The Annals of Statistics
2001, Vol. 29, No. 1, 153–193

GENERALIZED LIKELIHOOD RATIO STATISTICS AND WILKS PHENOMENON

BY JIANQING FAN,[1] CHUNMING ZHANG AND JIAN ZHANG[2]

Chinese University of Hong Kong and University of California at Los Angeles, University of Wisconsin–Madison, EURANDOM and Chinese Academy of Sciences

Likelihood ratio theory has had tremendous success in parametric inference, due to the fundamental theory of Wilks. Yet, there is no general applicable approach for nonparametric inferences based on function estimation. Maximum likelihood ratio test statistics in general may not exist in nonparametric function estimation setting. Even if they exist, they are hard to find and can not be optimal as shown in this paper. We introduce the generalized likelihood statistics to overcome the drawbacks of nonparametric maximum likelihood ratio statistics. A new Wilks phenomenon is unveiled. We demonstrate that a class of the generalized likelihood statistics based on some appropriate nonparametric estimators are asymptotically distribution free and follow χ^2-distributions under null hypotheses for a number of useful hypotheses and a variety of useful models including Gaussian white noise models, nonparametric regression models, varying coefficient models and generalized varying coefficient models. We further demonstrate that generalized likelihood ratio statistics are asymptotically optimal in the sense that they achieve optimal rates of convergence given by Ingster. They can even be adaptively optimal in the sense of Spokoiny by using a simple choice of adaptive smoothing parameter. Our work indicates that the generalized likelihood ratio statistics are indeed general and powerful for nonparametric testing problems based on function estimation.

1. Introduction.

1.1. *Background.* One of the most celebrated methods in statistics is maximum likelihood ratio tests. They form a useful principle that is generally applicable to most parametric hypothesis testing problems. An important fundamental property that contributes significantly to the success of the maximum likelihood ratio tests is that their asymptotic null distributions are independent of nuisance parameters. This property will be referred to as the "Wilks phenomenon" throughout this paper. A few questions arise naturally about how such a useful principle can be extended to infinite-dimensional problems,

Received July 1999; revised July 2000.

[1]Supported in part by NSF Grant DMS-01-96041 and a grant from University of California at Los Angeles.

[2]Supported in part by the National Natural Science Foundation of China and a grant from the research programme in EURANDOM, Netherlands.

AMS 2000 *subject classifications*. Primary 62G07; secondary 62G10, 62J12.

Key words and phrases. Asymptotic null distribution, Gaussian white noise models, nonparametric test, optimal rates, power function, generalized likelihood, Wilks thoerem.

whether the Wilks type of results continue to hold and whether the resulting procedures possess some optimal properties.

In an effort to extend the scope of the likelihood interference approach Owen (1988) introduced empirical likelihood. The empirical likelihood is applicable to a class of nonparametric functionals. These functionals are usually so smooth that they can be estimated at root-n rate. See also Owen (1990), Hall and Owen (1993), Chen and Qin (1993), Li, Hollander, McKeague and Yang (1996) for applications of the empirical likelihood. A further extension of the empirical likelihood, called the "random-sieve likelihood," can be found in Shen, Shi and Wong (1999). The random-sieve likelihood allows one to deal with the situations where stochastic errors and observable variables are not necessarily one-to-one. Nevertheless, it cannot be directly applied to a nonparametric function estimation setting. Zhang and Gijbels (1999) incorporated the idea of local modeling into the framework of empirical likelihood and proposed an approximate empirical likelihood, called the "sieve empirical likelihood." The sieve empirical likelihood can efficiently handle the estimation of nonparametric functions even with inhomogeneous error.

Nonparametric modeling techniques have developed rapidly due to the availability of modern computing power that permits statisticians to explore possible nonlinear relationships. This raises many important inference questions such as whether a parametric family adequately fits a given data set. Take for instance additive models [Hastie and Tibshrani (1990)]:

$$Y = m_1(X_1) + \cdots + m_p(X_p) + \varepsilon \tag{1.1}$$

or varying coefficient models [Cleveland, Grosse and Shyu (1992)]:

$$Y = a_1(U)X_1 + \cdots + a_p(U)X_p + \varepsilon, \tag{1.2}$$

where U and $X_1, \dots, X_p$ are covariates. After fitting these models, one often asks if certain parametric forms such as linear models fit the data adequately. This amounts to testing if each additive component is linear in the additive model (1.1) or if the coefficient functions in (1.2) are not varying. In both cases, the null hypothesis is parametric while the alternative is nonparametric. The empirical likelihood and random sieve likelihood methods cannot be applied directly to such problems. It also arises naturally if certain variables are significant in the models such as (1.1) and (1.2). This reduces to testing if certain functions in (1.1) or (1.2) are zero or not. For these cases, both null and alternative hypotheses are nonparametric. While these problems arise naturally in nonparametric modeling and appear often in model diagnostics, we do not yet have a generally acceptable method that can tackle such problems.

1.2. *Generalized likelihood ratios.* An intuitive approach to handling the aforementioned testing problems is based on discrepancy measures (such as the L_2 and L_∞ distances) between the estimators under null and alternative models. This is a generalization of the Kolmogorov–Smirnov and the Cramér–von Mises types of statistics. We contend that such a method is not as fundamental as likelihood ratio based tests. First, choices of measures and weights can be arbitrary. Take for example the problem of testing H_0: $m_1(\cdot) = m_2(\cdot) = 0$ in model (1.1). The test statistic based on a discrepancy method is $T = c_1\|\hat{m}_1\| + c_2\|\hat{m}_2\|$. One has not only to choose the norm $\|\cdot\|$ but also to decide the weights c_1 and c_2. Second, the null distribution of the test statistic T is in general unknown and depends critically on the nuisance functions $m_3, \dots, m_p$. This hampers the applicability of the discrepancy based methods.

To motivate the generalized likelihood ratio statistics, let us begin with a simple nonparametric regression model. Suppose that we have n data $\{(X_i, Y_i)\}$ sampled from the nonparametric regression model

$$Y_i = m(X_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.3}$$

where $\{\varepsilon_i\}$ are a sequence of i.i.d. random variables from $N(0, \sigma^2)$ and X_i has a density f with support $[0, 1]$. Suppose that the parameter space is

$$\mathscr{F}_k = \left\{ m \in L^2[0,1] : \int_0^1 m^{(k)}(x)^2\, dx \le C \right\}, \tag{1.4}$$

for a given C. Consider the testing problem

$$H_0\colon m(x) = \alpha_0 + \alpha_1 x \qquad \longleftrightarrow \qquad H_1\colon m(x) \ne \alpha_0 + \alpha_1 x. \tag{1.5}$$

Then, the conditional log-likelihood function is

$$\ell_n(m) = -n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - m(X_i))^2.$$

Let $(\hat{\alpha}_0,\ \hat{\alpha}_1)$ be the maximum likelihood estimator (MLE) under H_0, and $\hat{m}_{\mathrm{MLE}}(\cdot)$ be the MLE under the full model,

$$\min \sum_{i=1}^n (Y_i - m(X_i))^2, \quad \text{subject to } \int_0^1 m^{(k)}(x)^2 dx \le C.$$

The resulting estimator $\hat{m}_{\mathrm{MLE}}$ is a smoothing spline. Define the residual sum of squares RSS_0 and RSS_1 as follows:

$$\mathrm{RSS}_0 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2, \qquad \mathrm{RSS}_1 = \sum_{i=1}^n (Y_i - \hat{m}_{\mathrm{MLE}}(X_i))^2. \tag{1.6}$$

Then it is easy to see that the logarithm of the conditional maximum likelihood ratio statistic for problem (1.5) is given by

$$\lambda_n = \ell_n(\hat{m}_{\mathrm{MLE}}) - \ell_n(H_0) = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}.$$

Interestingly, the maximum likelihood ratio test is not optimal due to its restrictive choice of smoothing parameters. See Section 2.2. It is not technically convenient to manipulate either. In general, MLEs (if they exist) under nonparametric regression models are hard to obtain. To attenuate these difficulties, we replace the maximum likelihood estimator under the alternative nonparametric model by any reasonable nonparametric estimator, leading to the generalized likelihood ratio

$$\lambda_n = \ell_n(H_1) - \ell_n(H_0), \tag{1.7}$$

where $\ell_n(H_1)$ is the log-likelihood with unknown regression function replaced by a reasonable nonparametric regression estimator. A similar idea appears in Severini and Wong (1992) for construction of semi-parametric efficient estimators. Note that we do not require that the nonparametric estimator belong to $\mathscr{F}_k$. This relaxation extends the scope of applications and removes the impractical assumption that the constant C in (1.4) is known. Further, the smoothing parameter can now be selected to optimize the performance of the likelihood ratio test. For ease of presentation, we will call λ_n a "generalized likelihood ratio statistic."

The above generalized likelihood method can readily be applied to other statistical models such as additive models, varying-coefficient models, and any nonparametric regression model with a parametric error distribution. One needs to compute the likelihood function under null and alternative models, using suitable nonparametric estimators. We would expect that the generalized likelihood ratio tests are powerful for many nonparametric problems with proper choice of smoothing parameters. Yet we will only verify the claim based on the local polynomial fitting and some sieve methods, due to their technical tractability.

1.3. *Wilks phenomenon.* We will show in Section 3 that based on the local linear estimators [Fan (1993)], the asymptotic null distribution of the generalized likelihood ratio statistic is nearly χ^2 with large degrees of freedom in the sense that

$$r\lambda_n \overset{a}{\sim} \chi^2_{b_n} \tag{1.8}$$

for a sequence $b_n \to \infty$ and a constant r, namely, $(2b_n)^{-1/2}(r\lambda_n - b_n) \to_{\mathscr{L}} N(0,1)$. The constant r is shown to be near 2 for several cases. The distribution $N(b_n, 2b_n)$ is nearly the same as the χ^2 distribution with degrees of freedom b_n. This is an extension of the Wilks type of phenomenon, by which, we mean that the asymptotic null distribution is independent of the nuisance parameters α_0, α_1 and σ and the nuisance design density function f. With this, the advantages of the classical likelihood ratio tests are fully inherited; one makes a statistical decision by comparing likelihood under two competing classes of models and the critical value can easily be found based on the known null distribution $N(b_n, 2b_n)$ or $\chi^2_{b_n}$. Another important consequence of this result is that one does not have to derive theoretically the constants b_n and r in order to be able to use the generalized likelihood ratio test. As long as the Wilks type of results holds, one can simply simulate the null distributions and hence obtain the constants b_n and r. This is in stark contrast with other types of tests whose asymptotic null distributions depend on nuisance parameters. Another striking phenomenon is that the Wilks type of results holds in the nonparametric setting even though the estimators under alternative models are not MLE. This is not true for parametric likelihood ratio tests.

The above Wilks phenomenon does not hold by coincidence. It is not monopolized by the nonparametric model (1.3). In the exponential family of models with a growing number of parameters, Portnoy (1988) showed that the Wilks type of result continues to hold in the same sense as (1.8). Furthermore, Murphy (1993) demonstrated a similar type of result for the Cox proportional hazards model using a simple sieve method (piecewise constant approximation to a smooth function). We conjecture that it is valid for a large class of nonparametric models, including additive models (1.1). To demonstrate its versatility, we consider the varying-coefficient models (1.2) and the testing problem H_0: $a_1(\cdot) = 0$. Let $\hat{a}_2^0(\cdot), \dots, \hat{a}_p^0(\cdot)$ be nonparametric estimators based on the local linear method under the null hypothesis and let $\ell_n(H_0)$ be the resulting likelihood. Analogously, the generalized likelihood under H_1 can be formed. If one wishes to test if X_1 is significant, the generalized likelihood ratio test statistic is simply given by (1.7). We will show in Section 3 that the asymptotic null distribution is independent of the nuisance parameters and nearly χ^2-distributed. The result is striking because the null hypothesis involves many nuisance functions $a_2(\cdot), \dots, a_p(\cdot)$ and the density of U. This lends further support of the generalized likelihood ratio method.

The above Wilks phenomenon holds also for testing homogeneity of the coefficient functions in model (1.2), namely, for testing if the coefficient functions are really varying. See Section 4.

1.4. *Optimality.* Apart from the nice Wilks phenomenon it inherits, the generalized likelihood method based on some special estimator is asymptotically optimal in the sense that it achieves optimal rates for nonparametric hypothesis testing according to the formulations of Ingster (1993) and Spokoiny (1996). We first develop the theory under the Gaussian white noise model in Section 2. This model admits a simpler structure and hence allows one to develop a deeper theory. Nevertheless, this model is equivalent to the nonparametric regression model shown by Brown and Low (1996) and to the nonparametric density estimation model by Nussbaum (1996). Therefore, our minimax results and their understanding can be translated to nonparametric regression and density estimation settings. We also develop an adaptive version of the generalized likelihood ratio test, called the adaptive Neyman test by Fan (1996) and show that the adaptive Neyman test achieves minimax optimal rates adaptively. Thus, the generalized likelihood method is not only intuitive to use, but also powerful to apply.

The above optimality results can be extended to nonparametric regression and the varying coefficients models. The former is a specific case of the varying coefficient models with $p = 1$ and $X_1 = 1$. Thus, we develop the results under the latter multivariate models in Section 3. We show that under the varying coefficient models, the generalized likelihood method achieves the optimal minimax rate for hypothesis testing. This lends further support to the use of the generalized likelihood method.

1.5. *Related literature.* Recently, there have been many collective efforts on hypothesis testing in nonparametric regression problems. Most of them focus on one-dimensional nonparametric regression models. For an overview and references, see the recent book by Hart (1997).

An early paper on nonparametric hypothesis testing is Bickel and Rosenblatt (1973) where the asymptotic null distributions were derived. Azzalini, Bowman and Härdle (1989) and Azzalini and Bowman (1993) introduced the use of the F-type test statistic for testing parametric models. Bickel and Ritov (1992) proposed a few new nonparametric testing techniques. Härdle and Mammen (1993) studied nonparametric tests based on an L_2-distance. In the Cox's hazard regression model, Murphy (1993) derived a Wilks type of result for a generalized likelihood ratio statistic based on a simple sieve estimator. Various recent testing procedures are motivated by the seminal work of Neyman (1937). Most of them focus on selecting the smoothing parameters of the Neyman test and studying the properties of the resulting procedures. See, for example, Eubank and Hart (1992), Eubank and LaRiccia (1992), Inglot, Kallenberg and Ledwina (1994), Kallenberg and Ledwina (1997), Kuchibhatla and Hart (1996), among others. Fan (1996) proposed simple and powerful methods for constructing tests based on Neyman's truncation and wavelet thresholding. It was shown in Spokoiny (1996) that wavelet thresholding tests are nearly adaptively minimax. The asymptotic optimality of data-driven Neyman's tests was also studied by Inglot and Ledwina (1996).

Hypothesis testing for multivariate regression problems is difficult due to the curse of dimensionality. In bivariate regression, Aerts, Claeskens and Hart (1999) constructed tests based on orthogonal series. Fan and Huang (1998) proposed various testing techniques based on the adaptive Neyman test for various alternative models in a multiple regression setting. These problems become conceptually simple by using our generalized likelihood method.

