A Goodness-of-fit Test for Parametric and Semiparametric Models in Multiresponse Regression

Song Xi Chen[1] Ingrid Van Keilegom[2]

Iowa State University *Université catholique de Louvain*

August 23, 2006

Abstract

We propose an empirical likelihood test that is able to test the goodness-of-fit of a class of parametric and semiparametric multiresponse regression models. The class includes as special cases fully parametric models, semiparametric models, like the multi-index and the partially linear models, and models with shape constraints. Another feature of the test is that it allows both the response variable and the covariate be multivariate, which means that multiple regression curves can be tested simultaneously. The test also allows the presence of infinite dimensional nuisance functions in the model to be tested. It is shown that the empirical likelihood test statistic is asymptotically normally distributed under certain mild conditions and permits a wild bootstrap calibration. Despite that the class of models which can be considered is very large, the empirical likelihood test enjoys good power properties against departures from a hypothesized model within the class.

Key Words: Additive regression, bootstrap, empirical likelihood, goodness-of-fit, infinite dimensional parameter, kernel estimation, monotone regression, partially linear regression.

[1] Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, E-mail: songchen@iastate.edu.

[2] Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, E-mail : vankeilegom@stat.ucl.ac.be.

Both authors are grateful for financial support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy) and a National Science Foundation Grant (SES-0518904).

1 Introduction

Suppose $\{(X_i, Y_i)\}_{i=1}^n$ is an independent and identically distributed random vector, where Y_i is a k-variate response and X_i a d-variate covariate. Let $m(x) = E(Y_i|X_i = x) = (m_1(x), \ldots, m_k(x))$ be the conditional mean consisting of k regression curves on R^d and $\Sigma(x) = \mathrm{Var}(Y_i|X_i = x)$ be a $k \times k$ matrix whose values change along with the covariate.

Let $m(\cdot) = m(\cdot, \theta, g) = (m_1(\cdot, \theta, g), \ldots, m_k(\cdot, \theta, g))$ be a working regression model of which one would like to check its validity. The form of m is known up to a finite dimensional parameter θ and an infinite dimensional nuisance parameter g. The model $m(\cdot, \theta, g)$ includes a wide range of parametric and semiparametric regression models as special cases. In the absence of g, the model degenerates to a fully parametric model $m(\cdot) = m(\cdot, \theta)$, whereas the presence of g covers a range of semiparametric models including the single or multi-index models and partially linear single-index models. The class also includes models with qualitative constraints, like additive models and models with shape constraints. The variable selection problem, the comparison of regression curves and models for the variance function can be covered by the class of $m(\cdot, \theta, g)$ as well.

Multiresponse regression is frequently encountered in applications. In compartment analysis arising in biological and medical studies as well as chemical kinetics (Atkinson and Bogacka, 2002), a multivariate variable is described by a system of differential equations whose solutions satisfy multiresponse regression (Jacquez, 1996). In response surface designs, multivariate random vectors are collected as responses of some controlled variables (covariates) of certain statistical experiments. Khuri (2001) proposed using the generalized linear models for modeling such kind of data and Uciński and Bogacka (2005) studied the issue of optimal designs with an objective for discrimination between two multiresponse system models. The monographs by Bates and Watts (1988, chapter 4) and Seber and Wild (1989, chapter 11) contain more examples of multiresponse regression as well as their parametric inference.

The need for testing multiple curves occurs even in the context of univariate responses Y_i. Consider the following heteroscedastic regression model

$$Y_i = r(X_i) + \sigma(X_i)e_i,$$

where the e_i's are unit residuals such that $E(e_i|X_i) = 0$ and $E(e_i^2|X_i) = 1$, and $r(\cdot)$ and

$\sigma^2(\cdot)$ are respectively the conditional mean and variance functions. Suppose $r(x, \theta, g)$ and $\sigma^2(x, \theta, g)$ are certain working parametric or semiparametric models. In this case, the bivariate response vector is $(Y_i, Y_i^2)^T$ and the bivariate model specification $m(x, \theta, g) = (r(x, \theta, g), \sigma^2(x, \theta, g) + r^2(x, \theta, g))^T$.

The aim of the paper is to develop a nonparametric goodness-of-fit test for the hypothesis

$$H_0 : m(\cdot) = m(\cdot, \theta, g), \tag{1.1}$$

for some known k-variate function $m(\cdot, \theta, g)$, some finite dimensional parameter $\theta \in \Theta \subset R^p$ $(p \geq 1)$ and some function $g \in \mathcal{G}$ which is a complete metric space consisting of functions from R^d to R^q $(q \geq 1)$. We will use two pieces of nonparametric statistical hardware: the kernel regression estimation technique and the empirical likelihood technique to formulate a test for H_0.

In the case of a single regression curve (i.e. $k = 1$), the nonparametric kernel approach has been widely used to construct goodness-of-fit tests for the conditional mean or variance function. Eubank and Spiegelman (1990), Eubank and Hart (1992), Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995), Fan and Li (1996), Hart (1997), Hjellvik, Yao and Tjøstheim (1998) develop consistent tests for a parametric specification by employing the kernel smoothing method based on a fixed bandwidth. Horowitz and Spokoiny (2001) propose a test based on a set of smoothing bandwidths in the construction of the kernel estimator. Its extensions are considered in Chen and Gao (2006) for time series regression models and in Rodríguez-Póo, Sperlich and Vieu (2005) for semiparametric regression models. Other related references can be found in the books by Hart (1997) and Fan and Yao (2003).

The empirical likelihood (EL) (Owen, 1988, 1990) is a technique that allows the construction of a nonparametric likelihood for a parameter of interest in a nonparametric or semiparametric setting. Despite that it is intrinsically nonparametric, it possesses two important properties of a parametric likelihood: the Wilks' theorem and the Bartlett correction. Qin and Lawless (1994) establish EL for parameters defined by estimating equations, which is the widest framework for EL formulation. Chen and Cui (2006) show that the EL admits a Bartlett correction under this general framework. Hjort, McKeague and Van Keilegom (2005) consider the properties of the EL in the presence of both finite and infinite dimensional nuisance parameters as well as when the data dimension is high. See Owen (2001) for a comprehensive overview of the EL method and references therein.

Goodness-of-fit tests based on the EL have been proposed in the literature, which include Li (2003) and Li and Van Keilegom (2002) for survival data, Einmahl and McKeague (2003) for testing some characteristics of a distribution function, Chen, Härdle and Li (2003) for conditional mean functions with dependent data. 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). Tripathi and Kitamura (2003) propose an EL test for conditional moment restrictions.

One contribution of the present paper is the formulation of a test that is able to test a set of multiple regression functions simultaneously. Multiple regression curves exist when the response Y_i is genuinely multivariate, or when Y_i is in fact univariate but we are interested in testing the validity of a set of feature curves, for example the conditional mean and conditional variance, at the same time. EL is a natural device to formulate goodness-of-fit statistics to test multiple regression curves. This is due to EL's built-in feature to standardize a goodness-of-fit distance measure between a fully nonparametric estimate of the target functional curves and its hypothesized counterparts. The standardization carried out by the EL uses implicitly the true covariance matrix function, say $V(x)$ of the kernel estimator $m(\cdot)$, to studentize the distance between $\hat{m}(\cdot)$ and the hypothesized model $m(\cdot,\theta,g)$, so that the goodness-of-fit statistic is an integrated Mahalanobis distance between the two sets of multivariate curves $\hat{m}(\cdot)$ and $m(\cdot,\theta,g)$. This is attractive as we avoid estimating $V(x)$, which can be a daunting task when k is larger than 1. When testing multiple regression curves, there is an intrinsic issue regarding how much each component-wise goodness-of-fit measure contributes to the final test statistic. The EL distributes the weights naturally according to $V^{-1}(x)$. And most attractively, this is done without requiring extra steps of estimation since it comes as a by-product of the internal algorithm.

Another contribution of the proposed test is its ability to test a large class of regression models in the presence of both finite and infinite dimensional parameters. The class includes as special cases fully parametric models, semiparametric models, like the multi-index and the partially linear models, and models with shape constraints, like monotone regression models. It is shown that the EL test statistic is asymptotically normally distributed under certain mild conditions and permits a wild bootstrap calibration. Despite the fact that the class of models which can be considered by the proposed test is very large, the test enjoys good power properties against departures from a hypothesized model within the class.

The paper is organized as follows. In the next section we introduce some notations and formulate the EL test statistic. Section 3 is concerned with the main asymptotic results, namely the asymptotic distribution of the test statistic both under the null hypothesis and under a local alternative, and the consistency of the bootstrap approximation. In Section 4 we focus on a number of particular models and apply the general results of Section 3 to these models. Simulation results are reported in Section 5. We conclude the paper by giving in Section 6 the assumptions and the proofs of the main results.

2 The Test Statistic

Let $Y_i = (Y_{i1}, \ldots, Y_{ik})^T$ and $m(x) = (m_1(x), \ldots, m_k(x))^T$ where $m_l(x) = E(Y_{il}|X_i = x)$ is the l-th regression curve for $l = 1, \ldots, k$. Let $\epsilon_i = Y_i - m(X_i)$ be the i-th residual vector. Define $\sigma_{lj}(x) = \mathrm{Cov}(\epsilon_{il}, \epsilon_{ij}|X_i = x)$ which is the conditional covariance between the l-th and j-th component of the residual vector. Then, the conditional covariance matrix $\Sigma(x) = \mathrm{Var}(Y_i|X_i = x) = (\sigma_{lj}(x))_{k\times k}$.

Let K be a d-dimensional kernel with a compact support on $[-1, 1]^d$. Without loss of generality, K is assumed to be a product kernel based on a univariate kernel k, i.e. $K(t_1, \cdots, t_d) = \prod_{i=1}^d k(t_i)$ where k is a r-th order kernel supported on $[-1, 1]$ and

$$\int k(u)du = 1, \int u^l k(u)du = 0 \text{ for } l = 1, \cdots, r-1 \text{ and } \int u^r k(u)du = k_r \neq 0$$

for an integer $r \geq 2$. Define $K_h(u) = h^{-d}K(u/h)$. The Nadaraya-Watson (NW) kernel estimator of $m_l(x)$, $l = 1, \ldots, k$, is

$$\hat{m}_l(x) = \frac{\sum_{i=1}^n K_{h_l}(x - X_i)Y_{il}}{\sum_{t=1}^n K_{h_l}(x - X_t)},$$

where h_l is the smoothing bandwidth for curve l. Different bandwidths are allowed to smooth different curves which is sensible for multivariate responses. Then

$$\hat{m}(x) = (\hat{m}_1(x), \ldots, \hat{m}_k(x))^T$$

is the kernel estimator of the multiple regression curves. We assume throughout the paper that $h_l/h \to \beta_l$ as $n \to \infty$, where h represents a baseline level of the smoothing bandwidth and $c_0 \le \min_l\{\beta_l\} \le \max_l\{\beta_l\} \le c_1$ for finite and positive constants c_0 and c_1 free of n.

Under the null hypothesis (1.1),

$$Y_i = m(X_i, \theta_0, g_0) + \epsilon_i, \quad i = 1, \dots, n, \tag{2.1}$$

where θ_0 is the true value of θ in Θ, g_0 is the true function in $\mathcal{G}$, and $\epsilon_1, \dots, \epsilon_n$ are independent and identically distributed, so that $E(\epsilon_i|X_i = x) = 0$ and $\mathrm{Var}(\epsilon_i|X_i = x) = \Sigma(x)$.

Let $\hat\theta$ be a $\sqrt{n}$-consistent estimator of θ_0 and $\hat g$ be a consistent estimator of g_0 under a norm $\|\cdot\|_{\mathcal{G}}$ defined on the complete metric space $\mathcal{G}$. We suppose $\hat g$ is a kernel estimator based on a kernel L of order $s \ge 2$ and a bandwidth sequence b, most likely different from the bandwidth h used to estimate m. We will require that $\hat g$ converges to g_0 faster than $(nh^d)^{-1/2}$, the optimal rate in a completely d-dimensional nonparametric model. As demonstrated in Section 4, this can be easily satisfied since g is of lower dimensional than the saturated nonparametric model for m.

