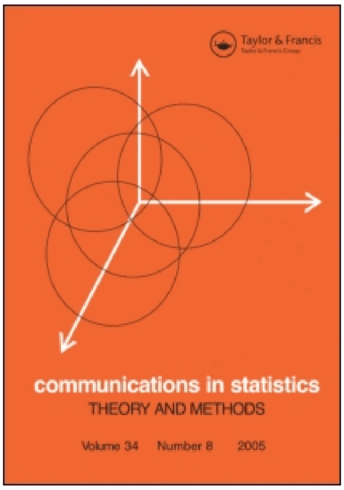

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Adaptive Nonparametric Comparison of Regression Curves

Changliang Zou[a]; Yukun Liu[a]; Zhaojun Wang[a]; Runchu Zhang[a]
[a] LPMC and School of Mathematical Sciences, Nankai University, Tianjin, China

Online publication date: 22 March 2010

To cite this Article Zou, Changliang , Liu, Yukun , Wang, Zhaojun and Zhang, Runchu(2010) 'Adaptive Nonparametric Comparison of Regression Curves', Communications in Statistics - Theory and Methods, 39: 7, 1299 — 1320
To link to this Article: DOI: 10.1080/03610920902878847
URL: http://dx.doi.org/10.1080/03610920902878847

Communications in Statistics—Theory and Methods, 39: 1299–1320, 2010

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920902878847

Adaptive Nonparametric Comparison of Regression Curves

CHANGLIANG ZOU, YUKUN LIU, ZHAOJUN WANG, AND RUNCHU ZHANG

LPMC and School of Mathematical Sciences, Nankai University, Tianjin, China

We propose a new test for comparison of two regression curves, which integrates generalized likelihood ratio (GLR) statistics (Fan et al., 2001) with the data-driven criterion of selecting the smoothing parameter proposed by Guerre and Lavergne (2005). The local linear nonparametric estimator is used to construct the GLR statistic. We prove that the corresponding test statistic is asymptotically normal and free of nuisance parameters and covariate designs under the null hypothesis. The test adapts to the unknown smoothness of the difference between two regression functions and can detect local alternatives converging to the null hypothesis at rate $(\ln\ln n/n)^{-\frac{4}{9}}$. *The wild bootstrap technique is used to approximate the critical values of the test for small samples. A simulation study is conducted to investigate the finite sample properties of the new adaptive test and to compare it with some other available procedures in the literature. The simulation results demonstrate the sensitivity and robustness of the proposed approach.*

Keywords Comparison of two regression curves; Data-driven criterion; Generalized likelihood ratio; Local linear smoother; Wild bootstrap.

Mathematics Subject Classification Primary 62G07; Secondary 62G09.

1. Introduction

The comparison of two (or more) regression curves is a widely discussed issue. In many cases of practical interest, we have a sample of n_i observations in the form $\{(y_{ij}, x_{ij}), j = 1, \ldots, n_i\}$, $i = 1, \ldots, k$ with

$$y_{ij} = f_i(x_{ij}) + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \ \ i = 1, \ldots, k,$$

where x_{ij}'s are independent covariate variables with positive density Γ_i having a common support Ω, and for $i = 1, \ldots, k$, ε_{ij} $(j = 1, \ldots, n_i)$ are independent random errors with mean zero and variance $\sigma_i(x_{ij})$. We are interested in testing the

Received January 14, 2009; Accepted March 9, 2009
Address correspondence to Changliang Zou, LPMC and School of Mathematical Sciences, Nankai University, Tianjin 300071, China; E-mail: chlzou@yahoo.com.cn

hypothesis

$$H_0 : f_1 = \cdots = f_k \quad \text{versus} \quad H_1 : f_i \neq f_j \quad \text{for some } i, j \in \{1, \ldots, k\}. \tag{1}$$

Many efforts have been devoted to this problem in a nonparametric setting in the literature. Earlier work focused on common designs (equal sample sizes and design points), such as Delgado (1993), Hall and Hart (1990), King et al. (1991), Young and Bowman (1995), etc. Some tests were proposed in Hall et al. (1997), Kulasekera (1995), and Kulasekera and Wang (1997) for the hypothesis (1) which applied under the assumption of unequal designs. Two tests based on weighted L^2 distance and marked empirical processes, respectively, were proposed in Dette and Munk (1998) and Neumeyer and Dette (2003). Dette and Neumeyer (2001) proved asymptotic normality of three different test statistics under the null hypothesis and local and fixed alternatives in the case of unequal designs and heteroscedasticity. A good and recent review on this topic can be found in Neumeyer and Dette (2003).

In this article, we propose a new test procedure based on the generalized likelihood ratio (GLR) statistics. It was demonstrated in Fan et al. (2001) that a class of the GLR statistics based on some appropriate nonparametric estimators are asymptotically of distribution free and have χ^2-distributions under null hypotheses for a variety of useful models. Our fundamental test statistic is derived by means of GLR for two-samples nonparametric regression model problem. Taking into account different variances of individual curves, we appropriately modify the GLR procedure and obtain the resulting test statistic with asymptotically normal distribution (χ^2-distribution in generalized sense) and free of nuisance parameters (variances) and covariate designs under the null hypothesis. This test method depends upon the smoothing technique and hence the power and level are sensitive to the choice of bandwidth. To overcome this shortcoming, we combine this testing method with a data-driven criterion of selecting the smoothing parameter proposed by Guerre and Lavergne (2005) recently. Guerre and Lavergne's (2005) method differs from most often-used adaptive rate-optimal lack-of-fit tests (specification tests) in the spirit of the maximum approach (see Fan, 1993; Hart, 1997; Horowitz and Spokoiny, 2001), and relies on a specific criterion tailored for testing statistics. By integrating GLR test and Guerre and Lavergne's (2005) criterion of selecting smoothing parameter, we finally obtain a test statistic which is still asymptotically normal and free of nuisance parameters and covariate designs under the null hypothesis. We further study the large-sample properties of the proposed testing procedure under alternatives and theoretically address the issue why such an adaptive test is useful for our considered hypothesis (1). A simulation study demonstrates the sensitivity and robustness of the proposed approach for the small sample sizes, and shows the superior power properties over a competing test in a variety of cases.

In the next section, we start by describing our testing methodology in the case of comparing two curves, and then state the main asymptotic results and give some discussions on the links and difference between the proposed procedure and other available approaches in the literature. In Sec. 3, we investigate the finite sample properties of the proposed approach and perform a comparison with some alternative procedures. A real example from semiconductor manufacturing is used to demonstrate the method. The proofs are given in the Appendix.

2. Methodology

2.1. *Testing Procedure*

To ease the exposition and facilitate the technical arguments, we elaborate on introducing test method for comparing two regression functions in the case of homoscedasticity within each curve. Suppose we have a sample of $n = n_1 + n_2$ observations in the form $\{(y_{1j}, x_{1j}), j = 1, \ldots, n_1\}$ and $\{(y_{2j}, x_{2j}), j = 1, \ldots, n_2\}$ with

$$y_{ij} = f_i(x_{ij}) + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \;\; i = 1, 2.$$

It is assumed that within each curve the errors are identically distributed with mean zero and σ_i^2, but the distributions of $\sigma_1^{-1}\varepsilon_{1j}$ and $\sigma_2^{-1}\varepsilon_{2j}$ may be different. We are interested in testing the hypothesis

$$H_0 : f_1 = f_2 \quad \text{versus} \quad H_1 : f_1 \neq f_2. \tag{2}$$

To motivate and derive the testing statistic, similar to Fan et al. (2001), we suppose that both the ε_{1j} and ε_{2j} come from normal distribution firstly (only to motivate our proposed method and is not necessary in asymptotic theory and practical use; see Secs. 2.2 and 3). According to this assumption on the errors, the logarithm of the likelihood function is given by

$$\sum_{i=1}^{2}\left[-n_i \ln\left(\sqrt{2\pi}\sigma_i\right) - \frac{1}{2\sigma_i^2}\sum_{j=1}^{n_i}\left(y_{ij} - f_i(x_{ij})\right)^2\right]. \tag{3}$$

To develop model specification test, Fan et al. (2001) proposed replacing the unknown functions under nonparametric alternative by some reasonable nonparametric estimators. Here, we need to use nonparametric smoothing estimators under both the null hypothesis and alternative since hypothesis (2) is fully nonparametric.

Under H_1, following Fan et al. (2001), we can construct a local linear estimators for each curve using the corresponding sample. Say for $i = 1, 2$,

$$\hat{f}_{i,h}(x) = \sum_{j=1}^{n_i} W_{i,j}(x) y_{ij}, \tag{4}$$

where

$$W_{i,j}(x) = U_{i,j}(x) \Big/ \sum_{j=1}^{n_i} U_{i,j}(x),$$

$$U_{i,j}(x) = K_h(x_{ij} - x)\left[m_{i,2}(x) - (x_{ij} - x)m_{i,1}(x)\right],$$

$$m_{i,l}(x) = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - x)^l K_h(x_{ij} - x), \quad l = 1, 2,$$

and $K_h(\cdot) = K(\cdot/h)/h$ with K being a symmetric probability density function and h a bandwidth.