1.6. *Outline of the paper.* We first develop the generalized likelihood ratio test theory under the Gaussian white noise model in Section 2. While this model is equivalent to a nonparametric regression model, it is not very convenient to translate the null distribution results and estimation procedures to the nonparametric regression model. Thus, we develop in Section 3 the Wilks type of results for the varying-coefficient model (1.2) and the nonparametric regression model (1.3). Local linear estimators are used to construct the generalized likelihood ratio test. We demonstrate the Wilks type of results in Section 4 for model diagnostics. In particular, we show that the Wilks type of results hold for testing homogeneity and for testing significance of variables. We also demonstrate that the generalized likelihood ratio tests are asymptotically optimal in the sense that they achieve optimal rates for nonparametric hypothesis testing. The results are also extended to generalized varying coefficient models in Section 5. The merits of the generalized likelihood method and its various applications are discussed in Section 6. Technical proofs are outlined in Section 7.

2. Maximum likelihood ratio tests in Gaussian white noise model. Suppose that we have observed the process $Y(t)$ from the following Gaussian white noise model

$$dY(t) = \phi(t)\,dt + n^{-1/2}\,dW(t), \qquad t \in (0,1), \tag{2.1}$$

where ϕ is an unknown function and $W(t)$ is the Wiener process. This ideal model is equivalent to models in density estimation and nonparametric regression [Nussbaum (1996) and Brown and Low (1996)] with n being the sample size. The minimax results under model (2.1) can be translated to these models for bounded loss functions.

By using an orthonormal series (e.g., the Fourier series), model (2.1) is equivalent to the following white noise model:

$$Y_i = \theta_i + n^{-1/2}\varepsilon_i, \qquad \varepsilon_i \sim_{\text{i.i.d.}} N(0,1), \qquad i = 1, 2, \ldots \tag{2.2}$$

where Y_i, θ_i and ε_i are the ith Fourier coefficients of $Y(t), \phi(t)$ and $W(t)$, respectively. For simplicity, we consider testing the simple hypothesis,

$$H_0\colon\ \theta_1 = \theta_2 = \cdots = 0, \tag{2.3}$$

namely, testing H_0: $\phi \equiv 0$ under model (2.1).

2.1. *Neyman test.* Consider the class of functions that are so smooth that the energy in high frequency components is zero, namely,

$$\mathscr{F} = \{\theta\colon\ \theta_{m+1} = \theta_{m+2} = \cdots = 0\},$$

for some given m. Then twice the log-likelihood ratio test statistic is

$$T_N = \sum_{i=1}^{m} nY_i^2. \tag{2.4}$$

Under the null hypothesis, this test has a χ^2 distribution with m degrees of freedom. Hence, $T_N \sim AN(m, 2m)$. The Wilks type of results holds trivially for this simple problem even when m tends to ∞. See Portnoy (1988) where he obtained a Wilks type of result for a simple hypothesis of some p_n-dimensional parameter in a regular exponential family with $p_n^{3/2}/n \to 0$.

By tuning the parameter m, the adaptive Neyman test can be regarded as a generalized likelihood ratio test based on the sieve approximation. We will study the power of this test in Section 2.4.

2.2. *Maximum likelihood ratio tests for Sobolev classes.* We now consider the parameter space $\mathscr{F}_k = \{\theta\colon\ \sum_{j=1}^{\infty} j^{2k}\theta_j^2 \le 1\}$ where $k > 1/2$ is a positive constant. By the Parseval identity, when k is a positive integer, this set in the frequency domain is equivalent to the Sobolev class of functions $\{\phi\colon \|\phi^{(k)}\| \le c\}$ for some constant c. For this specific class of parameter spaces, we can derive explicitly the asymptotic null distribution of the maximum likelihood ratio statistic. The asymptotic distribution is not exactly χ^2. Hence, the traditional

Wilks theorem does not hold for infinite-dimensional problems. This is why we need an enlarged view of the Wilks phenomenon.

It can easily be shown that the maximum likelihood estimator under the parameter space $\mathscr{F}_k$ is given by

$$\hat{\theta}_j = (1 + \hat{\xi} j^{2k})^{-1} Y_j,$$

where $\hat{\xi}$ is the Lagrange multiplier, satisfying the equation $\sum_{j=1}^{\infty} j^{2k} \hat{\theta}_j^2 = 1$. The function $F(\xi) = \sum_{j=1}^{\infty} j^{2k} (1 + \xi j^{2k})^{-2} Y_j^2$ is a decreasing function of ξ in $[0, \infty)$, satisfying $F(0) = \infty$ and $F(\infty) = 0$, almost surely. Thus, the solution $F(\hat{\xi}) = 1$ exists and is unique almost surely. The asymptotic expression for $\hat{\xi}$ depends on the unknown θ and is hard to obtain. However, for deriving the asymptotic null distribution of the maximum likelihood ratio test, we need only an explicit asymptotic expression of $\hat{\xi}$ under the null hypothesis (2.3).

LEMMA 2.1. *Under the null hypothesis* (2.3),

$$\hat{\xi} = n^{-2k/(2k+1)} \left\{ \int_0^{\infty} \frac{y^{2k}}{(1 + y^{2k})^2}\, dy \right\}^{2k/(2k+1)} \{1 + o_p(1)\}.$$

The maximum likelihood ratio statistic for the problem (2.3) is given by

$$\lambda_n^* = \frac{n}{2} \sum_{j=1}^{\infty} \left(1 - \frac{j^{4k} \hat{\xi}^2}{(1 + j^{2k} \hat{\xi})^2} \right) Y_j^2. \tag{2.5}$$

In Section 7 we show the following result.

THEOREM 1. *Under the null hypothesis* (2.3), *the normalized maximum likelihood ratio test statistic has an asymptotic* χ^2 *distribution with degrees of freedom* a_n *written as* $r_k \lambda_n^* \sim_a \chi^2_{a_n}$, *where*

$$r_k = \frac{4k + 2}{2k - 1}, \qquad a_n = \frac{(2k+1)^2}{2k - 1} \left[\frac{\pi}{4k^2 \sin(\pi/(2k))} \right]^{2k/(2k+1)} n^{1/(2k+1)}.$$

It is clear from Theorem 1 that the classical Wilks type of results does not hold for infinite-dimensional problems because $r_k \neq 2$. However, an extended version holds: asymptotic null distributions are independent of nuisance parameters and nearly χ^2-distributed. Table 1 gives numerical values for constant r_k and the degrees of freedom a_n. Note that as the degree of smoothness k tends to ∞, $r_k \to 2$.

Surprisingly, the maximum likelihood ratio test can not achieve the optimal rate for hypothesis testing (see Theorem 2 below). This is because the smoothing parameter $\hat{\xi}$ determined by $\sum_{j=1}^{\infty} j^{2k} \hat{\theta}_j^2 = 1$ is too restrictive. This is why we need generalized likelihood ratio tests which allow one the flexibility of choosing smoothing parameters.

TABLE 1
Constants r_k (r'_k in Theorem 3) and degrees of freedom in Theorem 1

k	1	2	3	4	5
r_k	6.0000	3.3333	2.8000	2.5714	2.4444
$a_n, n = 50$	28.2245	6.5381	3.8381	2.8800	2.4012
$a_n, n = 200$	44.8036	8.6270	4.6787	3.3596	2.7237
$a_n, n = 800$	71.1212	11.3834	5.7034	3.9190	3.0895
r'_k	3.6923	2.5600	2.3351	2.2391	2.1858

THEOREM 2. *There exists a $\theta \in \mathscr{F}_k$ satisfying $\|\theta\| = n^{-(k+d)/(2k+1)}$ with $d > 1/8$ such that the power function of the maximum likelihood ratio test at the point θ is bounded by α, namely,*

$$\limsup P_\theta\{r_k \lambda_n^* > a_n + z_\alpha (2a_n)^{1/2}\} \leq \alpha,$$

where z_α is the upper α quantile of the standard normal distribution.

Thus, the maximum likelihood ratio test λ_n^* can detect alternatives with a rate no faster than $n^{-(k+d)/(2k+1)}$. When $k > 1/4$, by taking d sufficiently close to 1/8, the rate $n^{-(k+d)/(2k+1)}$ is slower than the optimal rate $n^{-2k/(4k+1)}$ given in Ingster (1993).

2.3. *Generalized likelihood ratio tests.* As demonstrated in Section 2.2, maximum likelihood ratio tests are not optimal due to restrictive choice of smoothing parameters. Generalized likelihood tests remove this restrictive requirement and allow one to tune the smoothing parameter. For the testing problem (2.3), we take the generalized likelihood ratio test as

$$\lambda_n = \frac{n}{2} \sum_{j=1}^{\infty} \left(1 - \frac{j^{4k} \xi_n^2}{(1 + j^{2k} \xi_n)^2}\right) Y_j^2, \tag{2.6}$$

with $\xi_n = cn^{-4k/(4k+1)}$ for some $c > 0$. This ameliorated procedure achieves the optimal rate of convergence for hypothesis testing, which is stated as follows.

THEOREM 3. *Under the null hypothesis* (2.3), $r'_k \lambda_n \sim_a \chi^2_{a'_n}$, *where*

$$r'_k = \frac{2k+1}{2k-1} \frac{48k^2}{24k^2 + 14k + 1},$$

$$a'_n = \frac{(2k+1)^2}{2k-1} \frac{24k^2 c^{-1/(2k)}}{24k^2 + 14k + 1} \left[\frac{\pi}{4k^2 \sin(\pi/(2k))}\right] n^{2/(4k+1)}.$$

Furthermore, for any sequence $c_n \to \infty$, the power function of the generalized likelihood ratio test is asymptotically one,

$$\inf_{\theta\in\mathscr{F}_k:\ \|\theta\|\ge c_n n^{-2k/(4k+1)}} P_\theta\left\{\frac{r'_k\lambda_n - a'_n}{\sqrt{2a'_n}} > z_\alpha\right\} \to 1.$$

2.4. *Adaptive minimaxity.* The maximum likelihood ratio statistic (2.5) and the generalized likelihood statistic (2.6) depend critically on the value of k. Can we construct an adaptive version that achieves adaptively the optimal rates of convergence? The answer is affirmative and the construction is simple.

Based on power considerations, Fan (1996) proposed the following adaptive version of the generalized likelihood ratio statistic (2.4):

$$T^*_{\mathrm{AN}} = \max_{1\le m\le n} \sum_{i=1}^{m} (nY_i^2 - 1)/\sqrt{2m}. \tag{2.7}$$

He called the testing procedure the "adaptive Neyman test." Note that the adaptive Neyman test is simply the maximum of the normalized likelihood ratio statistic (2.4). It does not depend on the degree of smoothness. Following Fan (1996), we normalize the test statistic as

$$T_{\mathrm{AN}} = \sqrt{2\log\log n}\, T^*_{\mathrm{AN}} - \{2\log\log n + 0.5\log\log\log n - 0.5\log(4\pi)\}.$$

Then, under the null hypothesis (2.3), we have

$$P(T_{\mathrm{AN}} < x) \to \exp(-\exp(-x)) \quad \text{as } n\to\infty.$$

Thus, the critical region

$$T_{\mathrm{AN}} > -\log\{-\log(1-\alpha)\}$$

has asymptotic significance level α. The power of the adaptive Neyman test is given as follows. [A similar version was presented in Fan and Huang (1998).]

THEOREM 4. *The adaptive Neyman test can detect adaptively the alternatives with rates*

$$\delta_n = n^{-2k/(4k+1)}(\log\log n)^{k/(4k+1)}$$

when the parameter space is $\mathscr{F}_k$ with unknown k. More precisely, *for any sequence $c_n \to \infty$, the power function*

$$\inf_{\theta\in\mathscr{F}_k:\ \|\theta\|\ge c_n\delta_n} P_\theta[T_{\mathrm{AN}} > -\log\{-\log(1-\alpha)\}] \to 1.$$

The rate given in Theorem 4 is adaptively optimal in the sense that no testing procedure can detect adaptively the alternative with a rate faster than δ_n, according to Spokoiny (1996). Hence, the generalized likelihood ratio test achieves this adaptive optimality.

REMARK 2.1. By choosing the parameter $m = O(n^{2/(4k+1)})$ when the parameter space is $\mathscr{F}_k$, the Neyman test can also detect alternatives with the optimal rate $O(n^{-2k/(4k+1)})$. This follows from the proof of Theorem 4. By choosing m to maximize (2.7), we obtain an adaptive version of the Neyman test, which is independent of the degree of smoothness k. This test achieves the adaptive optimal rate because the maximum of the partial sum process in (2.7) grows very slowly. This is why we pay only a price of order (log log n) to achieve the adaptive minimax rate.

3. Generalized likelihood ratio tests in varying coefficient models. In this section we develop asymptotic theory for the generalized likelihood ratio statistics which are based on the local polynomial estimators and derive the optimal minimax rates of the corresponding tests under model (1.2). The Wilks phenomenon is unveiled in this general setting.

Suppose $\{(Y_i, \mathbf{X}_i, U_i)\}_{i=1}^n$ is a random sample from the varying-coefficient model (1.2). Namely,

$$Y = A(U)^\tau \mathbf{X} + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2),$$

with $\mathbf{X} = (X_1, \dots, X_p)^\tau$, $U = (U_1, \dots, U_q)^\tau$ and $A(U) = (a_1(U), \dots, a_p(U))^\tau$. For simplicity, we consider only $q = 1$. Extensions to the multidimensional case are similar. Consider the simple null hypothesis testing problem:

$$H_0\colon A = A_0 \qquad \longleftrightarrow \qquad H_1 : A \neq A_0. \tag{3.1}$$

We use the local linear approach to construct a generalized likelihood ratio statistic.

For each given u_0, let $\beta(u_0) = (A(u_0)^\tau, hA'(u_0)^\tau)^\tau$. Let $\beta = (A_*, hB^\tau)^\tau$, where A_* and B are vectors of p-dimensions. Then, the local log likelihood at the given point u_0 is given by

$$l(\beta) = -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta^\tau \mathbf{Z}_i)^2 K_h(U_i - u_0),$$

where $\mathbf{Z}_i = \mathbf{Z}_i(u_0) = (\mathbf{X}_i^\tau, (U_i - u_0)/h\mathbf{X}_i^\tau)^\tau$ and $K_h(\cdot) = K(\cdot/h)/h$ with K being a symmetric probability density function and h a bandwidth. Then, the local maximum likelihood estimator, denoted by $\hat\beta(u_0)$, is defined as argmax $l(\beta)$. The corresponding estimator of $A(u_0)$ is denoted by $\hat A(u_0)$. Using this nonparametric estimator, the likelihood under model (1.2) is

$$-n \log(\sqrt{2\pi}\sigma) - \mathrm{RSS}_1/(2\sigma^2),$$

where $\mathrm{RSS}_1 = \sum_{k=1}^n (Y_k - \hat A(U_k)^\tau \mathbf{X}_k)^2$. Maximizing over the parameter σ^2 leads to the generalized likelihood under model (1.2),

$$\ell_n(H_1) = -(n/2)\log(2\pi/n) - (n/2)\log(\mathrm{RSS}_1) - n/2.$$

Similarly, the maximum likelihood under H_0 can be expressed as

$$\ell_n(H_0) = -(n/2)\log(2\pi/n) - (n/2)\log(\mathrm{RSS}_0) - n/2,$$

where $\mathrm{RSS}_0 = \sum_{k=1}^{n}(Y_k - A_0(U_k)^\tau \mathbf{X}_k)^2$. Now, the generalized likelihood ratio statistic is

$$\lambda_n(A_0) = \left[\ell_n(H_1) - \ell_n(H_0)\right] = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} \tag{3.2}$$

In general, the above approach can often be extended to the composite null hypothesis testing problem,

$$H_0\colon A \in \mathscr{A}_0 \qquad \longleftrightarrow \qquad H_1\colon A \notin \mathscr{A}_0, \tag{3.3}$$

where $\mathscr{A}_0$ is a set of functions. As before, we can use the local linear estimator to construct the log-likelihood $\ell_n(H_1)$ for H_1. Assume that we can use MLE or some local linear estimators to build the log-likelihood $\ell_n(H_0)$. Let A_0' denote the true value of the parameter A. Then the generalized likelihood ratio $\lambda_n(\mathscr{A}_0)$ for the testing problem (3.3) can be decomposed as

$$\lambda_n(\mathscr{A}_0) = \lambda_n(A_0') - \lambda_n^*(A_0'), \tag{3.4}$$

where $\lambda_n(A_0') = \ell_n(H_1) - \ell_n(H_0')$ is the generalized likelihood ratio for the hypothesis testing problem

$$H_0'\colon A = A_0' \qquad \longleftrightarrow \qquad H_1\colon A \neq A_0'$$

and $\lambda_n^*(A_0') = \ell_n(H_0) - \ell_n(H_0')$ is the likelihood ratio for another hypothesis testing problem

$$H_0'\colon A = A_0' \qquad \longleftrightarrow \qquad H_1'\colon A \in \mathscr{A}_0.$$

The above two hypothesis problems are fabricated because A_0' is unknown. Therefore, the generalized likelihood ratio for the composite null hypothesis can be decomposed into two generalized likelihood ratios for two fabricated simple null hypothesis problems. As in the proof of Theorem 5, generally the asymptotic representation of the generalized likelihood ratio for the composite null hypothesis can be derived from those of the above fabricated simple null hypothesis problems. Then the asymptotic theory for composite null hypothesis can be easily obtained [see the proofs of Theorems 6 and 9, Remark 3.4 and the results in Fan and Zhang (1999)]. Thus we focus first on the simple null hypothesis testing problem (3.2). In order to include the above fabricated testing problems, we assume that A_0 is unknown. We should point out that for model (1.2), when A_0 is known, the testing problem (3.2) is equivalent to the problem $H_0\colon A = 0$ by a simple transform.

