The Canadian Journal of Statistics
Vol. 31, No. 2, 2003, Pages 151–171
La revue canadienne de statistique

Adaptive tests of regression functions via multiscale generalized likelihood ratios

Chunming M. ZHANG

Key words and phrases: Adaptive Neyman test; bandwidth selection; generalized likelihood ratio; local polynomial regression; nonparametric test.

MSC 2000: Primary 62G10; secondary 62G05, 62G20.

Abstract: Many applications of nonparametric tests based on curve estimation involve selecting a smoothing parameter. The author proposes an adaptive test that combines several generalized likelihood ratio tests in order to get power performance nearly equal to whichever of the component tests is best. She derives the asymptotic joint distribution of the component tests and that of the proposed test under the null hypothesis. She also develops a simple method of selecting the smoothing parameters for the proposed test and presents two approximate methods for obtaining its P-value. Finally, she evaluates the proposed test through simulations and illustrates its application to a set of real data.

Tests adaptatifs de fonctions de régression fondés sur des rapports de vraisemblances multi-échelle généralisés

Résumé : Moult applications des tests non paramétriques basés sur l'estimation de courbes font intervenir un paramètre de lissage. L'auteure propose un test adaptatif qui allie plusieurs tests du rapport de vraisemblances généralisés et rivalise de puissance avec le meilleur d'entre eux. Elle détermine la loi asymptotique conjointe des tests individuels et celle du test global sous l'hypothèse nulle. Elle montre aussi comment sélectionner facilement les paramètres de lissage du test global et propose deux méthodes de calcul approché de son seuil. Elle examine en outre le comportement du test proposé par voie de simulations et en illustre l'emploi dans un cas concret.

1. INTRODUCTION

Recently a series of model specification tests have been developed that combine nonparametric regression techniques, or "scatterplot smoothing." Without assuming any particular parametric form of the underlying regression function, nonparametric tests overcome the drawbacks of conventional parametric tests by broadening the scope of applications and guarding against modelling biases. Regardless of which nonparametric smoothing technique is employed, nonparametric tests demand suitable selection of the smoothing parameter. Examples of the smoothing parameter include the order in orthogonal series and thresholding approaches, the penalty factor in smoothing splines, and the bandwidth parameter in kernel and local polynomial regression, among others. Since the treatments can be adjusted analogously to tests based on other nonparametric estimation methods, in this paper we shall focus on bandwidth selection for local polynomial based tests.

There are many smoothing parameter selectors available in the literature. These include cross-validation (Stone 1974), generalized cross-validation (Wahba 1977), the pre-asymptotic substitution method (Fan & Gijbels 1995), the plug-in method (Ruppert, Sheather & Wand 1995), and many others. Unfortunately, straightforward applications of these types of approaches to the context of nonparametric hypothesis testing will encounter difficulties. This is mainly due to the following two reasons. First, as Ingster (1982) showed, the optimal rate for the smoothing parameter for nonparametric testing is different from that for nonparametric function estimation. In the latter context, the smoothing criterion is based on minimizing mean (integrated) squared errors, thus balancing the trade-off between bias and variance of the nonparametric estimator. In the former context, on the other hand, the optimal smoothing rule is defined so that contiguous alternatives with the fastest possible rate of convergence to the null can be detected consistently, thus rendering a most powerful test. In the practical applications of smoothing-based tests, however, Ingster's work has not been taken into account. Secondly, the best constant associated with the optimal rate of smoothing parameter for nonparametric test is generally not available, and cannot be determined from the formulation and derivation of the resulting optimal rate. This stands in stark contrast to nonparametric curve fitting, for which the best constant in the optimal rate for the smoothing parameter can be derived analytically and estimated consistently. A rate-optimal test, in the sense of Ingster (1982), was recently proposed in Fan, Zhang & Zhang (2001), based on the generalized likelihood ratio (GLR) statistic combined with the local polynomial smoother. While theoretical properties of the GLR test were justified asymptotically in the same paper for a variety of useful models, including the Gaussian white noise model, nonparametric regression model, varying coefficient model and generalized varying coefficient model, little attention has been focused on examining the finite-sample properties and smoothing parameter selection of the GLR test. These two reasons motivate our theoretical and methodological study here of multiscale generalized likelihood ratio.

An alternative approach for choosing the amount of smoothing consists of repeating the corresponding nonparametric test procedure across several selected values of a smoothing parameter. King, Hart & Wehrly (1991) and Azzalini & Bowman (1993) recommended plotting the observed significance against a range of smoothing parameters. They observed from the "significance trace" a remarkable stability of the P-values over a grid of smoothing parameters, except those over extremely small values, and concluded that the smoothing parameter is less important in nonparametric tests. Similar observations were reported in Härdle & Mammen (1993), Young & Bowman (1995), Bowman & Young (1996) and Kauermann & Tutz (1999), but no further convincing explanation has been given. On the other hand, Firth, Glosup & Hinkley (1991) discussed via asymptotic expansion how the smoothing parameter will affect the power of a test, and emphasized the need to incorporate an empirically chosen smoothing parameter into test procedures. Likewise, a simulation study which examines the effect of smoothing parameter on power was given in Raz (1990). Since one may not, in general, be informed of the kind of alternatives expected, it is desirable to develop nonparametric tests which have high power against a broad class of alternatives.

Our approach can briefly be described as follows. We first consider the GLR tests with multiple smoothing parameters, say $h_1, \ldots, h_J$, which span a wide range of reasonable smoothing parameters, yet are optimal in the rate of Ingster (1982) and Fan, Zhang & Zhang (2001). Depending on the nature of the alternative model [data], the power [P-value] of the GLR test will vary with different values of h_j. We shall take the maximum of the corresponding (normalized) GLR tests as a new test procedure, called a multiscale generalized likelihood ratio (MGLR) test. Our simulations show that the GLR test, with a single smoothing parameter, will usually suffer from power loss, while the discriminating power of the MGLR test is always close to that of the best GLR test, which uses the favorable but unknown scale of smoothing parameter. Namely, the MGLR test is nearly as powerful as if the GLR test with an unknown optimal smoothing parameter were used. This is the adaptive feature enjoyed by the MGLR test. Furthermore, the power of the MGLR test is competitive with the power of the adaptive Neyman test (Fan & Huang 2001), which does not depend on the degree of smoothness of the underlying regression function and theoretically achieves the adaptively optimal rate of convergence of nonparametric hypothesis testing. Compared with the adaptive Neyman test (ANT), the null distribution of the MGLR statistic can be approximated more accurately and thus enables one to obtain the P-value for the observed data more precisely. Moreover, our methodology of the MGLR test can be modified in a straightforward way to handle other more complicated models even with heteroscedasticity.

The rest of the article is organized as follows. In Section 2, we formulate the MGLR test statistic and derive its asymptotic null distribution. Specifically, we derive the asymptotic expressions for the correlation coefficients between the GLR statistics; our results explain the aforementioned stability in the significance trace. Issues about calibrating the P-value and level-α critical value associated with the MGLR test are studied in Section 3. In Section 4, we propose an empirical method of selecting smoothing parameters for the GLR and MGLR tests. In Section 5, we report on simulation evaluations of the powers of the MGLR test in comparison with the F-test and ANT. We apply the MGLR test to a real-data example in Section 6. A summary and concluding remarks are provided in Section 7. Technical proofs are postponed to the Appendix.

2. TESTS OF REGRESSION FUNCTION

2.1. *Background.*

We first briefly outline the GLR test proposed in Fan, Zhang & Zhang (2001). Suppose we are given independent observations, $(X_1, Y_1), \ldots, (X_n, Y_n)$, from a nonparametric regression model,

$$Y = m(X) + \varepsilon, \tag{1}$$

where, conditional on the predictor variable X, the error ε has a normal distribution with mean zero and unknown variance σ^2. The mean regression function, $m(x) = \mathrm{E}(Y \mid X = x)$, is assumed to belong to $\mathcal{M}$, a smooth functional space. Let $\Pi_k = \{\theta_0 + \theta_1 x + \cdots + \theta_k x^k : \boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)' \in \mathbb{R}^{k+1}\}$ denote the set of polynomial regression functions of degree k, where the superscript $'$ stands for the transpose of a vector or matrix. Suppose we are interested in testing

$$\mathcal{H}_0 : m(x) \in \Pi_k \quad \text{versus} \quad \mathcal{H}_1 : m(x) \in \mathcal{M}\backslash\Pi_k. \tag{2}$$

To derive the GLR statistic, consider the conditional log-likelihood function from (1), expressed as

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

Let $m_{\widehat{\boldsymbol{\theta}}}(\,\cdot\,)$ stand for the maximum likelihood estimator (MLE) under $\mathcal{H}_0$, in which $\widehat{\boldsymbol{\theta}}$ denotes the MLE of the unknown parameter $\boldsymbol{\theta}$. Usually, the MLE of $m(\cdot)$ will not exist under $\mathcal{H}_1$. In such instances, one could carry out a nonparametric fit, for example the pth degree local polynomial estimate (Fan & Gijbels 1996), which is denoted by $\widehat{m}_h(\,\cdot\,)$. That is, at a fitting point x, $\widehat{m}_h(x)$ is the estimated intercept $\widehat{\beta}_0$, where $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_p)'$ minimizes the weighted sum of squared residuals,

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(X_i - x) - \cdots - \beta_p(X_i - x)^p\}^2 K\{(X_i - x)/h\}.$$

Here, the smoothing parameter $h > 0$ is the bandwidth which governs the size of the local neighbourhood, and K is called the kernel function. Denote by RSS_0 the residual sum of squares under $\mathcal{H}_0$, and by $\mathrm{RSS}_1(h)$ under $\mathcal{H}_1$; that is, $\mathrm{RSS}_0 = \sum_{i=1}^n \{Y_i - m_{\widehat{\boldsymbol{\theta}}}(X_i)\}^2$, and $\mathrm{RSS}_1(h) = \sum_{i=1}^n \{Y_i - \widehat{m}_h(X_i)\}^2$. Then the logarithm of the conditional nonparametric likelihood ratio statistic for (2), given by

$$\lambda_n(h) = \ell_n(H_1) - \ell_n(H_0) = (n/2) \log\{\mathrm{RSS}_0/\mathrm{RSS}_1(h)\}, \tag{3}$$

is called a *GLR statistic.* It is asymptotically equivalent to the Azzalini & Bowman (1993) statistic. We make another remark here; that is, even if we drop the normality assumption in (1), $\lambda_n(h)$ itself as a valid statistic can still be utilized to assess the goodness of fit of a polynomial regression, under the general nonparametric regression model,

$$Y = m(X) + \varepsilon, \quad \text{where } \mathrm{E}(\varepsilon \mid X = x) = 0 \text{ and } \mathrm{E}(\varepsilon^2 \mid X = x) = \sigma^2(x). \tag{4}$$

In this case, the conditional variance function $\sigma^2(x)$ is assumed only to be smooth.

