The Annals of Statistics
1992, Vol. 20, No. 3, 1412–1425

TESTING GOODNESS-OF-FIT IN REGRESSION VIA ORDER SELECTION CRITERIA

BY R. L. EUBANK[1] AND JEFFREY D. HART[2]

Texas A & M University

A new test is derived for the hypothesis that a regression function has a prescribed parametric form. Unlike many recent proposals, this test does not depend on arbitrarily chosen smoothing parameters. In fact, the test statistic is itself a smoothing parameter which is selected to minimize an estimated risk function. The exact distribution of the test statistic is obtained when the error terms in the regression model are Gaussian, while the large sample distribution is derived for more general settings. It is shown that the proposed test is consistent against fixed alternatives and can detect local alternatives that converge to the null hypothesis at the rate $1/\sqrt{n}$, where n is the sample size. More importantly, the test is shown by example to have an ability to adapt to the alternative at hand.

1. Introduction. When one fits a parametric model to data it is always advisable to test the goodness-of-fit of the postulated model. Parametric goodness-of-fit tests are efficient in detecting lack of fit in certain specified directions, but are inconsistent against many alternatives [see, e.g., Eubank and Spiegelman (1990)]. In contrast, many nonparametric tests are consistent against virtually every alternative, but have poor power against all but the most innocent ones [see, e.g., Durbin and Knott (1972) and Eubank and LaRiccia (1990)]. In this paper we introduce and analyze a new goodness-of-fit test that adapts to the alternative at hand, and thereby tends to have good power against a broad variety of alternatives. The test is based on nonparametric regression methodology, but requires no arbitrary choice of smoothing parameters.

The setting we shall consider may be described as follows. The observed data $(x_1, Y_1), \dots, (x_n, Y_n)$ obey the regression model

$$Y_j = g(x_j) + \varepsilon_j, \qquad j = 1, \dots, n,$$

where $0 \le x_1 < x_2 < \cdots < x_n \le 1$ are fixed design points, the ε_j's are independent and identically distributed random variables with $E(\varepsilon_1) = 0$ and $\mathrm{Var}(\varepsilon_1) = \sigma^2 < \infty$, and g is a function that is essentially arbitrary. We wish to test a null hypothesis of the form

$$H_0\colon g(x) = \sum_{j=1}^{p} \beta_j t_j(x) \qquad \forall\, x \in [0, 1], \tag{1.1}$$

Received October 1990; revised June 1991.
[1]Research supported by NSF Grant DMS-89-02576.
[2]Research supported in part by ONR Contract N00014-85-K-0723.
AMS 1980 *subject classifications*. Primary 62G10; secondary 62E20, 62J99.
Key words and phrases. Nonparametric regression; smoothing parameter selection; risk estimation.

where the t_j's are known functions and the β_j's are unknown constants. The alternative hypothesis is

$$H_a: g(x) = \sum_{j=1}^{p} \beta_j t_j(x) + f(x) \qquad \forall\, x \in [0,1], \tag{1.2}$$

where f is not a linear combination of $t_1, \dots, t_p$.

Härdle and Mammen (1990) and Eubank and Spiegelman (1990) have proposed one means of testing (1.1). In their tests, least squares is used to estimate the parametric null model, and then residuals from the least-squares estimate are smoothed, yielding an estimator $\hat{f}_k$ (with smoothing parameter k) of f in (1.2). The test statistic is a standardized version of $\sum_{j=1}^{n} \hat{f}_k^2(x_j)$, which tends to be moderate under H_0 and large under H_a. A bothersome aspect of these procedures is that one must fix the smoothing parameter k in order to carry out the test. Corresponding to each k is a different test of H_0, and it is not clear how k should be chosen in practice.

A novel feature of the test proposed in this paper is that it uses a data-driven smoothing parameter as the test statistic. This eliminates the arbitrariness of fixing a smoothing parameter as in the procedures discussed above. In the next section we define estimates a_{jn}, $j = 1, \dots, n-p$, of Fourier coefficients a_j, $j = 1, 2, \dots,$ of the function f in H_a. The null hypothesis (1.1) is then equivalent to $a_j = 0$ for all j. Our test statistic is $\hat{k}$, the maximizer of $r(k)$ over $k = 0, 1, \dots, n-p$, where $r(0) \equiv 0$,

$$r(k) = \sum_{j=1}^{k} a_{jn}^2 - \frac{k c_\alpha \hat{\sigma}^2}{n}, \qquad k = 1, \dots, n-p, \tag{1.3}$$

$\hat{\sigma}^2$ is any consistent estimator of σ^2, and $c_\alpha > 1$ is a constant depending on the desired level α of the test. If H_0 is true, meaning that all the a_j's are 0, then $r(k)$ is likely to be maximized at zero. Thus, the null hypothesis is rejected at level α if and only if $\hat{k} \geq 1$.