Under H_0, taking into account the different variances in the two curves and utilizing the pooled samples, an ideal local linear estimator is given by

$$\hat{f}_{0,h}(x) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} W_{0,j}^{(i)}(x) y_{ij}, \tag{5}$$

where

$$W_{0,j}^{(i)}(x) = U_{0,j}^{(i)}(x) \Big/ \sum_{i=1}^{2}\sum_{j=1}^{n_i} U_{0,j}^{(i)}(x),$$

$$U_{0,j}^{(i)}(x) = \frac{1}{\sigma_i^2} K_h(x_{ij}-x)\left[m_{0,2}(x) - (x_{ij}-x)m_{0,1}(x)\right],$$

$$m_{0,l}(x) = \frac{1}{n}\sum_{i=1}^{2}\frac{1}{\sigma_i^2}\sum_{j=1}^{n_i}(x_{ij}-x)^l K_h(x_{ij}-x).$$

To complete the esimtator (5), we replace the two unknown parameters σ_1^2 and σ_2^2 by the following simple but consistent nonparametric estimators (Hall and Marron, 1990)

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \hat{f}_{i,h}(x_{ij}))^2 \quad i = 1, 2, \tag{6}$$

where $\hat{f}_{i,h}(x)$ are defined in (4).

By plugging in the nonparametric estimators of the curves and maximizing the likelihood over the parameters σ_i^2's, we get the generalized log-likelihood functions under H_0 and H_1, respectively:

$$l_0(h) = -\sum_{i=1}^{2}[(n_i/2)\ln(2\pi) + (n_i/2)\ln(\mathrm{RSS}_{0,i}/n_i) + n_i/2],$$

$$l_1(h) = -\sum_{i=1}^{2}[(n_i/2)\ln(2\pi) + (n_i/2)\ln(\mathrm{RSS}_i/n_i) + n_i/2],$$

where $\mathrm{RSS}_{0,i} = \sum_{j=1}^{n_i}(y_{ij} - \hat{f}_{0,h}(x_{ij}))^2$ and $\mathrm{RSS}_i = \sum_{j=1}^{n_i}(y_{ij} - \hat{f}_{i,h}(x_{ij}))^2$. Now, the generalized likelihood ratio statistic is

$$T_h = -2(l_0(h) - l_1(h)) = \sum_{i=1}^{2} n_i[\ln \mathrm{RSS}_{0,i} - \ln \mathrm{RSS}_i]. \tag{7}$$

Obviously, a large T_h leads to rejection of the null hypothesis.

Remark 2.1. By Taylor Expansion, it is easy to verify that

$$T_h \approx \sum_{i=1}^{2}\left[\mathrm{RSS}_{0,i} - \mathrm{RSS}_i\right]/\sigma_i^2.$$

Thus, from the viewpoint of asymptotics, T_h has a similar form to the test statistic based on the difference of variance estimators (Dette and Neumeyer, 2001). It was demonstrated that the latter statistic has excellent finite sample properties and is often remarkably more powerful than other tests in the literature. The test statistic T_h has two distinctive features. On the one hand, the asymptotical distribution of T_n is free of nuisance parameters σ_i^2's and covariate designs under H_0, as will be shown in Proposition 2.1. On the other hand, since we use the local linear smoother, a term of order nh^4 involving the covariate designs in asymptotic expansions of T_h under H_0 will vanish. The virtue of these features are two-fold. Theoretically speaking, T_h not only allows us to conveniently integrate the method of selecting h (at least from the asymptotic viewpoint; see Guerre and Lavergne, 2005) because its asymptotic null-distribution is free of nuisance parameters, but also yields relatively simple asymptotic quantities under H_1 so that it can be straightforward to analysis the power function and verify the effectiveness of the proposed adaptive data-based procedure. Our simulation indicates that when the error distributions of the two curves are different, T_h has more robust and sensitive performance than Dette and Neumeyer's (2001) test statistic. See Secs. 2.2 and 3 for detailed discussions.

Now, we turn to combining the GLR test statistic (7) with Guerre and Lavergne's (2005) method of selecting the smoothing parameter h. Consider a set of admissible smoothing parameters $\mathscr{H}_n$ as the following geometric grid:

$$\mathscr{H}_n = \{h_j = h_{\max} a^{-j} : h_j \geq h_{\min}, j = 0, \ldots, J_n\}, \tag{8}$$

where $0 < h_{\min} < h_{\max}$, and $a > 1$. In this case, $J_n \leq \log_a(h_{\max}/h_{\min})$.

Following Guerre and Lavergne (2005), we select h as

$$\tilde{h} = \arg\max_{h\in\mathscr{H}_n}\{(T_h - \mu_h) - (T_{h_0} - \mu_{h_0}) - \gamma_n v_{h,h_0}\},$$

where $\gamma_n > 0$ is a chosen penalty parameter, μ_h is the mean of T_h, and v_{h,h_0}^2 the variance of $T_h - T_{h_0}$ conditionally on the covariates of the pooled sample. The proposed testing statistic is

$$\widetilde{T} = (T_{\tilde{h}} - \mu_{\tilde{h}})/v_{h_0}, \tag{9}$$

where v_h^2 is the variance of T_h conditionally on the covariates of the pooled sample.

Remark 2.2. Here, we make several remarks on choosing γ_n, a, $h_{\max}$, and $h_{\min}$ in (9). Theoretically speaking, these quantities should satisfy certain conditions to obtain the corresponding asymptotic results; see Sec. 2.2 for detailed discussion. Based on simulations, we observe that performance of the proposed approach is hardly affected by these parameters, which is consistent with the findings in Guerre and Lavergne (2005). By both theoretical arguments and numerical studies, we recommend using the choices that $1 < a < 2$, $\gamma_n = 2.5\sqrt{\ln(J_n + 1)}$, J_n could be 4, 5, or 6, $h_{\max} = M(\max\{n_1, n_2\})^{-1/5}$, and $h_j = h_{\max}\gamma_n^{-j}$, for $j = 1, \ldots, J_n$, where $0.5 \leq M \leq 2$ is a constant.

Remark 2.3. In the definition of T_h, we assumed the equality of all bandwidths, which substantially simplifies the asymptotic results and their proofs. Actually, in practice it seems to be more reasonable to choose different bandwidths for $\hat{f}_{1,h}(x)$, $\hat{f}_{2,h}(x)$, and $\hat{f}_{0,h}(x)$, respectively, according to the sample size of the corresponding sample. This is a topic of ongoing research.

Remark 2.4. Note that Kulasekera and Wang (1997) proposed a method of selecting smoothing parameters by maximizing the empirical power to obtain the optimal power in tests of regression curves. Their method is somewhat related to Neyman-type smooth tests for lack-of-fit. However, how to utilize their method to select h for the GLR statistic T_h remains challenge because it is not easy to estimate the power function of T_h accurately; see Theorem 2.2 in Sec. 2.2. Compared with the maximizing method, the Guerre and Lavergne's (2005) selecting criterion has some nice features. First, the criterion favors a baseline statistic under H_0 which results in an asymptotic normal distribution of $\widetilde{T}$, the same as that for $(T_{h_0} - \mu_{h_0})/v_{h_0}$. Hence, this data-based smoothing parameters will not inflate the size of test; see Proposition 2.1 and Theorem 2.1 below. Second, this selection procedure allows us to use v_{h_0} in $\widetilde{T}$, which yields the increase in power at no cost of size of test from an asymptotic point.

To complete the testing procedure, it's necessary to evaluate the quantities μ_h, v_h^2, and v_{h,h_0}^2. However, these quantities are too complicated to be obtained. Moreover, even we know their explicit forms (functions of design points and unknown parameters), we still need to estimate them when we apply the above testing procedure. So we suggest using the following data-driven method to evaluate these quantities. Let

$$\mathbf{W}_k = \left(W_{k,j}(x_{ki})\right)_{n_k \times n_k}, \quad \mathbf{W}_0^{(k,l)} = \left(\frac{\hat{\sigma}_l}{\hat{\sigma}_k} W_{0,j}^{(l)}(x_{ki})\right)_{n_k \times n_l}, \quad k, l = 1, 2,$$

$$\mathbf{W}_h^{H_0} = \begin{pmatrix} \mathbf{W}_0^{(1,1)} & \mathbf{W}_0^{(1,2)} \\ \mathbf{W}_0^{(2,1)} & \mathbf{W}_0^{(2,2)} \end{pmatrix}, \quad \mathbf{W}_h^{H_1} = \begin{pmatrix} \mathbf{W}_1 & 0 \\ 0 & \mathbf{W}_2 \end{pmatrix},$$

$$\mathbf{V}_h = \mathbf{W}_h^{H_1} + (\mathbf{W}_h^{H_1})' - (\mathbf{W}_h^{H_1})'\mathbf{W}_h^{H_1} - \mathbf{W}_h^{H_0} - (\mathbf{W}_h^{H_0})' + (\mathbf{W}_h^{H_0})'\mathbf{W}_h^{H_0},$$

where $W_{k,j}(x_{ki})$ are defined in (4). Then, the proposed consistent data-driven estimators of μ_h, v_h^2, and v_{h,h_0}^2 are given by

$$\hat{\mu}_h = \sum_{i=1}^{n} V_h^{(ii)}, \quad \hat{v}_h^2 = 2\sum_{i=1}^{n}\sum_{j=1}^{n}\left[V_h^{(ij)}\right]^2,$$

$$\hat{v}_{h,h_0}^2 = 2\sum_{i=1}^{n}\sum_{j=1}^{n}\left[V_h^{(ij)} - V_{h_0}^{(ij)}\right]^2,$$

where $V_h^{(ij)}$ denotes the (i, j) element of the matrix $\mathbf{V}_h$. The validity of these estimators will be stated in Sec. 2.2. Our limited simulations demonstrate that for small or moderate samples, the test statistic $\widetilde{T}$ based on the above estimators performs better than that based on the corresponding asymptotic quantities.