The bandwidth h plays an important role in tuning the performances of the curve estimator $\widehat{m}_h(\cdot)$ and the resulting test statistic $\lambda_n(h)$. In case of extremely undersmoothing, i.e., $h \to 0$, it follows from (3) that $\lambda_n(h) \to +\infty$, whereas at the opposite extreme of oversmoothing, namely, $h \to +\infty$, it is seen that $\lambda_n(h) \to 0$. In general, $\lambda_n(h)$ with smaller h tends to be more powerful in detecting alternative regression functions containing higher-frequency components, while $\lambda_n(h)$ with larger h tends to have better powers against lower-frequency components of the alternatives (Hart 1997, p. 160). However, in the absence of prior knowledge about the alternative models, it is not clear how to determine a suitable value of $h \in (0, +\infty)$, for which the test is sensitive to as broad a set of alternatives as possible. This indicates that one particular choice of h either is subjective or suffers from power loss; see also the simulations in Section 5 below. Inspired by the idea of ANT proposed by Fan (1996), an alternative suggestion that takes into account multiple bandwidths is based on the maximum value of the normalized GLR statistics with J distinct bandwidths, $h_1, \ldots, h_J$. We shall refer to a test formed in this manner as a multiscale generalized likelihood ratio (MGLR) test. Nonetheless, the application of the "multi-scale" adaptive version is not restricted to a GLR statistic; it can be applied analogously to other types of smoothing-based test statistics, a comprehensive survey of which can be found in Hart (1997).

Traditional kernel regression, or local constant fit ($p = 0$), has also been employed frequently for model assessment. Due to the inherent boundary bias problem introduced by kernel estimators, kernel-based test statistics need bias corrections via either the boundary modifications of Rice (1984) or employing "boundary kernels" (Gasser & Müller 1979). This approach was utilized in Azzalini, Bowman & Härdle (1989) and Huang (1997). Using boundary kernels, Hart & Wehrly (1992) proposed the data-driven bandwidth as a test statistic for assessing the adequacy of polynomial regression models. Compared with kernel smoothing of regression functions, local polynomial estimators enjoy the theoretical advantages of design-adaptation, automatic boundary correction, and minimax efficiency.

A local polynomial estimator of degree p is unbiased for a kth degree polynomial function, if $k \le p$. This flexibility makes it attractive to develop tests based on local polynomial fit. For this reason, we shall assume $p \ge k$ when conducting the MGLR test combined with the pth degree local polynomial fit. (Setting $p = k$ is convenient to avoid stronger smoothness assumptions on $m(x)$.)

2.2. *Large-sample property of the* MGLR *test.*

A sequence of GLR statistics evaluated at multiple bandwidths no longer consists of mutually independent terms. When the null hypothesis in (2) holds, however, their asymptotic joint distribution will be derived in Theorem 1 below, under the general nonparametric regression model (4). After that, we will deduce the asymptotic null distribution of the MGLR statistic. For ease of presentation, we first introduce some necessary notations. We denote by $\mathcal{K}(t;p)$ the equivalent kernel function (defined in (18) of the Appendix), induced from the pth degree local polynomial smoother with a basic kernel function $K(t)$; the dependence of $\mathcal{K}(t)$ on p will be dropped wherever this simplification is clear from the context. Define $c_{\mathcal{K}} = \mathcal{K}(0) - 2^{-1}\mathcal{K} * \mathcal{K}(0)$ and

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

where $*$ denotes the convolution operator. For $C \neq 0$, set $\mathcal{K}_C(\cdot) = \mathcal{K}(\cdot/C)/C$. For an integer $J \ge 1$ and constants $C_{ij} > 0$, put

$$\gamma_{ij} = \frac{C_{ij}^{1/2}\int\{\mathcal{K}(t) - 2^{-1}\mathcal{K} * \mathcal{K}(t)\}\{\mathcal{K}_{C_{ij}}(t) - 2^{-1}\mathcal{K}_{C_{ij}} * \mathcal{K}_{C_{ij}}(t)\}\,dt}{\int\{\mathcal{K}(t) - 2^{-1}\mathcal{K} * \mathcal{K}(t)\}^2\,dt}, \quad 1 \le i \le j \le J. \quad (5)$$

THEOREM 1. *Assume the general nonparametric regression model* (4). *Assume*

(A1) *The marginal density* $f(x)$ *of* X *is Lipschitz continuous and bounded away from* 0. *The variable* X *has a bounded support* Ω.

(A2) *The mean regression function* $m(x)$ *has the continuous* $(p+1)$*th derivative and* $\sigma^2(x)$ *is continuous.*

(A3) *The kernel function* $K(t)$ *is a symmetric probability density function with bounded support, and is Lipschitz continuous.*

(A4) $0 < \mathrm{E}(\varepsilon^4) < \infty$.

For a finite integer $J \geq 1$, *let* $h_1 < \cdots < h_J$, *where* $h_1 = h > 0$, *and define* $C_{ij} = h_j/h_i$. *Assume that* C_{ij}, $1 \leq i < j \leq J$, *are constants. Then under the null hypothesis in* (2) *with* $k \leq p$, *as* $n \to \infty$, $h \to 0$, $nh^{3/2} \to \infty$, *it follows that*

$$\max_{1 \leq j \leq J} \frac{\Upsilon_{\mathcal{K}} \lambda_n(h_j) - \mathcal{D}_n(h_j)}{\sqrt{2\mathcal{D}_n(h_j)}} \xrightarrow{\mathcal{L}} \max_{1 \leq j \leq J} Z_j, \tag{6}$$

where $\xrightarrow{\mathcal{L}}$ *denotes converges in distribution,*

$$\Upsilon_{\mathcal{K}} = r_{\mathcal{K}} \frac{\int \sigma^2(x)\,dx}{\int \sigma^4(x)\,dx} \mathrm{E}\{\sigma^2(X)\}, \quad \mathcal{D}_n(h_j) = r_{\mathcal{K}} c_{\mathcal{K}} \frac{\{\int \sigma^2(x)\,dx\}^2}{\int \sigma^4(x)\,dx} h_j^{-1}, \tag{7}$$

and $(Z_1, \ldots, Z_J)'$ *is a mean-zero normal random vector with a correlation coefficient matrix* $\mathcal{R} = (\gamma_{ij})_{1 \leq i,j \leq J}$ *with* γ_{ij}, $1 \leq i \leq j \leq J$, *as given in* (5).

The detailed proof of this theorem is rather involved; a short sketch is given in the Appendix. For notational convenience, we denote $\{\Upsilon_{\mathcal{K}} \lambda_n(h) - \mathcal{D}_n(h)\}/\sqrt{2\mathcal{D}_n(h)}$ by GLR(h), and denote $\max_{1 \leq j \leq J}$ GLR(h_j) by MGLR. As a consequence, the MGLR test rejects $\mathcal{H}_0$ at a significance level α, if

$$\max_{1 \leq j \leq J} \mathrm{GLR}(h_j) = \max_{1 \leq j \leq J} \frac{\Upsilon_{\mathcal{K}} \lambda_n(h_j) - \mathcal{D}_n(h_j)}{\sqrt{2\mathcal{D}_n(h_j)}} \geq z_{\alpha;J},$$

where the critical value, $z_{\alpha;J}$, satisfies

$$\mathrm{P}\left\{\max(Z_1, \ldots, Z_J) \geq z_{\alpha;J}\right\} = \alpha. \tag{8}$$

Numerical calculation of $z_{\alpha;J}$ is described in Section 3. For the special case of $J = 1$ and $p = 1$, Theorem 1 reduces to a GLR test, proposed in Fan, Zhang & Zhang (2001), for assessing linearity, based on the local linear fit.

From a practical point of view, the variance function $\sigma^2(\cdot)$ is usually unknown, and thus the variance-dependent quantities in (7) need to be estimated. An appeal to Slutsky's theorem shows that $\sigma^2(\cdot)$ can be replaced by any consistent estimate of $\sigma^2(\cdot)$ without altering the conclusion of Theorem 1; see the local variance estimator of Fan & Yao (1998) based on local polynomial regression. Hence, the MGLR test can be used to make inference about a regression function in a heteroscedastic regression model.

Clearly, for a homoscedastic model (i.e., $\sigma^2(x) \equiv \sigma^2$), we obtain from (7) that $\Upsilon_{\mathcal{K}} = r_{\mathcal{K}}$ and $\mathcal{D}_n(h_j) = r_{\mathcal{K}} c_{\mathcal{K}} |\Omega| h_j^{-1}$, and therefore statement (6) simplifies to

$$\max_{1 \leq j \leq J} \frac{r_{\mathcal{K}} \lambda_n(h_j) - r_{\mathcal{K}} c_{\mathcal{K}} |\Omega| h_j^{-1}}{\sqrt{2 r_{\mathcal{K}} c_{\mathcal{K}} |\Omega| h_j^{-1}}} \xrightarrow{\mathcal{L}} \max_{1 \leq j \leq J} Z_j, \tag{9}$$

where $|\Omega|$ represents the length of Ω. Thus the null distribution of the MGLR test, in this case, is asymptotically independent of the nuisance parameters, $\boldsymbol{\theta}$ and σ^2, and the design density f.

As evidenced by Theorem 1, the normality assumption for the stochastic error ε in (4) is not required. Divergence from this assumption may deteriorate powers of some other test procedures, e.g., the normal-theory F-test and the adaptive Neyman test. On the other hand, when the errors are indeed normally distributed, the exact distributions of some proposed test statistics can be derived; see examples given in Diblasi & Bowman (1997).

TABLE 1: Kernel-dependent constants of $r_{\mathcal{K}}$ and $c_{\mathcal{K}}$ from the pth degree local polynomial fit.