3.1. *Asymptotic null distribution.* To derive the asymptotic distribution of $\lambda_n(A_0)$ under H_0, we need the following conditions.

CONDITION (A).

(A1) The marginal density $f(u)$ of U is Lipschitz continuous and bounded away from 0. U has a bounded support Ω.
(A2) $A(u)$ has a continuous second derivative.

(A3) The function $K(t)$ is symmetric and bounded. Further, the functions $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^4K(t)\,dt < \infty$.

(A4) $E|\varepsilon|^4 < \infty$.

(A5) $\mathbf{X}$ is bounded. The $p \times p$ matrix $E(\mathbf{X}\mathbf{X}^\tau|U = u)$ is invertible for each $u \in \Omega$. $(E(\mathbf{X}\mathbf{X}^\tau|U = u))^{-1}$ and $E(\mathbf{X}\mathbf{X}^\tau\sigma^2(\mathbf{X}, U)|U = u)$ are both Lipschitz continuous.

These conditions are imposed to facilitate the technical arguments. They are not the weakest possible. In particular, (A5) in Condition (A) can be relaxed by using the method in Lemma 7.4 in Zhang and Gijbels (1999). For example, we can replace the assumption that $\mathbf{X}$ is bounded in (A5) by the assumption that $E\exp(c_0\|\mathbf{X}\|) < \infty$ for some positive constant c_0. The following results continue to hold.

Note that in the above conditions, the normality of ε is not needed. Define

$$\Gamma(u) = E[\mathbf{X}\mathbf{X}^\tau|U = u]f(u), \quad w_0 = \iint t^2(s+t)^2K(t)K(s+t)\,dt\,ds.$$

Let $\varepsilon_i = Y_i - A_0(U)^\tau\mathbf{X}_i$. Set

$$R_{n10} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i A_0''(U_i)^\tau\mathbf{X}_i\int t^2K(t)\,dt(1 + O(h) + O(n^{-1/2})),$$

$$R_{n20} = \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\mathbf{X}_i^\tau\Gamma(U_i)^{-1}A_0''(U_i)^\tau E(\mathbf{X}_i|U_i)w_0,$$

$$R_{n30} = \frac{1}{8}EA_0''(U)^\tau\mathbf{X}\mathbf{X}^\tau A_0''(U)w_0(1 + O(n^{-1/2})),$$

$$\mu_n = \frac{p|\Omega|}{h}\left(K(0) - \frac{1}{2}\int K^2(t)\,dt\right),$$

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

$$d_{1n} = \sigma^{-2}\{nh^4R_{n30} - n^{1/2}h^2(R_{n10} - R_{n20})\} = O_p(nh^4 + n^{1/2}h^2),$$

where $K * K$ denotes the convolution of K. Note that both R_{n10} and R_{n20} are asymptotically normal and hence are stochastically bounded.

We now describe our generalized Wilks type of theorem as follows.

THEOREM 5. *Suppose Condition* (A) *holds. Then, under H_0, as $h \to 0$, $nh^{3/2} \to \infty$,*

$$\sigma_n^{-1}(\lambda_n(A_0) - \mu_n + d_{1n}) \xrightarrow{\mathscr{L}} N(0, 1).$$

Furthermore, if A_0 is linear or $nh^{9/2} \to 0$, then as $nh^{3/2} \to \infty$, $r_K\lambda_n(A_0) \sim_a \chi^2_{r_K\mu_n}$, where

$$r_K = \frac{K(0) - \frac{1}{2}\int K^2(t)\,dt}{\int(K(t) - \frac{1}{2}K * K(t))^2 dt}.$$

REMARK 3.1. As pointed out before, for model (1.2), when A_0 is known, the testing problem (3.2) is equivalent to the problem H_0: $A = 0 \longleftrightarrow H_1$: $A \neq 0$ by a simple transform. Hence, the condition in the second part of the theorem always holds and so does the Wilks phenomenon. Further, when $nh^5 \to 0$, the mean and variance of λ_n is free of nuisance parameters up to the first order because $d_{1n} = o(\mu_n)$. In this relaxed sense, even if A_0 is unknown, the Wilks phenomenon is valid when the condition $nh^{9/2} \to 0$ is relaxed as $nh^5 \to 0$.

REMARK 3.2. The degree of freedom in the asymptotic distribution depends on $p|\Omega|/h$. This can intuitively be understood as follows. If one partitions the support of U into intervals of length h and uses piecewise constant functions to model the functions in A, then we have total number of parameters $p|\Omega|/h$ under model (1.2). In this view, local linear fits can also be regarded as sieve approximation to nonparametric functions with effective number of parameters $r_K \mu_n$.

REMARK 3.3. If local polynomial estimators of degree v instead of the local linear estimators are used to construct the above generalized likelihood ratio, then the result holds when K is replaced by its equivalent kernel induced by the local polynomial fitting [Fan and Gijbels (1996)]. In this case, the second part of Theorem 5 is replaced by the condition that either A_0 is a polynomial of degree v or $nh^{(4v+5)/2} \to 0$.

REMARK 3.4. Suppose Condition (A) holds and the second term in (3.4) is $o_p(h^{-1/2})$ [e.g., in testing a parametric model, under some regularity conditions this term equals $O_p(1)$]. Then it follows directly from Theorem 5 that under the null hypothesis (3.3) the result in Theorem 5 continues to hold.

We now consider the more challenging and more interesting case where null hypotheses depend on many nuisance functions. Nevertheless, we will show that asymptotic null distributions are independent of the nuisance functions. Write

$$A_0(u) = \begin{pmatrix} A_{10}(u) \\ A_{20}(u) \end{pmatrix}, \quad A(u) = \begin{pmatrix} A_1(u) \\ A_2(u) \end{pmatrix}, \quad \mathbf{X}_k = \begin{pmatrix} \mathbf{X}_k^{(1)} \\ \mathbf{X}_k^{(2)} \end{pmatrix}, \quad \mathbf{Z}_k = \begin{pmatrix} \mathbf{Z}_k^{(1)} \\ \mathbf{Z}_k^{(2)} \end{pmatrix},$$

where $A_{10}(u)$, $A_1(u)$, $\mathbf{X}_k^{(1)}$ and $\mathbf{Z}_k^{(1)}$ are $p_1(< p)$-dimensional. Consider the testing problem,

$$H_{0u}: A_1 = A_{10} \qquad \longleftrightarrow \qquad H_{1u}: A_1 \neq A_{10} \tag{3.5}$$

with $A_2(\cdot)$ completely unknown. (3.5) is allowed to be a fictitious testing problem in which the function A_{10} is unknown. Following the same derivations, the logarithm of the generalized likelihood ratio statistic is given by

$$\lambda_{nu}(A_{10}) = \lambda_n(A_0) - \lambda_{n2}(A_{20}|A_{10})$$

with $\lambda_n(A_0)$ the full likelihood ratio defined in (3.2) and

$$\lambda_{n2}(A_{20}|A_{10}) = \frac{n}{2}\log\frac{\text{RSS}_0}{\text{RSS}_2},$$

where

$$\text{RSS}_2 = \sum_{k=1}^{n}\Big(Y_k - A_{10}(U_k)^\tau \mathbf{X}_k^{(1)} - \widetilde{A}_2(U_k)^\tau \mathbf{X}_k^{(2)}\Big)^2.$$

Here $\widetilde{A}_2(U_k)^\tau$ is the local linear estimator at U_k when A_{10} is given.

Recall that $\Gamma(u) = E[\mathbf{X}\mathbf{X}^\tau|U=u]f(u)$. Write

$$\Gamma = \begin{pmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{pmatrix} \quad\text{and}\quad \Gamma_{11,2} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21},$$

where $\Gamma_{11}, \Gamma_{12}, \Gamma_{21}, \Gamma_{22}$ are $p_1\times p_1$, $p_1\times p_2$, $p_2\times p_1$ and $p_2\times p_2$ matrices and $p_2 = p - p_1$. Define μ_{nu} and σ_{nu} the same as μ_n and σ_n except replacing p by p_1. Similarly, define d_{1nu} by replacing $\mathbf{X}$ and Γ, respectively, by $\mathbf{X}^{(1)} - \Gamma_{12}\Gamma_{22}\mathbf{X}^{(2)}$ and $\Gamma_{11,2}$ in the definition of d_{1n}.

THEOREM 6. *Suppose Condition* (A) *holds. Then, under H_{0u} in* (3.5), *as $nh^{3/2}\to\infty$ and $h\to 0$, we have*

$$\sigma_n^{-1}(\lambda_{nu}(A_0) - \mu_{nu} + d_{1nu}) \xrightarrow{\mathscr{L}} N(0,1).$$

In addition, if A_0 is linear or $nh^{9/2}\to 0$, then

$$r_K\lambda_{nu}(A_0) \overset{a}{\sim} \chi^2_{r_K\mu_{nu}}.$$

Theorem 6 provides convincing evidence that the Wilks type of phenomenon holds for generalized likelihood ratio tests with composite hypotheses.

3.2. *Power approximations and minimax rates.* We now consider the power of generalized likelihood ratio tests based on local linear fits. For simplicity of discussion, we focus only on the simple null hypothesis (3.1). As noted in Remark 3.1, one can assume without loss of generality that $A_0 = 0$. But, we do not take this option because we want to examine the impact of biases on generalized likelihood ratio tests. This has implications for the case of a composite hypothesis (3.5) because the biases inherited in that problem are genuine.

When A_0 is linear, the bias term in Theorem 5 will be zero. When A_0 is not linear, we will assume that $h_n = o(n^{-1/5})$ so that the second term in the definition of d_{1n} is of smaller order than σ_n. As will be seen in Theorem 8, the optimal choice of h for the testing problem (3.1) is $h = O(n^{-2/9})$, which satisfies the condition $h = o(n^{-1/5})$. Under these assumptions, if $nh^{3/2}\to\infty$, by Theorem 5, an approximate level α test based on the generalized likelihood ratio statistic is

$$\phi \equiv \phi_h = I\{\lambda_n(A_0) - \mu_n + \hat{v}_n \ge z_\alpha\sigma_n\},$$

where with $\hat{\sigma}^2 = \mathrm{RSS}_1/n$ and

$$\hat{v}_n = \tfrac{1}{8} n h^4 \hat{\sigma}^{-2} E A_0''(U)^\tau \mathbf{X}\mathbf{X}^\tau A_0''(U) \iint t^2 (s+t)^2 K(t) K(s+t)\, dt\, ds.$$

The power of the test under contiguous alternatives of form

$$H_{1n}\colon A(u) = A_0(u) + G_n(u)$$

can be approximated by using the following theorem, where $G_n(u) = (g_{1n}(u), \ldots, g_{pn}(u))^\tau)$ is a vector-valued function.

THEOREM 7. *Suppose that Condition* (A) *holds and that* A_0 *is linear or* $nh^5 \to 0$. *If*

$$nhEG_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U) \to C(G) \quad \text{and}$$
$$E(G_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U)\epsilon^2)^2 \;=\; O((nh)^{-3/2}),$$

for some constant $C(G)$, *then under* H_{1n},

$$(\lambda_n(A_0) - \mu_n + \hat{v}_n + v_{2n} - d_{2n})/\sigma_n^* \xrightarrow{\mathscr{L}} N(0,1),$$

where

$$d_{2n} = \frac{n}{2} EG_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U),$$

$$\sigma_n^* = \sqrt{\sigma_n^2 + n\sigma^{-2} EG_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U)},$$

$$v_{2n} = \frac{nh^4}{8\sigma^2} EG_n''(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n''(U) \iint t^2(s+t)^2 K(t)K(s+t)\, dt\, ds.$$

Theorem 7 can be extended readily to generalized likelihood ratio tests based on local polynomial estimators of degree v and to the case with nuisance parameter functions. It allows functions G_n of forms not only $g_n(u) = (nh)^{-1/2} g(u)$, but also $g_n(u) = a_n^{-2} g(a_n u)$ with $a_n = (nh)^{-1/5}$. The former function has a second derivative tending to zero, which is restrictive in nonparametric applications. The latter function has also a bounded second derivative, which does not always tend to zero, when g is twice differentiable. This is still not the hardest alternative function to be tested. A harder alternative can be constructed as follows. Let $\{u_j\}$ be a grid of points with distance a_n^{-1} apart and g be a twice differentiable function with support $[0,1]$. Then, Theorem 7 also allows functions of form $g_n(u) = a_n^{-2} \sum_j g(a_n(u - u_j))$ with $a_n = (nh)^{-1/4}$.

We now turn to studying the optimal property of the generalized likelihood ratio test. We first consider the class of functions $\mathscr{G}_n$, satisfying the following regularity conditions:

$$\begin{aligned}
&\operatorname{var}(G_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U)) \le M(EG_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U))^2,\\
&nEG_n^\tau(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n(U) > M_n \to \infty,\\
&EG_n''(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n''(U) \le M,
\end{aligned} \tag{3.6}$$

for some constants $M > 0$ and $M_n \to \infty$. For a given $\rho > 0$, let

$$\mathscr{G}_n(\rho) = \{G_n \in \mathscr{G}_n\colon EG_n^\tau(U)\mathbf{X}\mathbf{X}^\tau G_n(U) \geq \rho^2\}.$$

Then the maximum of the probabilities of type II errors is given by

$$\beta(\alpha, \rho) = \sup_{G_n \in \mathscr{G}_n(\rho)} \beta(\alpha, G_n),$$

where $\beta(\alpha, G_n) = P(\phi = 0 | A = A_0 + G_n)$ is the probability of a type II error at the alternative $A = A_0 + G_n$. The minimax rate of ϕ is defined as the smallest ρ_n such that:

1. For every $\rho > \rho_n$, $\alpha > 0$, and for any $\beta > 0$, there exists a constant c such that $\beta(\alpha, c\rho) \leq \beta + o(1)$.
2. for any sequence $\rho_n^* = o(\rho_n)$, there exist $\alpha > 0$, $\beta > 0$ such that for any $c > 0$, $P(\phi = 1 | A = A_0) = \alpha + o(1)$ and $\liminf_n \beta(\alpha, c\rho_n^*) > \beta$.