Formally, the proposed test is

$$\text{Reject } H_0 \quad \text{if } \widehat{T} \geq z_\alpha, \tag{10}$$

where $\widehat{T}$ is $\widetilde{T}$ with μ_h, v_h^2, and v_{h,h_0}^2 replaced by their estimators $\hat{\mu}_h$, $\hat{v}_h^2$, and $\hat{v}_{h,h_0}^2$, respectively, and z_α is the $(1-\alpha)$ quantile of the standard normal distribution.

Remark 2.5. In the case of comparing m regression functions, one can conveniently mimic the foregoing procedure of two curves and derive the corresponding test statistics. All the results shown in Sec. 2 can be generalized to the multiple curves case.

Remark 2.6. The preceding testing methodology can also be extended to the case of heteroscedasticity within each curve. One can modify the local linear smoother (5) by using a neighboring variance estimator to replace the global ones (6). For instance, the estimator given in Sec. 2.5 of Horowitz and Spokoiny (2001) or the estimator (3.6) in Guerre and Lavergne (2005) can be used.

2.2. *Main Results*

In this section, we study the asymptotic behavior of the proposed test. To be clear, a set of conditions for the results stated later are presented.

Conditions

(C1) The density functions Γ_1 and Γ_2 are bounded away from 0 and have bounded derivatives. Their common support Ω is bounded and compact.

(C2) $f_1(\cdot)$ and $f_2(\cdot)$ have s-order Lipschitz continuous second derivatives for some real $s > 0$.

(C3) The function $K(t)$ is symmetric and bounded. Further, the functions $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^{2i}K(t)dt < \infty$, for $i = 1, 2, \ldots$.

(C4) $E(|\varepsilon_{11}|^4) < \infty$ and $E(|\varepsilon_{21}|^4) < \infty$.

(C5) The bandwidth h satisfies that $h \to 0$, $nh^3 \to \infty$ and $nh^{\frac{9}{2}+\underline{s}} \to 0$, where $\underline{s} = \min\{s, 5/2\}$.

(C6) The penalty sequence γ_n is of order $\sqrt{2\ln\ln n}$.

Remark 2.7. Condition C1 implies that the density functions are positive, which ensures that the denominators of $\hat{f}_{1,h}(x)$, $\hat{f}_{2,h}(x)$, and $\hat{f}_{0,h}(x)$ are, with high probability, bounded away from 0. For the fixed design case, we can impose the assumption given in Dette and Neumeyer (2001) to replace Condition C1 and the arguments in the proof of theorems still hold. This condition is not the weakest possible. In fact, for Proposition 2.1 and Theorem 2.1, we only require Γ_1 and Γ_2 are Lipschitz continuous. Conditions C2 and C3 are commonly used smoothness condition. Condition C4 is necessary in asymptotic theory. Note that in this condition, the normality of the errors is not needed. Considering the range of h given by Condition C5, we restrict the set $\mathscr{H}_n$ of bandwidths as defined in (8) with

$$h_{\max} = O(n^{-\frac{2}{9}-s_1}) \quad h_{\min} = O(n^{-\frac{1}{3}+s_2}),$$

where s_1 and s_2 are two positive constants so that $h_{\max}$ and $h_{\min}$ satisfy Condition C5. Condition C6 assures γ_n diverges fast enough in order to obtain the asymptotic null distribution of $\widehat{T}$. Obviously, the conditions imposed here are mild.

Next, we begin by studying the asymptotic behavior of the GLR test statistic T_h under H_0.

Proposition 2.1. *Suppose Conditions* C1–C5 *hold. Then, under* H_0,

$$(T_h - \check{\mu}_h)/\check{\sigma}_h \xrightarrow{\mathscr{L}} N(0, 1),$$

where

$$\check{\mu}_h = \frac{2|\Omega|}{h}\left(K(0) - \frac{1}{2}\int K^2(t)dt\right) \quad \check{\sigma}_h^2 = \frac{8|\Omega|}{h}\int\left(K(t) - \frac{1}{2}\int K * K(t)\right)^2 dt.$$

The asymptotic null distribution of T_h is independent of the nuisance parameters σ_i^2 and the densities Γ_i. Furthermore, using a scale constant $r_K = \frac{1}{2}\frac{K(0)-\frac{1}{2}\int K^2(t)dt}{\int\left(K(t)-\frac{1}{2}\int K*K(t)\right)^2 dt}$, we can see $r_K T_h \overset{a}{\sim} \chi^2_{r_K \mu_h}$, where $\overset{a}{\sim}$ means approximation in a generalized sense (see Fan et al., 2001). Hence, this proposition can be seen as a generalization of Wilks phenomenon unveiled by Fan et al. (2001) to the two-samples case.

Remark 2.8. Note that, in Fan et al. (2001), the asymptotic conditional mean contains three additional terms: the asymptotically normal variables R_{n10} and R_{n20}, and a constant R_{n30} which is related to the second derivatives of the regression functions. The term R_{n30} vanishes in Proposition 2.1 because under H_0, the expansion of T_h yields two asymptotically equal R_{n30}-type bias terms for $l_0(h)$ and $l_1(h)$. This is a special feature in using GLR test for comparing the curves. We also prove that $(R_{n10} - R_{n20})$-type terms of T_h is asymptotically negligible under a slight modification of smoothness in Fan et al. (2001). See Lemma A.3 in the Appendix. The similar arguments can be applied for this term in Fan et al. (2001). Hence, as a by-product, we argue that in Theorem 5 of Fan et al. (2001), the assumption imposed on the bandwidth can be relaxed to some extent. To be specific, we only require $nh^{\frac{9}{2}+s} \to 0$ compared with $nh^{\frac{9}{2}} \to 0$ in Fan et al. (2001). This requirement guarantees that the asymptotic distribution of the GLR statistic does not depend on any stochastically bounded variables; and it may avoid under-smoothing and include the optimal bandwidth $O(n^{-\frac{1}{5}})$ of nonparametric estimating regression function $f_i(\cdot)$, as long as the smoothness of curves satisfies Condition C2.

The next proposition investigates the asymptotic behavior of T_h under local alternatives. Denote

$$\theta = n_1/n, \quad \delta(x) = f_2(x) - f_1(x), \quad \eta_1 = \int K(t)t^2 dt,$$

$$\eta_2 = 8|\Omega|\int\left(K(t) - \frac{1}{2}\int K * K(t)\right)^2 dt,$$

$$d_1(x) = \theta\Gamma_1(x)/\sigma_1^2, \quad d_2(x) = (1-\theta)\Gamma_2(x)/\sigma_2^2,$$

$$\Gamma_0(u) = d_1(u) + d_2(u), \quad \zeta_\delta = \int \Gamma_0^{-1}(u)d_1(u)d_2(u)\delta^2(u)du,$$

$$B(x) = \frac{\eta_1}{2}\Gamma_0^{-1}(x)[f_1''(x)d_1(x) + f_2''(x)d_2(x)] + \Gamma_0^{-3}(x)$$
$$\cdot [d_1'(x)d_2(x) - d_1(x)d_2'(x)][\delta'(x)\Gamma_0(x) - \delta(x)\Gamma_0'(x)],$$
$$\zeta_1 = \int B^2(u)\Gamma_0(u)du, \quad \zeta_2 = \frac{\eta_1^2}{4}\int [(f_1''(u))^2 d_1(u) + (f_2''(u))^2 d_2(u)]du.$$

Proposition 2.2. *Suppose Conditions* C1–C5 *hold and* $\zeta_1 < M_1$ *and* $\zeta_2 < M_2$ *for some constants* M_1 *and* M_2.

(i) *If* $nh\zeta_\delta \to M$ *for some constant M, then under* H_1,

$$\left[T_h - \check{\mu}_h - n\zeta_\delta + nh^4(\zeta_2 - \zeta_1)\right] / \sqrt{\check{\sigma}_h^2 + 4n\zeta_\delta} \xrightarrow{\mathscr{L}} N(0, 1).$$

(ii) *GLR test* T_h *can detect local alternative with rate* $\zeta_\delta = O_p(n^{-\frac{8}{9}})$ *using* $h = O(n^{-\frac{2}{9}})$, *provided* $\zeta_\delta > h^4(\zeta_2 - \zeta_1)$.

Based on this proposition, we can see that Kulasekera and Wang's (1997) maximizing empirical power function method is rather difficult to be applied for GLR test as many unknown functions need to be estimated. The complication of power function of T_h arises partially from uncommon covariate designs and the use of local linear smoother.