	$p=0$ and $p=1$		$p=2$ and $p=3$		$p=4$ and $p=5$	
Kernel K	$r_{\mathcal{K}}$	$c_{\mathcal{K}}$	$r_{\mathcal{K}}$	$c_{\mathcal{K}}$	$r_{\mathcal{K}}$	$c_{\mathcal{K}}$
Uniform	1.1999	0.2500	1.3964	0.5625	1.4998	0.8789
Epanechnikov	2.1153	0.4500	1.9755	0.7812	1.9336	1.1043
Biweight	2.3061	0.5804	2.1283	0.9370	2.0620	1.2715
Triweight	2.3797	0.6858	2.1946	1.0682	2.1219	1.4161
Gaussian $\mathsf{N}(0, 3^{-2})$	2.5375	0.7737	2.3569	1.0812	2.2849	1.3081

To facilitate implementing the MGLR test, we tabulate in Table 1 the constants $r_{\mathcal{K}}$ and $c_{\mathcal{K}}$ for several commonly used kernel functions belonging to the "symmetric Beta family" (Fan & Gijbels 1996, p. 15),

$$K(t) = \{\text{Beta}(1/2, \ell + 1)\}^{-1}(1 - t^2)^{\ell} I(|t| \le 1), \quad \ell = 0, 1, \ldots. \tag{10}$$

There, the Uniform, Epanechnikov, Biweight, and Triweight kernels correspond to the index ℓ, in (10), equal to 0, 1, 2, and 3, respectively. Note also that, in Theorem 1, the bounded support assumption on K is merely for technical simplicity; it can possibly be relaxed. Hence, for the sake of comparison, we include the Gaussian kernel as a limit of (10) when the index ℓ tends to infinity; in particular, the Gaussian kernel with mean 0 and standard deviation 3^{-1} can be regarded as having a support comparable with the support $[-1, 1]$ of the Beta-family. The figures in Table 1 also demonstrate that the constants $r_{\mathcal{K}}$ and $c_{\mathcal{K}}$ induced from the local polynomial estimation of an odd degree $p = 2\ell + 1$ coincide with those obtained from the local polynomial estimation of an even degree $p = 2\ell$. This relation can indeed be verified analytically, as shown in the following lemma.

LEMMA 1. *Assume that $K(t)$ is a symmetric probability density function. Then $\mathcal{K}(t; 2\ell + 1) = \mathcal{K}(t; 2\ell)$, for $t \in \mathbb{R}$, and $\ell = 0, 1, \ldots$.*

2.3. Correlation structure of MGLR.

In this section, we examine the structure of $\mathcal{R}$, the correlation coefficient matrix obtained in Theorem 1. Under the null hypothesis, for $1 \le i \le j \le J$, each entry γ_{ij} asymptotically measures the degree of linear association between $\text{GLR}(h_i)$ and $\text{GLR}(h_j)$. For brevity, let us first consider the case where $p = 0$ or $p = 1$, either of which by Lemma 1 implies $\mathcal{K} = K$. In this case, the positive correlation is established in Theorem 2, under general assumptions on K. For degrees $p \ge 2$, we conjecture that the γ_{ij} are positive, under the assumption on K given in Theorem 2; numerical evaluations of the γ_{ij}, when $p = 2, 3, 4, 5$, and K is given by (10) with $\ell = 0, 1, 2, 3, \infty$, lend support to this conjecture.

THEOREM 2. *For $p = 0, 1$, if K is a symmetric unimodal probability density, then the γ_{ij} as* *defined in* (5) *are strictly positive with a lower bound equal to*

$$\frac{C_{ij}^{1/2}[\mathcal{K} * \mathcal{K}_{C_{ij}}(0) + \int \{\mathcal{K}_{C_{ij}}(t) - \mathcal{K} * \mathcal{K}_{C_{ij}}(t)\}^2 \, dt]}{\int \{2\mathcal{K}(t) - \mathcal{K} * \mathcal{K}(t)\}^2 \, dt},$$

for any $C_{ij} \geq 1$.

Applying the Cauchy–Schwartz inequality to (5), we can only deduce $|\gamma_{ij}| \leq 1$. Moreover, the use of Theorem 2 enables us to verify the condition on C_{ij} under which γ_{ij} achieves its upper bound 1.

COROLLARY 1. *For* $p = 0, 1$, *if* K *is a symmetric unimodal probability density, then for the* γ_{ij}, *defined in* (5) *with* $C_{ij} \geq 1$, *it follows that* $\gamma_{ij} = 1$ *iff* $C_{ij} = 1$ *iff* $i = j$.

For numerical illustration, we evaluate below the matrix $\mathcal{R}$ for $J = 5$, in which $p = 0$ (or $p = 1$) is used. For simplicity of implementation we choose a geometric grid of bandwidths, $\{h_j = C^{j-1}h, \ j = 1, \ldots, J, \ C \geq 1\}$, which results in bandwidth ratios $C_{ij} = C^{j-i}$. The matrix $\mathcal{R}_{1.2}$ corresponds to $C = 1.2$, whereas $\mathcal{R}_{1.5}$ corresponds to $C = 1.5$.

(1) If K is the Epanechnikov kernel function,

$$\mathcal{R}_{1.2} = \begin{bmatrix} 1 & .9655 & .8927 & .8126 & .7388 \\ .9655 & 1 & .9655 & .8927 & .8126 \\ .8927 & .9655 & 1 & .9655 & .8927 \\ .8126 & .8927 & .9655 & 1 & .9655 \\ .7388 & .8126 & .8927 & .9655 & 1 \end{bmatrix},$$

$$\mathcal{R}_{1.5} = \begin{bmatrix} 1 & .8747 & .7084 & .5765 & .4703 \\ .8747 & 1 & .8747 & .7084 & .5765 \\ .7084 & .8747 & 1 & .8747 & .7084 \\ .5765 & .7084 & .8747 & 1 & .8747 \\ .4703 & .5765 & .7084 & .8747 & 1 \end{bmatrix}.$$

For instance, the figures, .8747, .7084, .5765, and .4703, in the first row of $\mathcal{R}_{1.5}$, represent the correlation coefficients between GLR(h) and GLR$(1.5h)$, GLR$(2.25h)$, GLR$(3.375h)$, and GLR$(5.0625h)$, respectively. The lower bounds for the distinct off-diagonal entries of $\mathcal{R}_{1.2}$, according to Theorem 2, are .7204, .6787, .6337, and .5879; in $\mathcal{R}_{1.5}$, the lower bounds are .6688, .5676, .4721, and .3889.

(2) If K is the Gaussian kernel function,

$$\mathcal{R}_{1.2} = \begin{bmatrix} 1 & .9887 & .9567 & .9089 & .8512 \\ .9887 & 1 & .9887 & .9567 & .9089 \\ .9567 & .9887 & 1 & .9887 & .9567 \\ .9089 & .9567 & .9887 & 1 & .9887 \\ .8512 & .9089 & .9567 & .9887 & 1 \end{bmatrix},$$

$$\mathcal{R}_{1.5} = \begin{bmatrix} 1 & .9472 & .8236 & .6861 & .5631 \\ .9472 & 1 & .9472 & .8236 & .6861 \\ .8236 & .9472 & 1 & .9472 & .8236 \\ .6861 & .8236 & .9472 & 1 & .9472 \\ .5631 & .6861 & .8236 & .9472 & 1 \end{bmatrix}.$$

The lower bounds for the distinct off-diagonal entries of $\mathcal{R}_{1.2}$ are .7168, .6868, .6538, and .6176; in $\mathcal{R}_{1.5}$, the lower bounds are .6797, .6005, .5127, and .4280.

In summary, the matrix $\mathcal{R}$, evaluated in the cases discussed in the paragraph before Theorem 2, leads to several interesting conclusions for the joint distribution of GLR(h_j), $1 \le j \le J$, under the null hypothesis. First, GLR(h) and GLR(Ch), for any factor $C \ge 1$, are positively correlated. This is consistent with the observation that the visual difference between the estimated regression curves, with bandwidths, say h, $1.2h$ and $1.5h$, does not appear substantial. Therefore, for the same set of observed data, the associated GLR test statistics are anticipated to produce consistently large or small P-values. Secondly, the larger C is, the further h deviates from Ch, and the smaller the correlation between GLR(h) and GLR(Ch) is. Thirdly, given the same factor $C \ge 1$, a Gaussian kernel yields a slightly larger correlation between GLR(h) and GLR(Ch) than the Beta-family kernels. These three conclusions justify the empirical findings (see Introduction) that P-values tend to be stable over a range of smoothing parameters, especially when a Gaussian kernel is employed in such numerical work. For the sake of computational expediency, we shall take the Epanechnikov kernel, throughout our subsequent simulations. Fourthly, $\mathcal{R}$ provides helpful guidance on how to select the bandwidth grid $\{h_j,\ 1 \le j \le J\}$ for GLR statistics, without introducing extreme undersmoothing and oversmoothing. Typically, $J = 3$ and $J = 5$ suffice for practical implementations. Further discussions on choosing J and C are addressed in Section 3.

3. COMPUTATION METHODS FOR LEVEL-α CRITICAL VALUE AND P-VALUE

3.1. Asymptotic method: large-sample sizes.

We now discuss the computation of $z_{\alpha;J}$, specified in (8), the theoretical $100(1-\alpha)$th percentile of the MGLR statistic under the null hypothesis. As in Section 2.3, we always consider the geometric style of bandwidth grid. Using local polynomial regression, of degrees $p = 0, 1, 2, 3, 4, 5$, the quantiles are listed in Table 2. For $J = 3$, $z_{\alpha;J}$ is evaluated using the method proposed in Yang & Zhang (1997); for $J = 5$, due to the lack of available numerical procedures, $z_{\alpha;J}$ is estimated based on Monte-Carlo percentage points, using 1,000,000 samples.

TABLE 2: The $100(1-\alpha)$ th percentile $z_{\alpha;J}$ as defined in (8) associated with the pth degree local polynomial regression method, using the Epanechnikov kernel function and the geometric bandwidths $h_j = C^{j-1}h, 1 \le j \le J$.