Each $m_l(x, \hat\theta, \hat g)$ is smoothed by the same kernel K and bandwidth h_l as in the kernel estimator $\hat m_l(x)$, in order to prevent the bias of the kernel regression estimators entering the asymptotic distribution of the test statistic (see also Härdle and Mammen, 1993):

$$\tilde m_l(x, \hat\theta, \hat g) = \frac{\sum_{i=1}^n K_{h_l}(x - X_i) m_l(X_i, \hat\theta, \hat g)}{\sum_{t=1}^n K_{h_l}(x - X_t)}.$$

for $l = 1, \dots, k$. Let $\tilde m(x, \hat\theta, \hat g) = (\tilde m_1(x, \hat\theta, \hat g), \dots, \tilde m_k(x, \hat\theta, \hat g))^T$.

We note in passing that the dimension of the response Y_i does not contribute to the curse of dimensionality. Rather, it is the dimension of the covariate X_i that contributes, since X_i is the direct target of smoothing. Hence, as far as the curse of dimensionality is concerned, testing multiple curves is the same as testing a single regression curve.

To formulate the empirical likelihood ratio test statistics, we first consider a fixed $x \in R^d$ and then globalize by integrating the local likelihood ratio over a compact set $S \subset R^d$ in the support of X. For each fixed $x \in S$, let

$$\hat Q_i(x, \hat\theta) = \Big(K_{h_1}(x - X_i)\left(Y_{i1} - \tilde m_1(x, \hat\theta, \hat g)\right), \dots K_{h_k}(x - X_i)\left(Y_{ik} - \tilde m_k(x, \hat\theta, \hat g)\right)\Big)^T \tag{2.2}$$

which is a vector of local residuals at x and its mean is approximately zero.

Let $\{p_i(x)\}_{i=1}^n$ be nonnegative real numbers representing empirical likelihood weights allocated to $\{(X_i, Y_i)\}_{i=1}^n$. The minus 2 log empirical likelihood ratio for the multiple conditional mean evaluated at $\tilde{m}(x, \hat{\theta}, \hat{g})$ is

$$\ell\{\tilde{m}(x, \hat{\theta}, \hat{g})\} = -2 \sum_{i=1}^{n} \log\{np_i(x)\}$$

subject to $p_i(x) \geq 0$, $\sum_{i=1}^n p_i(x) = 1$ and $\sum_{i=1}^n p_i(x)\hat{Q}_i(x, \hat{\theta}) = 0$. By introducing a vector of Lagrange multipliers $\lambda(x) \in R^k$, a standard empirical likelihood derivation (Owen, 1990) shows that the optimal weights are given by

$$p_i(x) = \frac{1}{n}\{1 + \lambda^T(x)\hat{Q}_i(x, \hat{\theta})\}^{-1}, \tag{2.3}$$

where $\lambda(x)$ solves

$$\sum_{i=1}^{n} \frac{\hat{Q}_i(x, \hat{\theta})}{1 + \lambda^T(x)\hat{Q}_i(x, \hat{\theta})} = 0. \tag{2.4}$$

Integrating $\ell\{\tilde{m}(x, \hat{\theta}, \hat{g})\}$ against a weight function π supported on S, gives

$$\Lambda_n(\vec{h}) = \int \ell\{\tilde{m}(x, \hat{\theta}, \hat{g})\}\pi(x)dx,$$

which is our EL test statistic based on the bandwidth vector $\vec{h} = (h_1, \ldots, h_k)^T$.

Let $\bar{\hat{Q}}(x, \hat{\theta}) = n^{-1} \sum_{i=1}^n \hat{Q}_i(x, \hat{\theta})$, $R(t) = \int K(u)K(tu)du$ and

$$V(x) = f(x) \left(\beta_j^{-d} R(\beta_l/\beta_j)\sigma_{lj}(x)\right)_{k \times k},$$

where $f(x)$ is the density of X. We note in particular that $R(1) = R(K) =: \int K^2(u)du$ and that $\beta_j^{-d} R(\beta_l/\beta_j) = \beta_l^{-d} R(\beta_j/\beta_l)$ indicating that $V(x)$ is a symmetric matrix.

Derivations given in Section 6 show that

$$\Lambda_n(\vec{h}) = nh^d \int \bar{\hat{Q}}^T(x, \theta_0)V^{-1}(x)\bar{\hat{Q}}(x, \theta_0)\pi(x)\, dx + o_p(h^{d/2}),$$

where $h^{d/2}$ is the stochastic order of the first term on the right hand side if $d < 4r$. Here r is the order of the kernel K. Since $\bar{\hat{Q}}(x, \theta_0) = f(x)\{\hat{m}(x) - \tilde{m}(x, \theta_0, \hat{g})\}\{1 + o_p(1)\}$, $\bar{\hat{Q}}(x, \theta_0)$ serves as a raw discrepancy measure between $\hat{m}(x) = (\hat{m}_1(x), \ldots, \hat{m}_k(x))$ and the hypothesized model $m(x, \theta_0, \hat{g})$. There is a key issue on how much each $\hat{m}_l(x) - \tilde{m}_l(x, \theta_0, \hat{g})$

contributes to the final statistic. The EL distributes the contributions according to $nh^d V^{-1}(x)$, the inverse of the covariance matrix of $\hat{\bar{Q}}(x, \theta_0)$ which is the most natural choice. The nice thing about the EL formulation is that this is done without explicit estimation of $V(x)$ due to its internal standardization. Estimating $V(x)$ when k is large can be challenging if not just tedious.

3 Main Results

Let

$$\omega_{l_1,l_2,j_1,j_2}(\beta, K) = \int\int\int \beta_{l_2}^{-d} K(u)K(v)K\{(\beta_{j_2} z + \beta_{l_1} u)/\beta_{l_2}\} K\left(z + \beta_{j_1} v/\beta_{j_2}\right) dudvdz,$$

$(\gamma_{lj}(x))_{k\times k} = \left((\beta_j^{-d} R(\beta_l/\beta_j)\sigma_{lj}(x))_{k\times k}\right)^{-1}$, and

$$\sigma^2(K, \Sigma) = 2\sum_{l_1,l_2,j_1,j_2}^{k} \beta_{l_2}^{-d} \omega_{l_1,l_2,j_1,j_2}(\beta, K) \int \gamma_{l_1 j_1}(x)\gamma_{l_2 j_2}(x)\sigma_{l_1 l_2}(x)\sigma_{j_1 j_2}(x)\pi^2(x)dx$$

which is a bounded quantity under assumption (A.1) and (A.4) given in Section 6.

Theorem 3.1 *Under the assumptions (A.1)-(A.6) and (B.1)-(B.5) given in Section 6, and under H_0,*

$$h^{-d/2}\{\Lambda_n(\vec{h}) - k\} \xrightarrow{d} N(0, \sigma^2(K, \Sigma))$$

as $n \to \infty$.

Remark 3.1 (equal bandwidths) If $h_1 = \ldots = h_k = h$, that is $\beta_1 = \ldots = \beta_k = 1$, then $\omega_{l_1,l_2,j_1,j_2}(\beta, K) = K^{(4)}(0)$ where $K^{(4)}$ is the convolution of $K^{(2)}$, and $K^{(2)}$ is the convolution of K, that is

$$K^{(2)}(u) = \int K(v)K(u+v)dv.$$

Since $V(x) = f(x)R(K)\Sigma(x)$ in the case of equal bandwidths, $\sum_{l=1}^{k} \gamma_{lj_1}\sigma_{lj_2}(x) = I(j_1 = j_2)R^{-1}(K)$ where I is the indicator function. Therefore, $\sigma^2(K, \Sigma) = 2kK^{(4)}(0)$ $R^{-2}(K)\int \pi^2(x)dx$, which is entirely known upon given the kernel function. Hence, the EL test statistic is asymptotically pivotal.

Remark 3.2 (unequal bandwidths) If the bandwidths are not all the same, the asymptotic variance of $\Lambda_n(\vec{h})$ may depend on $\Sigma(x)$, which means that the EL test statistic is no longer asymptotically pivotal. However, the distribution of $\Lambda_n(\vec{h})$ is always free of the design distribution of X_i.

Let $q_{n\alpha}$ be the upper α-quantile of the distribution of $h^{-d/2}\{\Lambda_n(\vec{h}) - k\}$ for a significance level $\alpha \in (0, 1)$. Theorem 3.1 assures that $q_{n\alpha} \to z_\alpha$, the upper α quantile of $N(0, \sigma^2(K, \Sigma))$. However, the convergence can be slow. There is also an issue of estimating $\sigma^2(K, \Sigma)$ when different bandwidths are used. For these reasons we prefer to use a bootstrap approximation to calibrate the quantile $q_{n\alpha}$.

Remark 3.3 (bootstrap) Let $\hat{\epsilon}_i = Y_i - \hat{m}(X_i)$ be the estimated residual vectors for $i = 1, \ldots, n$ and G be a multivariate k-dimensional random vector such that $E(G) = 0$, $\mathrm{Var}(G) = I_k$ and G has bounded fourth order moments. To facilitate simple construction of the test statistic, and faster convergence, we propose the following bootstrap estimate of $q_{n\alpha}$.

Step 1: For $i = 1, \ldots, n$, generate $\epsilon_i^* = \hat{\epsilon}_i G_i$ where $G_1, \ldots, G_n$ are independent and identical copies of G, and let $Y_i^* = m(X_i, \hat{\theta}, \hat{g}) + \epsilon_i^*$. Re-estimate θ and g based on $\{(X_i, Y_i^*)\}_{i=1}^n$ and denote them as $\hat{\theta}^*$ and $\hat{g}^*$.

Step 2: compute the EL ratio at $\tilde{m}(x, \hat{\theta}^*, \hat{g}^*)$ based on $\{(X_i, Y_i^*)\}_{i=1}^n$, denote it as $\ell^*\{\tilde{m}(x, \hat{\theta}^*, \hat{g}^*)\}$ and then obtain the bootstrap version of the test statistic $\Lambda_n^*(\vec{h}) = \int \ell^*\{\tilde{m}(x, \hat{\theta}^*, \hat{g}^*)\}\pi(x)dx$ and let $\xi^* = h^{-d/2}\{\Lambda_n^*(\vec{h}) - k\}$.

Step 3: Repeat Steps 1 and 2 N times, and obtain $\xi_1^* \leq \ldots \leq \xi_N^*$ without loss of generality.

The bootstrap estimate of $q_{n\alpha}$ is then $\hat{q}_{n\alpha} =: \xi^*_{[N\alpha]+1}$.

The proposed EL test with α-level of significance rejects H_0 if $h^{-d/2}\{\Lambda_n(\vec{h}) - k\} > \hat{q}_{n\alpha}$.

Remark 3.4 (bandwidth selection) Each bandwidth h_l used in the kernel regression estimator $\hat{m}_l(x)$ can be chosen by a standard bandwidth selection procedure for instance the cross-validation (CV) method. The range in term of order of magnitude for all the k bandwidths $\{h_l\}_{l=1}^k$ covers the order of $n^{-1/(d+2r)}$ which is the optimal order that minimizes the mean integrated squared error in the estimation of m_l and is also the asymptotic order of the bandwidth selected by the CV method. We also note that once $\{h_l\}_{l=1}^k$ are chosen, the same set of bandwidths will be used in formulating the bootstrap version of the test statistic $\Lambda_n^*(\vec{h})$.

Theorem 3.2 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5) given in Section 6, and under H_0,*

$$P\left(h^{-d/2}\{\Lambda_n(\vec{h}) - k\} \geq \hat{q}_{n\alpha}\right) \to \alpha,$$

as $\min(n, N) \to \infty$.

Theorem 3.2 maintains that the proposed test has asymptotically correct size.

We next consider the power of the test under a sequence of local alternatives. First, consider the following local alternative hypothesis:

$$H_{1n} : m(\cdot) = m(\cdot, \theta_0, g_0) + c_n \Gamma_n(\cdot), \tag{3.1}$$

where $c_n = n^{-1/2} h^{-d/4}$ and $\Gamma_n(x) = \left(\Gamma_{n1}(x), \cdots, \Gamma_{nk}(x)\right)^T$ for some bounded functions $\Gamma_{nl}(\cdot)$ $(l = 1, \ldots, k)$.