The statistic $\hat{k}$ may be regarded as the data-driven number of terms in a truncated Fourier series estimator of f. In fact, if $c_\alpha = 2$, then maximizing $r(k)$ is equivalent to minimizing an approximately unbiased estimator of the risk function

$$R(k) = E\left[\frac{1}{n} \sum_{j=1}^{n} \left(\hat{f}_k^2(x_j) - f(x_j)\right)^2\right],$$

where $\hat{f}_k$ is a Fourier series estimator of f with k terms. Smoothing parameter selectors based on criteria akin to $r(k)$ have been studied by Rice (1984) and Hart (1985).

Those familiar with time series will recognize a similarity between $r(k)$ and the various AIC criteria proposed for selecting the order of an autoregression [see, e.g., Akaike (1969) and Bhansali and Downham (1977)]. In fact, part of our asymptotic distribution theory for $\hat{k}$ parallels Shibata's (1976) theory for AIC. However, unlike the time series setting, we are able to develop our theory without the need for a fixed upper bound on $\hat{k}$.

The remainder of the paper is organized as follows. In Section 2 the test is defined and some examples of its applicability are discussed. Section 2 also provides the numerical values of c_α in (1.3) that yield tests of various levels. Section 3 details the distribution of $\hat{k}$ under the null hypothesis. The power of the test for both fixed and local alternatives is investigated in Section 4, where it is shown that the test can detect local alternatives converging to the null at the rate $1/\sqrt{n}$. Examples are also provided which demonstrate that our test has desirable power properties against "high frequency" alternatives. Proofs of all results are collected in Section 5.

2. The proposed test. Let $u_{jn}(\cdot)$, $j = 1, \dots, n-p$, be functions on $[0,1]$ that satisfy the orthogonality conditions

$$\sum_{r=1}^{n} u_{jn}(x_r) u_{ln}(x_r) = n\delta_{jl}, \qquad j, l = 1, \dots, n-p,$$

and

$$\sum_{r=1}^{n} u_{jn}(x_r) t_l(x_r) = 0, \qquad j = 1, \dots, n-p,\ l = 1, \dots, p.$$

Also assume that the $n \times p$ matrix $T = \{t_j(x_r)\}$ is of full column rank and define the sample "Fourier" coefficients

$$a_{jn} = \frac{1}{n} \sum_{r=1}^{n} u_{jn}(x_r) y_r. \tag{2.1}$$

To test H_0 we consider fitting the alternative "model"

$$y_r = \sum_{j=1}^{p} \beta_j t_j(x_r) + \sum_{j=1}^{k} a_j u_{jn}(x_r) + \varepsilon_r, \qquad r = 1, \dots, n. \tag{2.2}$$

For this purpose let $\mathbf{y}$ be the vector of responses, and define $\mathbf{b}_n = (b_{1n}, \dots, b_{pn})' = (T'T)^{-1}T'\mathbf{y}$, which is the least-squares estimator of $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ in (1.1). Assuming for the moment that σ^2 is known, an estimate of the risk or mean-squared error associated with (2.2) is

$$\begin{aligned} &\frac{1}{n} \sum_{r=1}^{n} \left(y_r - \sum_{j=1}^{p} b_{jn} t_j(x_r) - \sum_{j=1}^{k} a_{jn} u_{jn}(x_r) \right)^2 + \frac{2\sigma^2(p+k)}{n} \\ &\quad = \frac{1}{n} \sum_{r=1}^{n} \left(y_r - \sum_{j=1}^{p} b_{jn} t_j(x_r) \right)^2 - \sum_{j=1}^{k} a_{jn}^2 + \frac{2\sigma^2(p+k)}{n} \end{aligned} \tag{2.3}$$

[see, e.g., Rice (1984)]. An indication that H_0 is false is provided if (2.3) is minimized by any value of k other than zero.