		$p=0$ and $p=1$		$p=2$ and $p=3$		$p=4$ and $p=5$	
		$C=1.2$	$C=1.5$	$C=1.2$	$C=1.5$	$C=1.2$	$C=1.5$
J	α	$z_{\alpha;J}$	$z_{\alpha;J}$	$z_{\alpha;J}$	$z_{\alpha;J}$	$z_{\alpha;J}$	$z_{\alpha;J}$
3	.010	2.49	2.59	2.53	2.62	2.55	2.63
	.025	2.13	2.23	2.17	2.27	2.19	2.28
	.050	1.82	1.93	1.86	1.96	1.88	1.98
	.100	1.46	1.58	1.50	1.62	1.53	1.64
5	.010	2.59	2.72	2.64	2.76	2.67	2.77
	.025	2.23	2.38	2.29	2.42	2.32	2.43
	.050	1.93	2.09	1.99	2.13	2.02	2.14
	.100	1.58	1.75	1.64	1.79	1.67	1.81

Recall that in Theorem 1, the matrix $\mathcal{R} = (\gamma_{ij})$ is determined by only $J-1$ distinct entries, which are strictly positive in the cases discussed in the paragraph before Theorem 2. Hence following Slepian's theorem (Gupta 1963, p. 805), we can approximate the asymptotic P-value of the MGLR test, $\mathrm{P}\{\max(Z_1, \ldots, Z_J) > z\}$, by its lower and upper bounds given in the inequalities,

$$\begin{aligned} 1 - \int_{-\infty}^{\infty} \Phi^J\left(\frac{x\sqrt{\gamma_{\max}} + z}{\sqrt{1-\gamma_{\max}}}\right)\phi(x)\,dx &\le \mathrm{P}\left(\max_{1\le j\le J} Z_j > z\right) \\ &\le 1 - \int_{-\infty}^{\infty} \Phi^J\left(\frac{x\sqrt{\gamma_{\min}} + z}{\sqrt{1-\gamma_{\min}}}\right)\phi(x)\,dx. \end{aligned} \tag{11}$$

Here, $\gamma_{\min}$ and $\gamma_{\max}$ represent the smallest and largest off-diagonal values of γ_{ij}; $\phi(x)$ and $\Phi(x)$ denote the probability density and cumulative distribution functions of the standard normal distribution.

Plotting (omitted here) the Slepian bounds of $\mathrm{P}\{\max(Z_1, \ldots, Z_J) > z\}$ with respect to z shows that when values of z are large, the approximation of the exact tail probability by the Slepian bounds improves greatly without producing computational burden. For the same value of z, the approximation gets poorer for larger J; for the same level α of significance, the Slepian bounds when $C = 1.2$ are similar to those when $C = 1.5$. This observation also justifies the adoption of $C = 1.5$ and $J = 3$, by which the MGLR test can simultaneously adapt to broader alternatives and enhance the approximation accuracy of its quantiles.

3.2. Simulation method: small sample sizes.

For realistic finite sample sizes, the limiting normal distribution may not approximate well the null distribution of an individual GLR statistic, expressed in a quadratic form. Similarly, the distribution of the maximum of correlated normal random variables may not provide a good approximation for the null distribution of the MGLR statistic.

To deal with this problem, we propose a simulation method which consists of three steps. This method works well even for sample sizes equal to 50; see the simulation study in Section 4.

Step 1: For the original set of observations, $\{(X_i, Y_i),\ i = 1, \ldots, n\}$, obtain the bandwidths $h_1, \ldots, h_J$, using the method described in Section 4 below. Compute $\mathrm{MGLR_{obs}}$, the observed MGLR statistic.

Step 2: Choose regressor variables, X_i, $i = 1, \ldots, n$, equally spaced on the interval $[0, 1]$. (This choice can serve to ease a heavy computation.) Generate independent responses Y_i following a standard normal distribution. (This choice does not lose generality.) Based on the set of simulated data $\{(\mathsf{X}_i, \mathsf{Y}_i),\ i = 1, \ldots, n\}$, obtain $\mathrm{GLR}(h_1), \ldots, \mathrm{GLR}(h_J)$, and MGLR.

Step 3: Replicate Step 2 many times, say 1000, and obtain the estimated cutoff points of $z_{\alpha;J}$ for the MGLR statistic. Similarly, compute the proportion of times that the simulated MGLR statistics exceed $\mathrm{MGLR_{obs}}$; this yields the estimated P-value.

4. BANDWIDTH SELECTION

As discussed in the Introduction, the bandwidth h should be chosen to yield the most powerful test. In striking contrast to the extensive studies on optimal bandwidth selection in the areas of kernel density and local polynomial regression estimations, this equally important key issue in nonparametric testing has not received the attention it deserves, partly due to the difficult nature of this problem.

Recently, Fan, Zhang & Zhang (2001) showed that using the pth degree local polynomial estimation with the optimal bandwidth of rate $n^{-2/(4p+5)}$, the GLR test can detect alternatives converging to the null at the rate $n^{-2(p+1)/(4p+5)}$, which is the optimal rate of convergence of nonparametric testing (Ingster 1982; Lepski & Spokoiny 1999). Recall that $n^{-1/(2p+3)}$ is the optimal rate of bandwidth for nonparametric function estimation. Comparing the rates $n^{-2/(4p+5)}$ with $n^{-1/(2p+3)}$, we see that a powerful nonparametric test requires undersmoothing.

In order to develop a simple choice of bandwidth for conducting the (M)GLR tests, we shall first rescale the range of the observed regressor variable X to the interval $[0,1]$. Call X^* the rescaled variable. Based on the optimal rate of bandwidth above and the dispersion of X^*, we suggest a deterministic choice of bandwidth, calculated from an empirical formula,

$$h^* = \eta \times \operatorname{std}(X^*) \times n^{-2/(4p+5)}, \tag{12}$$

where η stands for a constant. According to Theorem 1, the null distribution of the (M)GLR statistic, under a homoscedastic regression model, is asymptotically independent of the nuisance parameters $\boldsymbol{\theta}$ and σ^2 and the design density f. This property enables us to simulate directly the null distribution of a GLR statistic, which incorporates h^* in (12), to compare this simulated distribution with reference to the asymptotic normal (or chi-squared) distribution, and to seek η resulting in close approximation, in terms of agreement of the type-I errors. For multiple bandwidths, we set as in Section 2.3 the geometric type of bandwidth grid, $\{\ldots, C^{-2}h^*, C^{-1}h^*, h^*, Ch^*, C^2h^*, \ldots\}$, for undersmoothing and oversmoothing, and in turn the MGLR test will be adaptive to differing patterns of the alternative models. In the simulations below, $J = 3$ and $C = 1.5$ are used.

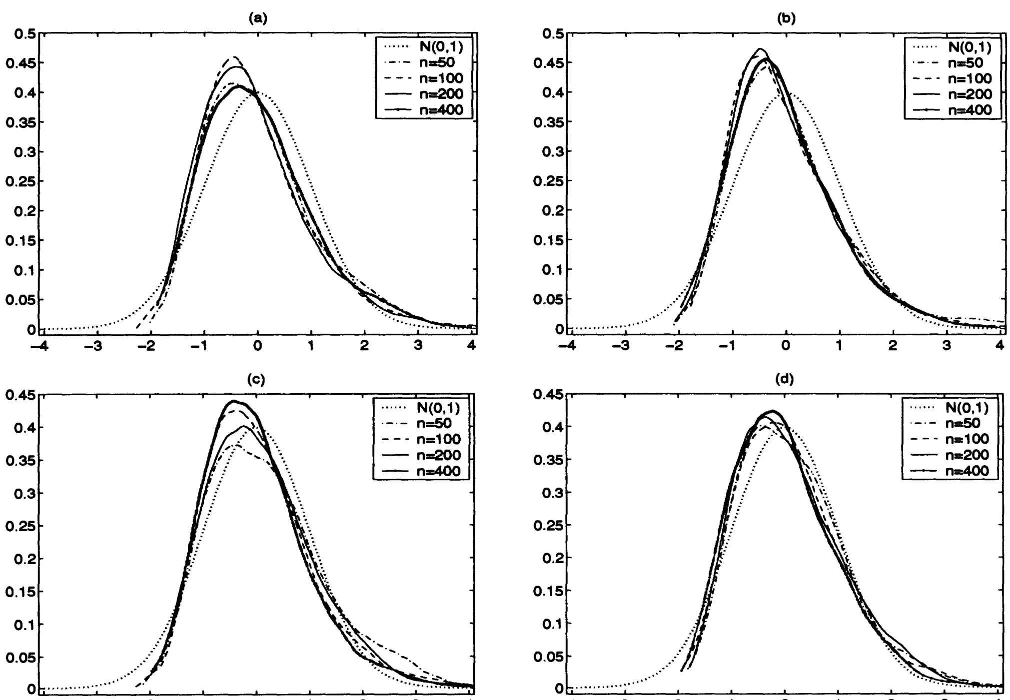

FIGURE 1: The kernel density estimate of GLR(h^*) when data are generated from model $Y = 1 + 2X + \varepsilon$, where X and ε are independent. Panel (a) – $X \sim U(-2,2)$ and $\varepsilon \sim \sigma\mathrm{N}(0,1)$; (b) – $X \sim U(-2,2)$ and $\varepsilon \sim \sigma\{\mathrm{Beta}(2,3) - .4\}$; (c) – $X \sim \mathrm{N}(0,1)$ and $\varepsilon \sim \sigma\mathrm{N}(0,1)$; (d) – $X \sim \mathrm{N}(0,1)$ and $\varepsilon \sim \sigma\{\mathrm{Beta}(2,3) - .4\}$. The variance of the error ε is determined such that the signal-to-noise ratio equals 4.

To illustrate this procedure, in a first series of simulations, 1000 replicates of observations $\{(X_i, Y_i),\ i = 1, \ldots, n\}$ are generated from a simple linear regression model,

$$Y = 1 + 2X + \varepsilon, \tag{13}$$

where X and ε are independent. In this example, the null distributions of the GLR(h^*) and MGLR tests for linearity are examined. Four values of the sample size n, 50, 100, 200, and 400, are considered. To conduct the GLR(h^*) test combined with local linear fitting, we set $\eta = 1.5$ and $p = 1$ in (12). Figure 1 presents the kernel density estimate (Fan & Gijbels 1996, p. 47) of the test statistic GLR(h^*). Among the four panels, case (a) represents $X \sim \mathsf{U}(-2, 2)$ and $\varepsilon \sim \sigma\mathsf{N}(0, 1)$; case (b) represents $X \sim \mathsf{U}(-2, 2)$ and $\varepsilon \sim \sigma\{\text{Beta}(2, 3) - .4\}$; case (c) represents $X \sim \mathsf{N}(0, 1)$ and $\varepsilon \sim \sigma\mathsf{N}(0, 1)$; case (d) represents $X \sim \mathsf{N}(0, 1)$ and $\varepsilon \sim \sigma\{\text{Beta}(2, 3) - .4\}$. In each case, σ is determined such that the signal-to-noise ratio, defined by $\text{var}\{m(X)\}/\text{var}(\varepsilon)$, equals 4. Based on the sample standard deviation of X^*, we take $\text{std}(X^*) = .29$, when $X \sim \mathsf{U}(-2, 2)$; when $X \sim \mathsf{N}(0, 1)$, we take $\text{std}(X^*)$ equal to .22, .20, .18 and .17 for sample size equal to 50, 100, 200 and 400, respectively. All plots show that, in the presence of either Gaussian or non-Gaussian errors, the simulated null distributions of GLR(h^*) are not well approximated by normal distributions. However, the approximation by a distribution, $(\chi^2_{\text{df}} - \text{df})/(2\text{df})^{1/2}$, is good, where χ^2_{df} has a chi-squared distribution with degrees of freedom equal to $r_K c_K / h^*$; see also (9).