It measures how close the alternatives are that can be detected by the generalized likelihood ratio test ϕ_h. The rate depends on the bandwidth h. To stress its dependence, we write it as $\rho_n(h)$.

THEOREM 8. *Under Condition* (A), *the generalized likelihood can detect alternatives with rate* $\rho_n(h) = n^{-4/9}$ *when* $h = c_* n^{-2/9}$ *for some constant* c_*.

REMARK 3.5. When $p = 1$ and $\mathbf{X} \equiv 1$, the varying-coefficient model becomes an ordinary nonparametric regression model. In this case, Lepski and Spokoiny (1999) proved the optimal rate for testing H_0 is $n^{-4/9}$. Thus the generalized likelihood ratio test is optimal in the sense that it achieves the optimal rate of convergence. Similarly, we can show that the generalized likelihood ratio test, constructed by using local polynomials of order v, can detect alternatives with rate $n^{-2(v+1)/(4v+5)}$, uniformly in the class of functions satisfying

$$E[G_n^{(v+1)}(U)^\tau \mathbf{X}]^2 < M,$$

for some $M < \infty$. The corresponding optimal bandwidth is $c_* n^{-2/(4v+5)}$ for some constant c_*.

REMARK 3.6. In the proof of Theorem 8, we in fact show that the bandwidth $h = c_* n^{-2/9}$ is optimal, optimizing the rate of $\rho_n(h)$, subject to the following constraints:

(a) $h \to 0$ and $nh^{3/2} \to \infty$, if A_0 is linear.

(b) $nh \to \infty$ and $nh^5 \to 0$, if A_0 is nonlinear with continuous second derivatives.

4. Model diagnostics. In this section, we demonstrate how the generalized likelihood ratio tests can be applied to check the goodness-of-fit for a family of parametric models. These kinds of problems occur very often in practice. Our results apply readily to these kinds of problems. We also note that the Wilks phenomenon continues to hold under general heteroscedastic regression models.

TABLE 2
Values of r_K and c_K in (4.1)

Kernel	**Uniform**	**Epanechnikov**	**Biweight**	**Triweight**	**Gaussian**
r_K	1.2000	2.1153	2.3061	2.3797	2.5375
c_K	0.2500	0.4500	0.5804	0.6858	0.7737

4.1. *Testing linearity.* Consider the nonparametric regression model (1.3) and the testing problem

$$H_0\colon m(x) = \alpha_0 + \alpha_1 x \quad \longleftrightarrow \quad H_1\colon m(x) \neq \alpha_0 + \alpha_1 x,$$

where α_0 and α_1 are unknown parameters. Following the same derivations as in Section 3, generalized likelihood ratio tests based on local linear fits are given by

$$\lambda_n = \ell_n(H_1) - \ell_n(H_0) = \frac{n}{2} \log \frac{\mathrm{RSS}_0}{\mathrm{RSS}_1},$$

where $\mathrm{RSS}_0 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2$ and $\mathrm{RSS}_1 = \sum_{i=1}^n (Y_i - \hat{m}_h(X_i))^2$. By using Remark 3.4, one can easily see that the Wilks type of results holds under the null hypothesis,

$$r_K \lambda_n \overset{a}{\sim} \chi^2_{r_K c_K |\Omega|/h}, \tag{4.1}$$

where Ω denotes the support of X and

$$c_K = K(0) - 2^{-1} \|K\|_2^2.$$

Note that when $K(0) = \max_x K(x)$, we have $K(0) \geq \|K\|_2^2$, $c_K \geq 2^{-1} K(0)$ and hence $r_K > 0$.

To help one determine the degree of freedom in (4.1), the values of r_K and c_K are tabulated in Table 2 for a few commonly used kernels. Among them, the Epanechnikov kernel has the closest r_K to 2.

Two interrelationships concerning the degrees of freedom will be exposed. If we define a "smoothing matrix" H based on local linear estimates just as a projection matrix P in the linear regression model, then under H_0, $\mathrm{RSS}_0 - \mathrm{RSS}_1 = \varepsilon^\tau (H^\tau + H - H^\tau H - P)\varepsilon$. Denoting the bracketed matrix as A, we have tr $(A) \approx 2c_K|\Omega|/h$ following the proof of Theorem 5. Thus, tr (A) is approximately the degrees of freedom only when $r_K \approx 2$. Secondly, note that $K(0) \geq K * K(0) = \|K\|_2^2$ implies approximately $\mathrm{tr}(H^\tau H) \leq \mathrm{tr}(H) \leq 2\,\mathrm{tr}(H) - \mathrm{tr}(H^\tau H)$, a property holding exactly for H based on smoothing splines in fixed designs [Hastie and Tibshirani (1990), Section 3.5].

REMARK 4.1. When one wishes to test parametric families other than the linear model such as H_0: $m(x) = m(x, \theta)$, then one can apply generalized likelihood ratio tests to the residuals $\{Y_i - m(X_i, \hat{\theta})\}$, where $m(X_i, \hat{\theta})$ is a fitted value under the null hypothesis. The Wilks type of result (4.1) continues to hold.

REMARK 4.2. For the more general regression model (1.3), where we assume only $E(\varepsilon|X = x) = 0$ and $E(\varepsilon^2|X = x) = \sigma^2(x)$, one can use the weighted residual sum of squares,

$$\mathrm{RSS}_0 = \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2 w(X_i), \quad \mathrm{RSS}_1 = \sum_{i=1}^{n}(Y_i - \hat{m}_h(X_i))^2 w(X_i).$$

If the weight function $w(\cdot)$ is continuous with a compact support contained in $\{x: f(x) > 0\}$, then we can show that under H_0, a generalized version of (4.1) holds:

$$r'_K \lambda_n \overset{a}{\sim} \chi^2_{a'_n},$$

where

$$r'_K = r_K[E\sigma^2(X)w(X)] \int \sigma^2(x)w(x)\,dx \left[\int \sigma^4(x)w^2(x)\,dx\right]^{-1},$$

$$a'_n = r_K c_K h^{-1}\left[\int \sigma^2(x)w(x)\,dx\right]^2 \left[\int \sigma^4(x)w^2(x)\,dx\right]^{-1}.$$

When $\sigma^2(x) = v(x)\sigma^2$ for a known function $v(x)$, the generalized likelihood ratio test corresponds to using $w(x) = v(x)^{-1}$. In this case, the Wilks type of result (4.1) continues to hold.

4.2. *Testing homogeneity.* Consider the varying-coefficient model defined in Section 3. A natural question arising in practice is if these coefficient functions are really varying. This amounts to testing the following problem:

$$H_0: a_1(U) = \theta_1, \dots, a_p(U) = \theta_p.$$

If the error distribution is homogeneous normal, then the generalized likelihood test based on local linear fits is given by (3.2) with $\mathrm{RSS}_0 = \sum_{i=1}^{n}(Y_i - \hat{\theta}^\tau \mathbf{X}_i)^2$ where $\hat{\theta}$ is the least-squares estimate under the null hypothesis.

To examine the property of the generalized likelihood ratio statistic (3.2) under the general heteroscedastic model, we now only assume that

$$E(\varepsilon|\mathbf{X} = \mathbf{x}, U = u) = 0, \qquad E(\varepsilon^2|\mathbf{X} = \mathbf{x}, U = u) = \sigma^2(\mathbf{x}, u),$$

for a continuous function $\sigma^2(\mathbf{x}, u)$. Strictly speaking, the statistic (3.2) is no longer a generalized likelihood ratio test under this heteroscedastic model. The generalized likelihood ratio test in this heteroscedastic case should involve weighted residual sum of squares when $\sigma^2(\mathbf{x}, u) = \sigma^2 v(\mathbf{x}, u)$ for a given v. See Remark 4.2. Let

$$\Gamma^*(u) = E[\mathbf{X}\mathbf{X}^\tau \sigma^2(\mathbf{X}, U)|U = u]f(u).$$

Then we have the following result.

THEOREM 9. *Assume Condition* (A). *Then under* H_0, *as* $h \to 0$, $nh^{3/2} \to \infty$,

$$r_K'' \lambda_n \overset{a}{\sim} \chi^2_{a_n''},$$

where

$$r_K'' = r_K [E\sigma^2(\mathbf{X}, U)] \int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})\, du \left[\int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})^2\, du\right]^{-1},$$

$$a_n'' = r_K c_K h^{-1} \left[\int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})\, du\right]^2 \left[\int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})^2 du\right]^{-1}.$$

It is clear that when $\sigma^2(\mathbf{x}, u) = \sigma^2$, Theorem 9 reduces to Theorem 5 and (3.2) is a generalized likelihood statistic. Hence the Wilks type of result continues to hold for testing homogeneity. It can also be shown that the Wilks phenomenon is still valid for the generalized likelihood ratio in the heteroscedastic model with $\sigma^2(\mathbf{x}, u) = \sigma^2 v(\mathbf{x}, u)$, bearing in mind that generalized likelihood ratio statistics are now based on the weighted residual sum of squares.

5. Extensions. The Wilks type of results does not hold only for the various problems that we have studied. They should be valid for nearly all regular nonparametric testing problems. In this section, we mention various possible extensions to indicate their versatility.

5.1. *Generalized varying coefficient models.* Inferences on generalized varying coefficient models have been empirically studied by Hastie and Tibshirani (1993) and Cai, Fan and Li (2000). The results in the previous sections can be directly extended to this setting.

Consider a generalized varying-coefficient model with the following log-likelihood function

$$l\{g^{-1}(\eta(x, u)), y\} = g_0(g^{-1}(\eta(x, u)))y - b(g_0(g^{-1}(\eta(x, u)))),$$

where $\eta(x, u) = g(m(x, u)) = A(u)^\tau x$, g is called a link function and $g_0 = b'$ is the canonical link. Poisson regression and logistic regression are two prototype examples.

Define

$$\begin{aligned}
l(g^{-1}(s), y) &= g_0(g^{-1}(s))y - b(g_0(g^{-1}(s))),\\
q_1(s, y) &= \frac{\partial l\{g^{-1}(s), y\}}{\partial s} = \frac{g_0'(s)}{g'(s)}(y - b'(s)),\\
q_2(s, y) &= \frac{\partial^2 l\{g^{-1}(s), y\}}{\partial s^2} = (g_0''/g' - g_0' g''/(g'^2))(y - g^{-1}(s)) - g_0'/(g')^2,\\
q_3(s, y) &= \frac{\partial^3 l\{g^{-1}(s), y\}}{\partial s^3} = (g_0'''/g' - g_0'' g''/g'^2 - (g_0'' g'' + g''' g_0')/g'^2\\
&\quad + 2g_0' g''^2/g'^3)(y - g^{-1}(s)) - 2g_0''/g'^2 - g_0' g''/g'^3.
\end{aligned}$$

In particular, when $g = g_0$ is the canonical link, we have

$$q_2(s, y) = -b''(s), \qquad q_3(s, y) = -b'''(s).$$

As in Section 3, we can define a local linear estimator $\hat{A}$ for A. Lemma 7.5 yields the following asymptotic representation for $\hat{A}$:

$$\begin{aligned}\hat{A}(u_0) - A(u_0)\\ &= r_n^2 \widetilde{\Gamma}(u_0)^{-1} \sum_{i=1}^n \varepsilon_i \mathbf{X}_i K((U_i - u_0)/h)(1 + o_p(1)) + H_n(u_0)(1 + o_p(1)),\end{aligned}$$

where

$$\widetilde{\Gamma}(u_0) = -E[q_2(A^\tau(u_0)\mathbf{X}, Y)\mathbf{X}\mathbf{X}^\tau | U = u_0] f(u_0), \quad \varepsilon_i = q_1(A(U_i)^\tau \mathbf{X}_i, Y_i),$$

$$\begin{aligned}H_n(u_0) = r_n^2 \widetilde{\Gamma}(u_0)^{-1} \sum_{i=1}^n &\big[q_1(\beta(u_0)^\tau \mathbf{Z}_i, Y_i) - q_1(A(U_i)^\tau \mathbf{X}_i, Y_i)\big]\\ &\times \mathbf{X}_i K((U_i - u_0)/h).\end{aligned}$$

The generalized likelihood ratio for testing the null hypothesis H_0: $A = A_0$ is defined as

$$\lambda_{ng}(A_0) = -\sum_{i=1}^n \big[l\{g^{-1}(\hat{A}(U_i)^\tau \mathbf{X}_i), Y_i\} - l\{g^{-1}(A_0(U_i)^\tau \mathbf{X}_i), Y_i\}\big].$$

Denote

$$q_{n*} = q_{n*}(U, X, Y) = \sup_{u_0, ||\alpha|| \le c_1 r_n} |q_2(\beta(u_0)^\tau Z(u_0) + \alpha^\tau Z(u_0), Y)| K\left(\frac{U - u_0}{h}\right)$$

where $r_n = 1/\sqrt{nh}$. For $j = 1, 2, 3$ and $c_1 > 0$, define

$$\begin{aligned}\Phi_{nj} &= \Phi_{nj}(U, X, Y)\\ &= \sup_{u_o, ||\alpha|| \le c_1 r_n} |q_2(\beta(u_0)^\tau Z(u_0) + \alpha^\tau Z(u_0), Y)| \left|\frac{U - u_0}{h}\right|^{j-1} K\left(\frac{U - u_0}{h}\right).\end{aligned}$$

The following technical conditions are needed.

CONDITION (B).

(B1) $E|q_1(A(U)^\tau \mathbf{X}, Y)|^4 < \infty$.

(B2) $E[q_2(A(U)^\tau \mathbf{X})\mathbf{X}\mathbf{X}^\tau | U = u]$ is Lipschitz continuous.

(B3) The function $q_2(s, y) < 0$ for $s \in R$ and y in the range of the response variable. For some function $q_*(y)$, $s_i \in C$, $i = 1, 2$, $|q_2(s_1, y) - q_2(s_2, y)| \le q_*(y)|s_1 - s_2|$. Further, for some constant $\xi > 2$,

$$E\{\Phi_{nj}(U, \mathbf{X}, Y)\, ||\mathbf{X}\mathbf{X}^\tau||\}^\xi = O(1), \quad j = 1, 2, 3,$$

$$Eq_{n*}(U, \mathbf{X}, Y)||\mathbf{X}||^2 = O(1), \quad Eq_*(Y)||\mathbf{X}||^3 < \infty,$$

$$\sup_{u_0, \|\alpha\| \le c_1 r_n} Eq_2^2(\beta(u_0)^\tau Z(u_0) + \alpha^\tau Z(u_0), Y) K^2((U - u_0)/h)/h \|XX^\tau\|^2$$
$$= O(1), \qquad j = 1, 2, 3.$$

Set

$$R_{n10g} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i A_0''(U_i)^\tau X_i \int t^2 K(t)\, dt (1 + O(h) + O(n^{-1/2})),$$

$$R_{n20g} = -\frac{1}{2} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \mathbf{X}_i^\tau \widetilde{\Gamma}(U_i)^{-1} A_0''(U_i)^\tau E(q_2(A_0^\tau(U)^\tau \mathbf{X})\mathbf{X}|U_i) w_0,$$

$$R_{n30g} = -\frac{1}{8} E A_0''(U)^\tau q_2(A_0(U)^\tau \mathbf{X}, Y)\mathbf{X}\mathbf{X}^\tau A_0''(U) w_0 (1 + O(n^{-1/2})),$$

where $w_0 = \iint t^2 (s+t)^2 K(t) K(s+t)\, dt\, ds$. Note that both R_{n10g} and R_{n20g} are asymptotic normal and hence stochastically bounded. Let $d_{1ng} = nh^4 R_{n30g} - n^{1/2} h^2 (R_{n10g} - R_{n20g})$. Then, $d_{1ng} = nh^4 R_{n30g}(1 + o_p(1))$ if $n^{1/2} h^2 \to \infty$. The following theorem shows that the Wilks type of results continues to hold for generalized varying coefficient models.