Minimizing (2.3) is equivalent to maximizing

$$\sum_{j=1}^{k} a_{jn}^2 - \frac{2\sigma^2 k}{n}. \tag{2.4}$$

Therefore, the preceding discussion suggests using the maximizer $\tilde{k}$ of (2.4) to test H_0. To achieve a prescribed level of significance α, one can use the test

$$\text{Reject } H_0 \text{ if } \tilde{k} \geq k_\alpha, \tag{2.5}$$

where k_α is the upper α percentage point for $\tilde{k}$. It can be shown that $P(\tilde{k} = 0) \to 0.71$ as $n \to \infty$. Therefore, if one desires a test with $\alpha < 0.29$, k_α must exceed 1. Unfortunately, test (2.5) is not consistent against alternatives such as (2.2) with $1 \leq k < k_\alpha$. This follows from the fact that the asymptotic distribution of $\tilde{k}$ has support $k, k+1, \ldots$ when (2.2) holds with $a_k \neq 0$. (The proof of this result is much like that of our Theorem 3.1.)

To avoid the problems with test (2.5) we propose using as test statistic the maximizer $\hat{k}$ of

$$r(k) = \begin{cases} 0, & k = 0, \\ \displaystyle\sum_{j=1}^{k} a_{jn}^2 - \frac{\hat{\sigma}^2 c_\alpha k}{n}, & k = 1, \ldots, n-p, \end{cases} \tag{2.6}$$

where $\hat{\sigma}^2$ is any consistent estimator of σ^2 and c_α is chosen so that $P(\hat{k} = 0) = 1 - \alpha$ under H_0. For $\hat{\sigma}^2$ we can use the estimators of Gasser, Sroka and Jennen-Steinmetz (1986) or Hall, Kay and Titterington (1990). The proposed test is formally given by

$$\text{Reject } H_0 \text{ if } \hat{k} \geq 1. \tag{2.7}$$

Using the rejection region $\hat{k} \geq 1$ insures that our test will be consistent under very general conditions whenever model (2.2) holds with at least one $a_j \neq 0$ (see Section 4).

It is important to note that one need not maximize (2.6) to obtain $\hat{k}$. One can alternatively minimize (2.3) with 2 replaced by c_α and σ by $\hat{\sigma}$. Since estimation is conducted by least squares, the u_{jn} and a_{jn} need not be computed explicitly in this expression. Instead the sum of squared residuals term can be computed using any convenient basis that has the same linear span as $t_1, \ldots, t_p, u_{1n}, \ldots, u_{kn}$ (see, e.g., Example 2). This can lead to considerable computational savings.

An attractive feature of the test (2.6) is that it leads immediately to a point estimate of the regression function in the event that H_0 is rejected. The data analyst generally desires such an estimate if there is reason to believe the null model is inadequate. In our setting, a natural estimator of $g(x)$ is

$$\hat{g}(x) = \sum_{j=1}^{p} b_{jn} t_j(x) + \sum_{j=1}^{\hat{k}} a_{jn} u_{jn}(x).$$

When the null hypothesis is not rejected $\hat{g}(x)$ is simply the least-squares estimate of the null model. Otherwise, $\hat{g}$ is a nonparametric estimate of the regression curve g.

TABLE 1

Values of c_α which make test (2.7) asymptotically valid at level α

α	**0.01**	**0.05**	**0.10**	**0.20**	**0.29**
c_α	6.74	4.18	3.22	2.38	2

Note: When $c_\alpha = 2$, test (2.7) is using the $\hat{k}$ which minimizes the estimated risk (2.3).

A good approximation to c_α in (2.6) is that value of c for which

$$1 - \alpha = \exp\left\{ - \sum_{j=1}^{\infty} \frac{P\left(\chi_j^2 > jc\right)}{j} \right\}, \tag{2.8}$$

where χ_j^2 is a random variable having the chi-squared distribution with j degrees of freedom. More precisely, if c_α is taken to be the solution of (2.8), then $P(\hat{k} = 0) \to 1 - \alpha$ as $n \to \infty$, at least under the conditions of Theorem 3.1. Equation (2.8) is a consequence of Theorem 3.1 and random walk theory in Spitzer (1956). Approximate solutions of (2.8) for various choices of α are given in Table 1. It is worth noting that when $c_\alpha = 2$, the asymptotic level of test (2.7) is about 0.29. This is the level which results from using test (2.7), rather than (2.5), with $\hat{k}$ the minimizer of the estimated risk (2.3).

We conclude this section with two examples of settings where our test is applicable.

EXAMPLE 1 (Testing for no effect). A simple case of (1.1) corresponds to testing for no effect in nonparametric regression [Raz (1990)]. In this setting the null hypothesis is $g(\cdot) \equiv \beta$ in (1.1), which is to be tested against the alternative that g is nonconstant.