Remark 2.9. From result (i) of Proposition 2.2, we observe that the asymptotic power of the test statistics T_h depends not only on the difference $f_2 - f_1$, but also on the first and second derivatives of regression functions. The term $nh^4(\zeta_2 - \zeta_1)$ intuitively explains why appropriately choosing a smoothing parameter h will gain increase in power of test. In particular, a smaller h is usually more effective in detecting the sharp or oscillating difference in the local area (corresponds to larger $\zeta_2 - \zeta_1$) and a larger h performs better when the difference is flat or smooth (corresponds to smaller $\zeta_2 - \zeta_1$). This motivates the use of adaptive selection method in conducting the GLR test; see Theorem 2.2 and the discussion that follows.

The next proposition establishes the validity of the data-based estimators $\hat{\mu}_h$, $\hat{v}_h^2$, and $\hat{v}_{h,h_0}^2$.

Proposition 2.3. *Suppose Conditions* C1–C5 *hold. Then* $\hat{\mu}_h = \mu_h + o_p(h^{-1/2})$, $\hat{v}_h^2 = v_h^2 + o_p(h^{-1})$, *and* $\hat{v}_{h,h_0}^2 = v_{h,h_0}^2 + o_p(h^{-1})$.

We now present a theorem to establish the null distribution of the proposed adaptive GLR test statistics $\widehat{T}$.

Theorem 2.1. *Suppose Conditions* C1–C6 *hold. If* $\gamma_n > (1 + c)\sqrt{2\ln J_n}$ *for some* $c > 0$, *then*

$$\widehat{T} \xrightarrow{\mathscr{L}} N(0, 1).$$

According to Theorem 2.1, the proposed test has bounded asymptotic critical values which significantly differs from the empirical smoothing parameter selection procedure given in Kulasekera and Wang (1997).

Theorem 2.2 below considers the consistency of $\widehat{T}$ under local alternatives.

Theorem 2.2. *Suppose Conditions* C1–C6 *hold and* $\zeta_1 < M_1$ *and* $\zeta_2 < M_2$ *for some constants* M_1 *and* M_2. *The test* (10) *at least has the asymptotic power*

$$\Phi\left[n^{8/9}\zeta_\delta\left(\frac{\gamma_n\eta_2^{1/2}}{8(\zeta_2-\zeta_1)}\right)^{1/9} - \frac{9}{8}\gamma_n\eta_2^{1/2}\right],$$

and thus is consistent when

$$\zeta_\delta \geq c(\zeta_2-\zeta_1)^{1/9}\left(\frac{\gamma_n\eta_2^{1/2}}{n}\right)^{8/9},$$

provided $c > 0$ *is large enough.*

In the proof of Theorem 2.2, we can find that to attain the above asymptotic power, the order of "optimal" h is

$$\min\left\{h_{\max}, \left[\frac{\gamma_n\eta_2^{1/2}}{n(\zeta_2-\zeta_1)}\right]^{2/9}\right\}. \tag{11}$$

Thus, the data-based test (10) adapts to different magnitudes of $\zeta_2 - \zeta_1$, a function that indicates the smoothness of individual curves themselves and local difference between the two curves. As a consequence, the test (10) has a more robust and sensitive power than the test based on T_h with a pre-chosen bandwidth does. For instance, consider the local alternative of rate $n^{-\frac{4}{9}}$ as in (ii) of Proposition 2.2. The test (10) will provide larger power than T_h with some $h = O_p(n^{-\frac{2}{9}})$ if only $\zeta_2 - \zeta_1$ is of smaller order than $(\ln\ln n)^{-4}$.

2.3. *Wild Bootstrap Implementation of the Test*

Based on Theorem 2.1, z_α is an asymptotically correct α-level critical value under null hypothesis. However, it is well known that in specification testing problem the rate of convergence of the distribution of the test statistic is usually rather slow; see, e.g., Hall and Hart (1990), Zhang (2003), and Fan and Zhang (2004). For this reason, we propose a wild bootstrap (see Wu, 1986) version of test (10) and prove its consistency. Some similar applications of wild bootstrap for present context and lack-of-fit tests can also be found in Hardle and Mammen (1993), Dette and Neumeyer (2001), Neumeyer and Dette (2003), and Guerre and Lavergne (2005).

To be precise, define nonparametric residuals by

$$\hat{e}_{ij} := y_{ij} - \hat{f}_{i,h_b}(x_{ij}), \quad j = 1, \ldots, n_i, \;\; i = 1, 2,$$

where h_b is a pre-specified bandwidth of order $h_{\max}$. Then the bootstrap residuals is given by $e^*_{ij} = \hat{e}_{ij}\omega_{ij}$, where ω_{ij}s are independent random variables generated by an arbitrary distribution so that $E(\omega_{ij}) = 0$, $E(\omega^2_{ij}) = 1$ and $E(\omega^4_{ij}) < \infty$. In this article, we follow the suggestion by Hardle and Mammen (1993) and use a two-point distribution with masses $(\sqrt{5}+1)/2\sqrt{5}$ and $(\sqrt{5}-1)/2\sqrt{5}$ at the points $(1-\sqrt{5})/2$ and $(1+\sqrt{5})/2$, respectively. We obtain the bootstrap sample

$$y_{ij}^* = \hat{f}_{0,h_b}(x_{ij}) + e_{ij}^*.$$

A bootstrap test statistic $\widehat{T}^*$ is built from the bootstrap sample as was the original test statistic in (10). When this procedure is repeated many times, the bootstrap critical value z_α^* is the empirical $1-\alpha$ quantile of the bootstrap test statistics. Finally, the hypothesis of equal regression curves is rejected if $\widehat{T} \geq z_\alpha^*$. The following theorem establishes the consistency of this procedure.

Theorem 2.3. *Under the assumption of Theorem* 2.1, *we have*:

$$\sup_{z\in R} \left| P(\widehat{T}^* \leq z | \{(x_{ij}, y_{ij}), j = 1, \ldots, n_i, i = 1, 2\}) - P(N(0,1) \leq z) \right| \xrightarrow{\mathscr{P}} 0.$$

We will use Monte Carlo simulations to show the effectiveness of this bootstrap version of the proposed test in next section.

3. Small-Sample Performance Assessment

3.1. *Simulation Study of Level*

We first investigate the approximation of the level by the wild bootstrap version of the test. We consider two regression functions, $f_1(x) = f_2(x) = x^2$ and $f_1(x) = f_2(x) = \cos(\pi x)$, and the sample size, $n_i = 25$ or 50. Four scenarios on distributions of errors considered are as follows:

$$\text{(I) } \varepsilon_{1i} \sim N(0,1), \varepsilon_{2i} \sim N(0,1); \text{ (II) } \varepsilon_{1i} \sim N(0,1), \varepsilon_{2i} \sim N(0,0.5);$$
$$\text{(III) } \varepsilon_{1i} \sim N(0,1), \varepsilon_{2i} \sim t(4); \text{ (IV) } \varepsilon_{1i} \sim \exp(1), \varepsilon_{2i} \sim t(4),$$

where $\exp(1)$ and $t(4)$ denote the centered standardized exponential distribution and Student-t distribution with four degrees of freedom, respectively. Both the random design Uniform ($U(0,1)$) and the following fixed designs are considered:

$$x_{1j} = \frac{j-1}{n_1 - 1} \quad \text{and} \quad x_{2j} = \frac{j}{n_2}, \quad j = 1, \ldots, n_i. \tag{12}$$

By the guidelines in Remark 2.2, we use $a = 5/4$, $h_{\max} = a(\max\{n_1, n_2\})^{-1/5}$, $\gamma_n = 2.5\sqrt{\ln(J_n + 1)}$, and J_n is fixed as 4 for the considered sample size. The Epanechnikov kernel is used. For each experiment we run 2,000 replications under null hypothesis and resample bootstrap 1,000 times. Moreover, we use the bootstrap bandwidth $h_b = h_{\max}$. The results for random design and fixed design are summarized, respectively, in Tables 1 and 2 which show the simulated rejection probabilities of the test with level 2.5% and 5% (In each entry, the upper and lower values, respectively). From these two tables, we observe a reasonable approximation of the level by the bootstrap procedure. The levels of the test are insensitive to the distribution of errors.

Table 1
Simulated level of test for random uniform design points

	$f_1(x) = f_2(x) = x^2$				$f_1(x) = f_2(x) = \cos \pi x$			
	(25, 25)	(25, 50)	(50, 25)	(50, 50)	(25, 25)	(25, 50)	(50, 25)	(50, 50)
(I)	2.1	2.4	2.4	2.7	2.1	2.3	2.3	2.5
	4.2	5.5	5.5	5.4	4.6	4.8	4.8	5.6
(II)	2.1	2.3	2.1	2.4	2.0	2.2	2.4	2.7
	3.8	4.2	4.0	4.9	3.1	4.0	4.1	5.1
(III)	2.2	2.4	2.3	2.1	2.5	2.7	2.5	2.3
	4.5	4.9	4.8	4.4	4.8	4.7	5.0	4.5
(IV)	1.8	2.2	1.8	2.5	1.7	2.2	1.7	2.9
	3.9	4.2	3.6	4.6	3.6	4.1	3.4	5.1

3.2. *Simulation Study of Power*

Now, we study the power of the proposed test. Dette and Neumeyer (2001) performed a comparison and demonstrated that the test based on the difference of variance estimators (hereafter we denote it as DN test for brevity) has excellent finite sample properties and is very often remarkably more powerful than several other tests proposed in the literature, such as Delgado (1993), Hall and Hart (1990), Kulasekera (1995), and Kulasekera and Wang (1997). Thus, here we use DN test as a benchmark for comparisons with our proposed adaptive test. Note that in Remark 2.1, we have given some discussions on the connection and difference between DN test and the GLR statistic (6). Here, to evaluate the gain of using the GLR statistic (7), our comparisons also involve the non adaptive GLR test based on the same bandwidth choice as that of DN test used in simulation study of Dette and Neumeyer (2001). That is, use different bandwidths of order $n^{-3/10}$ for the estimators $\hat{f}_1$, $\hat{f}_2$, and $\hat{f}_0$ (see (3.6) and (3.7) in Dette and Neumeyer, 2001).