Table 3 summarizes the proportion of rejections in 1000 samples. There, the cutoff points of GLR(h^*) are from those of the chi-squared distribution, whereas the cutoff points of MGLR use the approximate $z_{\alpha;3}$ described in Section 3.2. We observe from Table 3 that our suggested methods for approximating the cutoff points of GLR and MGLR tests perform reasonably well. On the contrary, the asymptotic null distribution (type I extreme-value distribution) of ANT cannot provide a good approximation to the finite-sample distribution of ANT even with sample sizes as large as $n = 800$ (Fan 1996; Fan & Huang 2001). This drawback hampers the applicability of ANT to real data inference problems.

To assess whether the choice $\eta = 1.5$ works well with other degrees p in the empirical formula (12), in the second simulation series we generate observations from a quadratic regression model,

$$Y = 1 + 2X + 3X^2 + \varepsilon,$$

where X and ε have the same distributional specifications as given in the four cases above. This time, local quadratic fits of degree $p = 2$ are conducted. Again we find that with $\eta = 1.5$, the χ^2 approximation to the null distribution of GLR(h^*) is satisfactory, and therefore this choice is adopted throughout the remaining simulations.

5. POWER COMPARISON

To investigate both the size and power of the (M)GLR tests in finite samples, we conduct a small sample simulation, based on three models studied in Fan & Huang (2001). They are

Example 1: $Y = 1 + \theta X^2 + \varepsilon$, $\theta \in [0, 1]$, where $X \sim \mathsf{U}(-2, 2)$ and $\varepsilon \sim \mathsf{N}(0, 1)$;

Example 2: $Y = 1 + \cos(\theta X \pi) + \varepsilon$, $\theta \in [0, 10]$, where $X \sim \mathsf{N}(0, 1)$ and $\varepsilon \sim \mathsf{N}(0, 1)$;

Example 3: $Y = 10\{1 + \theta \exp(-2X)\}^{-1} + \varepsilon$, $\theta \in [0, 2]$, where $X \sim \mathsf{N}(0, 1)$ and $\varepsilon \sim \mathsf{N}(0, 1)$.

In each of the three examples, we test the hypothesis that the real regression function $m(x)$ is linear, and it is assumed that X and ε are independent. We also include the parametric F-test for linearity versus quadratic nonlinearity. The paper by Fan & Huang (2001) has shown the adaptive optimality property of the ANT and demonstrated its power advantages over many other useful thresholding-style nonparametric tests. Regarding smoothing-based nonparametric tests, our present paper focuses more on examining the multiscale version of a given GLR test, for which a simple smoothing rule is developed in Section 4; however, a study of the amount of smoothing incorporated in other kernel-type nonparametric tests is not available in the literature.

TABLE 3: Simulated rejection probabilities of the GLR(h^*) and MGLR tests for linearity of model (13). In each case, 1000 replications are used. The cutoff points of the GLR(h^*) test are from those of $\chi^2_{r_K c_K / h^*}$, whereas the cutoff points of the MGLR test use the approximate $z_{\alpha;3}$ described in Section 3.2.

Case	Test statistic	n	$\alpha = .01$	$\alpha = .025$	$\alpha = .05$	$\alpha = .10$
(a)	GLR(h^*)	50	.011	.031	.058	.117
		100	.009	.022	.052	.096
		200	.012	.025	.045	.084
		400	.009	.023	.043	.083
	MGLR	50	.009	.017	.041	.099
		100	.008	.037	.061	.117
		200	.017	.027	.060	.113
		400	.012	.027	.048	.103
(b)	GLR(h^*)	50	.019	.032	.054	.105
		100	.003	.014	.038	.102
		200	.013	.025	.053	.091
		400	.004	.014	.042	.086
	MGLR	50	.029	.040	.058	.109
		100	.005	.023	.049	.133
		200	.013	.031	.057	.109
		400	.007	.027	.061	.114
(c)	GLR(h^*)	50	.015	.049	.084	.144
		100	.016	.039	.070	.134
		200	.009	.023	.048	.110
		400	.011	.025	.049	.106
	MGLR	50	.021	.042	.078	.135
		100	.013	.037	.080	.134
		200	.017	.032	.061	.153
		400	.013	.027	.062	.135
(d)	GLR(h^*)	50	.008	.025	.060	.114
		100	.010	.028	.053	.117
		200	.010	.026	.051	.113
		400	.006	.021	.053	.103
	MGLR	50	.008	.022	.042	.089
		100	.009	.030	.064	.119
		200	.009	.029	.057	.141
		400	.020	.044	.076	.144

For the purpose of illustration, we shall only make a power comparison between the (M)GLR, ANT, and F tests. To make the (M)GLR and ANT tests have significance level $\alpha = 5\%$ we generate 10,000 independent samples $\{(X_i, Y_i), i = 1, \ldots, n\}$ of sizes $n = 50$ and $n = 100$ from each null model, with the index $\theta = 0$; the critical values of these tests are determined by their 95th sample percentiles of the test statistics across 10,000 samples. For the (M)GLR test, three bandwidths $h_1 = 1.5^{-1}h^*$, $h_2 = h^*$, and $h_3 = 1.5h^*$ are employed in the local linear smoothing. The empirical powers are estimated by the proportion of observed rejections in 500 samples of size n and are plotted in Figure 2 (left panels for $n = 50$, and right panels for $n = 100$). In all cases, the proposed tests hold their nominal levels well. Our simulations reveal that in Examples 1 and 3, the GLR(h_2) test is about as powerful as the MGLR test, but in Example 2 falls far behind the MGLR test. Thus for brevity of exposition, power curves by the GLR(h_2) test are omitted.

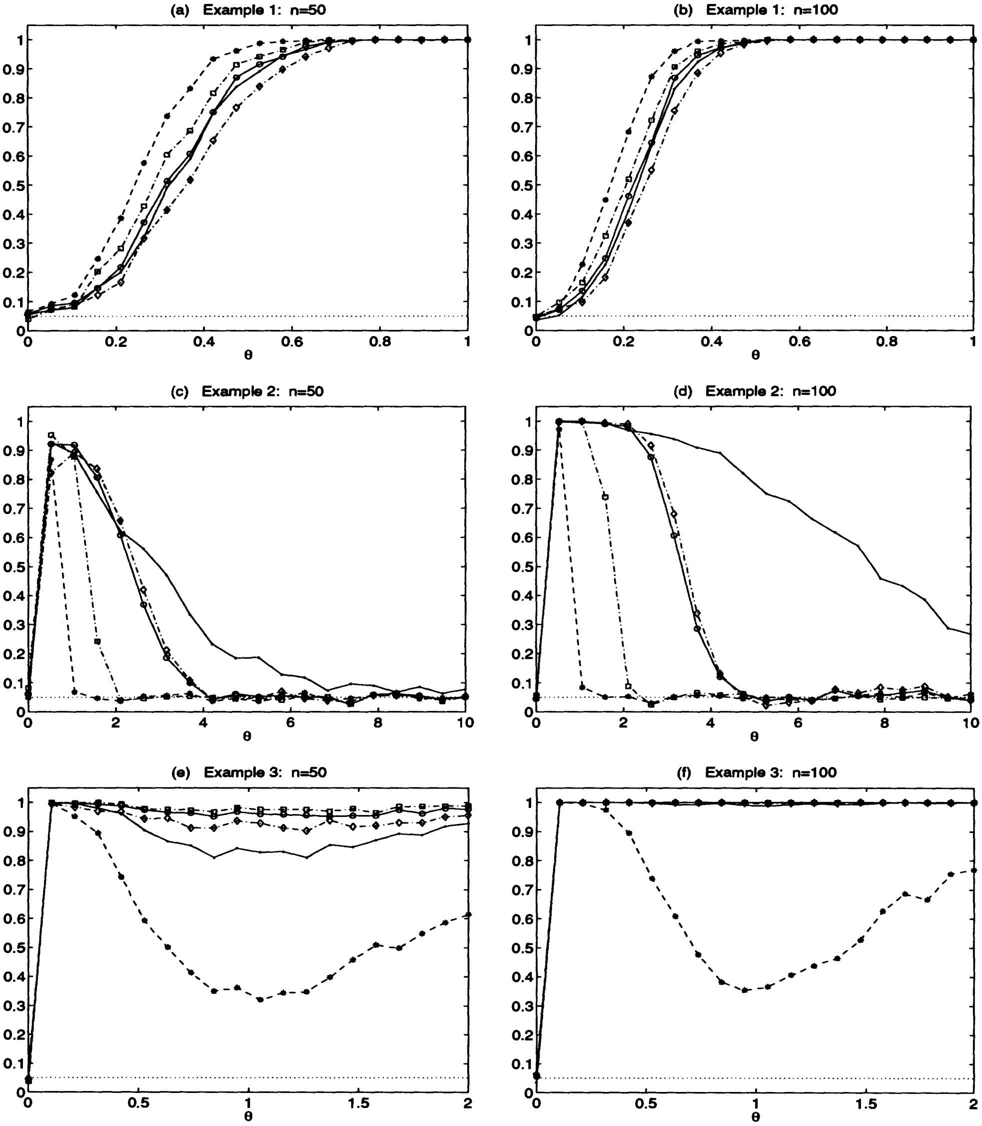

FIGURE 2: Comparison of Power Curves. Dashed curves with star $*$: F-test; solid curves with circle o : MGLR; solid curves with dot : adaptive Neyman test; dotdashed curves with ◇ : GLR(h_1); dotdashed curves with □ : GLR(h_3). The bottom dotted line represents the .05 significance level.