This model corresponds to $p = 1$, $t_1(\cdot) \equiv 1$ and, if the design points are $x_r = (r - 1/2)/n$, $r = 1, \dots, n$, we may choose $u_{jn}(x) = u_j(x) = \sqrt{2}\cos(\pi j x)$ for all n and $j \geq 1$. Thus, in this case, the choice of $\hat{k}$ is tantamount to selecting the order of a truncated Fourier series estimator for the regression function.

EXAMPLE 2 (Testing linearity). Suppose now that we wish to assess the goodness-of-fit of a linear model $y_r =' \beta_0 + \beta_1 x_r + \varepsilon_r$, $r = 1, \dots, n$. We could choose $t_1(x) \equiv 1$, $t_2(x) = x$ and select the u_{jn} by orthonormalizing the polynomials $1, x, x^2, \dots$ over the design points. The choice of $\hat{k}$ would then correspond to the selection of a polynomial regression estimator for g in (1.2). This basic approach can also be extended to assess the goodness-of-fit of a general polynomial model.

This example is an illustration of a case where we need not actually compute the u_{jn} or a_{jn}. Equivalently, one can minimize (2.3) with 2 replaced by c_α and σ by $\hat{\sigma}$ and the residuals computed by regression using any polynomial basis such as $1, t, \dots, t^k$.

3. Distribution theory. In this section we examine the distribution theory for $\hat{k}$ under the null model. We begin by dealing with the important special case when the errors in (1.1) are normal.

Let $q_0 = 1$ and $p_0 = 1$, and, for $s = 1, \dots, n - p$, set

$$q_s = \sum_{(\theta_1, \dots, \theta_s) \in C_s} \left\{ \prod_{r=1}^{s} \frac{1}{\theta_r!} \left(\frac{1 - P(\chi_r^2 > c_\alpha r)}{r} \right)^{\theta_r} \right\}$$

and

$$p_s = \sum_{(\theta_1, \dots, \theta_s) \in C_s} \left\{ \prod_{r=1}^{s} \frac{1}{\theta_r!} \left[\frac{P(\chi_r^2 > c_\alpha r)}{r} \right]^{\theta_r} \right\},$$

where χ_r^2 denotes a chi-squared random variable with r degrees of freedom and C_s is the set of all s-tuples $(\theta_1, \dots, \theta_s)$ of integers such that $\theta_1 + 2\theta_2 + \cdots + s\theta_s = s$. We then have the following result.

PROPOSITION 3.1. *Assume the ε_j are iid normal random variables in* (1.1) *and that σ^2 is known and used in place of $\hat{\sigma}^2$ in* (2.6). *Then*

$$P(\hat{k} = k) = p_k q_{n-p-k}, \qquad k = 0, 1, \dots, n - p. \tag{3.1}$$

To appreciate how (3.1) is derived, define the random walk $\{S_k : k = 0, 1, \dots\}$ by $S_0 \equiv 0$ and $S_k = \sum_{j=1}^{k} (Y_j - c_\alpha)$, $k = 1, 2, \dots$, where $Y_1, Y_2, \dots$ are iid random variables with the χ_1^2 distribution. Then Proposition 3.1 follows immediately from Spitzer (1956) upon observing that, when $\varepsilon_1, \dots, \varepsilon_n$ are iid $N(0, \sigma^2)$ and $\hat{\sigma} \equiv \sigma$, $nr(k)/\sigma^2$, $k = 0, \dots, n - p$, have the same distribution as $S_0, \dots, S_{n-p}$. As a result of Proposition 3.1 it is easy to determine the limiting distributions of $\hat{k}$ when the errors are normal and σ is known. The next result, which assumes that σ is estimated, says that this same large sample distribution holds as long as the errors have four moments.

THEOREM 3.1. *Assume that $\hat{\sigma} \to_p \sigma$, c_α is the solution of* (2.8), *the ε_j in* (1.1) *are iid with mean zero and finite fourth moments, and*

$$\max_{1 \le j \le n-p} \sup_x |u_{jn}(x)| \le C$$

for some constant C that is independent of n. Then,

$$\lim_{n \to \infty} P(\hat{k} = k) = p_k (1 - \alpha), \qquad k = 0, 1, 2, \dots .$$

Proposition 3.1 and Theorem 3.1 bear similarities to Theorem 1 of Shibata (1976). There are two essential differences. The first is the use of c_α rather than 2 in our criterion (2.6). The choice of c_α is used to insure that the test (2.7) has the desired level of significance. The other difference is that we do not restrict maximization of $r(k)$ to some bounded subset of the integers. Instead, the maximization region is taken to be as large as possible and grows at the same rate as the sample size. This result contrasts not only with Shibata's work, but also with optimality results as in Rice (1984) which place artificial bounds on data-driven smoothing parameters. In this regard Theorem 3.1 is the strongest result of its kind of which we are aware.