Theorem 3.3 *Under the assumptions (A.1)-(A.7) and (B.1)-(B.5) given in Section 6, and under H_{1n},*

$$h^{-d/2}\{\Lambda_n(\vec{h}) - k\} \xrightarrow{d} N(\beta(f, K, \Sigma, \Gamma), \sigma^2(K, \Sigma))$$

as $n \to \infty$, where

$$\begin{aligned} \beta(f, K, \Sigma, \Gamma) &= \sum_{l,j=1}^{k} \int \Gamma_l(x) \Gamma_j(x) \gamma_{lj}(x) f(x) \pi(x)\, dx \\ &= \int \Gamma^T(x) V^{-1}(x) \Gamma(x) f^2(x) \pi(x) dx \end{aligned}$$

and $\Gamma(x) = \lim_{n\to\infty} \Gamma_n(x)$ assuming such a limit does exists.

Remark 3.5 (power) The asymptotic mean of the EL test statistic is given by $\int \Gamma^T(x) V^{-1}(x) \Gamma(x) f^2(x) \pi(x) dx$, which is bounded away from zero since $V(x)$ is positive definite with smallest eigen-function uniformly bounded away from zero. As a result, the EL test has a non-trivial asymptotic power

$$\Phi\left[\{z_\alpha - \beta(f, K, \Sigma, \Gamma)\}/\sigma(K, \Sigma)\right]$$

where Φ is the distribution function of the standard normal distribution. We note here that the above power is attained for any $\Gamma(x)$ without requiring specific directions in which H_1 deviates from H_0. This indicates the proposed test is able to test consistently any departure from H_0.

By repeating the derivation of Theorem 3.3, it can be seen that if the order of c_n is larger than $n^{-1/2}h^{-d/4}$, then $\beta(f, K, \Sigma, \Gamma)$ will converge to infinity, which then implies that the power of the EL test will converge to 1. If otherwise c_n converges to zero faster than $n^{-1/2}h^{-d/4}$, then $\beta(f, K, \Sigma, \Gamma)$ will degenerate to zero and hence the power of the test will degenerate to the significance level α.

4 Examples

In this section we will apply the general results obtained in Section 3 on a number of particular models: partially linear models, single index models, additive models, monotone regression models, the selection of variables, and the simultaneous testing of the conditional mean and variance. These six examples form a representative subset of the more complete list of examples listed in the introduction section. For the other examples not treated here, the development is quite similar.

4.1 Partially linear models

Consider the model

$$\begin{aligned} Y_i &= m(X_i, \theta_0, g_0) + \epsilon_i \\ &= \theta_{00} + \theta_{01}X_{i1} + \ldots + \theta_{0,d-1}X_{i,d-1} + g_0(X_{id}) + \epsilon_i, \end{aligned} \quad (4.1)$$

where Y_i is a one-dimensional response variable ($k = 1$), $d > 1$, $E(\epsilon_i|X_i = x) = 0$ and $\mathrm{Var}(\epsilon_i|X_i = x) = \Sigma(x)$ ($1 \le i \le n$). For identifiability reasons we assume that $E(g_0(X_{id})) = 0$. This testing problem has been studied in Yatchew (1992), Whang and Andrews (1993) and Rodríguez-Póo, Sperlich and Vieu (2005), among others. For any $\theta \in R^d$ and $x \in R$, let

$$\hat{h}(x, \theta) = \sum_{i=1}^{n} W_{in}(x, b)[Y_i - \theta_0 - \theta_1 X_{i1} - \ldots - \theta_{d-1}X_{i,d-1}] \quad (4.2)$$

$$\hat{g}(x,\theta) \;=\; \hat{h}(x,\theta) - \frac{1}{n}\sum_{i=1}^{n} \hat{h}(X_{id},\theta), \tag{4.3}$$

where

$$W_{in}(x,b) = \frac{L\left(\frac{x-X_{id}}{b}\right)}{\sum_{j=1}^{n} L\left(\frac{x-X_{jd}}{b}\right)},$$

b is a univariate bandwidth sequence and L a kernel function. Next, define

$$\hat{\theta} = \mathrm{argmin}_{\theta\in R^d} \sum_{i=1}^{n} \Big[Y_i - \theta_0 - \theta_1 X_{i1} - \ldots - \theta_{d-1}X_{i,d-1} - \hat{g}(X_{id},\theta)\Big]^2.$$

Then, $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$, see Härdle, Liang and Gao (2000, Chapter 2), and

$$\begin{aligned} |m(X_i,\theta_0,\hat{g}) - m(X_i,\theta_0,g_0)| &= |\hat{g}(X_i,\theta_0) - g_0(X_i)| \\ &= O_p\{(nb)^{-1/2}\log(n)\} = o_p\{(nh^d)^{-1/2}\log(n)\}, \end{aligned}$$

uniformly in $1 \le i \le n$, provided $h^d/b \to 0$. This is the case when $h \sim n^{-1/(d+4)}$ and $b \sim n^{-1/5}$. Hence, condition (B.1) is satisfied. Conditions (B.2) and (B.3) obviously hold, since

$$\frac{\partial m(X_i,\theta_0,g)}{\partial\theta} = (1, X_{i1},\ldots,X_{i,d-1})^T \quad \text{and} \quad \frac{\partial^2 m(X_i,\theta_0,g)}{\partial\theta\partial\theta^T} = 0$$

for any g. Finally, when the order of the kernel L equals 2,

$$E\{\hat{g}(x,\theta_0)\} = g_0(x) + O(b^2),$$

uniformly in x; and $O(b^2)$ is $o(h^2)$ provided $b/h \to 0$, which is satisfied for the above choices of h and b. Hence, (B.4) is satisfied for $r = 2$.

4.2 Single index models

In single index models it is assumed that

$$Y_i \;=\; m(X_i,\theta_0,g_0) + \epsilon_i = g_0(\theta_0^T X_i) + \epsilon_i, \tag{4.4}$$

where k (the dimension of Y_i) equals 1, $\theta_0 = (\theta_{01},\ldots,\theta_{0d})^T$, $X_i = (X_{i1},\ldots,X_{id})^T$ for some $d > 1$, $E(\epsilon_i|X_i = x) = 0$ and $\mathrm{Var}(\epsilon_i|X_i = x) = \Sigma(x)$ $(1 \le i \le n)$. In order to identify the model, set $\|\theta_0\| = 1$. See e.g. Xia, Li, Tong and Zhang (2004), Stute and Zhu (2005) and

Rodríguez-Póo, Sperlich and Vieu (2005) for procedures to test this single index model. For any $\theta \in \Theta$ and $u \in R$, let

$$\hat{g}(u,\theta) = \sum_{i=1}^{n} \frac{L_b(u - \theta^T X_i)}{\sum_{j=1}^{n} L_b(u - \theta^T X_j)} Y_i.$$

Then, the estimator of θ_0 is defined by

$$\hat{\theta} = \text{argmin}_{\theta:\|\theta\|=1} \sum_{i=1}^{n} [Y_i - \hat{g}(\theta^T X_i, \theta)]^2.$$

Härdle, Hall and Ichimura (1993) showed that $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$. Obviously, from standard kernel regression theory we know that

$$\begin{aligned} \max_i |m(X_i,\theta_0,\hat{g}) - m(X_i,\theta_0,g_0)| &\leq \sup_u |\hat{g}(u,\theta_0) - g_0(u)| \\ &= O_p\{(nb)^{-1/2}\log(n)\} = o_p\{(nh^d)^{-1/2}\log(n)\}, \end{aligned}$$

$$\begin{aligned} \max_i \Big|\frac{\partial}{\partial\theta} m(X_i,\theta_0,\hat{g}) - \frac{\partial}{\partial\theta} m(X_i,\theta_0,g_0)\Big| &\leq C \sup_u |\hat{g}'(u,\theta_0) - g_0'(u)| \\ &= O_p\{(nb^3)^{-1/2}\log(n)\} = o_p(1), \end{aligned}$$

$$\begin{aligned} \max_i \Big|\frac{\partial^2}{\partial\theta\partial\theta^T} m(X_i,\theta_0,\hat{g})\Big| &\leq C \sup_u |\hat{g}''(u,\theta_0)| \\ &= C \sup_u |g_0''(u)| + O_p\{(nb^5)^{-1/2}\log(n)\} = o_p(n^{1/2}), \end{aligned}$$

and

$$\sup_u |E\{\hat{g}(u,\theta_0)\} - g_0(u)| = O(b^2) = o(h^2),$$

for some $C > 0$, provided $h^d/b \to 0$ and $nb^3 \log^{-2}(n) \to \infty$, which is the case, as for the partially linear model, when e.g. $h \sim n^{-1/(d+4)}$ and $b \sim n^{-1/5}$.

4.3 Additive models

We suppose now that the model is given by

$$Y_i = m_{00} + m_{10}(X_{i1}) + \ldots + m_{d0}(X_{id}) + \epsilon_i, \tag{4.5}$$

where $k = 1$, $d > 1$, $E(\epsilon_i|X_i = x) = 0$, $\mathrm{Var}(\epsilon_i|X_i = x) = \Sigma(x)$ and $E(m_{j0}(X_{ij})) = 0$ ($1 \le i \le n$; $1 \le j \le d$). The estimation of the parameter m_{00} and of the functions $m_{j0}(\cdot)$ ($1 \le j \le d$) has been considered in e.g. Linton and Nielsen (1995) (marginal integration), Opsomer and Ruppert (1997) (backfitting), Mammen, Linton and Nielsen (1999) (smooth backfitting). Using e.g. the covering technique to extend pointwise convergence results to uniform results (see e.g. Bosq, 1998), it can be shown that the estimators $\hat{m}_j(\cdot)$ ($j = 1, \ldots, d$) considered in these papers satisfy the following properties:

$$\sup_x |\hat{m}_j(x) - m_{j0}(x)| = O_p\{(nb)^{-1/2} \log(n)\}$$
$$\sup_x |E\{\hat{m}_j(x)\} - m_{j0}(x)| = O(b^2),$$

where b is the bandwidth used for either of these estimators. Hence, assumptions (B.1)-(B.5) hold true provided $h^d/b \to 0$ and $b/h \to 0$, which is the case when e.g. h and b equal the optimal bandwidths for kernel estimation in dimension d respectively 1, namely $h \sim n^{-1/(d+4)}$ and $b \sim n^{-1/5}$ (take $r = s = 2$).

4.4 Monotone regression

Consider now the following model

$$Y_i = m_0(X_i) + \epsilon_i, \tag{4.6}$$

where X_i and Y_i are one-dimensional, and where we assume that m_0 is monotone. An overview of nonparametric methods for estimating a monotone regression function, as well as testing for monotonicity is given in Gijbels (2005). Let $\hat{m}(x)$ be an estimator of $m_0(x)$ under the assumption of monotonicity, that is based on a bandwidth sequence b and a kernel L of order s, and that satisfies

$$\sup_x |\hat{m}(x) - m_0(x)| = O_p\{(nb)^{-1/2} \log(n)\}$$
$$\sup_x |E\{\hat{m}(x)\} - m_0(x)| = O(b^s)$$

(as for the additive model, the uniformity in x can be obtained by using classical tools based on e.g. the covering technique). Then, the required regularity conditions on $\hat{m}(x)$ are satisfied provided $h/b \to 0$ and $b^s/h^r \to 0$, i.e. when e.g. $s = 3$, $r = 2$, $b = Kn^{-1/5}$ and $h = b \log^{-1}(n)$.