Example 1 attempts to compare the (M)GLR tests and ANT in a case where the parametric F-test performs best. Figures 2(a)–2(b) show that the GLR(h_3) test is more powerful than the GLR(h_1) test, and outperforms ANT. The power of the MGLR test falls between powers of the GLR(h_1) and GLR(h_3) tests, and also slightly outperforms ANT. As sample size increases, the power of MGLR is closer to that of GLR(h_3). Example 2 intends to examine how powerful each testing procedure is in detecting alternatives with different frequency components; the larger the θ, the higher the frequency. Since ANT is constructed specifically to detect high frequency alternatives, it is anticipated to be superior in this example to other tests. Figure 2(c) shows that the GLR(h_3) test is more powerful in detecting alternatives of lower frequency, whereas GLR(h_1) outperforms higher-frequency alternatives. Figure 2(d) shows that the MGLR test apparently performs much closer to the GLR(h_1) test. The alternative model in Example 3 is ideal for the (M)GLR tests to detect. Figure 2(e) shows that GLR(h_3) is more powerful than GLR(h_1), and MGLR is closer to GLR(h_3). Both GLR(h_3) and MGLR are superior to ANT.

In summary, the power of the GLR test depends on the choice of bandwidth parameter. Nonetheless, the MGLR test performs always close to the best of the three GLR tests, which is GLR(h_3) in Example 1, GLR(h_3) against lower-frequency (GLR(h_1) against higher-frequency) alternatives in Example 2, and GLR(h_3) in Example 3. Unlike the individual GLR tests, the adaptive feature enjoyed by the MGLR test makes it more desirable to be used in nonparametric testing.

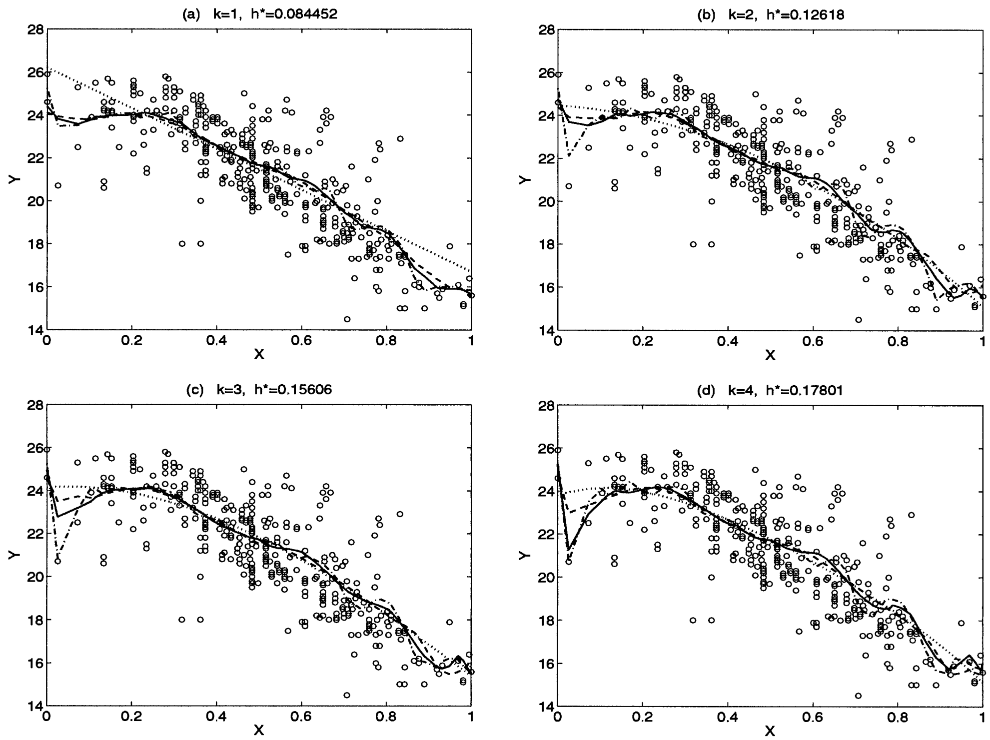

FIGURE 3: Fits to the Babinet Data. Circle : response; dotted line : the least-squares regression fit; dashed dotted curve, solid curve and dashed curve : the kth degree local polynomial fit with bandwidths $1.5^{-1}h^*$, h^* and $1.5h^*$, respectively.

6. REAL DATA EXAMPLE

As an illustration, we apply our MGLR test to a data set of moderate sample size found in Cleveland (1993). The data consist of 355 observations resulting from an experiment on the scattering of sunlight in the atmosphere (Bellver 1987). The response variable Y is the *Babinet point*, the scattering angle at which the polarization of sunlight vanishes, while the explanatory variable X is the cube root of a measure of particulate concentration in the atmosphere. These data have been previously analyzed in Hart (1997) to test for the linearity of the underlying regression, the procedure of which essentially relies on the assumption of homoscedasticity. However, a residual plot considered by Hart (1997, p. 258) showed a certain amount of heteroscedastic pattern, which can also be revealed from the nonparametric variance function estimation of Fan & Yao (1998). Naturally, our (M)GLR tests can take into account this structural variability. Figure 3 displays a scatter plot of the data (X has been re-scaled to the interval $[0, 1]$), super-imposed with the kth degree polynomial regression line, $k = 1, 2, 3, 4$, and the kth degree local polynomial fits with three bandwidths, $h_1 = 1.5^{-1}h^*$, $h_2 = h^*$ and $h_3 = 1.5h^*$. In each panel of Figure 3, the bandwidth h^* calculated from the empirical formula (12) for nonparametric testing, seems to work well for nonparametric curve fitting.

Table 4 gives P-values of the MGLR and GLR tests for the polynomial regression of degree k. Notice that the P-values of the MGLR tests, based on the Slepian bounds (11) and the simulation method described in Section 3.2, are similar. The similarity can also be observed in the GLR tests, the P-values of which are calculated from the chi-squared approximation and the normal approximation. All four tests clearly indicate that a linear model for the observed data is not appropriate, which agrees with the conclusion based on the F-test and reached by Hart (1997) under the assumption of homoscedasticity. Furthermore for $k = 2$, the MGLR and GLR(h_1) tests report evidence that a quadratic model does not describe the data well. However, all four tests give no evidence against a cubic model with $k = 3$. For $k = 4$, unlike the GLR(h_2) and GLR(h_3) tests, the MGLR and GLR(h_1) tests are significant. Considerations of parsimony suggest the cubic nature of the regression.

TABLE 4: Babinet data: the estimated P-values.

Test statistic	Method	$k=1$	$k=2$	$k=3$	$k=4$
MGLR	simulation-based	.001	.040	.181	.041
	Slepian bounds	$[9.8, 11.2] \times 10^{-7}$	[.037, .043]	[.267, .307]	[.045, .052]
GLR(h_1)	χ^2-based	.0002	.033	.163	.038
	N(0, 1)-based	3.8×10^{-7}	.018	.169	.023
GLR(h_2)	χ^2-based	.003	.667	.373	.411
	N(0, 1)-based	8.7×10^{-5}	.701	.431	.466
GLR(h_3)	χ^2-based	.001	.781	.647	.558
	N(0, 1)-based	10^{-6}	.789	.696	.621

7. CONCLUSION

In this article, we have focused on the smoothing parameter selection in nonparametric tests. We offer a simple empirical rule of bandwidth for performing the GLR test (Fan, Zhang & Zhang 2001), an optimal nonparametric test under the formulation of Ingster (1982). Based on power considerations, we further proposed the MGLR test employing multiple bandwidths, the rates of which are optimal in the sense of Ingster (1982) and Fan, Zhang & Zhang (2001). Simulations have shown that the MGLR test is nearly as powerful as the GLR test with an unknown optimal bandwidth. Furthermore, the power of the MGLR test is competitive with that of ANT; but compared with ANT, the finite-sample null distribution of MGLR can be approximated more accurately. Development of multiscale versions of other nonparametric tests will be straightforward. Although our current work focuses on the local polynomial smoother, (M)GLR tests can be carried out in similar fashion for other nonparametric function estimation techniques, such as smoothing splines (see Zhang 2001) and wavelets. As demonstrated in Zhang (2000) and Zhang & Cheng (2003), the MGLR test can easily be extended to goodness of fit, partially linear models, multiple regression models, generalized varying coefficient models, etc. With the automatic and optimal smoothing parameter selection, and the well controlled type-I error and P-value, our test is ready to be used in practice as a useful diagnostic tool. Theoretically, the finite-dimensional weak convergence of $(\mathrm{GLR}(h_1), \ldots, \mathrm{GLR}(h_J))'$ in (6) to a multivariate Gaussian random vector with mean zero and nontrivial covariance function also indicates the possibility of deriving a version of (6), with continuous scales of bandwidth in terms of the Gaussian random fields. Likewise, approximate formulas for the level, P-value, or more generally the tail probability, of this proposed statistic might be obtained from results of Adler (1990) and Sun (1993).