Concerning the conditions of Theorem 3.1 we note that consistent estimators of σ^2 are easy to construct [Gasser, Sroka and Jennen-Steinmetz (1986) and Hall, Kay and Titterington (1990)]. The boundedness condition on the u_{jn} is satisfied by the functions in Example 1 but not by those in Example 2.

4. Power of the proposed test. In this section we focus on the power of our test against both fixed and local alternatives. Conditions are established which guarantee the consistency of our test for a fixed alternative. In addition, we show that our test can detect local alternatives of the form

$$g(x) = \sum_{j=1}^{p} \beta_j t_j(x) + \frac{1}{\sqrt{n}} f_0(x),$$

as $n \to \infty$. Examples are also provided which demonstrate how the test can adapt to the particular alternative at hand.

Concerning power of the test against fixed alternatives we have the following theorem.

THEOREM 4.1. *Define, for* $j = 1, \dots, n-p$, $\alpha_{jn} = n^{-1}\sum_{r=1}^{n} f(x_r)u_{jn}(x_r) = E(a_{jn})$. *Assume that the* ε_j*'s are iid with finite variance* σ^2, *that* $\sup_{1 \le i \le n}|u_{jn}(x_i)| = o(\sqrt{n})$ *for each fixed* j *and that there exists a* j *such that either* $\liminf_{n\to\infty} \alpha_{jn} > 0$ *or* $\limsup_{n\to\infty} \alpha_{jn} < 0$. *Then the power of the test* (2.7) *tends to* 1 *as* $n \to \infty$.

The conditions in the theorem appear to be the weakest that insure consistency. They may be shown to hold under various conditions on f, the design, and the u_{jn}, as we now illustrate.

EXAMPLE 1 (Continued). Assume that we are in the setting of Example 1 and that $x_j = x_{jn} = (j - 1/2)/n$. Then, $\alpha_{jn} \to \alpha_j = \sqrt{2}\int_0^1 f(x)\cos(\pi j x)\,dx$, for each $j = 1, 2, \dots$, if f is Riemann integrable. So the conditions of Theorem 4.1 are satisfied provided that $0 < \int_0^1 f^2(x)\,dx < \infty$.

EXAMPLE 2 (Continued). Assume that the $x_j = x_{jn}$ in Example 2 are the n-tiles of a positive, continuous density h on $[0, 1]$. Let $L_2(h)$ be the set of all functions g such that $\int_0^1 g^2(t)h(t)\,dt < \infty$ with the inner product $\int_0^1 g_1(t)g_2(t)h(t)\,dt$ for $g_1, g_2 \in L_2(h)$. Now construct orthogonal polynomials $u_1, u_2, \dots$ by orthonormalizing the power functions with respect to this inner product.

Arguments similar to those in Jayasuriya (1990) can be used to show that, if f is Riemann integrable, $|\alpha_{jn} - \alpha_j| = O(n^{-1})$, for any fixed j, where $\alpha_j = \int_0^1 u_j(x)f(x)h(x)\,dx$. An induction argument can be employed to establish that, for each j, $\sup_x|u_{jn}(x)| = O(1)$ in this case. Thus the conditions of the theorem are satisfied if $f \in L_2(h)$.

The next result establishes that our test can detect alternatives which differ from the null by the order $1/\sqrt{n}$.

THEOREM 4.2. *Let the function f in* (1.2) *be of the form $f_0/\sqrt{n}$ for some fixed function f_0, and define $\alpha_{jn} = n^{-1}\sum_{r=1}^{n} f_0(x_r)u_{jn}(x_r)$, $j = 1, \dots, n-p$. Assume that all the conditions of Theorem* 3.1 *are satisfied and also that the following hold*:

(i) *There exist a_j's and ϕ_j's such that, for all n and $j = 1, \dots, n-p$, $|a_j - \alpha_{jn}| \le \phi_j b_n$, where $b_n \to 0$ as $n \to \infty$.*
(ii) $|a_j| < C < \infty\ \forall\, j$.
(iii) $\sum_{j=1}^{k} a_j^2/k \to 0$ *as* $k \to \infty$.
(iv) $\sum_{j=1}^{k} \alpha_{jn}^2/k \to 0$ *as* $n, k \to \infty$.