4.5 Selection of variables

In this example we apply the general testing procedure on the problem of selecting explanatory variables in regression. Let $X_i = (X_i^{(1)T}, X_i^{(2)T})^T$ be a vector of $d = d_1 + d_2$ ($d_1, d_2 \geq 1$) explanatory variables. We like to test whether the vector $X_i^{(2)}$ should or should not be included in the model. See Delgado and González Manteiga (2001) for other nonparametric approaches to this problem. Our null model is

$$Y_i = m_0(X_i^{(1)}) + \epsilon_i. \tag{4.7}$$

Hence, under the hypothesized model the regression function $m(x^{(1)}, x^{(2)})$ is equal to a function $m_0(x^{(1)})$ only. In our testing procedure we estimate the regression function $m_0(\cdot)$ by

$$\hat{m}(x^{(1)}) = \sum_{i=1}^{n} \frac{L_b(x^{(1)} - X_i^{(1)})}{\sum_j L_b(x^{(1)} - X_j^{(1)})} Y_i,$$

where L is a d_1-dimensional kernel function of order $s = 2$ and b a bandwidth sequence. It is easily seen that this estimator satisfies the regularity conditions provided $h^d/b^{d_1} \to 0$ and $b/h \to 0$ (take $r = 2$). As before, the optimal bandwidths for estimation, namely $h \sim n^{-1/(d+4)}$ and $b \sim n^{-1/(d_1+4)}$ satisfy these constraints.

4.6 Simultaneous testing of the conditional mean and variance

Let $Z_i = r(X_i) + \Sigma^{1/2}(X_i)e_i$ where Z_i is a k_1-dimensional response variable of a d-dimensional covariate X_i, and $r(x) = E(Z_i|X_i = x)$ and $\Sigma(x) = Var(Z_i|X_i = x)$ are respectively the conditional mean and variance functions. This is a standard multivariate nonparametric regression model. Suppose that $r(x, \theta, g)$ and $\Sigma(x, \theta, g)$ are certain working models for the conditional mean and variance respectively. Hence, the hypothesized regression model is

$$Z_i = r(X_i, \theta, g) + \Sigma^{1/2}(X_i, \theta, g)e_i, \tag{4.8}$$

where the standardized residuals $\{e_i\}_{i=1}^n$ satisfy $E(e_i|X_i) = 0$ and $Var(e_i|X_i) = I_d$. Here, I_d is the d-dimensional identity matrix. Clearly, the parametric (without g) or semi-parametric (with g) model specification of (4.8) consists of two components of specifications: one for the regression part $r(X_i, \theta, g)$ and the other is the conditional variance part $\Sigma(X_i, \theta, g)$. The model (4.8) is valid if and only if both components of the specifications are valid simultaneously. Hence, we need to test the goodness-of-fit of both $r(x, \theta, g)$ and $\Sigma(x, \theta, g)$ simultaneously.

To use the notations of this paper, we have

$$m(x, \theta, g) = (r(x, \theta, g), vec\{\Sigma(x, \theta, g) + r(x, \theta, g) r^T(x, \theta, g)\})^T$$

and the multivariate "response" $Y_i = \Big(Z_i, vec(Z_i Z_i^T)\Big)^T$. Here $vec(A)$ denotes the operator that stacks columns of a matrix A into a vector.

5 Simulations

Consider the following model :

$$Y_i = 1 + 0.5X_{i1} + a g_1(X_{i1}) + g_2(X_{i2}) + \varepsilon_i \tag{5.1}$$

$(i = 1, \ldots, n)$. Here, the covariates X_{ij} $(j = 1, 2)$ follow a uniform distribution on $[0, 1]$, the error ε_i is independent of $X_i = (X_{i1}, X_{i2})$, and follows a normal distribution with mean zero and variance given by $\mathrm{Var}(\varepsilon_i|X_i) = (1.5 + X_{i1} + X_{i2})^2/100$. Several choices are considered for the constant $a \geq 0$ and the functions g_1 and g_2. We are interested in testing whether the data (X_i, Y_i) $(i = 1, \ldots, n)$ follow a partially linear model, in the sense that the regression function is linear in X_{i1}, and (possibly) non-linear in X_{i2}.

We will compare our empirical likelihood based test with the test considered by Rodríguez-Póo, Sperlich and Vieu (2005) (RSV hereafter), which is based on the L_∞-distance between a completely nonparametric kernel estimator of the regression function and an estimator obtained under the assumption that the model is partially linear.

The simulations are carried out for samples of size 100 and 200. The significance level is $\alpha = 0.05$. A total of 300 samples are selected at random, and for each sample 300 random resamples are drawn. We use a triangular kernel function $K(u) = (1 - |u|) I(|u| \leq 1)$ and we determine the bandwidth b by using a cross-validation procedure. For the bandwidth h, we follow the procedure used by Rodríguez-Póo, Sperlich and Vieu (2005), i.e. we consider the test statistic $\sup_{h_0 \leq h \leq h_1} [h^{-d/2}\{\Lambda_n(\vec{h}) - k\}]$, where h_0 and h_1 are chosen in such a way that the bandwidth obtained by cross-validation is included in the interval. For $n = 100$, we take $h_0 = 0.22$ and $h_1 = 0.28$, and for $n = 200$ we select $h_0 = 0.18$ and

$h_1 = 0.24$. The critical values for this test statistic are obtained from the distribution of the bootstrap statistic, given by $\sup_{h_0 \le h \le h_1}[h^{-d/2}\{\Lambda_n^*(\vec{h}) - k\}]$.

The results are shown in Table 1. The table shows that the level is well respected for both sample sizes, and for both choices of the function g_2. Under the alternative hypothesis, all the considered models demonstrate that the power increases with increasing sample size and increasing value of a. The empirical likelihood test is in general more powerful than the RSV test when c_n is small ($c_n = 0.5$ and 1.0). For the largest c_n considered, i.e. $c_n = 3$, the RSV test is slightly more powerful. However, this happens when both tests enjoy a large amount of power.

6 Assumptions and Proofs

Assumptions:

(A.1) K is a d-dimensional product kernel of the form $K(t_1, \dots, t_d) = \prod_{j=1}^d k(t_j)$, where k is a r-th order ($r \ge 2$) univariate kernel (i.e. $k(t) \ge 0$ and $\int k(t)\,dt = 1$) supported on $[-1, 1]$, k is symmetric, bounded and Lipschitz continuous.

(A.2) The baseline smoothing bandwidth h satisfies $nh^{d+2r} \to K$ for some $K \ge 0$, $nh^{3d/2}\log^{-4}(n) \to \infty$, and $h_l/h \to \beta_l$ as $n \to \infty$, where $c_0 \le \min_{1\le l\le k}\{\beta_l\} \le \max_{1\le l\le k}\{\beta_l\} \le c_1$ for finite and positive constants c_0 and c_1. Moreover, $d < 4r$ and the weight function π is bounded, Lipschitz continuous on its compact support S and satisfies $\int \pi(x)dx = 1$.

(A.3) Let $\epsilon_i = Y_i - m(X_i, \theta_0, g_0) = (\epsilon_{i1}, \dots, \epsilon_{ik})^T$. $E(|\prod_{j=1}^6 \epsilon_{il_j}|\,|X_i = x)$ is uniformly bounded for all $l_1, \dots, l_6 \in \{1, \dots, k\}$ and all $x \in S$.

(A.4) $f(x)$ and all the $\sigma_{lj}^2(x)$'s have continuous derivatives up to the second order in S, $\inf_{x\in S} f(x) > 0$ and $\min_l \inf_{x\in S}\sigma_{ll}^2(x) > 0$. Let $\xi_1(x)$ and $\xi_k(x)$ be the smallest and largest eigenvalues of $V(x)$. We assume that $c_2 \le \inf_{x\in S}\xi_1(x) \le \sup_{x\in S}\xi_k(x) \le c_3$ for finite and positive constants c_2 and c_3.

(A.5) Θ is a compact subspace of R^p, $P(\hat\theta \in \Theta) \to 1$ as $n \to \infty$, and $\hat\theta$ satisfies $\hat\theta - \theta_0 = O_p(n^{-1/2})$.

(A.6) $m(x, \theta, g)$ is twice continuously partially differentiable with respect to the components of θ and x for all g, and the derivatives are bounded uniformly in $x \in S, \theta \in \Theta$ and

			$n=100$		$n=200$	
$g_1(x_1)$	$g_2(x_2)$	a	RSV	EL	RSV	EL
x_1^2	$\exp(x_2)$	0	.047	.053	.040	.043
		0.5	.123	.153	.160	.193
		1	.377	.420	.653	.683
		1.5	.787	.743	.973	.980
	$\frac{2}{x_2+1}$	0	.033	.037	.043	.053
		0.5	.110	.120	.153	.177
		1	.373	.397	.667	.657
		1.5	.753	.733	.977	.963
$\log(x_1+0.5)$	$\exp(x_2)$	0	.047	.053	.040	.043
		1	.123	.147	.127	.160
		2	.387	.400	.657	.660
		3	.747	.723	.973	.960
	$\frac{2}{x_2+1}$	0	.033	.037	.043	.053
		1	.107	.133	.113	.147
		2	.407	.440	.660	.713
		3	.797	.763	.990	.983

Table 1: *Rejection probabilities under the null hypothesis ($a=0$) and under the alternative hypothesis ($a>0$). The test of Rodríguez-Póo, Sperlich and Vieu (2005) is indicated by 'RSV', the new test is indicated by 'EL'.*

$g \in \mathcal{G}$.

(A.7) Each of the functions $\Gamma_{nl}(x)$ $(l=1,\dots,k)$ appearing in the local alternative hypothesis converges to $\Gamma_l(x)$ as $n \to \infty$, and $\Gamma_l(x)$ is uniformly bounded with respect to x.

Let $\hat{\Delta}_l(x,\theta) = \tilde{m}_l(x,\theta,\hat{g}) - \tilde{m}_l(x,\theta,g_0)$ for $l=1,\dots,k$, $\hat{\Delta}(x,\theta) = (\hat{\Delta}_l(x,\theta))_{l=1}^k$,

$$\hat{Q}_i^{(2)}(x,\theta) = \left(K_{h_1}(x-X_i)\hat{\Delta}_1(x,\theta),\dots,K_{h_k}(x-X_i)\hat{\Delta}_k(x,\theta)\right)^T,$$

and let $\|\cdot\|$ be the Euclidean norm.

The following conditions specify stochastic orders for some quantities involving $\hat{Q}_i^{(2)}(x,\theta_0)$:

(B.1) $\max_{i,l} |m_l(X_i,\theta_0,\hat{g}) - m_l(X_i,\theta_0,g_0)| = o_p\{(nh^d)^{-1/2}\log(n)\}$.

(B.2) $\max_{i,l} |\frac{\partial m_l(X_i,\theta_0,\hat{g})}{\partial\theta} - \frac{\partial m_l(X_i,\theta_0,g_0)}{\partial\theta}| = o_p(1)$.

(B.3) $\max_{i,l} \|\frac{\partial^2 m_l(X_i,\theta_0,\hat{g})}{\partial\theta\partial\theta^T}\| = o_p(n^{1/2})$.

(B.4) $\sup_{x\in S} \|E\{m(x,\theta_0,\hat{g})\} - m(x,\theta_0,g_0)\| = o(h^r)$.

(B.5) $P(\hat{g}\in\mathcal{G}) \to 1$ as $n\to\infty$.

Lemma 6.1 *Assumption (B.1) implies (B.1a), (B.1b) and (B.1c), given by*

(B.1a) $\sup_{x\in S}\left[n^{-1}\sum_{i=1}^n \hat{Q}_i^{(2)}(x,\theta_0)\hat{Q}_i^{(2)T}(x,\theta_0)\right] = o_p\{n^{-1}h^{-2d}\log^2(n)\}$.

(B.1b) $\sup_{x\in S}\max_{1\le i\le n}\|\hat{Q}_i^{(2)}(x,\theta_0)\| = o_p\{n^{-1/2}h^{-3d/2}\log(n)\}$.

(B.1c) $\sup_{x\in S}\left[n^{-1}\sum_{i=1}^n \hat{Q}_i^{(2)}(x,\theta_0)\right] = o_p\{(nh^d)^{-1/2}\log(n)\}$.