THEOREM 10. *Under Conditions* (A1)–(A3) *and* (B1)–(B3), *as* $h \to 0$, $nh^{3/2} \to \infty$ *and* $n^{(\xi-1)/\xi} h \ge c_0 (\log n)^\delta$ *for some* $\delta > (\xi - 1)/(\xi - 2)$, *we have the following asymptotic null distribution:*

$$\sigma_n^{-1}(\lambda_{ng}(A_0) - \mu_n + d_{1ng}) \xrightarrow{\mathscr{L}} N(0, 1).$$

Furthermore, if A is linear or $nh^{9/2} \to 0$, *then as* $nh \to \infty$, $r_K \lambda_{ng}(A_0) \overset{a}{\sim} \chi^2_{r_K \mu_n}$, *where* μ_n *and* r_K *are given in Theorem* 5.

Extensions of the other theorems and the remarks in Section 3 are similar. In particular the optimal minimax rate and the optimal bandwidth are the same as those in Section 3. The generalized likelihood ratio tests can be employed to check the inhomogeneity of the coefficient functions and significance of variables in the generalized varying-coefficient models. The related theorems in Section 4 hold true after some mild modifications. The details are omitted.

5.2. *Empirical likelihoods.* As pointed out in the introduction, neither Owen's empirical likelihood nor its extension, random sieve likelihood [Shen, Shi and Wong (1999)] can be directly used to make inferences on a nonparametric regression function. However, the idea of sieve empirical likelihood [Zhang and Gijbels (1999)] can be effective in this situation. In an unpublished manuscript, Fan and Zhang (1999) have developed the corresponding theory. Advantages of sieve empirical likelihood ratios are that no parametric models are needed for stochastic errors and that it adapts automatically for inhomogeneous stochastic errors. The main disadvantage is that it requires intensive computation.

6. Discussion.

6.1. *Other tests.* There are many nonparametric tests designed for specific problems. Most of them are in the univariate nonparametric regression setting. See Section 1.5 for an overview of the literature. While they can be powerful for problems for which the tests were designed, extensions of these tests to multivariate settings can pose some challenges. Further, these tests are usually not distribution free, when null hypotheses involve nuisance functions. This would hamper their applicability.

Nonparametric maximum likelihood ratio tests are a natural alternative. Usually, they do not exist. If they do, they are hard to find. Further, as shown in Section 2.2, they are not optimal. For this reason, they cannot be a generic and powerful method.

6.2. *Conclusions.* The generalized likelihood method is widely applicable. It applies not only to univariate settings, but also to multivariate nonparametric problems. It is ready to use because of the Wilks phenomenon. It is powerful since it achieves optimal rates of convergence. It can also be adaptively minimax when tuning parameters are properly tuned (Section 2.4). The tuning method for a local polynomial based generalized likelihood ratio test can be surprisingly simple. Motivated by the adaptive Neyman test constructed in Fan (1996), when the null hypothesis is linear, an adaptive construction of the generalized likelihood would naturally be

$$T^*_{\mathrm{ASL}} = \max_{h\in[n^{-a},\, n^{-b}]} \frac{r\lambda_n(h) - d(h)}{\sqrt{2d(h)}} \quad \text{for some } a, b > 0, \tag{6.1}$$

where r is the normalizing constant, $\lambda_n(h)$ is the generalized likelihood ratio test and $d(h)$ is the degrees of freedom. Therefore, the generalized likelihood is a very useful principle for all nonparametric hypothesis testing problems.

While we have observed the Wilks phenomenon and demonstrated it for a few useful cases, it is impossible for us to verify the phenomenon for all nonparametric hypothesis testing problems. The Wilks phenomenon needs to be checked for other problems that have not been covered in this paper. More work is needed in this direction.

7. Proofs.

Proof of Lemma 2.1. For each given $\xi_{n,c} = cn^{-2k/(2k+1)}$ ($c > 0$), under the null hypothesis (2.3), by using the mean-variance decomposition, we have

$$\begin{aligned} F(\xi_{n,c}) &= n^{-1}\sum j^{2k}(1 + j^{2k}\xi_{n,c})^2 \\ &\quad + O_p\left[n^{-1}\left\{\sum j^{4k}(1 + j^{2k}\xi_{n,c})^{-4}\right\}^{1/2}\right]. \end{aligned} \tag{7.1}$$

Note that $g_n(x) = x^{2k}/(1 + x^{2k}\xi_{n,c})^2$ is increasing for $0 \le x \le \xi_{n,c}^{-1/(2k)}$ and decreasing for $x \ge \xi_{n,c}^{-1/(2k)}$. By using the unimodality of g_n and approximating discrete sums by their corresponding integrals, one can show that

$$n^{-1}\sum j^{2k}(1+j^{2k}\xi_{n,c})^2 = c^{-(2k+1)/(2k)}\int_0^\infty \frac{y^{2k}}{(1+y^{2k})^2}\,dy + O(n^{-1/(2k+1)}). \tag{7.2}$$

Using the same arguments as those used in obtaining (7.2), we have

$$n^{-1}\{\sum j^{4k}(1+j^{2k}\xi_{n,c})^{-4}\}^{1/2} = O[n^{-1/\{2(2k+1)\}}].$$

This together with (7.1) and (7.2) yields

$$F(\xi_{n,c}) = (c_0/c)^{(2k+1)/(2k)} + O_p(n^{-1/\{2(2k+1)\}}), \tag{7.3}$$

where $c_0 = (\int_0^\infty y^{2k}(1+y^{2k})^{-2}\,dy)^{2k/(2k+1)}$.

For any $\varepsilon > 0$, since the function $F(x)$ is strictly decreasing,

$$P(|n^{2k/(2k+1)}(\hat{\xi}-\xi_{n,c_0})| > \varepsilon) = P(F(\hat{\xi}) < F(\xi_{n,c_0+\varepsilon})) + P(F(\hat{\xi}) > F(\xi_{n,c_0-\varepsilon})) = o(1),$$

which implies $\hat{\xi} - \xi_{n,c_0} = o_p(n^{-2k/(2k+1)})$. $\square$

PROOF OF THEOREM 1. Define the jth coefficients in $F(\xi)$ and λ_n^* as

$$F(j;\xi) = \frac{j^{2k}}{(1+j^{2k}\xi)^2}, \qquad \lambda(j;\xi) = \frac{1+2j^{2k}\xi}{(1+j^{2k}\xi)^2}.$$

Then

$$F'(j;\xi) = -\frac{2j^{4k}}{(1+j^{2k}\xi)^3}, \qquad \lambda'(j;\xi) = -\frac{2j^{4k}\xi}{(1+j^{2k}\xi)^3} = \xi F'(j;\xi). \tag{7.4}$$

Let c_0 be defined the same as in Lemma 2.1. For any $\eta_{n,j}$ between $\hat{\xi}$ and ξ_{n,c_0}, it can easily be shown that

$$\sup_{j\geq 1}\left|\frac{F'(j;\eta_{n,j}) - F'(j;\xi_{n,c_0})}{F'(j;\xi_{n,c_0})}\right| = o_p(1) \tag{7.5}$$

and that for any $\zeta_{n,j}$ between $\hat{\xi}$ and ξ_{n,c_0},

$$\sup_{j\geq 1}\left|\frac{\lambda'(j;\zeta_{n,j}) - \lambda'(j;\xi_{n,c_0})}{\lambda'(j;\xi_{n,c_0})}\right| = o_p(1). \tag{7.6}$$

Let $\lambda_n(\xi) = \frac{1}{2}\sum_{j=1}^{\infty}(1+2j^{2k}\xi)/(1+j^{2k}\xi)^2\varepsilon_j^2$. By using Taylor's expansion together with (7.4), (7.5) and (7.6), under the null hypothesis (2.3),

$$
\begin{aligned}
\lambda_n^* &= \frac{1}{2}\sum_{j=1}^{\infty}\left[\lambda(j;\xi_{n,c_0}) + (\hat{\xi}-\xi_{n,c_0})\lambda'(j;\zeta_{n,j})\right]\varepsilon_j^2 \\
&= \lambda_n(\xi_{n,c_0}) + [F(\hat{\xi}) - F(\xi_{n,c_0})]\frac{\frac{1}{2}\sum_{j=1}^{\infty}\lambda'(j;\xi_{n,c_0})\varepsilon_j^2}{1/n\sum_{j=1}^{\infty}F'(j;\xi_{n,c_0})\varepsilon_j^2}(1+o_p(1)) \\
&= \lambda_n(\xi_{n,c_0}) + [1-F(\xi_{n,c_0})]\frac{n}{2}\xi_{n,c_0} + o_p(n^{1/(2(2k+1))}) \\
&= \frac{1}{2}\sum_{j=1}^{\infty}\frac{1}{(1+j^{2k}\xi_{n,c_0})}\varepsilon_j^2 + \frac{1}{2}c_0 n^{1/(2k+1)} + o_p(n^{1/(2(2k+1))}).
\end{aligned}
\tag{7.7}
$$

Define $\lambda_{n,1} = \frac{1}{2}\sum_{j=1}^{\infty}\{1+j^{2k}\xi_{n,c_0}\}^{-1}\varepsilon_j^2$ in (7.7) and $V_n = \frac{1}{2}\sum_{j=1}^{n}\{1+j^{2k}\xi_{n,c_0}\}^{-1}\times \varepsilon_j^2$, we have

$$
\frac{\max_{1\le j\le n}\{1+j^{2k}\xi_{n,c_0}\}^{-1}}{\sqrt{\sum_{j=1}^{n}\{1+j^{2k}\xi_{n,c_0}\}^{-2}}} \le \left\{\sum_{j=1}^{n}(1+j^{2k}\xi_{n,c_0})^{-2}\right\}^{-1/2} = O(\xi_{n,c_0}^{1/(4k)}) \to 0,
$$

which implies that $(V_n - E(V_n))/\sqrt{\mathrm{var}(V_n)} \xrightarrow{\mathscr{L}} N(0,1)$ by Lemma 2.1 of Huber (1973). Note that

$$
\mathrm{var}(\lambda_{n,1} - V_n) \le \frac{1}{2}\int_n^{\infty}\frac{dx}{(1+x^{2k}\xi_{n,c_0})^2} \le \frac{1}{2}\int_n^{\infty}\frac{dx}{x^{4k}\xi_{n,c_0}^2} = O(\xi_{n,c_0}^{-2}n^{-(4k-1)}).
$$

Hence

$$
\frac{\mathrm{var}(\lambda_{n,1} - V_n)}{\mathrm{var}(\lambda_{n,1})} = O(\xi_{n,c_0}^{-2}n^{-(4k-1)}/\xi_{n,c_0}^{-1/(2k)}) \to 0.
$$

This implies that

$$
\frac{\lambda_{n,1} - E(\lambda_{n,1})}{\sqrt{\mathrm{var}(\lambda_{n,1})}} \xrightarrow{\mathscr{L}} N(0,1)
$$

[by Theorem 3.2.15 of Randle and Wolfe (1979)], where

$$
E(\lambda_{n,1}) = 2^{-1}\xi_{n,c_0}^{-1/(2k)}\int_0^{\infty}\frac{dy}{(1+y^{2k})} + O(1),
$$

$$
\mathrm{var}(\lambda_{n,1}) = 2^{-1}\xi_{n,c_0}^{-1/(2k)}\int_0^{\infty}\frac{dy}{(1+y^{2k})^2} + O(1).
$$

This together with (7.7) yields

$$
\frac{\lambda_n^* - 2^{-1}c_0^{-1/(2k)}n^{1/(2k+1)}\int_0^{\infty}(1+2y^{2k})/(1+y^{2k})^2\,dy}{\sqrt{2^{-1}c_0^{-1/(2k)}n^{1/(2k+1)}\int_0^{\infty}dy/(1+y^{2k})^2}} \xrightarrow{\mathscr{L}} N(0,1).
$$

Namely, $r_k \lambda_n^* \sim_a \chi^2_{a_n}$, where

$$r_k = 2\int_0^\infty \frac{1+2y^{2k}}{(1+y^{2k})^2}\,dy\left(\int_0^\infty \frac{1}{(1+y^{2k})^2}\,dy\right)^{-1},$$

$$a_n = 2^{-1} r_k c_0^{-1/(2k)} \int_0^\infty \frac{1+2y^{2k}}{(1+y^{2k})^2}\,dy\; n^{1/(2k+1)}.$$

Finally, by using

$$\int_0^\infty \frac{dy}{(1+y^{2k})} = \frac{1}{2k\sin(\pi/(2k))}\pi, \qquad \int_0^\infty \frac{dy}{(1+y^{2k})^2} = \frac{(2k-1)}{4k^2\sin(\pi/(2k))}\pi,$$

$$\int_0^\infty \frac{dy}{(1+y^{2k})^3} = \frac{(2k-1)(4k-1)}{16k^3\sin(\pi/(2k))}\pi,$$

$$\int_0^\infty \frac{dy}{(1+y^{2k})^4} = \frac{(2k-1)(4k-1)(6k-1)}{96k^4\sin(\pi/(2k))}\pi,$$

we obtain

$$r_k = \frac{4k+2}{2k-1}, \qquad a_n = \frac{(2k+1)^2}{2k-1}\left[\frac{\pi}{4k^2\sin(\pi/(2k))}\right]^{2k/(2k+1)} n^{1/(2k+1)}. \qquad \square$$

PROOF OF THEOREM 2. Take $j_n^{-k} = n^{-(k+d)/(2k+1)}$. Let θ be a vector whose j_nth position is j_n^{-k} and the rest are zero. Then, $\theta \in \mathscr{F}_k$ and $\|\theta\| = n^{-(k+d)/(2k+1)}$. For $\xi_{n,c} = cn^{-2k/(2k+1)}$, we have

$$j_n^{2k}\xi_{n,c} = cn^{2d/(2k+1)}.$$

Under this specific alternative, by using model (2.2), we have, for $d > 1/8$,

$$\begin{aligned} F(\xi_{n,c}) &= F(\xi_{n,c}|H_0) + \frac{j_n^{2k}}{(1+j_n^{2k}\xi_{n,c})^2}(2j_n^{-k}n^{-1/2}\varepsilon_{j_n} + j_n^{-2k}) \\ &= F(\xi_{n,c}|H_0) + o_p(n^{-1/\{2(2k+1)\}}), \end{aligned}$$

where $F(\xi_{n,c}|H_0) = n^{-1}\sum_{j=1}^\infty \frac{j^{2k}}{(1+j^{2k}\xi_{n,c})^2}\varepsilon_j^2$. By arguments such as those in the proof of Lemma 2.1, one can see that

$$\hat{\xi} = \xi_{n,c_0}(1+o_p(1)),$$

where $\hat{\xi}$ solves $F(\hat{\xi}) = 1$.