Define the random function $\tilde{r}$ by $\tilde{r}(0) \equiv 0$ and $\tilde{r}(k) = \sum_{j=1}^{k}(Z_j + a_j/\sigma)^2 - kc_\alpha$, $k = 1, 2, \dots$, where $Z_1, Z_2, \dots$ are independent and identically distributed standard normal random variables. Let $\tilde{k}$ be the maximizer of $\tilde{r}$, and define $\gamma = 1 - P(\tilde{k} = 0)$. Under the above conditions the power of the test (2.7) *tends to γ as $n \to \infty$.*

One consequence of Theorem 4.2 is that $\gamma > \alpha$ if one of the a_j's is nonzero, which proves our claim about detection of alternatives tending to the null at rate $1/\sqrt{n}$.

Theorem 4.2 also provides us with a convenient way of comparing the power of test (2.7) with that of other goodness-of-fit tests. One example of a competing test statistic is

$$T_n = n\int_0^1 \hat{F}^2(x)\,dx/\hat{\sigma}^2,$$

where

$$\hat{F}(x) = \int_0^x \sum_{j=1}^{n} \alpha_{jn}u_{jn}(t)\,dt.$$

The statistic T_n is an analog of the Cramér–von Mises statistic. The latter statistic is one of the more frequently used in testing the goodness-of-fit of a probability distribution. Note that $\hat{F}(x)$ estimates the function $F(x) = \int_0^x f(t)\,dt$, which is identically zero if and only if f vanishes on $[0, 1]$. Thus, another test is to reject H_0 for large values of T_n. It is clear that this test will be consistent under very general conditions on the alternative hypothesis.

To compare test (2.7) with the test based on T_n, we will consider the setting of Example 1. In this case, under the conditions of Theorem 4.2 one can show

TABLE 2

Estimated limiting power of test (2.7) and test based on T_n for local alternatives of form $f(x) = 5\sqrt{2}\cos(\pi jx)/\sqrt{n}$, $j = 1, \dots, 5$ ($\alpha = 0.05$)

	j = 1	2	3	4	5
Estimated power of test (2.7)	0.998	0.984	0.957	0.890	0.807
Estimated power of test based on T_n	0.998	0.875	0.388	0.152	0.103

Notes: Estimated power based on 1000 replications of the limiting distributions. The value of σ is 1.

that T_n converges in distribution to

$$\frac{1}{\pi^2}\sum_{j=1}^{\infty}\frac{(Z_j + a_j/\sigma)^2}{j^2}, \tag{4.1}$$

where the Z_j's are iid standard normal random variables. Using this fact along with Theorem 4.2, we have computed the limiting power of the two tests, with $\alpha = 0.05$ and $\sigma = 1$, for each of the five local alternatives $f(x) = 5\sqrt{2}\cos(\pi jx)/\sqrt{n}$, $j = 1, \dots, 5$.

For the five alternatives of interest, simulation (using 1000 replications) was used to approximate the limiting distribution of T_n and of $\hat{k}$ (given in Theorem 4.2). Our results also give very good approximations to the power in finite sample cases in which the error terms are Gaussian with $\sigma = 1$, $n \geq 10$, and the (fixed) alternative at a given n has the form $5\sqrt{2}\cos(\pi jx)/\sqrt{n}$.

The results of the power study are summarized in Table 2. The test based on $\hat{k}$ does remarkably better than the one based on T_n for all but the lowest frequency alternative. The discrepancy between the two becomes more marked as the alternative becomes higher frequency. The reason for the poor performance of T_n is evident from (4.1) in the downweighting of $(Z_j + a_j/\sigma)^2$ by the factor j^{-2}. Because of this weighting scheme, T_n tends to have much better power at low than at high frequency alternatives. The good performance of test (2.7) is explained by Table 3, which shows the distribution of $\hat{k}$ in the various cases. The criterion $r(k)$ tends to have its maximum at the value of k where $a_k \neq 0$. Only when the alternative becomes very high frequency does it become difficult for $\hat{k}$ to detect the presence of a nonzero Fourier coefficient. This is to be expected since a very high frequency function sampled at a small number of points will be hard to distinguish from noise.

We have also considered in detail the situation where the alternative is $f(x) = \sqrt{2}\,a\cos(\pi x)/\sqrt{n}$, a low frequency case well-suited for the Cramér–von Mises type statistic. Our finding here is that for values of a/σ ranging from 0.25 to 5 the power of T_n and test (2.7) (at level 0.05) never differ by more than 2.5%.