Proof. First note that

$$\begin{aligned}\sup_x \max_i \|\hat{Q}_i^{(2)}(x,\theta_0)\| &\le \sup_x\left[\max_{i,l} K_{h_l}(x-X_i)\|\hat{\Delta}(x,\theta_0)\|\right]\\ &= O(h^{-d})\sup_x\max_l |\tilde{m}_l(x,\theta_0,\hat{g}) - \tilde{m}_l(x,\theta_0,g_0)|\\ &\le O(h^{-d})\max_{i,l}|m_l(X_i,\theta_0,\hat{g}) - m_l(X_i,\theta_0,g_0)|\\ &= o_p\{n^{-1/2}h^{-3d/2}\log(n)\}.\end{aligned}$$

Therefore,

$$\sup_x \|n^{-1}\sum_i \hat{Q}_i^{(2)}(x,\theta_0)\| \le h^d \sup_i \max_i \|\hat{Q}_i^{(2)}(x,\theta_0)\| = o_p\{(nh^d)^{-1/2}\log(n)\}.$$

Finally,

$$\sup_x \|n^{-1}\sum_i \hat{Q}_i^{(2)}(x,\theta_0)\hat{Q}_i^{(2)T}(x,\theta_0)\| \le h^d \sup_x\max_i \|\hat{Q}_i^{(2)}(x,\theta_0)\|^2 = o_p\{n^{-1}h^{-2d}\log^2(n)\}.$$

□

Lemma 6.2 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under* H_0,

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

Proof. Write $\lambda(x) = \rho(x)\eta(x)$ where $\rho(x) = \|\lambda(x)\|$ and $\eta(x) \in R^k$ satisfying $\|\eta(x)\| = 1$. And define $\hat{B}_n(x, \hat{\theta}) = \max_{1\le i\le n} \|\hat{Q}_i(x, \hat{\theta})\|$ and $\hat{S}_n(x, \hat{\theta}) = n^{-1}\sum_{i=1}^n \hat{Q}_i(x, \hat{\theta})\hat{Q}_i^T(x, \hat{\theta})$. Then, following Owen (1990),

$$\rho(x) \le \frac{\{1 + \rho(x)\hat{B}_n(x, \hat{\theta})\}\eta^T(x)\bar{\hat{Q}}(x, \hat{\theta})}{\eta^T(x)\hat{S}_n(x, \hat{\theta})\eta(x)}. \tag{6.1}$$

We want to show first that

$$\sup_{x\in S} \hat{B}_n(x, \hat{\theta}) = o_p(n^{1/2}h^{-d/2}/\log(n)), \tag{6.2}$$

$$\sup_{x\in S} \|h^d\hat{S}_n(x, \hat{\theta}) - V(x)\| = O_p\{(nh^d)^{-1/2}\log(n)\}. \tag{6.3}$$

Let $\hat{Q}_i^{(1)}(x, \hat{\theta}) = \hat{Q}_i(x, \hat{\theta}) + \hat{Q}_i^{(2)}(x, \hat{\theta})$. From (2.2),

$$\hat{Q}_i^{(1)}(x, \hat{\theta}) = \Big(K_{h_1}(x - X_i)(Y_{i1} - \tilde{m}_1(x, \hat{\theta}, g_0)), \ldots, K_{h_k}(x - X_i)(Y_{ik} - \tilde{m}_k(x, \hat{\theta}, g_0))\Big)^T. \tag{6.4}$$

Furthermore for $s, t = 1, 2$, let

$$\hat{B}_n^{(s)}(x, \hat{\theta}) = \max_{1\le i\le n} \|\hat{Q}_i^{(s)}(x, \hat{\theta})\|,$$

$$\hat{S}_n^{(st)}(x, \hat{\theta}) = n^{-1}\sum_{i=1}^n \hat{Q}_i^{(s)}(x, \hat{\theta})\hat{Q}_i^{(t)T}(x, \hat{\theta})$$

$$\bar{\hat{Q}}^{(s)}(x, \hat{\theta}) = n^{-1}\sum_{i=1}^n \hat{Q}_i^{(s)}(x, \hat{\theta}).$$

Then, $\hat{B}_n(x, \hat{\theta}) \le \hat{B}_n^{(1)}(x, \hat{\theta}) + \hat{B}_n^{(2)}(x, \hat{\theta})$ and $\hat{S}_n(x, \hat{\theta}) = \sum_{s,t=1}^2 (-1)^{s+t}\hat{S}_n^{(st)}(x, \hat{\theta})$.

To obtain (6.2), first note that

$$\sup_{x\in S} \hat{B}_n(x, \hat{\theta}) = \sup_{x\in S} \hat{B}_n(x, \theta_0) + O_p(n^{-1/2}h^{-d}). \tag{6.5}$$

This is because by conditions (A.5) and (A.6) and the boundedness of K,

$$\begin{aligned}\sup_x \max_i \|\hat{Q}_i(x, \hat{\theta}) - \hat{Q}_i(x, \theta_0)\| &\le O_p(n^{-1/2})\sup_x \max_i \sqrt{\sum_{l=1}^k K_{h_l}^2(x - X_i)} \\ &= O_p(n^{-1/2}h^{-d}).\end{aligned} \tag{6.6}$$

Next, note that

$$\begin{aligned}h^d\|\hat{Q}_i^{(1)}(x, \hat{\theta})\| &= \Big[\sum_{l=1}^k \beta_l^{-2d}K^2\left(\frac{x - X_i}{h_l}\right)\{\epsilon_{il} - [\tilde{m}_l(x, \hat{\theta}, g_0) - m_l(X_i, \theta_0, g_0)]\}^2\Big]^{1/2} \\ &\le 2c_0^{-d}\Big[\sum_{l=1}^k K^2\left(\frac{x - X_i}{h_l}\right)\{\epsilon_{il}^2 + o(1)\}\Big]^{1/2}.\end{aligned}$$

Since each component of ϵ_i has finite conditional second moments, we have uniformly in $x \in S$,

$$n^{-1} \sum_i \|\hat{Q}_i^{(1)}(x, \hat{\theta})\| = O_p(1), \tag{6.7}$$

and similarly for $\hat{Q}_i^{(2)}(x, \hat{\theta})$.

Let's show now that

$$\sup_{x \in S} h^d \|\hat{S}_n(x, \hat{\theta}) - \hat{S}_n(x, \theta_0)\| = O_p(n^{-1/2}), \tag{6.8}$$

$$\sup_{x \in S} \|\hat{\bar{Q}}(x, \hat{\theta}) - \hat{\bar{Q}}(x, \theta_0)\| = O_p(n^{-1/2}). \tag{6.9}$$

We will prove (6.8), the proof of (6.9) being similar if not easier.

$$\begin{aligned}
&\|\hat{S}_n(x, \hat{\theta}) - \hat{S}_n(x, \theta_0)\| \\
&\leq n^{-1} \sum_i [\|\hat{Q}_i(x, \hat{\theta})\| \, \|\hat{Q}_i(x, \hat{\theta}) - \hat{Q}_i(x, \theta_0)\| + \|\hat{Q}_i(x, \hat{\theta}) - \hat{Q}_i(x, \theta_0)\| \, \|\hat{Q}_i(x, \theta_0)\|] \\
&= O_p(n^{-1/2} h^{-d}),
\end{aligned}$$

uniformly in x, which follows from (6.6) and (6.7).

It can be shown that by following the argument of Owen (1990),

$$\sup_{x \in S} \hat{B}_n^{(1)}(x, \theta_0) = o_p(n^{1/2} h^{-d/2} / \log(n)). \tag{6.10}$$

Assumption (B.1b) implies that $\sup_{x \in S} \hat{B}_n^{(2)}(x, \theta_0) = o_p(n^{1/2} h^{-d/2} / \log(n))$. This together with (6.5) and (6.10) leads to (6.2).

To prove (6.3), one can show that

$$\sup_{x \in S} \|h^d \hat{S}_n^{(11)}(x, \theta_0) - V(x)\| = O_p\{(nh^d)^{-1/2} \log(n)\}. \tag{6.11}$$

Condition (B.1a) implies that

$$\sup_{x \in S} \|h^d \hat{S}_n^{(22)}(x, \theta_0)\| = O_p((nh^d)^{-1/2} \log(n)), \tag{6.12}$$

while

$$\begin{aligned}
&\sup_{x \in S} \|h^d \hat{S}_n^{(12)}(x, \theta_0)\| \\
&\leq \sup_x \Big[\max_i \|h^d \hat{Q}_i^{(2)}(x, \theta_0)\| \, n^{-1} \sum_i \|\hat{Q}_i^{(1)}(x, \theta_0)\| \Big] \\
&= O_p\{(nh^d)^{-1/2} \log(n)\}
\end{aligned} \tag{6.13}$$

and similarly for $\hat{S}_n^{(21)}(x,\theta_0)$. Combining (6.8), (6.11) - (6.13), (6.3) is derived.

From Condition (B.1c) and (6.9),

$$\sup_{x\in S}\|\hat{\tilde{Q}}(x,\hat{\theta})\| = \sup_{x\in S}\|\hat{\tilde{Q}}^{(1)}(x,\theta_0)\| + o_p\{(nh^d)^{-1/2}\log(n)\} = O_p\{(nh^d)^{-1/2}\log(n)\}, \quad (6.14)$$

since

$$\begin{aligned}\hat{\tilde{Q}}_l^{(1)}(x,\theta_0) &= (nh_l^d)^{-1}\sum_i K\Big(\frac{x-X_i}{h_l}\Big)(Y_{il}-\tilde{m}_l(x,\theta_0,g_0)) \\ &= (nh_l^d)^{-1}\sum_i K\Big(\frac{x-X_i}{h_l}\Big)\epsilon_{il} \\ &= O((nh^d)^{-1/2}\log(n)),\end{aligned}$$

uniformly in x. This and (6.2) imply that

$$\sup_{x\in S} h^d\|\lambda(x)\|\hat{B}_n(x,\hat{\theta})\,|\eta^T(x)\hat{\tilde{Q}}(x,\hat{\theta})| = o_p\{\sup_{x\in S}\|\lambda(x)\|\}. \quad (6.15)$$

By condition (A.4),

$$\inf_{x\in S}\eta^T(x)V(x)\eta(x) \geq \inf_{x\in S}\xi_1(x) \geq c_2 > 0. \quad (6.16)$$

Hence, returning to (6.1) and using (6.3), (6.14)-(6.16), we have

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

Hence, this completes the proof. □

The following lemma gives a one-step expansion for $\lambda(x)$.

Lemma 6.3 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under H_0,*

$$\begin{aligned}\lambda(x)h^{-d} &= (h^d\hat{S}_n(x,\hat{\theta}))^{-1}\hat{\tilde{Q}}(x,\hat{\theta}) + O_p\{(nh^d)^{-1}\log^3(n)\} \\ &= V^{-1}(x)\hat{\tilde{Q}}(x,\hat{\theta}) + O_p\{(nh^d)^{-1}\log^3(n)\}, \quad (6.17)\end{aligned}$$

uniformly with respect to $x \in S$.

Proof. As Lemma 6.2 implies that $\sup_{x\in S}\max_i|\lambda^T(x)\hat{Q}_i(x,\hat{\theta})| = o_p(1)$, we can safely expand (2.4) to

$$\hat{\tilde{Q}}(x,\hat{\theta}) - \hat{S}_n(x,\hat{\theta})\lambda(x) + \hat{A}_n(x) = 0, \quad (6.18)$$

where $\hat{A}_n(x) = n^{-1}\sum_{i=1}^n \frac{\{\lambda^T(x)\hat{Q}_i(x,\hat{\theta})\}^2\hat{Q}_i(x,\hat{\theta})}{\{1+\xi^T(x)\hat{Q}_i(x,\hat{\theta})\}^3}$ for some $\xi(x) = (\xi_l(x))_{l=1}^k$, with each $\xi_l(x)$ between 0 and $\lambda_l(x)$ $(1 \le l \le k)$. As

$$\|\hat{A}_n(x)\| \le \|\lambda(x)h^{-d}\|^2 n^{-1}\sum_{i=1}^n \frac{h^{2d}\|\hat{Q}_i(x,\hat{\theta})\|^2\hat{Q}_i(x,\hat{\theta})}{|1+\xi^T(x)\hat{Q}_i(x,\hat{\theta})|^3} = O_p\{(nh^d)^{-1}\log^3(n)\}, \quad (6.19)$$

(6.17) is reached by inverting (6.18) while using (6.3). □

We next derive an expansion of the EL ratio statistic.

