(B.44)}$$

On the other hand, condition (B.42) implies that as $n \to \infty$

$$P_*\left(\frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} < 2\tilde{l}_{n\alpha}^*\right) \to 0. \quad \textbf{(B.45)}$$

Observe that

$$\begin{aligned} P_*(L_{2n}(h) > l_{n\alpha}^{0*}) &= P_*\left(L_{2n}(h) > l_{n\alpha}^{0*}, \frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^*\right) \\ &\quad + P_*\left(L_{2n}(h) > l_{n\alpha}^{0*}, \frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} < 2\tilde{l}_{n\alpha}^*\right) \\ &\equiv I_{1n} + I_{2n}. \end{aligned}$$

It follows from (B.42)–(B.45) that as $n \to \infty$

$$\begin{aligned} I_{1n} &= P_*\left(L_{0n}^*(h) + \frac{Q_n(\tilde{\theta}_1) + \Pi_n(\tilde{\theta}_1)}{h_1^{d/2}} > l_{n\alpha}^{0*} \mid \frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^*\right) P_*\left(\frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^*\right) \\ &\geq P_*\left(L_{0n}^*(h) > l_{n\alpha}^{0*} - 4\tilde{l}_{n\alpha}^* \mid \frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^*\right) P_*\left(\frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} \geq 2\tilde{l}_{n\alpha}^*\right) \to 1 \end{aligned}$$

because $L_{0n}^*(h)$ is asymptotically normal and $l_{n\alpha}^{0*} - 4\tilde{l}_{n\alpha}^* \to -\infty$ as $n \to \infty$.

Because of (B.45), as $n \to \infty$

$$I_{2n} \leq P_*\left(\frac{Q_n(\tilde{\theta}_1)}{2h_1^{d/2}} < 2\tilde{l}_{n\alpha}^*\right) \to 0.$$

This finishes the proof. ■

B.3. Proofs of Theorems 3.1–3.4

Proof of Theorem 3.1. It is directly implied by Lemma B.7, as $L_n = \max_{h\in\mathcal{H}_n} L_{1n}(h) + o_p(1)$ and $L_n^* = \max_{h\in\mathcal{H}_n} L_{1n}^*(h) + o_p(1)$.

In order to prove Theorems 3.2–3.3, in view of Lemma B.10, it suffices to verify (B.42). Using Lemma B.1(ii), it suffices to verify

$$\lim_{n\to\infty} P_*\left(h_1^d \lambda(\theta)^\tau \lambda(\theta) \geq 4\tilde{l}_{n\alpha}^* h_1^{d/2}\right) = 1. \quad \textbf{(B.46)}$$

■

Proof of Theorem 3.2. In view of the definition of $\tilde{l}^*_{n\alpha}$, equation (B.46) follows from the fact that as $n \to \infty$,

$$\frac{1}{2n}\lambda(\theta)^\tau \lambda(\theta) - \rho_1^2(m_1, \mathcal{M}_{1\Theta}) - \rho_2^2(m_2, \mathcal{M}_{2\Theta}) \to 0 \tag{B.47}$$

holds in probability, and $nh_1^d \geq C_0 \tilde{l}^*_{n\alpha} h_1^{d/2}$ for some constant $0 < C_0 < \infty$ and n large enough. ■

Proof of Theorem 3.3. Using the definition of $\tilde{l}^*_{n\alpha}$, (B.47),

$$\frac{1}{2n}\sum_{t=1}^n \Delta_1^2(X_t) + \frac{1}{2n}\sum_{t=1}^n \Delta_2^2(X_t) \to \frac{1}{2}\left(\mathrm{E}\left[\Delta_1^2(X_1)\right] + \mathrm{E}\left[\Delta_2^2(X_1)\right]\right) \tag{B.48}$$

as $n \to \infty$, and

$$\frac{1}{2n}\lambda(\theta)^\tau \lambda(\theta) = \frac{C_{1n}^2}{2n}\sum_{t=1}^n \Delta_1^2(X_t) + \frac{C_{2n}^2}{2n}\sum_{t=1}^n \Delta_2^2(X_t)$$
$$\geq C_1 C_{1n}^2 + C_2 C_{2n}^2 \tag{B.49}$$

holds in probability, one can see that (B.46) holds when $h_1 = h_{1\max} = (\log\log(n))^{-1/d}$. This finishes the proof of Theorem 3.3. ■

Proof of Theorem 3.4. In order to verify (B.42), we need to introduce the notation:

$$h_1^* = \left(n^{-1}\tilde{l}^*_{n\alpha}\right)^{2/4s_1+d}.$$

This implies $n\left(h_1^*\right)^{4s_1+d/2} = \tilde{l}^*_{n\alpha}$. Choose $h_1 \in \mathcal{H}_{1n}$ such that $h_1^* \leq h_1 < 2h_1^*$. We then have

$$4h_1^{d/2}\tilde{l}^*_{n\alpha} = 4nh_1^{d/2}\left(h_1^*\right)^{4s_1+d/2} \leq 4nh_1^{4s_1+d/2+d/2} = 4nh_1^{2s_1+d}. \tag{B.50}$$

Thus, in order to verify (B.42), it suffices to show that

$$Q_n(\tilde{\theta}_1) \geq 4nh_1^{2s_1+d} \tag{B.51}$$

holds in probability for the selected $h_1 \in \mathcal{H}_{1n}$ and $\theta_1 \in \Theta$.

The verification of (B.51) can be done using similar techniques as in the proof of Lemma B.1(ii). Alternatively, one may follow the proof of (A13) of Horowitz and Spokoiny (2001) by noting that all their derivations below (A13) hold with probability one and therefore in probability when all X_i are random variables. For this case, one needs to notice that the additional factor h_1^d is involved in (B.51) because the form of $Q_n(\tilde{\theta}_1)$ involves an integral of $K_h(\cdot)$. This can be seen from the proof of (B.7). We finally finish the outline of the proof of Theorem 3.4. ■