5. Proofs. In this final section we give the proofs for the major results in our paper. We begin by establishing Theorem 3.1.

TABLE 3

Estimates of limiting values of $P(\hat{k} = k)$ under null hypothesis and local alternatives of form $f(x) = 5\sqrt{2}\cos(\pi jx)/\sqrt{n}$, $j = 1,\dots,5$ ($\alpha = 0.05$)

	k									
Model	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Null case	0.944	0.042	0.011	0.003	0	0	0	0	0	0
$j = 1$	0.002	0.932	0.054	0.009	0.003	0	0	0	0	0
$j = 2$	0.016	0	0.932	0.038	0.013	0.001	0	0	0	0
$j = 3$	0.043	0	0	0.900	0.048	0.008	0	0.001	0	0
$j = 4$	0.110	0.002	0	0	0.860	0.021	0.005	0.002	0	0
$j = 5$	0.193	0.006	0	0	0	0.758	0.027	0.012	0.003	0.001

Note: The estimated probabilities are based on 1000 replications of the limiting distribution of $\hat{k}$. The value of σ is 1.

PROOF OF THEOREM 3.1. We first prove the result assuming σ is known. Let $Z_{jn} = na_{jn}^2/\sigma^2$, $j = 1,\dots,n-p$, and set $S_{0n} \equiv 0$,

$$S_{rn} = \sum_{j=1}^{r} (Z_{jn} - c), \qquad r = 1,2,\dots,$$

for some constant $c > 1$. Write

$$P(\hat{k} = k) = P(A_n \cap B_n),$$

where

$$A_n = \{S_{kn} - S_{rn} \geq 0,\, r = 0,\dots,k-1;\, S_{rn} - S_{kn} \leq 0,\, r = k+1,\dots,m_n\},$$
$$B_n = \{S_{rn} - S_{kn} \leq 0,\, r = m_n + 1,\dots,n-p\}$$

and m_n is an unbounded sequence of integers to be defined subsequently. The first part of the proof involves showing that

$$P(\hat{k} = k) - P(A_n) \to 0 \tag{5.1}$$

as $n \to \infty$. The second task is to show that

$$P(A_n) - P(A_n^*) \to 0, \tag{5.2}$$

where $A_n^* = \{S_k - S_r \geq 0,\ r = 0,\dots,k-1;\ S_k - S_r \leq 0,\ r = k+1,\dots,m_n\}$ for $S_0 \equiv 0$, $S_r = \sum_{j=1}^{r}(Y_j^2 - c)$, $r = 1,2,\dots,$ with the Y_j's iid standard normal random variables. Having proven (5.1) and (5.2), Theorem 3.1 will follow from the fact that $P(A_n^*) \to p_k(1-\alpha)$ as $n \to \infty$ [see Proposition 3.1 and Spitzer (1956), formula 4.7].

To simplify notation, the proof will be given only for the case $\hat{k} = 0$. The other cases follow in a completely analogous manner. To verify (5.1), it suffices to show that $P(B_n) \to 1$. For this purpose define the subsequence of integers $n_j = j^2$, $j = 1,2,\dots$. Let j_1 be the largest integer j such that $j^2 < m_n$ and let j_2 be the largest j such that $j^2 \leq n - p$. Define $n_{j_2+1} = n - p$, and for each n and $j = j_1, \dots, j_2$ set

$$\xi_{jn} = \max_{1 \le i \le n_{j+1} - n_j} \left| \sum_{r = n_j + 1}^{n_j + i} (Z_{rn} - 1) \right|.$$

Note that if n_j is such that $j_1^2 \le n_j < n - p$, and if $n_j < r \le n_{j+1}$, then

$$\frac{|\sum_{l=1}^{r}(Z_{ln} - 1)|}{r} \le \frac{|\sum_{l=1}^{n_j}(Z_{ln} - 1)|}{n_j} + \frac{\xi_{jn}}{n_j}.$$

It follows that

$$B_n \supset \bigcap_{r = m_n}^{n-p} \left\{ \frac{|\sum_{l=1}^{r}(Z_{ln} - 1)|}{r} \dot{\le} (c - 1) \right\}$$

$$\supset \bigcap_{j=j_1}^{j_3} \left[\left\{ \left| \sum_{l=1}^{n_j} \frac{(Z_{ln} - 1)}{n_j} \right| \le \frac{c-1}{2} \right\} \cap \left\{ \frac{\xi_{jn}}{n_j} \le \frac{c-1}{2} \right\} \right],$$

with $j_3 = j_2$ if $j_2^2 < n - p$, and $j_3 = j_2 - 1$ otherwise.