Lemma 6.4 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under* H_0,

$$\ell\{\tilde{m}(x,\hat{\theta},\hat{g})\} = nh^d\hat{\bar{Q}}^T(x,\hat{\theta})V^{-1}(x)\hat{\bar{Q}}(x,\hat{\theta}) + \hat{q}_n(x,\hat{\theta}) + o_p(h^{d/2}),$$

uniformly with respect to $x \in S$, *where*

$$\hat{q}_n(x,\hat{\theta}) \quad = \quad nh^d\hat{\bar{Q}}^T(x,\hat{\theta})\{(h^d\hat{S}_n(x,\hat{\theta}))^{-1} - V^{-1}(x)\}\hat{\bar{Q}}(x,\hat{\theta}) + \tfrac{2}{3}nh^d\hat{D}_n(x).$$

Proof. From (2.3) and a Taylor expansion,

$$\begin{aligned}\ell\{\tilde{m}(x,\hat{\theta},\hat{g})\} &= 2\sum_{i=1}^n \log\{1+\lambda^T(x)\hat{Q}_i(x,\hat{\theta})\} \\ &= 2n\lambda^T(x)\hat{\bar{Q}}(x,\hat{\theta}) - n\lambda^T(x)\hat{S}_n(x,\hat{\theta})\lambda(x) + \tfrac{2}{3}nh^d\hat{D}_n(x), \quad (6.20)\end{aligned}$$

where

$$\hat{D}_n(x) = (nh^d)^{-1}\sum_{i=1}^n \frac{\{\lambda^T(x)\hat{Q}_i(x,\hat{\theta})\}^3}{\{1+\eta(x)\lambda^T(x)\hat{Q}_i(x,\hat{\theta})\}^3}$$

for some $|\eta(x)| \le 1$.

Now, substitute (6.17) into (6.20),

$$\begin{aligned}\ell\{\tilde{m}(x,\hat{\theta},\hat{g})\} &= n\hat{\bar{Q}}^T(x,\hat{\theta})\hat{S}_n^{-1}(x,\hat{\theta})\hat{\bar{Q}}(x,\hat{\theta}) + \tfrac{2}{3}nh^d\hat{D}_n(x) - n\hat{A}_n^T(x)\hat{S}_n^{-1}(x,\hat{\theta})\hat{A}_n(x). \\ &= nh^d\hat{\bar{Q}}^T(x,\hat{\theta})V^{-1}(x)\hat{\bar{Q}}(x,\hat{\theta}) + \hat{q}_n(x,\hat{\theta}) + O_p\{(nh^d)^{-1}\log^4(n)\},\end{aligned}$$

where the last equality follows from (6.3) and (6.19). Hence the claim of the lemma is reached, since $O_p\{(nh^d)^{-1}\log^4(n)\} = o_p(h^{d/2})$. □

Applying Lemma 6.4 and (6.9), the EL test statistic can be written as

$$\begin{aligned}\Lambda_n(\vec{h}) &= nh^d \int \hat{\tilde{Q}}^T(x,\hat{\theta})V^{-1}(x)\hat{\tilde{Q}}(x,\hat{\theta})\pi(x)dx + R_{n1} + o_p(h^{d/2}) \\ &= nh^d \int \hat{\tilde{Q}}^T(x,\theta_0)V^{-1}(x)\hat{\tilde{Q}}(x,\theta_0)\pi(x)dx + R_{n1} + R_{n2} + o_p(h^{d/2}), \quad (6.21)\end{aligned}$$

where

$$\begin{aligned}R_{n1} &= \int \hat{q}_n(x,\hat{\theta})\pi(x)dx \quad \text{and} \\ R_{n2} &= 2nh^d \int \hat{\tilde{Q}}^T(x,\theta_0)V^{-1}(x)\{\hat{\tilde{Q}}(x,\hat{\theta}) - \hat{\tilde{Q}}(x,\theta_0)\}\pi(x)dx \\ &\quad + nh^d \int \{\hat{\tilde{Q}}(x,\hat{\theta}) - \hat{\tilde{Q}}(x,\theta_0)\}^T V^{-1}(x)\{\hat{\tilde{Q}}(x,\hat{\theta}) - \hat{\tilde{Q}}(x,\theta_0)\}\pi(x)dx.\end{aligned}$$

Let us consider the orders of R_{n1} and R_{n2}. From (6.20), $R_{n1} = R_{n11} + R_{n12}$, where

$$\begin{aligned}R_{n11} &= nh^d \int \hat{\tilde{Q}}^T(x,\hat{\theta})\{(h^d\hat{S}_n(x,\hat{\theta}))^{-1} - V^{-1}(x)\}\hat{\tilde{Q}}(x,\hat{\theta})\pi(x)dx \\ R_{n12} &= \tfrac{2}{3}nh^d \textstyle\int \hat{D}_n(x)\pi(x)dx.\end{aligned}$$

Lemma 6.5 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under H_0, $R_{n1} = o_p(h^{d/2})$.*

Proof. To obtain the order for R_{n1}, we analyze only R_{n12} as that for R_{n11} is similar and easier. From the proof of Lemma 6.2 we know that $\sup_{x\in S}\max_i |\lambda^T(x)\hat{Q}_i(x,\hat{\theta})| = o_p(1)$. Hence, it follows from (6.6) that

$$\begin{aligned}\hat{D}_n(x) &= (nh^d)^{-1}\sum_{i=1}^n \{\lambda^T(x)\hat{Q}_i^{(1)}(x,\theta_0) + \lambda^T(x)\hat{Q}_i^{(2)}(x,\theta_0)\}^3\{1+o_p(1)\} \\ &= (nh^d)^{-1}\sum_{i=1}^n\sum_{j=0}^3 C_j\{\lambda^T(x)\hat{Q}_i^{(1)}(x,\theta_0)\}^{3-j}\{\lambda^T(x)\hat{Q}_i^{(2)}(x,\theta_0)\}^j\{1+o_p(1)\} \\ &=: \sum_{j=0}^3 \hat{D}_{nj}(x)\{1+o_p(1)\}\end{aligned}$$

where $C_0 = C_3 = 1$ and $C_1 = C_2 = 3$. We will evaluate each of $\hat{D}_{nj}(x)$.

Starting from $\hat{D}_{n3}(x)$, we note that

$$\begin{aligned}|\hat{D}_{n3}(x)| &\le \|\lambda(x)h^{-d}\|^3(nh^d)^{-1}\sum_{i=1}^n \|h^d\hat{Q}_i^{(2)}(x,\theta_0)\|^3 \\ &= \|\lambda(x)h^{-d}\|^3(nh^d)^{-1}\sum_{i=1}^n \Big\{\sum_{l=1}^k \beta_l^{-2d}K^2\left(\frac{x-X_i}{h_l}\right)\hat{\Delta}_l^2(x,\theta_0)\Big\}^{3/2}.\end{aligned}$$

From (B.0) and Lemma 6.2, $\sup_{x\in S}|\hat{D}_{n3}(x)| = o_p\{(nh^d)^{-3}\log^{9/2}(n)\}$, and using exactly the same argument, combined with (6.7), $\sup_{x\in S}|\hat{D}_{nj}(x)| = o_p\{(nh^d)^{-(3+j)/2}\log^{(6+j)/2}(n)\}$ for $j = 1$ and 2. Hence,

$$\sup_{x\in S}|\sum_{j=1}^{3}\hat{D}_{nj}(x)| = o_p\{(nh^d)^{-2}\log^4(n)\},$$

which means that

$$nh^d\int\sum_{j=1}^{3}\hat{D}_{nj}(x)\pi(x)dx = o_p\{(nh^d)^{-1}\log^4(n)\} = o_p(h^{d/2}).$$

It remains to work on

$$\begin{aligned}\hat{D}_{n0}(x) &= (nh^d)^{-1}\sum_{i=1}^{n}\{\lambda^T(x)\hat{Q}_i^{(1)}(x,\theta_0)\}^3\\ &= (nh^d)^{-1}\sum_{i=1}^{n}\{h^d\hat{Q}_i^{(1)T}(x,\theta_0)V^{-1}(x)\hat{\bar{Q}}^{(1)}(x,\theta_0)\}^3\{1+o_p(1)\}.\end{aligned}$$

Without loss of generality assume $h_1 = \ldots = h_k = h$. Recall that

$$V^{-1}(x) = f^{-1}(x)\left(\gamma_{lj}(x)\right)_{k\times k} \qquad (6.22)$$

and let

$$\begin{aligned}\phi_{i_1,\ldots,i_4}(x) &= \sum_{l_1,\ldots,l_6=1}^{k}K^3\left(\frac{x-X_{i_1}}{h}\right)K\left(\frac{x-X_{i_2}}{h}\right)K\left(\frac{x-X_{i_3}}{h}\right)K\left(\frac{x-X_{i_4}}{h}\right)f^{-3}(x)\\ &\quad\times\gamma_{l_1l_2}(x)\gamma_{l_3l_4}(x)\gamma_{l_5l_6}(x)\tilde{\epsilon}_{i_1l_1}(x)\tilde{\epsilon}_{i_1l_3}(x)\tilde{\epsilon}_{i_1l_5}(x)\tilde{\epsilon}_{i_2l_2}(x)\tilde{\epsilon}_{i_3l_4}(x)\tilde{\epsilon}_{i_4l_6}(x),\end{aligned}$$

where $\tilde{\epsilon}_{il}(x) = Y_{il} - \tilde{m}_l(x,\theta_0,g_0)$. Then,

$$\begin{aligned}\int\hat{D}_{n0}(x)\pi(x)dx &= (nh^d)^{-1}\int\sum_{i=1}^{n}\{h^d\hat{Q}_i^{(1)T}(x,\theta_0)V^{-1}(x)\hat{\bar{Q}}^{(1)}(x,\theta_0)\}^3\pi(x)\,dx\,\{1+o_p(1)\}\\ &= (nh^d)^{-4}\int\sum_{i_1,\ldots,i_4}^{n}\phi_{i_1,\ldots,i_4}(x)\pi(x)\,dx\,\{1+o_p(1)\}\\ &= (nh^d)^{-4}\int\Big\{\sum_{i_1=i_2=i_3=i_4}^{n}+\sum_a+\sum_b+\sum_c\Big\}\phi_{i_1,\ldots,i_4}(x)\pi(x)\,dx\,\{1+o_p(1)\}\\ &= \int\{I_{n1}(x)+I_{n2}(x)+I_{n3}(x)+I_{n4}(x)\}\pi(x)\,dx\,\{1+o_p(1)\},\end{aligned}$$

where $\sum_a$ denotes the sum over all terms for which the set $\{i_1,i_2,i_3,i_4\}$ contains two distinct indices in total, $\sum_b$ for three distinct indices, and $\sum_c$ for all indices different.