Because the levels of bootstrap versions of the considered tests are different under each comparison scenario, for the sake of a fair comparison, the critical values of the tests under different comparison setting are obtained by simulations in order

Table 2
Simulated level of test for fixed design points (12)

	$f_1(x) = f_2(x) = x^2$				$f_1(x) = f_2(x) = \cos \pi x$			
	(25, 25)	(25, 50)	(50, 25)	(50, 50)	(25, 25)	(25, 50)	(50, 25)	(50, 50)
(I)	2.6	2.4	2.4	2.6	2.3	2.4	2.4	2.5
	5.0	5.0	5.0	5.1	4.8	5.3	4.3	5.1
(II)	2.8	2.9	2.7	2.8	2.5	2.1	2.9	2.9
	4.7	5.5	5.3	5.1	4.8	4.1	4.9	4.9
(III)	1.7	3.1	2.5	2.0	1.6	2.8	2.5	2.2
	4.0	5.0	5.2	5.1	4.0	5.6	5.4	5.0
(IV)	2.0	2.0	2.2	2.3	1.9	2.0	2.3	2.4
	4.5	4.7	4.5	4.9	4.0	4.6	4.9	5.2

to attain more precise false probability. For simplicity, we only consider nominal level of 5% in each case and fixed design points (12) as in Dette and Neumeyer (2001). The following comparison scenarios are considered:

$$\begin{aligned}
&\text{(I)}\quad f_1(x) = f_2(x) - 1 = \cos(\pi x), && \text{(II)}\quad f_1(x) = f_2(x) - x = \cos(\pi x),\\
&\text{(III)}\ f_1(x) = f_2(x) - \cos(3\pi x) = \cos(\pi x), && \text{(IV)}\ f_1(x) = f_2(x) - \sin(2\pi x) = \exp(x),\\
&\text{(V)}\quad f_1(x) = f_2(x) - \sin(3\pi x) = \exp(x), && \text{(VI)}\ f_1(x) = f_2(x) - 2\sin(4\pi x) = \exp(x).
\end{aligned} \tag{13}$$

These scenarios cover various smoothness of f_1, f_2, and $f_1 - f_2$ which can give a limited but illustrative explanation of the effect of choice of bandwidth. Table 3 shows the simulation results of comparisons of three tests for various sample sizes under the above scenarios with standard normal errors. The symbols "GLR" and "AGLR" denote the non adaptive and adaptive GLR tests, (7) and (10), respectively.

When the difference between two functions is smooth or flat, such as in the cases (I), (II), and (IV), these three tests yield very similar performance. Note that compared with DN test, the GLR test has a very slight disadvantage. This is not surprising to us because as pointed in Remark 2.1, the form of DN test can be derived by using Nadaraya–Watson estimator and the GLR procedure as in Sec. 2.1 under the assumption that the two distributions of errors has the same variance which is just the case considered in Table 3. When $f_1 - f_2$ becomes more oscillating, such as in the cases (III), (V), and (VI), the data-based AGLR test generally provides more robust rejection probabilities than those of DN and GLR tests. Especially in the case (VI), we observe the proposed adaptive test yields a substantial improvement with respect to power for all the sample sizes considered. This demonstrates that the AGLR test can adapt to the unknown smoothness of the difference between two regression functions and can pick up smaller bandwidth to detect more irregular alternatives. Another noteworthy point is that in the case (III),

Table 3
Power comparisons of the tests for various sample sizes with standard normal errors under (13)

		$n_2 = 25$			$n_2 = 50$		
		DN	GLR	AGLR	DN	GLR	AGLR
$n_1 = 25$	(I)	0.850	0.812	0.848	0.935	0.903	0.928
	(II)	0.392	0.339	0.362	0.477	0.413	0.441
	(III)	0.044	0.202	0.281	0.170	0.420	0.401
	(IV)	0.386	0.356	0.358	0.519	0.513	0.502
	(V)	0.183	0.182	0.238	0.356	0.380	0.387
	(VI)	0.062	0.090	0.543	0.254	0.596	0.930
$n_1 = 50$	(I)	0.929	0.900	0.918	0.991	0.987	0.990
	(II)	0.494	0.432	0.482	0.679	0.627	0.646
	(III)	0.060	0.242	0.354	0.306	0.608	0.596
	(IV)	0.489	0.457	0.456	0.709	0.722	0.683
	(V)	0.244	0.239	0.334	0.534	0.553	0.560
	(VI)	0.073	0.096	0.832	0.477	0.792	0.993

DN test is not unbiased for small sample sizes, which can be partially explained by the fact that a large negative bias term in the variance estimators (larger than the magnitude of difference between two functions) vanishes very slowly as sample size n increase due to using the Nadaraya–Watson smoother (see Theorem 2.1 in Dette and Neumeyer, 2001) and Proposition 2.2 in Sec. 2.2). This phenomenon is more remarkable in the next comparison example.

To check whether the foregoing conclusions are affected by the distribution of errors and to study the effect of unequal variances of two errors on the performance of DN test, next we consider ε_{1i} and ε_{2i} are distributed as $N(0, 0.5)$ and $t(4)$, respectively. Table 4 summarizes the simulation results of powers of the considered tests for various sample sizes under alternatives (13). From this table, we can observe a better power for the AGLR test in most of cases. Moreover, compared with DN test, GLR test performs more robust for various functions and sample sizes which is due to taking unequal variances into account.

3.3. *A Real-Data Application*

Here, we apply the proposed AGLR test to a dataset obtained from a deep reactive ion etching (DRIE) process in semiconductor manufacturing industry. In the DRIE process, the desired curve is the one with smooth and straight sidewalls and flat bottoms, and ideally the sidewalls of a trench are perpendicular to the bottom of the trench with certain degree of smoothness around the corners. Various curve shapes, such as positive and negative ones due to underetching and overetching, are considered to be unacceptable. More detailed discussion about the DRIE example can be found in Zou et al. (2007, 2008) and the references cited there. In the quality control of the DRIE process, a critical step is comparing a new curve with some (or one) desired curves (reference sample) to check if the new curve has good shape. Then we would update the reference sample with the new curve without

Table 4
Power comparisons of the tests for various sample sizes with unequal variances under (13)

		$n_2 = 25$			$n_2 = 50$		
		DN	GLR	AGLR	DN	GLR	AGLR
$n_1 = 25$	(I)	0.732	0.700	0.763	0.665	0.877	0.925
	(II)	0.212	0.292	0.335	0.045	0.410	0.485
	(III)	0.017	0.108	0.237	0.040	0.341	0.403
	(IV)	0.234	0.301	0.328	0.159	0.574	0.520
	(V)	0.252	0.142	0.215	0.622	0.466	0.412
	(VI)	0.017	0.064	0.441	0.060	0.382	0.929
$n_1 = 50$	(I)	0.798	0.695	0.743	0.928	0.929	0.973
	(II)	0.353	0.298	0.335	0.268	0.491	0.596
	(III)	0.050	0.121	0.236	0.138	0.334	0.484
	(IV)	0.384	0.298	0.348	0.441	0.605	0.634
	(V)	0.197	0.134	0.258	0.695	0.450	0.480
	(VI)	0.058	0.061	0.681	0.249	0.389	0.964

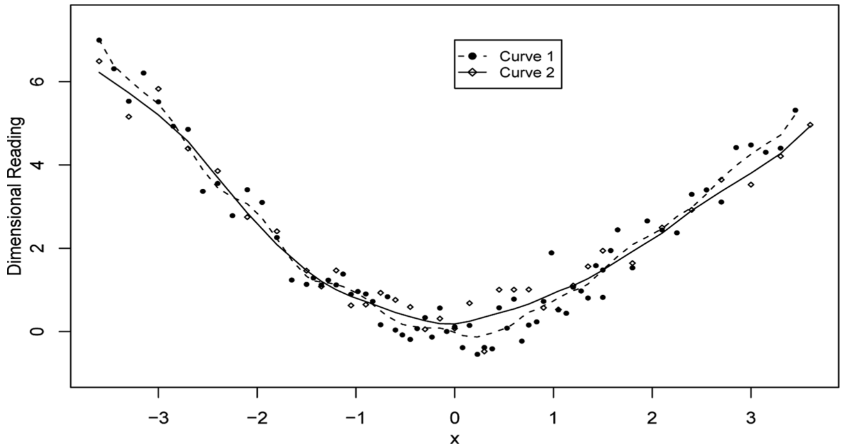

Figure 1. The original observations and local linear fits of two DRIE curves.

significant difference. Thus, it requires carefully examination and powerful testing approach because once an inferior DRIE sample is regarded as a good one, the reference sample will be potentially (highly) affected. To make the curve convenient to describe by a mathematical function, it is rotated by 45° along a reference point in a pre-specified coordinate system. The dimensional readings (the responses) of the profile are then collected by the scanning electron microscope at some given design points (the covariates). Figure 1 shows the observations of two considered curves and the corresponding nonparametric regression curves with bandwidths selected by simple cross-validation.