Next, consider the likelihood ratio statistic λ_n^* under the alternative hypothesis. Let

$$\lambda_{n,0} = \frac{1}{2}\sum_j \left(1 - \frac{j^{4k}\hat{\xi}^2}{(1+j^{2k}\hat{\xi})^2}\right)\varepsilon_j^2.$$

Then for $d > 1/8$,

$$\begin{aligned}\lambda_n^* &= \lambda_{n,0} + \frac{n}{2}\left(1 - \frac{j_n^{4k}\hat{\xi}^2}{(1 + j_n^{2k}\hat{\xi})^2}\right)(2 j_n^{-k} n^{-1/2} \varepsilon_{j_n} + j_n^{-2k}) \\ &= \lambda_{n,0} + o_p(n^{1/\{2(2k+1)\}}).\end{aligned}$$

By a proof similar to that in Theorem 1, $r_k \lambda_{n,0} \overset{a}{\sim} \chi^2_{a_n}$, which entails that

$$P_\theta\{r_k \lambda_n^* > a_n + z_\alpha (2a_n)^{1/2}\} = \alpha + o(1). \qquad \square$$

Proof of Theorem 3. The first part of the result follows directly from the central limit theory using similar arguments to those in the proof of Theorem 1 for $\lambda_{n,1}$. We now establish the power of the test. Under the alternative hypothesis,

$$E_\theta(r'_k \lambda_n) = a'_n + O(1) + r'_k \sum_{j=1}^{\infty} \left(1 - \frac{j^{4k}\xi_n^2}{(1 + j^{2k}\xi_n)^2}\right) n\theta_j^2/2$$

and

$$\operatorname{var}_\theta(r'_k \lambda_n) = 2a'_n + b'_n + O(1),$$

where $b'_n = r_k'^2 \sum_{j=1}^\infty (1 - j^{4k}\xi_n^2/(1 + j^{2k}\xi_n)^2)^2 n\theta_j^2$. Thus, it follows from Chebyshev's inequality that

$$\begin{aligned}P_\theta(r'_k \lambda_n > a'_n + z_\alpha (2a'_n)^{1/2}) &= P_\theta\left\{\frac{r'_k \lambda_n - r'_k E_\theta(\lambda_n)}{\operatorname{var}_\theta(r'_k\lambda_n)^{1/2}} \geq (2a'_n + b'_n + O(1))^{-1/2} \right. \\ &\qquad \left. \times \{a'_n + z_\alpha(2a'_n)^{1/2} - r'_k E_\theta(\lambda_n)\}\right\} \\ &\geq 1 - d_n^{-2},\end{aligned}$$

if $(2a'_n + b'_n + O(1))^{-1/2}\{a'_n + z_\alpha(2a'_n)^{1/2} - r'_k E_\theta(\lambda_n)\} \leq -d_n$ for some $d_n > 0$. Thus, Theorem 3 holds, if we show that

$$\inf_{\theta \in \mathscr{F}_k:\ \|\theta\| \geq c_n n^{-2k/(4k+1)}} n^{-1/(4k+1)} \sum_{j=1}^{\infty} \left(1 - \frac{j^{4k}\xi_n^2}{(1 + j^{2k}\xi_n)^2}\right) n\theta_j^2 \to \infty \tag{7.8}$$

and

$$\inf_{\theta \in \mathscr{F}_k:\ \|\theta\| \geq c_n n^{-2k/(4k+1)}} b_n'^{-1/2} \sum_{j=1}^{\infty} \left(1 - \frac{j^{4k}\xi_n^2}{(1 + j^{2k}\xi_n)^2}\right) n\theta_j^2 \to \infty. \tag{7.9}$$

Note that for each $\theta \in \mathscr{F}_k$, $|\theta|\xi c_n n^{-2k/(4k+1)}$

$$\begin{aligned}
&\sum_{j=1}^{\infty}\left(1-\frac{j^{4k}\xi_n^2}{(1+j^{2k}\xi_n)^2}\right)\theta_j^2 \\
&\qquad \geq c_n^2 n^{-4k/(4k+1)} - \xi_n \max_{x\geq 0}\frac{x}{(1+x)^2}\sum_{j=1}^{\infty} j^{2k}\theta_j^2 \\
&\qquad \geq c_n^2 n^{-4k/(4k+1)}/2.
\end{aligned} \tag{7.10}$$

Hence, (7.8) holds.

To show (7.9), we note that $(1 - j^{4k}\xi_n^2/(1+j^{2k}\xi_n)^2) \in (0,1)$. It follows from (7.10) that

$$\begin{aligned}
b_n'^{-1/2}\sum_{j=1}^{\infty}\left(1-\frac{j^{4k}\xi_n^2}{(1+j^{2k}\xi_n)^2}\right)n\theta_j^2 &\geq r_k'^{-1} n^{1/2}\left(\sum_{j=1}^{\infty}\left(1-\frac{j^{4k}\xi_n^2}{(1+j^{2k}\xi_n)^2}\right)\theta_j^2\right)^{1/2} \\
&\geq r_k'^{-1} n^{1/2} c_n n^{-2k/(4k+1)}/2,
\end{aligned}$$

which tends to ∞. $\square$

Proof of Theorem 4. For any given m, when n is sufficiently large, we have

$$\begin{aligned}
&P_\theta[T_{\mathrm{AN}} > -\log\{-\log(1-\alpha)\}] \\
&\quad \geq P_\theta\{T_{\mathrm{AN}}^* > 2(\log\log n)^{1/2}\} \\
&\quad \geq P_\theta\left\{\sum_{j=1}^{m}(nY_j^2-1)/\sqrt{2m} \geq 2(\log\log n)^{1/2}\right\}.
\end{aligned} \tag{7.11}$$

Note that the sequence of random variables

$$\left\{\sum_{j=1}^{m}(nY_j^2 - 1 - n\theta_j^2)\bigg/\left(2m + 4n\sum_{j=1}^{m}\theta_j^2\right)^{1/2}\right\}$$

have mean zero and variance one. By normalizing the random variables in (7.11), one can easily see that the power of the adaptive Neyman test is at least

$$\begin{aligned}
P_\theta\Bigg\{&\sum_{j=1}^{m}(nY_j^2 - 1 - n\theta_j^2)\bigg/\left(2m + 4n\sum_{j=1}^{m}\theta_j^2\right)^{1/2} \\
&\geq \left\{2\sqrt{2m}\sqrt{\log\log n} - n\sum_{j=1}^{m}\theta_j^2\right\}\bigg/\left(2m + 4n\sum_{j=1}^{m}\theta_j^2\right)^{1/2}\Bigg\}.
\end{aligned}$$

Thus Theorem 4 holds via the Chebyshev inequality if we show that

$$\inf_{\theta\in\mathscr{F}_k:\,\|\theta\|\geq c_n\delta_n} m^{-1/2}\left\{n\sum_{j=1}^{m}\theta_j^2 - 2\sqrt{2m}\sqrt{\log\log n}\right\} \to \infty \tag{7.12}$$

and

$$\inf_{\theta\in\mathscr{F}_k:\ \|\theta\|\ge c_n\delta_n}\left(n\sum_{j=1}^{m}\theta_j^2\right)^{-1/2}\left\{n\sum_{j=1}^{m}\theta_j^2-2\sqrt{2m}\sqrt{\log\log n}\right\}\to\infty \tag{7.13}$$

for some choice of m.

Note that for any $\theta\in\mathscr{F}_k$,

$$\sum_{j=m+1}^{\infty}\theta_j^2\le m^{-2k}\sum_{j=m+1}^{\infty}j^{2k}\theta_j^2\le m^{-2k}.$$

Thus,

$$m^{-1/2}\sum_{j=1}^{m}\theta_j^2\ge m^{-1/2}(c_n\delta_n)^2-m^{-2k-1/2}.$$

Maximizing the above expression with respect to m leads to the choice of $m=O((c_n\delta_n)^{-1/k})$; we have

$$m^{-1/2}\sum_{j=1}^{m}\theta_j^2\ge O\{c_n^{(4k+1)/(2k)}n^{-1}(\log\log n)^{1/2}\}, \tag{7.14}$$

and

$$n\sum_{j=1}^{m}\theta_j^2\ge n((c_n\delta_n)^2-m^{-2k})=O\{nc_n^2n^{-4k/(4k+1)}(\log\log n)^{2k/(4k+1)}\}. \tag{7.15}$$

Since $c_n\to\infty$, the conclusion (7.12) holds from (7.14). And (7.13) follows from

$$\left(n\sum_{j=1}^{m}\theta_j^2\right)^{-1/2}\left\{n\sum_{j=1}^{m}\theta_j^2-2\sqrt{2m}\sqrt{\log\log n}\right\}=\left(n\sum_{j=1}^{m}\theta_j^2\right)^{1/2}(1+o(1))$$

and (7.15). □

The following four lemmas are used in the proofs for the theorems in Sections 3, 4 and 5.

LEMMA 7.1. *Suppose the matrix $\Psi=(\psi_{ij})_{i,\,j=1}^n$ is symmetric, $w_1,\dots,w_n$ are independent random variables with moments $E(w_i)=0$, $E(w_i^2)=u_2(i)$, $E(w_i^3)=u_3(i)$, $E(w_i^4)=u_4(i)$. Let $\mathbf{W}=(w_1,\dots,w_n)^\tau$. Then*

$$E(\mathbf{W}^\tau\Psi\mathbf{W})^2=\sum_{i=1}^{n}\psi_{ii}^2[u_4(i)-3u_2^2(i)]+\left[\sum_{i=1}^{n}\psi_{ii}u_2(i)\right]^2+2\sum_{i,j=1}^{n}\psi_{ij}^2u_2(i)u_2(j).$$

PROOF. This can be shown by modifying the proof of Theorem 1.8 in Seber (1977), where only $u_l(i)\equiv u_l$ ($i=1,\dots,n$; $l=1,2,3,4$) were considered. □

Let $r_n = 1/\sqrt{nh}$. Denote by

$$\alpha_n(u_0) = r_n^2 \Gamma(u_0)^{-1} \sum_{i=1}^{n} \varepsilon_i \mathbf{X}_i K((U_i - u_0)/h), \tag{7.16}$$

$$R_n(u_0) = r_n^2 \sum_{i=1}^{n} \Gamma(u_0)^{-1} (A(U_i)^\tau \mathbf{X}_i - \beta(u_0)^\tau \mathbf{Z}_i) \mathbf{X}_i K((U_i - u_0)/h), \tag{7.17}$$

$$R_{n1} = \sum_{k=1}^{n} \varepsilon_k R_n(U_k)^\tau \mathbf{X}_k,$$

$$R_{n2} = \sum_{k=1}^{n} \alpha_n(U_k)^\tau \mathbf{X}_k \mathbf{X}_k^\tau R_n(U_k),$$

$$R_{n3} = \tfrac{1}{2} \sum_{k=1}^{n} R_n(U_k)^\tau \mathbf{X}_k \mathbf{X}_k^\tau R_n(U_k).$$

LEMMA 7.2. *Under Condition* (A), *as* $h \to 0$, $nh \to \infty$,

$$R_{n1} = n^{1/2} h^2 R_{n10} + O(n^{-1/2} h),$$
$$R_{n2} = n^{1/2} h^2 R_{n20} + O(n^{-1/2} h),$$
$$R_{n3} = n h^4 R_{n30} + O(h^3).$$

Furthermore, for any $\delta > 0$, *there exists* $M > 0$ *such that*

$$\sup_{G_n \in \mathscr{G}_n} P(|(n^{1/2} h^2)^{-1} R_{nj}| > M) \le \delta, \qquad j = 1, 2,$$

$$\sup_{G_n \in \mathscr{G}_n} P(|(nh^4)^{-1} R_{n3}| > M) \le \delta.$$

The proof follows from some direct but tedious calculations.

Using Lemma 7.5, we can easily show the following lemma.

LEMMA 7.3. *Let* $\hat{A}$ *be the local linear estimator define in Section* 3. *Then, under Condition* (A), *uniformly for* $u_0 \in \Omega$,

$$\hat{A}(u_0) - A(u_0) = (\alpha_n(u_0) + R_n(u_0))(1 + o_p(1))$$

where $\alpha_n(u_0)$ *and* $R_n(u_0)$ *are defined in (7.16) and (7.17).*

Denote by

$$T_n = r_n^2 \sum_{k,i} \varepsilon_k \varepsilon_i \mathbf{X}_i^\tau \Gamma(U_k)^{-1} \mathbf{X}_k K((U_i - U_k)/h),$$

$$S_n = r_n^4 \sum_{i,j} \varepsilon_i \varepsilon_j \mathbf{X}_i^\tau \Bigg\{ \sum_{k=1}^{n} \Gamma(U_k)^{-1} \mathbf{X}_k \mathbf{X}_k^\tau \Gamma(U_k)^{-1} K((U_i - U_k)/h) \times K((U_j - U_k)/h) \Bigg\} \mathbf{X}_j.$$

LEMMA 7.4. *Under Condition* (A), *as* $h \to 0$, $nh^{3/2} \to \infty$,

$$T_n = \frac{1}{h} pK(0)\sigma^2 Ef(U)^{-1} + \frac{1}{n}\sum_{k\neq i} \varepsilon_k\varepsilon_i \mathbf{X}_i^\tau \Gamma(U_k)^{-1}\mathbf{X}_k K_h(U_k - U_i) + o_p(h^{-1/2}),$$

$$S_n = \frac{1}{h} p\sigma^2 Ef^{-1}(U)\int K^2(t)\,dt$$
$$+ \frac{2}{nh}\sum_{i<j} \varepsilon_i\varepsilon_j \mathbf{X}_i^\tau \Gamma^{-1}(U_i) K * K((U_i - U_j)/h)\mathbf{X}_j + o_p(h^{-1/2}),$$

with $K_h(\cdot) = K(\cdot/h)/h$.