By noting that

$$\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\tilde{\epsilon}_{ij}(x) = \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\epsilon_{ij},$$

it can be readily shown that

$$\begin{aligned} E\{\textstyle\int I_{n1}(x)\pi(x)dx\} &= O\{(nh^d)^{-3}\}, \quad E\{\textstyle\int I_{n2}(x\pi(x)dx)\} = O\{(nh^d)^{-2}\}, \\ E\{\textstyle\int I_{n3}(x)\pi(x)dx\} &= 0, \quad\quad\quad\quad\;\; E\{\textstyle\int I_{n4}(x)\pi(x)dx\} = 0 \quad \text{and} \\ \text{Var}\{\textstyle\int I_{n1}(x)\pi(x)dx\} &= O\{(nh^d)^{-7}\}, \quad \text{Var}\{\textstyle\int I_{n2}(x)\pi(x)dx\} = O\{(nh^d)^{-6}\}, \\ \text{Var}\{\textstyle\int I_{n3}(x)\pi(x)dx\} &= O\{(nh^d)^{-5}\}. \end{aligned} \tag{6.23}$$

Therefore, for $j=1,2$ and 3, $\int I_{nj}(x)\pi(x)dx = O_p\{(nh^d)^{-2}\}$. To finish the analysis, we are to derive

$$\begin{aligned} &\text{Var}\{\int I_{n4}(x)\pi(x)dx\} \\ &= (nh^d)^{-8}\int\int \sum_{\{i_1,\ldots,i_4\}\cap\{j_1,\ldots,j_4\}\neq\phi} \text{Cov}\left(\phi_{i_1,\ldots,i_4}(x), \phi_{j_1,\ldots,j_4}(x')\right)\pi(x)\pi(x')dxdx' \\ &= (nh^d)^{-8}\int\int \sum_d \text{Cov}\left(\phi_{i_1,\ldots,i_4}(x), \phi_{j_1,\ldots,j_4}(x')\right)\pi(x)\pi(x')dxdx', \end{aligned}$$

where ϕ is the empty set and $\sum_d$ is the sum over all the cases where there are four distinct pairs formed between a i_l and a j_m. Note that all $\{i_1,\ldots i_4\}$ and $\{j_1,\ldots,j_4\}$ are respectively all different among themselves due to the definition of $I_{n4}(x)$. As ϵ_i has bounded sixth conditional moments, it is readily seen that

$$\text{Var}\{\int I_{n4}(x)\pi(x)dx\} = O\{(nh^d)^{-4}\},$$

which together with (6.23) leads to

$$\int I_{n4}(x)\pi(x)dx = O_p\{(nh^d)^{-2}\}.$$

In summary of these results, we have $R_{n12} = \frac{2}{3}nh^d\int \hat{D}_n(x)\pi(x)dx = O_p\{(nh^d)^{-1}\} = o_p(h^{d/2})$. □

Lemma 6.6 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under* H_0, $R_{n2} = o_p(h^{d/2})$.

Proof. Note that

$$\begin{aligned}
R_{n2} &= 2nh^d \int \hat{\tilde{Q}}^T(x,\theta_0)V^{-1}(x)\{\hat{\tilde{Q}}(x,\hat\theta) - \hat{\tilde{Q}}(x,\theta_0)\}\pi(x)dx \\
&\quad + nh^d \int \{\hat{\tilde{Q}}(x,\hat\theta) - \hat{\tilde{Q}}(x,\theta_0)\}^T V^{-1}(x)\{\hat{\tilde{Q}}(x,\hat\theta) - \hat{\tilde{Q}}(x,\theta_0)\}\pi(x)dx \\
&=: 2R_{n21} + R_{n22}.
\end{aligned}$$

We will only show the case for R_{n21}, as that for R_{n22} is similar if not easier.

Without loss of generality, we assume $h_1 = \ldots = h_k = h$ in order to simplify notation. From a Taylor expansion and (B.2) and (B.3),

$$\begin{aligned}
\tilde{m}(x,\theta_0,\hat{g}) - \tilde{m}(x,\hat\theta,\hat{g}) &= -\frac{\partial \tilde{m}(x,\theta_0,\hat{g})}{\partial\theta^T}(\hat\theta-\theta_0) + o_p(n^{-1/2}) \\
&= -\frac{\partial \tilde{m}(x,\theta_0,g_0)}{\partial\theta^T}(\hat\theta-\theta_0) + o_p(n^{-1/2}),
\end{aligned}$$

uniformly with respect to all x. This leads to

$$\begin{aligned}
\hat{\tilde{Q}}(x,\hat\theta) - \hat{\tilde{Q}}(x,\theta_0) &= n^{-1}\sum_{i=1}^n K_h(x-X_i)\{\tilde{m}(x,\theta_0,\hat{g}) - \tilde{m}(x,\hat\theta,\hat{g})\} \\
&= -n^{-1}\sum_{i=1}^n K_h(x-X_i)\frac{\partial \tilde{m}(x,\theta_0,g_0)}{\partial\theta^T}(\hat\theta-\theta_0)\{1+o_p(1)\} \\
&= -n^{-1}\sum_{i=1}^n K_h(x-X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}(\hat\theta-\theta_0)\{1+o_p(1)\}.
\end{aligned}$$

Hence,

$$R_{n21} = T_n(\hat\theta-\theta_0)\{1+o_p(1)\}, \tag{6.24}$$

where

$$\begin{aligned}
T_n &= -n^{-1}h^{-d}\sum_{i,j}\int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)\{\tilde\epsilon_i(x) - \hat\Delta(x,\theta_0)\}^T \\
&\qquad \times V^{-1}(x)\frac{\partial m(X_j,\theta_0,g_0)}{\partial\theta^T}\pi(x)dx \\
&= n^{-1}\sum_{i,j}K^{(2)}\left(\frac{X_i-X_j}{h}\right)\{\epsilon_i - \hat\Delta(x,\theta_0)\}^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i)\{1+o_p(1)\} \\
&= R(K)n^{-1}\sum_{i=1}^n\{\epsilon_i - \hat\Delta(x,\theta_0)\}^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i)\{1+o_p(1)\} \\
&\quad + n^{-1}\sum_{i\neq j}K^{(2)}\left(\frac{X_i-X_j}{h}\right)\{\epsilon_i - \hat\Delta(x,\theta_0)\}^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i)\{1+o_p(1)\} \\
&=: \{T_{n1}+T_{n2}\}\{1+o_p(1)\},
\end{aligned}$$

where $\tilde{\epsilon}_i(x) = Y_i - \tilde{m}(x, \theta_0, g_0)$,

$$\begin{aligned} T_{n1} &= R(K)n^{-1}\sum_{i=1}^{n}\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i) \\ &\quad -R(K)n^{-1}\sum_{i=1}^{n}\{\hat{\Delta}(x,\theta_0)\}^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i) \\ &=: T_{n11} - T_{n12} \end{aligned}$$

and

$$\begin{aligned} T_{n2} &= n^{-1}\sum_{i\neq j}K^{(2)}\left(\frac{X_i - X_j}{h}\right)\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i) \\ &\quad -n^{-1}\sum_{i\neq j}K^{(2)}\left(\frac{X_i - X_j}{h}\right)\{\hat{\Delta}(x,\theta_0)\}^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i) \\ &=: T_{n21} - T_{n22}. \end{aligned}$$

Condition (A.3), (A.4) and the boundedness of π give that $T_{n11} = O_p(1)$. At the same time (B.1) implies that $T_{n12} = o_p\{(nh^d)^{-1/2}\log(n)\} = o_p(1)$ as $E\|V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\| < C$. Thus,

$$T_{n1} = O_p(1). \tag{6.25}$$

Next, note that $E(T_{n21}) = 0$ and

$$\begin{aligned} &\mathrm{Var}(T_{n21}) \\ &= n^{-2}\sum_{i\neq j}\mathrm{Var}\left\{K^{(2)}\left(\frac{X_i - X_j}{h}\right)\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i)\right\} \\ &+ n^{-2}\sum_{i\neq j_1\neq j_2}\mathrm{Cov}\left\{K^{(2)}\left(\frac{X_i - X_{j_1}}{h}\right)\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i),\right. \\ &\quad \left. K^{(2)}\left(\frac{X_i - X_{j_2}}{h}\right)\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T}\pi(X_i)\right\} \\ &\leq \sup_x|\pi(x)|^2\{h^d K^{(4)}(0) + O(nh^{2d})\}E\{\epsilon_i^T V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial\theta^T} \\ &\quad \times\frac{\partial m^T(X_i,\theta_0,g_0)}{\partial\theta}V^{-1}(X_i)\epsilon_i\}\{1 + O(h^2)\} \\ &= O(nh^{2d}). \end{aligned}$$

Hence,

$$T_{n21} = O_p(n^{1/2}h^d). \tag{6.26}$$

Finally, consider T_{n22}.

$$\|T_{n22}\| \le \sup_x \|\hat{\Delta}(x,\theta_0)\| n^{-1} \sum_{i\neq j} \left| K^{(2)}\left(\frac{X_i - X_j}{h}\right)\right| \|V^{-1}(X_i)\frac{\partial m(X_i,\theta_0,g_0)}{\partial \theta^T}\| \sup_x |\pi(x)|.$$

By repeating the above variance derivation of T_{n21}, we have from (B.1) that

$$T_{n22} = o_p\{(nh^d)^{1/2}\}, \tag{6.27}$$

since by condition (B.4), $E(T_{n22}) = o\{(nh^d)^{1/2}\}$. Combining (6.25), (6.26) and (6.27), we arrive at $T_n = o_p\{(nh^d)^{1/2}\} + O_p(1)$. Substituting this into (6.24), we have $R_{n21} = o_p(h^{d/2})$. □

Lemma 6.7 *Under assumptions (A.1)-(A.6) and (B.1)-(B.5), and under H_0,*

$$\Lambda_n(\vec{h}) = \Lambda_{n1}(\vec{h}) + o_p(h^{d/2}), \tag{6.28}$$

where $\Lambda_{n1}(\vec{h}) = nh^d \int \hat{\tilde{Q}}^{(1)T}(x,\theta_0)V^{-1}(x)\,\hat{\tilde{Q}}^{(1)}(x,\theta_0)\pi(x)dx$.

Proof. Lemma 6.5 and (6.21) lead to

$$\begin{aligned}\Lambda_n(\vec{h}) &= \Lambda_{n1}(\vec{h}) + 2nh^d \int \hat{\tilde{Q}}^{(1)T}(x,\theta_0)V^{-1}(x)\,\hat{\tilde{Q}}^{(2)}(x,\theta_0)\pi(x)dx \\ &\quad + nh^d \int \hat{\tilde{Q}}^{(2)T}(x,\theta_0)V^{-1}(x)\,\hat{\tilde{Q}}^{(2)}(x,\theta_0)\pi(x)dx + o(h^{d/2}).\end{aligned}$$

Applying the same analysis to the term $\hat{D}_{n3}(x)$ in the proof of Lemma 6.5, we have

$$nh^d \int \hat{\tilde{Q}}^{(2)T}(x,\theta_0)V^{-1}(x)\,\hat{\tilde{Q}}^{(2)}(x,\theta_0)\pi(x)dx = o_p\{(nh^d)^{-1}\log^2(n)\} = o_p(h^{d/2}).$$

It remains to check the order of $\Lambda_{n2}(\vec{h}) = nh^d \int \hat{\tilde{Q}}^{(1)T}(x,\theta_0)V^{-1}(x)\,\hat{\tilde{Q}}^{(2)}(x,\theta_0)\pi(x)dx$. Applying the same style of derivation as for $\hat{D}_{n1}(x)$, it can be shown that $\Lambda_{n2}(\vec{h}) = o_p(h^{d/2})$. This finishes the proof. □

Proof of Theorem 3.1. Recalling (6.22),

$$\Lambda_{n1}(\vec{h}) = n^{-1}h^d \sum_{i,j}^{n}\sum_{l,t}^{k} \tilde{\epsilon}_{il}(x)\tilde{\epsilon}_{jt}(x) \int K_{h_l}(x - X_i)K_{h_t}(x - X_j)\gamma_{lt}(x)f^{-1}(x)\pi(x)dx,$$

where $\tilde{\epsilon}_{il}(x) = Y_{il} - \tilde{m}_l(x,\theta_0,g_0)$. Let $K^{(2)}(\beta_l,\beta_t,u) = \beta_t^{-d}\int K(z)K\left(\frac{\beta_l z}{\beta_t} + u\right)dz$, which is a generalization of the standard convolution of K to accommodate different bandwidths and is symmetric with respect to β_l and β_t . By a change of variable and noticing that K is a compact kernel supported on $[-1,1]^d$,

$$\Lambda_{n1}(\vec{h}) = \Lambda_{n11}(\vec{h})\{1 + O_p(h^2)\},$$

where

$$\begin{aligned}
\Lambda_{n11}(\vec{h}) &= n^{-1}\sum_{i,j}^{n}\sum_{l,t}^{k}\epsilon_{il}\epsilon_{jt}K^{(2)}\left(\beta_l,\beta_t,\frac{X_i-X_j}{h_t}\right)\sqrt{\frac{\pi(X_i)\pi(X_j)\gamma_{lt}(X_i)\gamma_{lt}(X_j)}{f(X_i)f(X_j)}} \\
&= n^{-1}\sum_{i\neq j}^{n}\sum_{l,t}^{k}\epsilon_{il}\epsilon_{jt}K^{(2)}\left(\beta_l,\beta_t,\frac{X_i-X_j}{h_t}\right)\sqrt{\frac{\pi(X_i)\pi(X_j)\gamma_{lt}(X_i)\gamma_{lt}(X_j)}{f(X_i)f(X_j)}} \\
&+ n^{-1}\sum_{i=1}^{n}\sum_{l,t}^{k}\epsilon_{il}\epsilon_{it}\beta_t^{-d}R\left(\beta_l/\beta_t\right)\frac{\pi(X_i)\gamma_{lt}(X_i)}{f(X_i)} \\
&=: \Lambda_{n111}(\vec{h}) + \Lambda_{n112}(\vec{h}). \qquad (6.29)
\end{aligned}$$