APPENDIX: DERIVATIONS

Proof of Theorem 1. Tedious calculations (Zhang 2000) show that, under the null hypothesis,

$$\begin{aligned} & h^{1/2}\{\mathrm{RSS}_0 - \mathrm{RSS}_1(h)\} \\ = \;& h^{-1/2}\left[\{2\mathcal{K}(0) - \mathcal{K} * \mathcal{K}(0)\} \int \sigma^2(x)\, dx - \frac{\mathcal{K}^2(0)}{nh} \int \sigma^2(x)/f(x)\, dx\right] \\ & + n^{-1}h^{1/2} \sum_{j \neq l} \left\{2\mathcal{K}_h(X_l - X_j)/f(X_j) - \int \mathcal{K}_h(X_l - x)\mathcal{K}_h(X_j - x)/f(x)\, dx\right\} \varepsilon_j \varepsilon_l \\ & + o_P(1). \end{aligned} \tag{14}$$

Now denote the quadratic form in (14) by

$$\begin{aligned} T_n(h) \;=\; & n^{-1}h^{1/2} \sum_{j \neq l} \Big\{2\mathcal{K}_h(X_l - X_j)/f(X_j) \\ & - \int \mathcal{K}_h(X_l - x)\mathcal{K}_h(X_j - x)/f(x)\, dx\Big\} \varepsilon_j \varepsilon_l. \end{aligned} \tag{15}$$

Slight modifications of the arguments used in Theorem 5 of Fan, Zhang & Zhang (2001) guarantee that

$$(T_n(h_1), \ldots, T_n(h_J))' \xrightarrow{\mathcal{L}} N_J(0, \Sigma). \tag{16}$$

To determine entries of the covariance matrix $\Sigma = (\sigma_{ij})$ explicitly, we only need to evaluate the covariance σ_{12} between $T_n(h_1)$ and $T_n(h_2)$. For any constants l_1 and l_2, we have

$$\begin{aligned} \ell_1 T_n(h_1) + \ell_2 T_n(h_2) &= n^{-1} \sum_{j \neq \ell} \Bigg[2 \frac{\left\{\ell_1 h_1^{1/2} \mathcal{K}_{h_1}(X_\ell - X_j) + \ell_2 h_2^{1/2} \mathcal{K}_{h_2}(X_\ell - X_j)\right\}}{f(X_j)} \\ & - \int \frac{\{\ell_1 h_1^{1/2} \mathcal{K}_{h_1}(X_\ell - x)\mathcal{K}_{h_1}(X_j - x) + \ell_2 h_2^{1/2} \mathcal{K}_{h_2}(X_\ell - x)\mathcal{K}_{h_2}(X_j - x)\}}{f(x)}\, dx \Bigg] \varepsilon_j \varepsilon_\ell \\ &= \sum_{j < \ell} W_{j\ell}, \end{aligned}$$

where $W_{j\ell} = n^{-1}\{a_1(j, \ell) + a_2(j, \ell) - a_3(j, \ell)\} \varepsilon_j \varepsilon_\ell$, $1 \leq j < \ell \leq n$, with

$$a_1(j, \ell) = 2\{\ell_1 h_1^{1/2} \mathcal{K}_{h_1}(X_\ell - X_j) + \ell_2 h_2^{1/2} \mathcal{K}_{h_2}(X_\ell - X_j)\}/f(X_j), \quad a_2(j, \ell) = a_1(\ell, j),$$

$$a_3(j, \ell) = 2 \int \frac{\{\ell_1 h_1^{1/2} \mathcal{K}_{h_1}(X_\ell - x)\mathcal{K}_{h_1}(X_j - x) + \ell_2 h_2^{1/2} \mathcal{K}_{h_2}(X_\ell - x)\mathcal{K}_{h_2}(X_j - x)\}}{f(x)}\, dx.$$

It follows that

$$\mathrm{var}\{\ell_1 T_n(h_1) + \ell_2 T_n(h_2)\} = 2^{-1} \mathrm{E}\left[\{2a_1(1,2) - a_3(1,2)\}^2 \varepsilon_1^2 \varepsilon_2^2\right] + o(1),$$

where

$$\{2a_1(1,2) - a_3(1,2)\}^2$$
$$= 4\ell_1^2 h_1 \Big\{ 2\mathcal{K}_{h_1}(X_1 - X_2)/f(X_1) - \int \mathcal{K}_{h_1}(X_1 - u)\mathcal{K}_{h_1}(X_2 - u)/f(u)\,du \Big\}^2$$
$$+4\ell_2^2 h_2 \Big\{ 2\mathcal{K}_{h_2}(X_1 - X_2)/f(X_1) - \int \mathcal{K}_{h_2}(X_1 - v)\mathcal{K}_{h_2}(X_2 - v)/f(v)\,dv \Big\}^2$$
$$+8\ell_1\ell_2(h_1h_2)^{1/2}\Big\{ \frac{2\mathcal{K}_{h_1}(X_1 - X_2)}{f(X_1)} - \int \frac{\mathcal{K}_{h_1}(X_1 - u)\mathcal{K}_{h_1}(X_2 - u)}{f(u)}\,du \Big\}$$
$$\times \Big\{ 2\mathcal{K}_{h_2}(X_1 - X_2)/f(X_1) - \int \mathcal{K}_{h_2}(X_1 - v)\mathcal{K}_{h_2}(X_2 - v)/f(v)\,dv \Big\}.$$

Algebraic manipulations yield the following four expressions,

$$\mathrm{E}\left[\{\mathcal{K}_{h_1}(X_1 - X_2)\mathcal{K}_{h_2}(X_1 - X_2)/f^2(X_1)\}\,\varepsilon_1^2\varepsilon_2^2\right]$$
$$= h_2^{-1}\int \sigma^4(x)\,dx \Big\{ \int \mathcal{K}(y)\mathcal{K}\Big(\frac{y}{C_{12}}\Big)\,dy + O(h_1) \Big\},$$

$$\mathrm{E}\left[\Big\{\mathcal{K}_{h_1}(X_1 - X_2)/f(X_1)\int \mathcal{K}_{h_2}(X_1 - v)\mathcal{K}_{h_2}(X_2 - v)/f(v)\,dv\Big\}\varepsilon_1^2\varepsilon_2^2\right]$$
$$= h_1h_2^{-2}\int \sigma^4(x)\,dx \Big\{ \int \mathcal{K}\Big(\frac{y}{C_{12}}\Big)\int \mathcal{K}(v)\mathcal{K}\Big(\frac{y-v}{C_{12}}\Big)\,dv\,dy + O(h_1) \Big\},$$

$$\mathrm{E}\left[\Big\{\mathcal{K}_{h_2}(X_1 - X_2)/f(X_1)\int \mathcal{K}_{h_1}(X_1 - u)\mathcal{K}_{h_1}(X_2 - u)/f(u)\,du\Big\}\varepsilon_1^2\varepsilon_2^2\right]$$
$$= h_2^{-1}\int \sigma^4(x)\,dx \Big\{ \int \mathcal{K}(y)\int \mathcal{K}(v)\mathcal{K}\Big(\frac{y-v}{C_{12}}\Big)\,dv\,dy + O(h_1) \Big\},$$

$$\mathrm{E}\left[\Big\{\int \mathcal{K}_{h_1}(X_1 - u)\mathcal{K}_{h_1}(X_2 - u)/f(u)\,du\Big\}\Big\{\int \mathcal{K}_{h_2}(X_1 - v)\mathcal{K}_{h_2}(X_2 - v)/f(v)\,dv\Big\}\varepsilon_1^2\varepsilon_2^2\right]$$
$$= h_1h_2^{-2}\int \sigma^4(x)\,dx \left[\int \Big\{ \int \mathcal{K}(u)\mathcal{K}\Big(\frac{y-u}{C_{12}}\Big)\,du \Big\}^2 dy + O(h_1)\right].$$

These lead to

$$\lim_{n\to\infty} \mathrm{var}\{\ell_1 T_n(h_1) + \ell_2 T_n(h_2)\}$$
$$= 2^{-1}\int \sigma^4(x)\,dx \Big[4\ell_1^2 \int \{2\mathcal{K}(y) - \mathcal{K}*\mathcal{K}(y)\}^2\,dy + 4\ell_2^2\int \{2\mathcal{K}(y) - \mathcal{K}*\mathcal{K}(y)\}^2\,dy$$
$$+ 8\ell_1\ell_2 C_{12}^{-1/2}\int \Big\{ 2\mathcal{K}(y) - C_{12}^{-1}\int \mathcal{K}(v)\mathcal{K}\Big(\frac{y-v}{C_{12}}\Big)\,dv \Big\}$$
$$\times \Big\{ 2\mathcal{K}\Big(\frac{y}{C_{12}}\Big) - \int \mathcal{K}(v)\mathcal{K}\Big(\frac{y-v}{C_{12}}\Big)\,dv \Big\}\,dy \Big].$$

Using the Cramér–Wold device, we deduce

$$(T_n(h_1), T_n(h_2))' \xrightarrow{\mathcal{L}} N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \right),$$

where

$$\sigma_{11} = 2\int \sigma^4(x)\,dx \int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}(y)\}^2 dy,$$
$$\sigma_{12} = 2C_{12}^{1/2} \int \sigma^4(x)\,dx \int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}_{C_{12}}(y)\}\{2\mathcal{K}_{C_{12}}(y) - \mathcal{K} * \mathcal{K}_{C_{12}}(y)\}\,dy.$$

For σ_{12}, using properties of the convolution operator, we see that

$$\int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}_{C_{12}}(y)\}\{2\mathcal{K}_{C_{12}}(y) - \mathcal{K} * \mathcal{K}_{C_{12}}(y)\}\,dy = \int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}(y)\}\{2\mathcal{K}_{C_{12}}(y) - \mathcal{K}_{C_{12}} * \mathcal{K}_{C_{12}}(y)\}\,dy.$$

Similar arguments show

$$\sigma_{ii} = 2\int \sigma^4(x)\,dx \int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}(y)\}^2 dy, \quad 1 \le i \le J,$$
$$\sigma_{ij} = 2C_{ij}^{1/2} \int \sigma^4(x)\,dx \int \{2\mathcal{K}(y) - \mathcal{K} * \mathcal{K}(y)\}\{2\mathcal{K}_{C_{ij}}(y) - \mathcal{K}_{C_{ij}} * \mathcal{K}_{C_{ij}}(y)\}\,dy, \quad 1 \le i < j \le J.$$

Combining $n^{-1}\mathrm{RSS}_1(h_j) = \mathrm{E}\{\sigma^2(X)\} + O_P(n^{-1/2}) + O_P\{(nh_j)^{-1}\}$ (Fan, Zhang & Zhang 2001) with (14), (15) and (16), we could then show that the joint distribution of

$$\frac{\frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1(h_j)}{\mathrm{RSS}_1(h_j)}\,\mathrm{E}\{\sigma^2(X)\} - h_j^{-1}\{\mathcal{K}(0) - 2^{-1}\mathcal{K} * \mathcal{K}(0)\}\int \sigma^2(x)\,dx}{\sqrt{\sigma_{11}(4h_j)^{-1}}}, \quad 1 \le j \le J, \tag{17}$$

converges in law to $\mathsf{N}_J(0, \mathcal{R})$, where $\mathcal{R} = (\gamma_{ij})$ with entries $\gamma_{ij} = \sigma_{ij}/\sigma_{ii}$. To derive the distribution of $\lambda_n(h_j)$, $j = 1, \ldots, J$, in (3), we apply the inequality, $x(1+x)^{-1} \le \log(1+x) \le x$ for $x > -1$, which implies that

$$\lambda_n(h_j) = \frac{n}{2}\left\{\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1(h_j)}{\mathrm{RSS}_1(h_j)} + O_P(n^{-2}h_j^{-2})\right\} = \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1(h_j)}{\mathrm{RSS}_1(h_j)} + O_P(n^{-1}h_j^{-2}).$$

This combined with (17) implies the asymptotic joint distribution; that is,

$$\left(\frac{\Upsilon_{\mathcal{K}}\lambda_n(h_1) - \mathcal{D}_n(h_1)}{\sqrt{2\mathcal{D}_n(h_1)}}, \ldots, \frac{\Upsilon_{\mathcal{K}}\lambda_n(h_J) - \mathcal{D}_n(h_J)}{\sqrt{2\mathcal{D}_n(h_J)}}\right)' \xrightarrow{\mathcal{L}} \mathsf{N}_J(0, \mathcal{R}),$$

which in turn leads to (6).