In this example, one curve has $n_1 = 35$ observations and the other has $n_2 = 70$ observations. We can obtain the estimates of the error variances to be 0.17 and 0.28 for the two curves, respectively, using formula (6). This demonstrates that the proposed approach which takes into account the different variances in the two curves, is likely to be more appropriate and efficient. Following the guidelines in Remark 2.2, we choose $a = 5/4$, $h_{\max} = \frac{5}{4}(\max\{n_1, n_2\})^{-1/5}$, $\gamma_n = 2.5\sqrt{\ln(J_n + 1)}$, and $J_n = 5$. The resulting test statistic $\widehat{T}$ is 8.811. By using wild bootstrap approximation of critical value, we find it shows strong evidence against equality at a 0.002 level. As a comparison, when using the GLR test with the bandwidths less than 0.18, we cannot reject the null hypothesis at 0.05 level.

Appendix

Here, we only present a sketch of proof of theorems. A detailed technical report is available from the authors. Throughout the Appendix, for notation convenience, we need the following additional definitions. Again we omit all indices referring to the bandwidth. For $i = 1, 2$, let

$$\phi_i(x) = \frac{1}{n}\sum_{j=1}^{n_i} K_h(x_{ij} - x)\varepsilon_{ij}/\sigma_i^2,$$

$$\phi_{2+i}(x) = \frac{1}{nh}\sum_{j=1}^{n_i} K_h(x_{ij}-x)(x_{ij}-x)\varepsilon_{ij}/\sigma_i^2.$$

Obviously, ϕ_i $(i = 1, \ldots, 4)$ are all of order $O_p(nh)^{-1/2}$. Moreover, define

$$\alpha_i(x) = \phi_i(x)/d_i(x),$$
$$R_i(x) = \frac{f_i''(x)}{2nd_i(x)}\sum_{j=1}^{n_i} K_h(x_{ij}-x)(x_{ij}-x)^2/\sigma_i^2,$$
$$\alpha_0(x) = \frac{1}{\Gamma_0(x)}[d_1(x)\alpha_1(x) + d_2(x)\alpha_2(x)],$$
$$R_0(x) = \frac{1}{\Gamma_0(x)}[d_1(x)R_1(x) + d_2(x)R_2(x)],$$

and the following vector-matrix notation:

$$\mathbf{D} = \operatorname{diag}\left\{\frac{1}{\Gamma_0(x_{11})}, \ldots, \frac{1}{\Gamma_0(x_{1n_1})}, \frac{1}{\Gamma_0(x_{21})}, \ldots, \frac{1}{\Gamma_0(x_{2n_2})}\right\},$$
$$\mathbf{D}_1 = \operatorname{diag}\left\{\frac{1}{d_1(x_1)}, \ldots, \frac{1}{d_1(x_{n_1})}, \mathbf{0}^\tau_{n_2}\right\}, \quad \mathbf{D}_2 = \operatorname{diag}\left\{\mathbf{0}^\tau_{n_1}, \frac{1}{d_2(x_{21})}, \ldots, \frac{1}{d_2(x_{2n_2})}\right\},$$
$$\mathbf{J}_1 = \frac{1}{\sigma_1^2}\operatorname{diag}\{\mathbf{1}^\tau_{n_1}, \mathbf{0}^\tau_{n_2}\}, \quad \mathbf{J}_2 = \frac{1}{\sigma_2^2}\operatorname{diag}\{\mathbf{0}^\tau_{n_1}, \mathbf{1}^\tau_{n_2}\},$$
$$\boldsymbol{\xi} = \{\varepsilon_{11}, \ldots, \varepsilon_{1n_1}, \varepsilon_{21}, \ldots, \varepsilon_{2n_2}\}^\tau = \{\xi_1, \ldots, \xi_{n_1}, \xi_{n_1+1}, \ldots, \xi_n\}^\tau,$$
$$\mathbf{K}_h = \begin{pmatrix} \left(K_h(x_{1i}-x_{1j})\right)_{n_1\times n_1} & \left(K_h(x_{1i}-x_{2j})\right)_{n_1\times n_2} \\ \left(K_h(x_{2i}-x_{1j})\right)_{n_2\times n_1} & \left(K_h(x_{2i}-x_{2j})\right)_{n_2\times n_2} \end{pmatrix},$$
$$\mathbf{K}_h * \mathbf{K}_h = \begin{pmatrix} \left(K_h * K_h(x_{1i}-x_{1j})\right)_{n_1\times n_1} & \left(K_h * K_h(x_{1i}-x_{2j})\right)_{n_1\times n_2} \\ \left(K_h * K_h(x_{2i}-x_{1j})\right)_{n_2\times n_1} & \left(K_h * K_h(x_{2i}-x_{2j})\right)_{n_2\times n_2} \end{pmatrix}.$$

First, we state the following two necessary lemmas without giving their proofs.

Lemma A.1. *Under the assumption of Proposition* 2.1, *we have*:

$$\frac{1}{n_i}\mathrm{RSS}_i = \hat{\sigma}_i^2 = \sigma_i^2(1 + O_p(n^{-1/2}) + O_p(nh)^{-1}),$$
$$\frac{1}{n_i}\mathrm{RSS}_{0,i} = \sigma_i^2(1 + O_p(n^{-1/2}) + O_p(nh)^{-1}), \quad i = 1, 2.$$

Lemma A.2. *Under the assumption of Proposition* 2.1, *uniformly for* $x \in \Omega$ *and* $k = 0, 1, 2$, *we have*:

$$\hat{f}_{k,h}(x) - f_1(x) = (\alpha_k(x) + R_k(x))\,(1 + o_p(1)).$$

Lemma A.3. *Under the assumption of Proposition* 2.1, *we have*:

$$\sum_{i=1}^{n_k} R_k(x_{ki})\alpha_k(x_{ki}) = \sum_{i=1}^{n_k} R_k(x_{ki})\varepsilon_{ki} + o_p(h^{-1/2}), \quad k = 1, 2,$$

$$\sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} R_0(x_{ij})\alpha_0(x_{ij}) = \sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} R_0(x_{ij})\varepsilon_{ij} + o_p(h^{-1/2}).$$

Proof. We sketch a proof of the first part of the assertion; the others follow by similar arguments. By using the fact that

$$R_1(x) = \frac{h^2}{2} f_1''(x)\eta_1(1 + O(h^s) + O_p(nh)^{-1/2}),$$

we have

$$\sum_{i=1}^{n_1} R_1(x_{1i})\varepsilon_{1i} = \frac{h^2}{2}\eta_1 \sum_{i=1}^{n_1} f_1''(x_{1i})\varepsilon_{1i}(1 + O(h^s) + O_p(nh)^{-1/2})$$

and

$$\begin{aligned}\sum_{i=1}^{n_1} R_1(x_{1i})\alpha_1(x_{1i}) &= \frac{h}{2n_1} \sum_{i=j} K(0)\varepsilon_{1i}\Gamma_1^{-1}(x_{1i}) f_1''(x_{1i})\eta_1(1 + O(h^s) + O_p(nh)^{-1/2}) \\ &\quad + \sum_{i \neq j} \frac{1}{n_1}\varepsilon_{1i}\Gamma_1^{-1}(x_{1j})K_h(x_{1i} - x_{1j}) \\ &\quad \times \frac{h^2}{2} f_1''(x_{1j})\eta_1(1 + O(h^s) + O_p(nh)^{-1/2}) \\ &= O_p(n^{-1/2}h) + \frac{h^2}{2}\eta_1 \sum_{i=1}^{n_1} \varepsilon_{1i} f_1''(x_{1i})(1 + O(h^s) + O_p(nh)^{-1/2}),\end{aligned}$$

which implies that $\sum_{i=1}^{n_1} R_1(x_{1i})\alpha_1(x_{1i}) = \sum_{i=1}^{n_1} R_1(x_{1i})\varepsilon_{1i} + O(h^s) \cdot (n^{\frac{1}{2}} h^2 \frac{1}{\sqrt{n}} \sum_{i=1}^{n_1} \varepsilon_{1i} f_1''(x_{1i}))$. Note that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n_1} \varepsilon_{1i} f_1''(x_{1i})$ is stochastically bounded. Thus, using Condition C5 yields the assertion.

Lemma A.4. *Under the assumption of Proposition* 2.1, *we have*:

$$\sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} R_0^2(x_{ij}) = \sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} R_i^2(x_{ij}) + o_p(h^{-1/2}).$$

The proof is similar to that of Lemma A.3 and hence is omitted here.