It is straightforward to show that

$$\Lambda_{n112}(\vec{h}) = k + o_p(h^{d/2}).$$

Thus, it contributes only to the mean of the test statistic. As $\Lambda_{n111}(\vec{h})$ is a degenerate U-statistic with kernel depending on n, straightforward but lengthy calculations lead to

$$h^{-d/2}\Lambda_{n111}(\vec{h}) \xrightarrow{d} N(0,\sigma^2(K,\Sigma))$$

The establishment of the above asymptotic normality can be achieved by either the approach of martingale central limit theorem (Hall and Heyde, 1980) as demonstrated in Hall (1984) or the approach of the generalized quadratic forms (de Jong, 1987) as demonstrated in Härdle and Mammen (1993). Note that $(nh^d)^{-1}\log^4(n) = o(h^{d/2})$. Applying Slutsky's Theorem leads to the result. □

Proof of Theorem 3.2. It can be checked that given the original sample $\chi_n = \{(X_i,Y_i)\}_{i=1}^n$, versions of assumptions (B.1)-(B.5) are true for the bootstrap resample. And hence, Lemmas 6.2-6.7 are valid for the resample given χ_n. In particular, let $\hat{\bar{Q}}^*(x,\hat{\theta})$ be the bootstrap version of $\hat{\bar{Q}}(x,\theta_0)$, let $\hat{V}(x) = \hat{f}(x)\left(\beta_j^{-d}R(\beta_l/\beta_j)\hat{\sigma}_{lj}(x)\right)_{k\times k}$, where $\hat{\sigma}_{lj}(x) = \hat{f}^{-1}(x)n^{-1}\sum_{i=1}^n K_h(x-X_i)\hat{\epsilon}_{il}\hat{\epsilon}_{ij}$, $\hat{f}(x) = n^{-1}\sum_i K_h(x-X_i)$, and let $(\hat{\gamma}_{lj}(x))_{k\times k} = \hat{f}(x)\hat{V}^{-1}(x)$. Then, conditional on χ_n,

$$\ell^*\{\tilde{m}(x,\hat{\theta}^*,\hat{g}^*)\} = nh^d\hat{\bar{Q}}^{*T}(x,\hat{\theta})\hat{V}^{-1}(x)\hat{\bar{Q}}^*(x,\hat{\theta}) + o_p(h^{d/2})$$

and $\Lambda_n^*(\vec{h}) = \Lambda_{n11}^*(\vec{h}) + o_p(h^{d/2})$, where

$$\Lambda_{n11}^*(\vec{h}) \quad = \quad n^{-1} \sum_{i,j}^{n} \sum_{l,t}^{k} \epsilon_{il}^* \epsilon_{jt}^* K^{(2)} \left(\beta_l, \beta_t, \frac{X_i - X_j}{h_t} \right) \sqrt{\frac{\pi(X_i)\pi(X_j)\hat{\gamma}_{lt}(X_i)\hat{\gamma}_{lt}(X_j)}{\hat{f}(X_i)\hat{f}(X_j)}},$$

which are respectively the bootstrap versions of (6.28) and (6.29).

Then apply the central limit theorem for degenerate U-statistics as in the proof of Theorem 3.1, conditional on χ_n,

$$h^{-d/2} \left(\Lambda_{n11}^* - k \right) \xrightarrow{d} N(0, \sigma^2(K, \hat{\Sigma})),$$

where $\sigma^2(K, \hat{\Sigma})$ is $\sigma^2(K, \Sigma)$ with $\Sigma(x)$ replaced by $\hat{\Sigma}(x) = (\hat{\sigma}_{lj}(x))_{k \times k}$. This implies that

$$h^{-d/2} \left(\Lambda_n^* - k \right) \xrightarrow{d} N(0, \sigma^2(K, \hat{\Sigma})). \tag{6.30}$$

Let $\hat{Z} \stackrel{d}{=} N(0, \sigma^2(K, \hat{\Sigma}))$ and $Z \stackrel{d}{=} N(0, \sigma^2(K, \Sigma))$, and $\hat{z}_\alpha$ and z_α be the upper-α quantiles of $N(0, \sigma^2(K, \hat{\Sigma}))$ and $N(0, \sigma^2(K, \Sigma))$ respectively. Recall that $\hat{q}_{n\alpha}$ and $q_{n\alpha}$ are respectively the upper-α quantile of

$$h^{-d/2} \left(\Lambda_n^* - k \right) \quad \text{given } \chi_n \text{ and} \quad h^{-d/2} \left(\Lambda_n - k \right).$$

As (6.30) implies that

$$1 - \alpha = P\left(h^{-d/2} \left(\Lambda_n^* - k \right) < \hat{q}_{n\alpha} | \chi_n \right) = P\left(\hat{Z} < \hat{q}_{n\alpha} \right) + o(1),$$

it follows that $\hat{q}_{n\alpha} = \hat{z}_\alpha + o(1)$ conditionally on χ_n. A similar argument by using Theorem 3.1 leads to $q_{n\alpha} = z_\alpha + o(1)$. As $\hat{\Sigma}(x) \xrightarrow{p} \Sigma(x)$ uniformly in $x \in S$, then $\sigma^2(K, \hat{\Sigma}) \xrightarrow{p} \sigma^2(K, \Sigma)$, and hence $\hat{z}_\alpha = z_\alpha + o(1)$. Therefore, $\hat{q}_{n\alpha} = q_{n\alpha} + o(1)$ and this completes the proof. □

Proof of Theorem 3.3. It can be shown that Lemmas 6.2-6.7 continue to hold true when we work under the local alternative H_{1n}. In particular, (6.28) is still valid. By using a derivation that resembles very much that for obtaining (6.29), we have

$$\Lambda_n(\vec{h}) = \{\Lambda_{n11}(\vec{h}) + \Lambda_{n112}^a(\vec{h}) + \Lambda_{n113}^a(\vec{h})\}\{1 + O_p(h^2)\} + o_p(h^{d/2})$$

where $\Lambda_{n11}(\vec{h})$ is defined in (6.29),

$$\begin{aligned} \Lambda_{n112}^a(\vec{h}) \quad &= \quad n^{-1} h^d c_n \sum_{l,t}^{k} \int \sum_{i,j}^{n} K_{h_l}(x - X_i) K_{h_t}(x - X_j) \gamma_{lt}(x) \pi(x) f^{-1}(x) \\ &\qquad \times \{\epsilon_{jt} \Gamma_{nl}(X_i) + \epsilon_{il} \Gamma_{nt}(X_j)\} dx \end{aligned}$$

and

$$\Lambda^a_{n113}(\vec{h}) = n^{-1}h^d c_n^2 \sum_{l,t}^{k} \int \sum_{i,j}^{n} K_{h_l}(x-X_i)K_{h_t}(x-X_j)\gamma_{lt}(x)\Gamma_{nl}(X_i)\Gamma_{nt}(X_j)\pi(x)f^{-1}(x)dx.$$

It can be shown that $E\{\Lambda^a_{n112}(\vec{h})\} = 0$ and that

$$\begin{aligned} E\{\Lambda^a_{n113}(\vec{h})\} &= (n-1)h^d c_n^2 \int \Gamma_n^T(x)V^{-1}(x)\Gamma_n(x)f^2(x)\pi(x)dx \\ &+ c_n^2\beta_l^{-d} \int \sum_{l,t}^{k} R(\beta_l/\beta_t)\Gamma_{nl}(x)\gamma_{lt}(x)\Gamma_{nt}(x)\pi(x)dx\{1+O(h^2)\} \\ &= nh^d c_n^2 \int \Gamma_n^T(x)V^{-1}(x)\Gamma_n(x)f^2(x)\pi(x)dx + O(c_n^2 + nh^{d+2}c_n^2) \\ &= h^{d/2}\beta(f,K,\Sigma,\Gamma) + O(c_n^2 + nh^{d+2}c_n^2) + o(h^{d/2}). \end{aligned} \tag{6.31}$$

It is fairly easy to see that $\Lambda^a_{n112}(\vec{h}) = o_p(h^{d/2})$ and

$$\Lambda^a_{n113}(\vec{h}) = h^{d/2}\beta(f,K,\Sigma,\Gamma) + o_p(h^{d/2}).$$

From Lemma 6.7,

$$h^{-d/2}[\Lambda_{n11}(\vec{h}) - k\}] \xrightarrow{d} N(0, \sigma^2(K,\Sigma)). \tag{6.32}$$

The theorem now follows after combining these results. ▫

References

Atkinson, A. C. and Bogacka, B. (2002). Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. *Chemometr. Intell. Lab. Syst.*, **61**, 17-33.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications.* Wiley, New York.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

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

Chen, S. X. and Gao, J. (2006). An adaptive empirical likelihood test for parametric time series regression models. *J. Econometrics* (under revision).

Chen, S. X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. R. Statist. Soc. - Series B*, **65**, 663–678.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms, *Probab. Theory Rel. Fields*, **75**, 261–277.

Delgado, M. A., González Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.*, **29**, 1469–1507.

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

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection. *Ann. Statist.*, **20**, 1412–1425.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.*, **85**, 387–392.

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

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, **29**, 153–193.

Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.*, **32**, 1858–1907.

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

Gijbels, I. (2005). Monotone regression. In Kotz, S., Johnson, N.L., Read, C. B., Balakrishnan, N. and Vidakovic B. (Eds), *Encyclopedia of Statistical Sciences*, New York: Wiley, pp 4951-4968.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multiv. Anal.*, **14**, 1–16.

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

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157–178.

Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models.* Physica-Verlag, Heidelberg.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.

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

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

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Infer.*, **68**, 295–321.

Hjort, N. L., McKeague, I. W. and Van Keilegom, I. (2005). Extending the scope of empirical likelihood. *Ann. Statist.* (under revision).

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

Jacquez, J. A. (1996). *Compartmental Analysis in Biology and Medicine.* BioMedware, Ann Arbor.

Khuri, A. I. (2001). An overview of the use of generalized linear models in response surface methodology. *Nonlinear Anal.*, **47**, 2023–2034.

Li, G. (2003). Nonparametric likelihood ratio goodness-of-fit tests for survival data. *J. Multiv. Anal.*, **86**, 166–182.

Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in non-parametric regression with censored data. *Scand. J. Statist.*, **29**, 547–562.

Linton, O., Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**, 93–100.

Mammen, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.*, **19**, 724–740.

Mammen, E., Linton, O. B. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.*, **27**, 1443–1490.

Opsomer, J.-D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.*, **25**, 186–211.

Owen, A. (1988). Empirical likelihood confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. (2001). *Empirical Likelihood*. Chapman and Hall, New York.

Qin, J., and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Ann. Statist.*, **22**, 300–325.

Rodríguez-Póo, J. M., Sperlich, S. and Vieu, P. (2005). An adaptive specification test for semiparametric models (submitted).

Seber, G. A. and Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

Stute, W. and Zhu, L.-X. (2005). Nonparametric checks for single-index models. *Ann. Statist.*, **33**, 1048–1083.

Tripathi, G. and Kitamura, Y. (2003). Testing conditional moment restrictions. *Ann. Statist.*, **31**, 2059–2095.

Uciñski and Bogacka (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *J. R. Statist. Soc. - Series B*, **67**, 3-18.

Whang, Y. and Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models. *J. Econometrics*, **57**, 277-318.

Xia, Y., Li W. K., Tong, H. and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statist. Sinica*, **14**, 1–39.

Yatchew, A. (1992). Nonparametric regression tests based on least squares. *Econometric Theory*, **8**, 435–451.