Proof of Lemma 1. Let $e_{1,p+1} = (1, 0, \ldots, 0)'$ denote a $(p+1) \times 1$ vector, and $\boldsymbol{S}_p = (\mu_{j+\ell})_{0 \le j,\ell \le p}$ a $(p+1) \times (p+1)$ matrix, where $\mu_j = \int t^j K(t)\,dt$. Then the equivalent kernel function $\mathcal{K}(t;p)$ is expressed as

$$\mathcal{K}(t;p) = e_{1,p+1}'\boldsymbol{S}_p^{-1}(1, t, \ldots, t^p)'K(t), \quad t \in \mathbb{R}, \tag{18}$$

(for details, see Fan & Gijbels 1996, p. 64). For $p = 2\ell + 1$, we can write $\boldsymbol{S}_p$ in the form,

$$\boldsymbol{S}_p = \begin{bmatrix} \boldsymbol{S}_{p-1} & \boldsymbol{q} \\ \boldsymbol{q}' & \mu_{2p} \end{bmatrix},$$

where $q = (0, \mu_{p+1}, 0, \mu_{p+3}, \ldots, 0, \mu_{2(p-1)}, 0)'$. The zero entries of q result from the assumption on the kernel K. Since $\mu_{2p} \neq 0$, we have

$$S_p^{-1} = \begin{bmatrix} (S_{p-1} - \mu_{2p}^{-1} qq')^{-1} & -\mu_{2p}^{-1}(S_{p-1} - \mu_{2p}^{-1} qq')^{-1} q \\ \star & \star \end{bmatrix}, \tag{19}$$

where the symbol $\star$ denotes an entry whose explicit expression is not required in the following derivations. Applying the Sherman–Morrison–Woodbury formula (Golub & Van Loan 1996, p. 50) gives

$$(S_{p-1} - \mu_{2p}^{-1} qq')^{-1} = S_{p-1}^{-1} + \frac{\mu_{2p}^{-1} S_{p-1}^{-1} qq' S_{p-1}^{-1}}{1 - \mu_{2p}^{-1} q' S_{p-1}^{-1} q}. \tag{20}$$

Observing that the zero entries of S_r^{-1}, for any integer $r \geq 0$, occur at the same locations as those of S_r, we can easily show that $e'_{1,p} S_{p-1}^{-1} q = 0$. Using this identity and putting (20) into (19), we obtain $e'_{1,p+1} S_p^{-1} = [e'_{1,p} S_{p-1}^{-1}, 0]$, which leads to

$$e'_{1,p+1} S_p^{-1} (1, t, \ldots, t^{p-1}, t^p)' = e'_{1,p} S_{p-1}^{-1} (1, t, \ldots, t^{p-1})', \quad p = 2\ell + 1.$$

This, combined with (18), yields $\mathcal{K}(t; 2\ell + 1) = \mathcal{K}(t; 2\ell)$.

Proof of Theorem 2. Writing $g = \{4\mathcal{K} * \mathcal{K}_{C_{ij}} - 2\mathcal{K} * (\mathcal{K} * \mathcal{K}_{C_{ij}}) - 2\mathcal{K}_{C_{ij}} * (\mathcal{K} * \mathcal{K}_{C_{ij}}) + (\mathcal{K} * \mathcal{K}_{C_{ij}}) * (\mathcal{K} * \mathcal{K}_{C_{ij}})\}(0)$, the numerator in (5) is equal to $4^{-1} C_{ij}^{1/2} g$. It suffices to check the expression, $g - \mathcal{K} * \mathcal{K}_{C_{ij}}(0) - (\mathcal{K}_{C_{ij}} - \mathcal{K} * \mathcal{K}_{C_{ij}}) * (\mathcal{K}_{C_{ij}} - \mathcal{K} * \mathcal{K}_{C_{ij}})(0)$, which equals

$$3\mathcal{K} * \mathcal{K}_{C_{ij}}(0) - 2\mathcal{K} * (\mathcal{K} * \mathcal{K}_{C_{ij}})(0) - \mathcal{K}_{C_{ij}} * \mathcal{K}_{C_{ij}}(0). \tag{21}$$

For $p = 0$ and $p = 1$, we have $\mathcal{K} \equiv K$. In this case, since both $\mathcal{K}$ and $\mathcal{K}_{C_{ij}}$ are symmetric unimodal probability densities, the convolution $\mathcal{K} * \mathcal{K}_{C_{ij}}$ is unimodal (Feller 1966, p. 164) and symmetric. It follows that

$$\mathcal{K} * (\mathcal{K} * \mathcal{K}_{C_{ij}})(0) = \int \mathcal{K}(t)\, \mathcal{K} * \mathcal{K}_{C_{ij}}(t)\, dt \leq \sup_{t \in \mathbf{R}} \mathcal{K} * \mathcal{K}_{C_{ij}}(t) \int \mathcal{K}(t)\, dt = \mathcal{K} * \mathcal{K}_{C_{ij}}(0). \tag{22}$$

Again since $\mathcal{K}$ is symmetric and unimodal we deduce $\mathcal{K}(t) \leq \mathcal{K}(t/C_{ij})$, if $C_{ij} \geq 1$, and thus

$$\mathcal{K}_{C_{ij}} * \mathcal{K}_{C_{ij}}(0) = \int \mathcal{K}(t)\mathcal{K}(t)/C_{ij}\, dt \leq \int \mathcal{K}(t)\mathcal{K}(t/C_{ij})/C_{ij}\, dt = \mathcal{K} * \mathcal{K}_{C_{ij}}(0). \tag{23}$$

Applying (22) and (23) to (21) indicates the desired lower bound.

ACKNOWLEDGEMENTS

The author wishes to thank the Editor, Professor Richard A. Lockhart, and a referee for their constructive suggestions that greatly improved the presentation of this paper.

REFERENCES

R. J. Adler (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes.* Institute of Mathematical Statistics Lecture Notes–Monograph Series 12, Institute of Mathematical Statistics, Hayward, California.

A. Azzalini & A. W. Bowman (1993). On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society Series B*, 55, 549–557.

A. Azzalini, A. W. Bowman & W. Härdle (1989). On the use of nonparametric regression for model checking. *Biometrika*, 76, 1–11.

C. Bellver (1987). Influence of particulate pollution on the positions of neutral points in the sky in Seville (Spain). *Atmospheric Environment*, 21, 699–702.

A. W. Bowman & S. G. Young (1996). Graphical comparison of nonparametric curves. *Applied Statistics*, 45, 83–98.

W. S. Cleveland (1993). *Visualizing Data*. Hobart Press, Summit, NJ.

A. Diblasi & A. W. Bowman (1997). Testing for constant variance in a linear model. *Statistics & Probability Letters*, 33, 95–103.

J. Fan (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91, 674–688.

J. Fan & I. Gijbels (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society Series B*, 57, 371–394.

J. Fan & I. Gijbels (1996). *Local Polynomial Modeling and Its Applications*. Chapman & Hall, London.

J. Fan & L. S. Huang (2001). Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association*, 96, 640–652.

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

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

W. Feller (1966). *An Introduction to Probability and Its Applications*, Volume II. Wiley, New York.

D. Firth, J. Glosup & D. Hinkley (1991). Model checking with nonparametric curves. *Biometrika*, 78, 245–252.

T. Gasser & H. G. Müller (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.), Springer, New York, pp. 23–68.

G. H. Golub & C. F. Van Loan (1996). *Matrix Computations*, Third edition. Johns Hopkins University Press, Baltimore, MD.

S. S. Gupta (1963). Probability integrals of multivariate normal and multivariate t. *The Annals of Mathematical Statistics*, 34, 792–828.

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

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

J. D. Hart & T. S. Wehrly (1992). Kernel regression when the boundary region is large, with application to testing the adequacy of polynomial model. *Journal of the American Statistical Association*, 87, 1018–1024.

L. S. Huang (1997). Testing goodness-of-fit based on a roughness measure. *Journal of the American Statistical Association*, 92, 1399–1402.

Yu. I. Ingster (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems of Information Transmission*, 18, 130–140.

G. Kauermann & G. Tutz (1999). On model diagnostics using varying coefficient models. *Biometrika*, 86, 119–128.

E. King, J. D. Hart & T. E. Wehrly (1991). Testing the equality of two regression curves using linear smoothers. *Statistics & Probability Letters*, 12, 239–247.

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

J. Raz (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *Journal of the American Statistical Association*, 85, 132–137.

J. Rice (1984). Boundary modification in kernel regression. *Communications in Statistics, Theory and Methods*, 13, 893–900.

D. Ruppert, S. J. Sheather & M. P. Wand (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90, 1257–1270.

M. Stone (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society Series B*, 36, 111–147.

J. Y. Sun (1993). Tail probabilities of the maxima of Gaussian random fields. *The Annals of Probability*, 21, 34–71.

G. Wahba (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (P. R. Krisnaiah, ed.), North Holland, Amsterdam, pp. 507–523.

Z. Q. Yang & C. M. Zhang (1997). Dimension reduction and L_1 approximation for evaluation of multivariate normal integrals. *Mathematica Numerica Sinica*, 19, 91–102.

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

C. M. Zhang (2000). *Topics in Generalized Likelihood Ratio Test.* Unpublished doctoral thesis, Department of Statistics, University of North Carolina, Chapel Hill, NC.

C. M. Zhang (2001). Equivalent nonparametric regression tests based on spline and local polynomial smoothers. *Technical Report*, Department of Statistics, University of Wisconsin, Madison, WI.

C. M. Zhang & B. Cheng (2003). Binning methodology for nonparametric goodness-of-fit test. *Journal of Statistical Computation and Simulation*, 73, 71–82.

Received 3 September 2001
Accepted 23 January 2003

Chunming M. ZHANG: cmzhang@stat.wisc.edu
Department of Statistics, University of Wisconsin
Madison, WI 53706, USA