Proof of Proposition 2.1. By Lemmas A.1–A.4 and first-order Taylor expansion, we can show

$$\begin{aligned}T_h &= \sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} [2\alpha_i(x_{ij})\varepsilon_{ij} - \alpha_i^2(x_{ij})] - \sum_{i=1}^{2} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \left[2\alpha_0(x_{ij})\varepsilon_{ij} - \alpha_0^2(x_{ij})\right] \\ &\quad + o_p(h^{-1/2}).\end{aligned}$$

Now we rewrite T_h as

$$\begin{aligned}
T_h = \frac{2}{n}\Bigg\{ & \sum_{i=1}^{2}\left(\boldsymbol{\xi}^\tau \mathbf{J}_i\mathbf{D}_i\mathbf{K}_h\mathbf{J}_i\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\tau\mathbf{J}_i\mathbf{D}_i\mathbf{K}_h * \mathbf{K}_h\mathbf{J}_i\boldsymbol{\xi}\right) \\
& - \left[\boldsymbol{\xi}^\tau(\mathbf{J}_1+\mathbf{J}_2)\mathbf{D}\mathbf{K}_h(\mathbf{J}_1+\mathbf{J}_2)\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{\xi}^\tau(\mathbf{J}_1+\mathbf{J}_2)\mathbf{D}\mathbf{K}_h * \mathbf{K}_h(\mathbf{J}_1+\mathbf{J}_2)\boldsymbol{\xi}\right]\Bigg\} \\
& + o_p(h^{-1/2}) \\
\equiv 2h^{-1/2} & \left[\sum_{1\le i\le j\le n} v_{ij}\xi_i\xi_j + o_p(1)\right] \equiv 2h^{-1/2}[\boldsymbol{\xi}^\tau \boldsymbol{v}_h\boldsymbol{\xi}^\tau + o_p(1)], \qquad \text{(A.1)}
\end{aligned}$$

where we have used the fact that

$$\frac{1}{n^2}\boldsymbol{\xi}^\tau\mathbf{J}_i\mathbf{K}_h\mathbf{D}_i\mathbf{J}_i\mathbf{D}_i\mathbf{K}_h\mathbf{J}_i\boldsymbol{\xi} = \frac{1}{n}\boldsymbol{\xi}^\tau\mathbf{J}_i\mathbf{D}_i\mathbf{K}_h * \mathbf{K}_h\mathbf{J}_i\boldsymbol{\xi}(1+o_p(1)),$$

and $\boldsymbol{v}_h = (v_{ij})$ is defined as follows:

$$\begin{aligned}
v_{ij} &= \frac{h^{1/2}}{n\sigma_1^4}\left(\frac{d_2(x_{1i})}{\Gamma_0(x_{1i})d_1(x_{1i})} + \frac{d_2(x_{1j})}{\Gamma_0(x_{1j})d_1(x_{1j})}\right) \\
&\quad \times \left(K_h(x_{1i}-x_{1j}) - \frac{1}{2}K_h * K_h(x_{1i}-x_{1j})\right), \quad 1\le i<j\le n_1; \\
v_{ij} &= \frac{h^{1/2}}{n\sigma_1^4}\frac{d_2(x_{1i})}{\Gamma_0(x_{1i})d_1(x_{1i})}\left(K_h(0) - \frac{1}{2}K_h * K_h(0)\right), \quad 1\le i\le n_1; \\
v_{ii} &= \frac{h^{1/2}}{n\sigma_2^4}\left(\frac{d_1(x_{2k})}{\Gamma_0(x_{2k})d_2(x_{2k})} + \frac{d_1(x_{2l})}{\Gamma_0(x_{2l})d_2(x_{2l})}\right) \\
&\quad \times \left(K_h(x_{2k}-x_{2l}) - \frac{1}{2}K_h * K_h(x_{2k}-x_{2l})\right), \\
&\quad n_1 < i < j \le n \quad \text{with } k = i-n_1 \text{ and } l = j-n_1; \\
v_{jj} &= \frac{h^{1/2}}{n\sigma_2^4}\frac{d_1(x_{2l})}{\Gamma_0(x_{2l})d_2(x_{2l})}\left(K_h(0) - \frac{1}{2}K_h * K_h(0)\right), \quad n_1 < j \le n; \\
v_{ij} &= -\frac{h^{1/2}}{n\sigma_1^2\sigma_2^2}\left(\frac{1}{\Gamma_0(x_{1i})} + \frac{1}{\Gamma_0(x_{2l})}\right)\left(K_h(x_{1i}-x_{2l}) - \frac{1}{2}K_h * K_h(x_{1i}-x_{2l})\right), \\
&\quad 1\le i\le n_1 < j\le n.
\end{aligned}$$

Note that $\mathrm{Var}(2h^{-1/2}\sum_{i=1}^n v_{ii}\xi_i^2) = o_p(h^{-1})$, and

$$\begin{aligned}
E\left(2h^{-1/2}\sum_{i=1}^n v_{ii}\xi_i^2\right) &= 2h^{-1/2}\Bigg\{\sigma_1^2\sum_{i=1}^{n_1}\frac{h^{1/2}}{n\sigma_1^4}\frac{d_2(x_{1i})}{\Gamma_0(x_{1i})d_1(x_{1i})}\left(K_h(0) - \frac{1}{2}K_h * K_h(0)\right) \\
&\quad + \sigma_2^2\sum_{i=1}^{n_2}\frac{h^{1/2}}{n\sigma_2^4}\frac{d_1(x_{2i})}{\Gamma_0(x_{2i})d_2(x_{2i})}\left(K_h(0) - \frac{1}{2}K_h * K_h(0)\right)\Bigg\} + o_p(h^{-1/2}) \\
&= \frac{2|\Omega|}{h}\left(K(0) - \frac{1}{2}\int K^2(t)dt\right) + o_p(h^{-1/2}).
\end{aligned}$$

As a consequence, we have:

$$2h^{-1/2}\sum_{i=1}^{n} v_{ii}\xi_i = \check{\mu}_h + o_p(h^{-1/2}).$$

Now, it remains to show the asymptotic normality of $W_h = 2h^{-\frac{1}{2}}\sum_{1\le i<j\le n} v_{ij}\xi_i\xi_j$. Since it can be written as a symmetric quadratic form with vanishing diagonal elements, we can apply Theorem 5.2 in De Jong (1987). Obviously, the expectation of W_h is zero, while

$$\begin{aligned}
\mathrm{Var}\Bigg(\sum_{1\le i<j\le n} v_{ij}\xi_i\xi_j\Bigg) &= E\Bigg(\mathrm{Var}\Bigg(\sum_{1\le i<j\le n} v_{ij}\xi_i\xi_j \,|\, x_{11},\ldots,x_{1n_1},x_{21},\ldots,x_{2n_2}\Bigg)\Bigg)\\
&= E\Bigg(\sum_{1\le i<j\le n_1} v_{ij}^2\sigma_1^4 + \sum_{1\le i\le n_1<j\le n} v_{ij}^2\sigma_1^2\sigma_2^2 + \sum_{n_1<i<j\le n} v_{ij}^2\sigma_2^4\Bigg)\\
&= \frac{1}{2}\sum_{i=1}^{2}\iint g_i^2(x,y)d_i(x)d_i(y)G^2\left(\frac{x-y}{h}\right)\frac{1}{h}dx\,dy\\
&\quad + \iint g_3^2(x,y)d_1(x)d_2(y)G^2\left(\frac{x-y}{h}\right)\frac{1}{h}dx\,dy + O_p(n^{-1})\\
&= 2\int\left(\frac{d_2^2(y)}{\Gamma_0^2(y)} + \frac{d_1^2(y)}{\Gamma_0^2(y)} + \frac{2d_1(y)d_2(y)}{\Gamma_0^2(y)}\right)dy\int G^2(x)dx + o(1)\\
&= 2|\Omega|\int G^2(x)dx + o(1),
\end{aligned}$$

where we define

$$\begin{aligned}
g_1(x,y) &= \frac{d_2(x)}{\Gamma_0(x)d_1(x)} + \frac{d_2(y)}{\Gamma_0(y)d_1(y)}, \quad g_2(x,y) = \frac{d_1(x)}{\Gamma_0(x)d_2(x)} + \frac{d_1(y)}{\Gamma_0(y)d_2(y)},\\
g_3(x,y) &= \frac{1}{\Gamma_0(x)} + \frac{1}{\Gamma_0(y)}, \quad G(x) = K(x) - \frac{1}{2}K*K(x).
\end{aligned}$$

Thus, the asymptotic variance of T_h is $\check{\sigma}^2$. Finally, by tedious but straightforward algebra, we can verify that v_{ij}s satisfy all the conditions given in Theorem 2 of De Jong (1987) and thus we complete the proof.