PROOF. The first equality is obvious. Here we focus on the second one. We use the following decomposition: $S_n = S_{n1} + S_{n2}$ with

$$S_{n1} = \frac{1}{(nh)^2}\sum_{i=1}^n \varepsilon_i^2 \mathbf{X}_i^\tau \left\{\sum_{k=1}^n \Gamma(U_k)^{-1}\mathbf{X}_k\mathbf{X}_k^\tau \Gamma(U_k)^{-1} K^2((U_i - U_k)/h)\right\}\mathbf{X}_i$$

$$S_{n2} = \frac{1}{n^2}\sum_{i\neq j} \varepsilon_i\varepsilon_j \mathbf{X}_i^\tau \left\{\sum_{k=1}^n \Gamma(U_k)^{-1}\mathbf{X}_k\mathbf{X}_k^\tau \Gamma(U_k)^{-1} K_h(U_k - U_i) K_h(U_k - U_j)\right\}\mathbf{X}_j.$$

It is easy to see that as $h \to 0$,

$$S_{n1} = o_p(h^{-1/2}) + O_p(n^{-3/2}h^{-2}) + \frac{1}{2}V_n(1 + o(1)) + O_p\left(\frac{1}{nh^2}\right), \tag{7.18}$$

where

$$V_n = \frac{2}{n(n-1)}\sum_{1\le i<k\le n} \sigma^2 (\mathbf{X}_i^\tau \Gamma(U_k)^{-1}\mathbf{X}_k\mathbf{X}_k^\tau \Gamma(U_k)^{-1}\mathbf{X}_i$$
$$+ \mathbf{X}_k^\tau \Gamma(U_i)^{-1}\mathbf{X}_i\mathbf{X}_i^\tau \Gamma(U_i)^{-1}\mathbf{X}_k) K_h^2(U_k - U_i).$$

Using Hoeffding's decomposition for the variance of U-statistics [see, e.g., Koroljuk and Borovskich (1994)] we obtain

$$\operatorname{var}(V_n) = O\left(\frac{1}{n}\right)\sigma_n^2$$

with

$$\sigma_n^2 \le E\{E[(\mathbf{X}_1^\tau \Gamma(U_2)^{-1}\mathbf{X}_2\mathbf{X}_2^\tau \Gamma(U_2)^{-1}\mathbf{X}_1$$
$$+ \mathbf{X}_2^\tau \Gamma(U_1)^{-1}\mathbf{X}_1\mathbf{X}_1^\tau \Gamma(U_1)^{-1}\mathbf{X}_2)\ K_h^2(U_2 - U_1)|(\mathbf{X}_1, U_1)])^2\}^2$$
$$= O(h^{-2}).$$

Thus, $V_n = EV_n + o_p(h^{-1/2})$ as $nh \to \infty$ and $h \to 0$. Consequently,

$$S_{n1} = \frac{1}{h} p\sigma^2 Ef^{-1}(U)\int K^2(t)\,dt + o_p(h^{-1/2}). \tag{7.19}$$

We now deal with the term S_{n2}. Decompose $S_{n2} = S_{n21} + S_{n22}$ with

$$S_{n21} = \frac{2}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \mathbf{X}_i^\tau \frac{1}{n} \times \sum_{k \ne i,j} \{\Gamma(U_k)^{-1} \mathbf{X}_k \mathbf{X}_k^\tau \Gamma^{-1}(U_k) K_h(U_k - U_i) K_h(U_k - U_j)\} \mathbf{X}_j,$$

$$S_{n22} = \frac{K(0)}{n^2 h} \sum_{i \ne j} \varepsilon_i \varepsilon_j \{\mathbf{X}_i^\tau \Gamma(U_i)^{-1} \mathbf{X}_i \mathbf{X}_i^\tau \Gamma(U_i)^{-1} \mathbf{X}_j + \mathbf{X}_i^\tau \Gamma(U_j)^{-1} \mathbf{X}_j \mathbf{X}_j^\tau \Gamma(U_j)^{-1} \mathbf{X}_j\} K_h(U_i - U_j).$$

It can easily be shown that

$$\operatorname{var}(S_{n22}) = O(1/(n^2 h^3)) = o(1/h)$$

which implies

$$S_{n22} = o_p(h^{-1/2}). \tag{7.20}$$

Let

$$Q_{ijkh} = \Gamma^{-1}(U_k) \mathbf{X}_k \mathbf{X}_k^\tau \Gamma(U_k)^{-1} K_h(U_k - U_i) K_h(U_k - U_j).$$

Note that

$$E\left[\mathbf{X}_i^\tau \frac{1}{n} \sum_{k \ne i,j} (Q_{ijkh} - E(Q_{ijkh} | (u_i, u_j))) \mathbf{X}_j\right]^2 \le \operatorname{trace}\left\{n^{-2} \sum_{k \ne 1,2}^{n} E(Q_{12kh} \mathbf{X}_2 \mathbf{X}_2^\tau Q_{12kh} \mathbf{X}_1 \mathbf{X}_1^\tau)\right\} = O(1/(nh^2)),$$

which leads to

$$S_{n21} = \frac{2(n-2)}{n^2} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \mathbf{X}_i^\tau E(Q_{ijkh} | (U_i, U_j)) \mathbf{X}_j + o_p(h^{-1/2}). \tag{7.21}$$

Combining (7.18)–(7.21), we complete the proof. $\square$

PROOF OF THEOREM 5. Note that

$$\frac{\mathrm{RSS}_1}{n} = \sigma^2 (1 + O_p(n^{-1/2}) + O_p(h^{-1})).$$

It then follows from the definition that

$$-\lambda_n(A_0)\sigma^2 = -r_n^2 \sum_{k=1}^{n} \varepsilon_k \left\{\sum_{i=1}^{n} \varepsilon_i \mathbf{X}_i^\tau \Gamma(U_k)^{-1}\right\} \mathbf{X}_k K((U_i - u_0)/h) + \frac{1}{2} r_n^4 \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j \mathbf{X}_i^\tau \Gamma(U_k)^{-1} \mathbf{X}_k \mathbf{X}_k^\tau \mathbf{X}_j \Gamma(U_k)^{-1}$$

$$\times K((U_i - U_k)/h)K((U_j - U_k)/h)$$
$$- R_{n1} + R_{n2} + R_{n3} + O_p\left(\frac{1}{nh^2}\right).$$

Applying Lemmas 7.2, 7.3 and 7.4, we get

$$-\lambda_n(A_0) = -\mu_n + d_{1n} - W(n)h^{-1/2}/2 + o_p(h^{-1/2}),$$

where

$$W(n) = \frac{\sqrt{h}}{n\sigma^2}\sum_{j\neq l}\varepsilon_j\varepsilon_l[2K_h(U_j - U_l) - K_h * K_h(U_j - U_l)]\mathbf{X}_j^\tau\Gamma(U_l)^{-1}\mathbf{X}_l.$$

It remains to show that

$$W(n) \xrightarrow{\mathscr{L}} N(0, v)$$

with $v = 2\|2K - K * K\|_2^2 pEf^{-1}(U)$.

Define $W_{jl} = (\sqrt{h}/n)\, b_n(j,l)\varepsilon_j\varepsilon_l/\sigma^2$ ($j < l$), where $b_n(j,l)$ is written in a symmetric form

$$b_n(j,l) = a_1(j,l) + a_2(j,l) - a_3(j,l) - a_4(j,l),$$

with

$$a_1(j,l) = 2K_h(U_j - U_l)\mathbf{X}_j^\tau\Gamma(U_l)^{-1}\mathbf{X}_l, \qquad a_2(j,l) = a_1(l,j),$$
$$a_3(j,l) = K_h * K_h(U_j - U_l)\mathbf{X}_j^\tau\Gamma(U_l)^{-1}\mathbf{X}_l, \quad a_4(j,l) = a_3(l,j).$$

Then $W(n) = \sum_{j<l} W_{jl}$. To apply Proposition 3.2 in de Jong (1987), we need to check:

(i) $W(n)$ is clean [see de Jong (1987) for the definition];
(ii) $\operatorname{var}(W(n)) \to v$;
(iii) G_{I} is of smaller order than $\operatorname{var}(W(n))$;
(iv) G_{II} is of smaller order than $\operatorname{var}(W(n))$;
(v) G_{IV} is of smaller order than $\operatorname{var}(W(n))$;

where

$$G_{\mathrm{I}} = \sum_{1\le i<j\le n} E(W_{ij}^4),$$
$$G_{\mathrm{II}} = \sum_{1\le i<j<k\le n}\{E(W_{ij}^2W_{ik}^2) + E(W_{ji}^2W_{jk}^2) + E(W_{ki}^2W_{kj}^2)\},$$
$$G_{\mathrm{IV}} = \sum_{1\le i<j<k<l\le n}\{E(W_{ij}W_{ik}W_{lj}W_{lk}) + E(W_{ij}W_{il}W_{kj}W_{kl}) + E(W_{ik}W_{il}W_{jk}W_{jl})\}.$$

We now check each of the following conditions. Condition (i) follows directly from the definition.

To prove (ii), we note that

$$\operatorname{var}(W(n)) = \sum_{j<l} E(W_{jl}^2).$$

Denote $K(t,m) = K * \cdots * K(t)$ as the mth convolution of $K(\cdot)$ at t for $m = 1, 2, \ldots$. Then it follows that

$$E[b_n^2(j,l)\varepsilon_j^2\varepsilon_l^2]$$
$$= \frac{\sigma^4}{h}[16K(0,2) - 16K(0,3) + 4K(0,4)]\, pEf^{-1}(U)(1+O(h))$$

which entails

$$v = 2\int \bigl[2K(x) - K * K(x)\bigr]^2 dx \;\, pEf^{-1}(U).$$

Condition (iii) is proved by noting that

$$E\bigl[a_1(1,2)\varepsilon_1\varepsilon_2\bigr]^4 = O(h^{-3}), \quad E[a_3(1,2)\varepsilon_1\varepsilon_2]^4 = O(h^{-2}),$$

which implies that $E(W_{12}^4) = (h^2/n^4)O(h^{-3}) = O(n^{-4}h^{-1})$. Hence $G_{\mathrm{I}} = O(n^{-2}h^{-1}) = o(1)$.

Condition (iv) is proved by the following calculation:

$$E(W_{12}^2 W_{13}^2) = O(EW_{12}^4) = O(n^{-4}h^{-1}),$$

which implies that $G_{\mathrm{II}} = O(1/(nh)) = o(1)$.

To prove (v), it suffices to calculate the term $E(W_{12}W_{23}W_{34}W_{41})$. By straightforward calculations,

$$E\{a_1(1,2)a_1(2,3)a_1(3,4)a_1(4,1)\varepsilon_1^2\varepsilon_2^2\varepsilon_3^2\varepsilon_4^2\} = O(h^{-1}),$$
$$E\{a_1(1,2)a_1(2,3)a_1(3,4)a_3(4,1)\varepsilon_1^2\varepsilon_2^2\varepsilon_3^2\varepsilon_4^2\} = O(h^{-1}),$$
$$E\{a_1(1,2)a_1(2,3)a_3(3,4)a_3(4,1)\varepsilon_1^2\varepsilon_2^2\varepsilon_3^2\varepsilon_4^2\} = O(h^{-1}),$$
$$E\{a_1(1,2)a_3(2,3)a_3(3,4)a_3(4,1)\varepsilon_1^2\varepsilon_2^2\varepsilon_3^2\varepsilon_4^2\} = O(h^{-1}),$$
$$E\{a_3(1,2)a_3(2,3)a_3(3,4)a_3(4,1)\varepsilon_1^2\varepsilon_2^2\varepsilon_3^2\varepsilon_4^2\} = O(h^{-1}),$$

and similarly for the other terms. So

$$E(W_{12}W_{23}W_{34}W_{41}) = n^{-4}h^2 O(h^{-1}) = O(n^{-4}h)$$

which yields

$$G_{\mathrm{IV}} = O(h) = o(1). \qquad \square$$

PROOF OF THEOREM 6. Analogously to the arguments for $\hat{A}$, we get

$$(\widetilde{A}_2(u_0) - A_2(u_0)) = r_n^2\Gamma_{22}^{-1}(u_0)\sum_{k=1}^n \{Y_k - A_1(U_k)^\tau \mathbf{X}_k^{(1)} - \overline{\eta}_2(u_0, \mathbf{X}_k^{(2)}, U_k)\}$$
$$\times \mathbf{X}_k^{(2)} K((U_k - u_0)/h)(1 + o_p(1)),$$

where $\overline{\eta}_2(u_0, \mathbf{X}_k^{(2)}, U_k) = A_2(u_0)^\tau \mathbf{X}_k^{(2)} + A_2'(u_0)^\tau \mathbf{X}_k^{(2)}(U_k - u_0)$. Note that

$$\lambda_{nu}(A_{10}) = \lambda_n(A_0) - \lambda_{n2}(A_{20}|A_{10}).$$

Similarly to the proof of Theorem 5, under H_{0u}, we have

$$\begin{aligned}
\lambda_{n2}(A_{20}|A_{10})\sigma^2 = {} & r_n^2 \sum_{k=1}^n \sum_{i=1}^n \varepsilon_i K((U_i - U_k)/h)\mathbf{X}_i^{(2)} \Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)} \varepsilon_k \\
& - \tfrac{1}{2} r_n^4 \sum_{k=1}^n \left[\sum_{i=1}^n \varepsilon_i K((U_i - U_k)/h)\mathbf{X}_i^{(2)\tau}\right] \\
& \times \left(\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)}\mathbf{X}_k^{(2)\tau}\Gamma_{22}^{-1}(U_k)\right) \\
& \times \left[\sum_{i=1}^n \varepsilon_i K((U_i - u_0)/h)\mathbf{X}_i^{(2)}\right] + o_p(h^{-1/2}) - d_{1n*},
\end{aligned}$$

where d_{1n*} is defined by replacing X and Γ by $X^{(2)}$ and Γ_{22} in d_{1n}. Consequently,

$$\begin{aligned}
-\lambda_{nu}(A_{10})\sigma^2 = {} & -r_n^2 \sum_{k,i} \varepsilon_k \varepsilon_i (\mathbf{X}_i^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)X_i^{(2)})^\tau \Gamma_{11,2}^{-1} \\
& \times (U_k)(\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)})K((U_i - U_k)/h) \\
& + \frac{r_n^4}{2} \sum_{i,j} \varepsilon_i \varepsilon_j \sum_{k=1}^n (\mathbf{X}_i^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_i^{(2)})^\tau \\
& \times \Gamma_{11,2}^{-1}(U_k)(\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)}) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_j^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_j^{(2)}) \\
& + R_{n4} + R_{n5} + o_p(h^{-1/2}) + d_{1n} - d_{1n*},
\end{aligned}$$

where

$$\begin{aligned}
R_{n4} = {} & \frac{r_n^4}{2} \sum_{i,j}^n \varepsilon_i \varepsilon_j \sum_{k=1}^n (\mathbf{X}_i^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_i^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)})\mathbf{X}_k^{(2)\tau}\Gamma_{22}^{-1}(U_k)\mathbf{X}_j^{(2)} \\
& \times K((U_i - U_k)/h)K((U_j - U_k)/h), \\
R_{n5} = {} & \frac{r_n^4}{2} \sum_{i,j}^n \varepsilon_i \varepsilon_j \sum_{k=1}^n (\mathbf{X}_j^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_j^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)})\mathbf{X}_k^{(2)\tau}\Gamma_{22}^{-1}(U_k)\mathbf{X}_i^{(2)} \\
& \times K((U_i - U_k)/h)K((U_j - U_k)/h).
\end{aligned}$$

A simple calculation shows that as $nh^{3/2} \to \infty$,

$$ER_{n4}^2 = O\left(\frac{1}{n^2h^4}\right) = o(h^{-1}),$$

which yields $R_{n4} = o_p(h^{-1/2})$. Similarly, we can show $R_{n5} = o_p(h^{-1/2})$. Therefore,

$$\begin{aligned}
-\lambda_{nu}(A_{10})\sigma^2 = & -r_n^2 \sum_{k,i} \varepsilon_k \varepsilon_i (\mathbf{X}_i^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_i^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)}) K((U_i - U_k)/h) + o_p(h^{-1/2}) \\
& + \frac{r_n^4}{2} \sum_{i,j} \varepsilon_i \varepsilon_j \sum_{k=1}^{n} (\mathbf{X}_i^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_i^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)}) \\
& \times (\mathbf{X}_k^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_k^{(2)})^\tau \Gamma_{11,2}^{-1}(U_k) \\
& \times (\mathbf{X}_j^{(1)} - \Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)\mathbf{X}_j^{(2)}) \\
& \times K((U_i - U_k)/h) K((U_j - U_k)/h) + d_{1nu} + o_p(h^{-1/2}).
\end{aligned}$$

The remaining proof follows the same lines as those in the proof of Theorem 5. □

PROOF OF THEOREM 7. Under H_{n1} and Condition (B), applying Theorem 5, we have

$$\begin{aligned}
-\lambda_n(A_0) = & -\mu_n + v_n + v_{2n} - d_{2n} \\
& - \left[W(n)h^{-1/2}/2 + \sum_{k=1}^{n} c_n G_n^\tau(U_k)\mathbf{X}_k \varepsilon_k/\sigma^2 \right] + o_p(h^{-1/2}),
\end{aligned}$$

where $W(n)$ is defined in the proof of Theorem 5. The rest of the proof is similar to the proof of Theorem 5. The details are omitted. □