Proof of Proposition A.2. The proof of this proposition is analogous to that of Proposition 2.1. Here we only highlight the differences between them. By some algebra, we can obtain

$$\begin{aligned}
\hat{f}_{0,h} &= \Gamma_0^{-1}(\phi_1+\phi_2) + \Gamma_0^{-1}(f_1d_1+f_2d_2) + \frac{\eta_1}{2}h^2\Gamma_0^{-1}(f_1''d_1+f_2''d_2)\\
&\quad + h^2\Gamma_0^{-3}(d_1'd_2 - d_1d_2')[\delta'\Gamma_0 - \delta\Gamma_0'] + o(h^2).
\end{aligned}$$

Therefore, under H_1,

$$\begin{aligned}
\hat{f}_{0,h}(x) - f_1(x) &= \Gamma_0^{-1}(x)d_2(x)\delta(x) + \Gamma_0^{-1}[\phi_1(x)+\phi_2(x)] + h^2B(x) + o(h^2);\\
\hat{f}_{0,h}(x) - f_2(x) &= -\Gamma_0^{-1}(x)d_1(x)\delta(x) + \Gamma_0^{-1}[\phi_1(x)+\phi_2(x)] + h^2B(x) + o(h^2).
\end{aligned}$$

Then substituting these expressions into T_h and by lengthy algebra, we have:

$$T_h = 2h^{-1/2} \sum_{1\le i\le j\le n} v_{ij}\xi_i\xi_j + 2\sum_{i=1}^{2}\sum_{j=1}^{n_i}(-1)^i\Gamma_0^{-1}(x_{ij})\delta(x_{ij})d_{3-i}(x_{ij})\varepsilon_{ij}/\sigma_i^2$$
$$+ n\zeta_\delta + nh^4\zeta_1 - nh^4\zeta_2 + o_p(h^{-1/2}),$$

where $v_{ij}\xi_i\xi_j$'s are given by (A.1). Using Proposition 2.1 and calculating the expectation and variance of the second term in the latter equation, we complete the proof.

Proof of Proposition A.3. Using similar arguments to those in the proof of Proposition 2.1, we can show that

$$T_h = \boldsymbol{\xi}_1^\tau \mathbf{V}_h \boldsymbol{\xi}_1 + o_p(h^{-1/2}), \quad T_h - T_{h_0} = \boldsymbol{\xi}_1^\tau(\mathbf{V}_h - \mathbf{V}_{h_0})\boldsymbol{\xi}_1 + o_p(h^{-1/2} - h_0^{-1/2}),$$

where $\boldsymbol{\xi}_1 = (\sigma_1^{-1}\varepsilon_{11}, \ldots, \sigma_1^{-1}\varepsilon_{1n_1}, \sigma_2^{-1}\varepsilon_{21}, \ldots, \sigma_2^{-1}\varepsilon_{2n_2})^\tau$. Hence, conditionally on the pool sample, a direct calculation yields the assertions.

Denote $T_h^c = T_h - \hat{\mu}_h$ which is the asymptotic centered GLR statistic. Moreover, define $\boldsymbol{v}_h^c$ as a $n \times n$ matrix with zero diagonal elements and the rest elements v_{ij}^c equal to v_{ij}. The following lemma is helpful in proving Theorem 2.1.

Lemma A.5. *Under the assumption of Theorem* 2.1, *for all* $h \in \mathscr{H}_n\backslash\{h_0\}$, *we have*:

(i) $\hat{v}_{h,h_0} = O_p(h^{-1} - h_0^{-1})^{1/2}$;

(ii) $\frac{\xi^\tau(\boldsymbol{v}_h^c - \boldsymbol{v}_{h_0}^c)\xi}{v_{h,h_0}} \xrightarrow{\mathscr{L}} N(0,1)$;

(iii) $\max_{h\in\mathscr{H}_n\backslash\{h_0\}} \left|\frac{T_h^c - T_{h_0}^c}{\hat{v}_{h,h_0}}\right| = (1 + o_p(1)) \max_{h\in\mathscr{H}_n\backslash\{h_0\}} \left|\frac{\xi^\tau(\boldsymbol{v}_h^c - \boldsymbol{v}_{h_0}^c)\xi}{v_{h,h_0}}\right| + o_p(1)$.

Proof. By (A.1), we have

$$T_h^c - T_{h_0}^c = \boldsymbol{\xi}^\tau(\boldsymbol{v}_h^c - \boldsymbol{v}_{h_0}^c)\boldsymbol{\xi} + o_p(h^{-1/2} - h_0^{-1/2}).$$

Results (i) and (ii) can be shown by analogous arguments in the proof of Proposition 2.1. Result (iii) directly follows (i) and the above equation.

Proof of Theorem 2.1. By using Lemma A.5(i) and (iii),

$$Pr(\tilde{h} \ne h_0) = Pr\left(\max_{h\in\mathscr{H}_n\backslash\{h_0\}} \left|\frac{T_h^c - T_{h_0}^c}{\hat{v}_{h,h_0}}\right| > \gamma_n\right)$$
$$\le \Pr\left(\max_{h\in\mathscr{H}_n\backslash\{h_0\}} \left|\frac{\boldsymbol{\xi}^\tau(\boldsymbol{v}_h^c - \boldsymbol{v}_{h_0}^c)\boldsymbol{\xi}}{v_{h,h_0}}\right| \ge \frac{\gamma_n}{1+c}\right) + o_p(1).$$

It can be easily checked that $\boldsymbol{v}_h^c - \boldsymbol{v}_{h_0}^c$ satisfies the conditions of Lemma A.2 for general symmetric matrix in Guerre and Lavergne (2005). Applying Lemma A.2(ii) of Guerre and Lavergne (2005), Lemma A.5 and Proposition 2.1, the remaining proofs can be completed by using the same arguments as in the proof of Theorem 1 of Guerre and Lavergne (2005). Hence, details are omitted here.

Proof of Theorem 2.2. The proof is based on the following lower bound for the power of the test:

$$P(\widehat{T} \geq v_{h_0} z_\alpha) \geq P(T_h - \hat{\mu}_h \geq v_{h_0} z_\alpha + \gamma_n \hat{v}_{h,h_0}). \tag{A.2}$$

That is, the adaptive test (10) inherits the power properties of each of the T_h up to $\gamma_n \hat{v}_{h,h_0}$; see the discussions in Guerre and Lavergne (2005).

Under a local alternative δ, by Proposition 2.2, collecting the leading terms we have:

$$T_h - \hat{\mu}_h = h^{-\frac{1}{2}} w + n\zeta_\delta(1 + o_p(1)) - nh^4(\zeta_2 - \zeta_1), \tag{A.3}$$

where w is an asymptotically normal random variable with variance η_2. Upon remembering that $\hat{v}_{h,h_0}$ and $\hat{v}_h$ is of order $h^{-\frac{1}{2}}$, we can find an appropriate h in the set of $\mathscr{H}_n$, say h_n, such that

$$n\zeta_\delta - nh^4(\zeta_2 - \zeta_1) - \gamma_n \hat{v}_{h,h_0}$$

attains its maximum value asymptotically. That is, take $h_n = h_0 a^{-j_n}$, where j_n is the integer part of

$$\frac{2}{9 \ln a} \ln\left[\frac{8n(\zeta_2 - \zeta_1)}{\gamma_n \eta_2^{1/2}}\right].$$

Note that h_n is in $\mathscr{H}_n$ when $n^{s_1}(\frac{\zeta_2 - \zeta_1}{\gamma_n})^{\frac{2}{9}}$ is large enough. Then, substituting h_n into (A.3), Theorem 2 now follows from (A.2).

Proof of Theorem 2.3. Denote E^* as the conditional expectation given the total sample $\{(x_{ij}, y_{ij}), j = 1, \ldots, n_i, i = 1, 2\}$. Under H_0, we can easily verify the following moment condition for bootstrap residuals:

$$E^*[\varepsilon^*_{ij}] = 0, \quad E^*[\varepsilon^*_{ij}]^2 = \sigma_i^2(1 + O(h_b^4) + O(nh_b)^{-1} + O(n^{-\frac{1}{2}})), \quad E^*[\varepsilon^*_{ij}]^4 < \infty,$$

where we refer to Lemma 3 in Zhu and Xue (2006) for the second moment condition. Using these conditions, Theorem 2.3 can be established by mimicking the proof of Theorem 2.1.

Acknowledgments

The authors thank the Editor and an anonymous referee for many constructive comments that greatly improved the article. This work was supported by the NSF of Tianjin grant No. 07JCYBJC04300, the NNSF of China grant No. 10571093, 10771107, 10671099, 10711120448, and the SRFDP of China grant No. 20050055038.

References

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.* 17:199–204.

De Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theor. Relat. Field* 75:261–277.

Dette, H., Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* 60:693–708.

Dette, H., Neumeyer, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* 29:1361–1400.

Fan, J. (1993). Local linear regression smoothers and their minima efficiencies. *Ann. Statist.* 21:196–216.

Fan, J., Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.* 5:1858–1907.

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

Guerre, E., Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* 33:840–870.

Hall, P., Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* 85:1039–1049.

Hall, P., Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* 77:415–419.

Hall, P., Huber, C., Speckman, P. L. (1997). Covariate-matched one-sided tests for the difference between functional means. *J. Amer. Statist. Assoc.* 92:1074–1083.

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

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

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

King, E. C., Hart, J. D., Wehrly, T. E. (1991). Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.* 12:239–247.

Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Statist. Assoc.* 90:1085–1093.

Kulasekera, K. B., Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.* 92:500–511.

Neumeyer, N., Dette, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *Ann. Statist.* 31:880–920.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* 14:1261–1350.

Young, S. G., Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* 51:920–931.

Zhang, C. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *J. Amer. Statist. Assoc.* 98:609–628.

Zhu, L., Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *Journal of Royal Statistical Society B* 68:549–570.

Zou, C., Tsung, F., Wang, Z. (2007). Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics* 49:395–408.

Zou, C., Tsung, F., Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics* 50:512–526.