PROOF OF THEOREM 8. For brevity, we present only case I in Remark 3.5. To begin with, we note that under H_{1n}: $A = A_0 + G_n$ and under Condition (C), it follows from the Chebyshev inequality that uniformly for $h \to 0$, $nh^{3/2} \to \infty$,

$$\begin{aligned}
-\lambda_n(A_0)\sigma^2 = & -\mu_n\sigma^2 - \sigma^2 W(n)h^{-1/2}/2 - \sqrt{nEG_n^\tau(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n(U)}\, O_p(1) \\
& - \frac{n}{2} EG_n^\tau(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n(U)(1 + o_p(1)) - R_{n1} + R_{n2} + R_{n3},
\end{aligned}$$

where μ_n, $W(n)$, R_{ni}, $i = 1, 2, 3$ are defined in the proof of Theorem 5 and its associated lemmas, and $o_p(1)$ and $O_p(1)$ are uniform in $G_n \in \mathscr{G}_n$ in a sense similar to that in Lemma 7.2. Thus,

$$\begin{aligned}\beta(\alpha, G_n) &= P\{\sigma_n^{-1}(-\lambda_n(A_0) + \mu_n) \geq c(\alpha)\}\\ &= P\bigg\{\sigma_n^{-1}\bigg[-W(n)h^{-1/2}/2 - \\ &\qquad\qquad \bigg(R_{n1} - R_{n2} - R_{n3}\\ &\qquad\qquad\qquad + \frac{n}{2}EG_n^\tau(U)^\tau \mathbf{X}\mathbf{X}^\tau G_n(U)(1+o_p(1))\bigg)/\sigma^2\bigg] \geq c(\alpha)\bigg\}\\ &= P_{1n} + P_{2n}\end{aligned}$$

with

$$\begin{aligned}P_{1n} &= P\{\sigma_n^{-1}(-W(n)h^{-1/2}/2) + n^{1/2}h^{5/2}b_{1n} + nh^{9/2}b_{2n} - nh^{1/2}b_{3n} \geq c(\alpha),\\ &\qquad\qquad |b_{1n}| \leq M, |b_{2n}| \leq M\},\\ P_{2n} &= P\{\sigma_n^{-1}(-W(n)h^{-1/2}/2) + n^{1/2}h^{5/2}b_{1n} + nh^{9/2}b_{2n} - nh^{1/2}b_{3n} \geq c(\alpha),\\ &\qquad\qquad |b_{1n}| > M, |b_{2n}| > M\}\end{aligned}$$

and

$$\begin{aligned}b_{1n} &= (n^{1/2}h^{5/2}\sigma_n\sigma^2)^{-1}(-R_{n1} + R_{n2}),\\ b_{2n} &= (nh^{9/2}\sigma_n\sigma^2)^{-1}R_{n3},\\ b_{3n} &= (h^{1/2}\sigma_n\sigma^2)^{-1}\tfrac{1}{2}EG_n^\tau(U)^\tau\mathbf{X}\mathbf{X}^\tau G_n(U)(1+o_p(1)).\end{aligned}$$

When $h \leq c_0^{-1/2}n^{-1/4}$, we have

$$n^{1/2}h^{5/2} \geq c_0 nh^{9/2}, \quad n^{1/2}h^{5/2} \to 0, \quad nh^{9/2} \to 0.$$

Thus for $h \to 0$ and $nh \to \infty$, it follows from Lemma 7.2 that $\beta(\alpha, \rho) \to 0$ only when $nh^{1/2}\rho^2 \to -\infty$. It implies that $\rho_n^2 = n^{-1}h^{-1/2}$ and the possible minimum value of ρ_n in this setting is $n^{-7/16}$. When $nh^4 \to \infty$, for any $\delta > 0$, applying Lemma 7.2, we find a constant $M > 0$ such that $P_{2n} < \delta/2$ uniformly in $G_n \in \mathscr{G}_n$. Then

$$\beta(\alpha, \rho) \leq \delta/2 + P_{1n}.$$

Note that $\sup_{\mathscr{G}_n(\rho)} P_{1n} \to 0$ only when $B(h) = nh^{9/2}M - nh^{1/2}\rho^2 \to -\infty$. $B(h)$ attains the minimum value $-\frac{8}{9}(9M)^{-1/8}n\rho^{9/4}$ at $h = (\rho^2/(9M))^{1/4}$. Now it is easily shown that in this setting the corresponding minimum value of ρ_n is $n^{-4/9}$ with $h = c_* n^{-2/9}$ for some constant c_*. $\square$

PROOF OF THEOREM 9. Let c denote a generic constant. Then, under H_0,

$$\mathrm{RSS}_0 - \mathrm{RSS}_1 = -D_1 - D_2,$$

where $D_1 = \varepsilon^\tau P_{\mathbf{X}_D} \varepsilon$, $\mathbf{X}_D$ is the design matrix with the ith row $\mathbf{X}_i^\tau$ ($i = 1, \dots, n$), $P_{\mathbf{X}_D}$ is the projection matrix of $\mathbf{X}_D$ and

$$D_2 = \sum_{i=1}^n (A(U_i) - \hat{A}(U_i))^\tau \mathbf{X}_i \mathbf{X}_i^\tau (A(U_i) - \hat{A}(U_i)) + 2\sum_{i=1}^n \varepsilon_i (A(U_i) - \hat{A}(U_i))^\tau \mathbf{X}_i.$$

The proof will be completed by showing the following four steps.

(i) $D_1 = O_p(1)$,

(ii) $-\sqrt{h} D_2 = (D/\sqrt{h}) + W(n) + o_p(1)$,

(iii) $W(n) = (\sqrt{h}/n) \sum_{j \neq l} \varepsilon_j \varepsilon_l [2K_h(U_j - U_l) - K_h * K_h(U_j - U_l)] \times \mathbf{X}_j^\tau \Gamma(U_l)^{-1} \mathbf{X}_l \xrightarrow{\mathscr{L}} N(0, V)$,

(iv) $\mathrm{RSS}_1/n = E\sigma^2(\mathbf{X}, U) + O_p(1/\sqrt{n}) + O_p(1/(nh))$,

with

$$D = [2K(0) - K * K(0)] \int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})\, du - \frac{1}{nh} K^2(0) E[(\mathbf{X}^\tau \Gamma(U)^{-1}\mathbf{X})^2 \sigma^2(\mathbf{X}, U)],$$

$$V = 2\int [2K(x) - K * K(x)]^2 dx \int_\Omega \mathrm{tr}(\Gamma^*(u)\Gamma(u)^{-1})^2\, du.$$

It follows from Lemma 7.1 that

$$E[(\varepsilon^\tau P_{\mathbf{X}_D}\varepsilon)^2 | (\mathbf{X}_1, U_1), \dots, (\mathbf{X}_n, U_n)] \le c\ \mathrm{tr}(P^2_{\mathbf{X}_D}) + c[\mathrm{tr}(P_{\mathbf{X}_D})]^2 = p(p+1)c,$$

which implies (i). The proofs of (ii) and (iii) are the same as the proof of Theorem 5. The details are omitted. The last step follows from $\mathrm{RSS}_1 = \sum_{i=1}^n \varepsilon_i^2 + D_2$. Using the inequality $x/(1+x) \le \log(1+x) \le x$ for $x > -1$, we have

$$\lambda_n = \frac{n}{2}\left[\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} + O_p(n^{-2}h^{-2})\right] = \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} + O_p(n^{-1}h^{-2}). \quad \square$$

Before proving Theorem 10, we introduce the following lemma.

LEMMA 7.5. *Under Condition* (A1)–(A3) *and* (B1)–(B3), $n^{(\xi-1)/\xi} h \ge c_0 \times (\log n)^\delta$ *and* $\delta > (\xi - 1)/(\xi - 2)$, *we have*

$$\hat{A}(u_0) - A(u_0) = r_n^2 \widetilde{\Gamma}(u_0)^{-1} \sum_{i=1}^n q_1(A(U_i)^\tau \mathbf{X}_i, Y_i) \mathbf{X}_i K((U_i - u_0)/h)(1 + o_p(1)) + H_n(u_0),$$

where $r_n = 1/\sqrt{nh}$,

$$H_n(u_0) = r_n^2 \widetilde{\Gamma}(u_0)^{-1} \sum_{i=1}^{n} [q_1(\beta(u_0)^\tau \mathbf{Z}_i, Y_i) - q_1(A(U_i)^\tau \mathbf{X}_i, Y_i)] \mathbf{X}_i K((U_i - u_0)/h)(1 + o_p(1))$$

and $o_p(1)$ *is uniform with respect to* u_0.

PROOF. It follows from some arguments similar to Carroll, Fan, Gijbels and Wand (1997) and Zhang and Gijbels (1999). □

PROOF OF THEOREM 10. Let $\varepsilon_i = q_1(A_0(U_i)^\tau \mathbf{X}_i, Y_i)$. Using the Taylor expansion of $\lambda_{ng}(A_0)$ and Lemma 7.5, we obtain

$$\begin{aligned}\lambda_{ng}(A_0) = &-r_n^2 \sum_{k=1}^{n} \sum_{i=1}^{n} \varepsilon_k \varepsilon_i \mathbf{X}_i^\tau \widetilde{\Gamma}(u_k)^{-1} \mathbf{X}_k - R_{n1g} \\ &- \frac{r_n^4}{2} \sum_{k=1}^{n} \sum_{i,j} q_2(A_0(U_k)^\tau \mathbf{X}_k, Y_k) \varepsilon_i \varepsilon_j \widetilde{\Gamma}(U_k)^{-1} \mathbf{X}_i \mathbf{X}_k \mathbf{X}_k^\tau \widetilde{\Gamma}(U_k)^{-1} \\ &\times \mathbf{X}_j K((U_i - U_k)/h) K((U_j - U_k)/h) + R_{n2g} + R_{n3g},\end{aligned}$$

where

$$R_{n1g} = r_n^2 \sum_{k=1}^{n} \varepsilon_k H_n(U_k) \mathbf{X}_k,$$

$$R_{n2g} = -r_n^2 \sum_{k=1}^{n} \sum_{i=1}^{n} \varepsilon_i \mathbf{X}_i^\tau \widetilde{\Gamma}(U_k)^{-1} \mathbf{X}_k \mathbf{X}_k^\tau H_n(U_k),$$

$$R_{n3g} = -\frac{r_n^4}{2} \sum_{k=1}^{n} q_2(A_0(U_k)^\tau \mathbf{X}_k, Y_k) H_n(U_k)^\tau \mathbf{X}_k \mathbf{X}_k^\tau H_n(U_k).$$

The remaining proof is almost the same as that of Theorem 5 if we invoke the following equalities:

$$E[\varepsilon_i | (\mathbf{X}_i, U_i)] = 0, \quad E[\varepsilon_i^2 | (\mathbf{X}_i, U_i)] = -E[q_2(A_0(U_i)^\tau \mathbf{X}_i), Y_i) | (\mathbf{X}_i, U_i)]. \quad \square$$

Acknowledgments. The authors thank an Associate Editor and an anonymous referee for their constructive comments. The work of C. Zhang was based on part of her dissertation at the University of North Carolina–Chapel Hill under J. Fan's direction. She appreciates his kind guidance and encouragement. J. Zhang's research was partially conducted while visiting the Department of Statistics, University of California at Los Angeles. He is grateful for Professor Wing-Hung Wong's support.

REFERENCES

AERTS, M., CLAESKENS, G. and HART, J. D. (1999). Testing the fit of a parametric function. *J. Amer. Statist. Assoc.* **94** 869–879.

AZZALINI, A. and BOWMAN, A. N. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55** 549–557.

AZZALINI, A., BOWMAN, A. N. and Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76** 1–11.

BICKEL, P. J. and RITOV, Y. (1992). Testing for goodness of fit: a new approach. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.) 51–57. North-Holland, New York.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviation of density function estimates. *Ann. Statist.* **1** 1071–1095.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

CAI, Z., FAN, J. and LI, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* To appear.

CARROLL, R. J., FAN, J., GIJBELS, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489.

CHEN, J. H. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80** 107–116.

CLEVELAND, W. S. and DEVLIN, S. J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 597–610.

CLEVELAND, W. S., GROSSE, E. and SHYU, W. M. (1992). Local regression models. In *Statistical Models in S* (J. M. Chambers and T. J. Hastie, eds.) 309–376. Wadsworth and Brooks Cole, Pacific Grove, CA.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

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

EUBANK, R. L. and LARICCIA, V. M. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20** 2071–2086.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.* **21** 196–216.

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.

FAN, J. and HUANG, L. (1998). Goodness-of-fit test for parametric regression models. Technical Report, Dept. Statistics, Univ. California, Los Angeles.

FAN, J. and ZHANG, J. (1999). Sieve empirical likelihood ratios for nonparametric functions. Unpublished manuscript.

HALL, P. and OWEN, A. B. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Statist.* **2** 273–289.

HÄRDLE, W. and 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*. Springer, New York.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

HASTIE, T. J. and TIBSHIRANI, R. J. (1993). Varying-coefficient models (with discussion). *J. Royal Statist. Soc. Ser. B* **55** 757–796.

HUBER, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

INGLOT, T., KALLENBERG, W. C. M. and LEDWINA, T. (1994). Power approximations to and power comparison of smooth goodness-of-fit tests. *Scand. J. Statist.* **21** 131–145.

INGLOT, T. and LEDWINA, T. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *Ann. Statist.* **24** 1982–2019.

INGSTER, YU. I. (1993). Asymptotic minimax hypothesis testing for nonparametric alternatives I-III. *Math. Methods Statist.* **2** 85–114; **3** 171–189; **4** 249–268.

KALLENBERG, W. C. M. and LEDWINA, T. (1997). Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* **92** 1094–1104.

KOROLJUK, V. S. and BOROVSKICH, YU. V. (1994). *Theory of U- Statistics.* Kluwer, Amsterdam.

KUCHIBHATLA, M. and HART, J. D. (1996). Smoothing-based lack-of-fit tests: variations on a theme. *J. Nonparameter. Statist.* **7** 1–22.

LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli* **5** 333–358.

LI, G., HOLLANDER, M., MCKEAGUE, I. W. and YANG, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Ann. Statist.* **24** 628–640.

MURPHY, S. A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scand. J. Statist.* **20** 35–50.

NEYMAN, J. (1937). Smooth test for goodness of fit. *Skand. Aktuar. J.* **20** 149–199.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

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

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

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366.

RANDLE, D. H. and WOLFE, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.

SEBER, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

SHEN, X., SHI, J. and WONG, W. H. (1999). Random sieve likelihood and general regression models. *J. Amer. Statist. Assoc.* **94** 835–846.

SEVERINI, T. A. and WONG, W. H. (1992). Generalized profile likelihood and conditional parametric models. *Ann. Statist.* **20** 1768–1802.

SILVERMAN, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** 898–916.

SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498.

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.

ZHANG, J. and GIJBELS, I. (1999). Sieve empirical likelihood and extensions of generalized least squares. Discussion paper, Institute of Statistics, Univ. catholique de Louvain.

J. FAN
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN
HONG KONG
E-MAIL: jfan@sta.cuhk.edu.hk

C. ZHANG
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706
E-MAIL: cmzhang@stat.wisc.edu

J. ZHANG
EURANDOM
DEN DOLECH 2
5612 AZ, EINDHOVEN
THE NETHERLANDS
E-MAIL: jzhang@euridice.tue.nl