By Markov's inequality

$$P\left(\bigcup_{j=j_1}^{j_3} \left\{ \left| \sum_{l=1}^{n_j} \frac{(Z_{ln} - 1)}{n_j} \right| > \frac{(c-1)}{2} \right\} \right) \le \frac{4}{(c-1)^2} \sum_{j=j_1}^{j_3} \frac{g(n_j, n)}{j^4} \tag{5.3}$$

for

$$g(r,n) = \operatorname{Var}\left[\sum_{l=1}^{r} (Z_{ln} - 1) \right] = 2r + \left(\frac{E(\varepsilon_1^4)}{\sigma^4} - 3 \right) \frac{1}{n^2} \sum_{i=1}^{n} \left(\sum_{j=1}^{r} u_{jn}^2(x_i) \right)^2.$$

Assume without loss of generality that the uniform bound on the u_{jn}'s is 1. Then $g(n_j, n) \le 2j^2 + |E\varepsilon_1^4/\sigma^4 - 3| j^4/n$, and hence the bound (5.3) is of the order $\sum_{j=j_1}^{j_3} j^{-2}$, which tends to 0 as $m_n \to \infty$.

To deal with the ξ_{jn} we use a result of Serfling (1970). For any set of jointly distributed random variables $Y_1, \dots, Y_l$ with joint distribution function F, let L be the functional $L(F) = \sum_{i=1}^{l} E(Y_i + D)$ with $D = (2 + |E\varepsilon_1^4/\sigma^4 - 3|)$. Defining $F_{a,l}$ to be the joint distribution of $(Z_{a+1,n} - 1), \dots, (Z_{a+l,n} - 1)$ for all a and l, it is clear that $L(F_{a,l}) = Dl$, $L(F_{a,l}) + L(F_{a+l,m}) = L(F_{a,l+m})$ and $E(\sum_{j=a+1}^{a+l}(Z_{jn} - 1))^2 \le Dl = L(F_{a,l})$. Now an application of Theorem A of Serfling (1970) gives

$$E\xi_{jn}^2 \le D[\log(4j + 2)]^2 \frac{(2j+1)}{(\log 2)^2}.$$

Consequently,

$$P\left(\bigcap_{j=j_1}^{j_3} \left\{ \frac{\xi_{jn}}{n_j} \le \frac{c-1}{2} \right\} \right)$$

$$\ge 1 - \sum_{j=j_1}^{j_3} 4D[\log(4j+2)]^2 (2j+1) j^{-4} [(c-1)\log 2]^{-2},$$

which tends to 1 as $n \to \infty$. Combining the preceding results yields $P(B_n) \to 1$.

To prove (5.2), we apply Theorem 13.3 of Bhattacharya and Ranga Rao (1976) to the vectors $\mathbf{v}_{in} = (\varepsilon_i/\sigma)(u_{1n}(x_i), \dots, u_{m_n n}(x_i))'$, $i = 1, \dots, n$. We note that $E\mathbf{v}_{in} = 0$, $n^{-1}\sum_{i=1}^{n} \operatorname{Var} \mathbf{v}_{in}$ is the n-dimensional identity and $n^{-1}\sum_{j=1}^{n} E\|\mathbf{v}_{jn}\|^4 \le m_n^2 E(\varepsilon_1/\sigma)^4$, where $\|\cdot\|$ is the Euclidean norm and we have again assumed that the u_{jn} are uniformly bounded by 1. Using these facts it follows from Theorem 13.3 of Bhattacharya and Ranga Rao (1976) that

$$|P(A_n) - P(A_n^*)| \le a(m_n) m_n^2 \frac{E(\varepsilon_1/\sigma)^4}{\sqrt{n}},$$

where $a(m)$ is a positive constant that depends only on m. Since one can always choose m_n to grow sufficiently slowly that $m_n^2 a(m_n)/\sqrt{n} \to 0$, the proof of (5.2) is complete.

To finish the proof of Theorem 3.1 we need to verify that σ may be replaced by a consistent estimator $\hat{\sigma}$. Let $\tilde{k}$ be $\operatorname{argmin}(R(k))$, where $R(0) = 0$ and $R(k) = \sum_{j=1}^{k} Z_{jn} - kc(\hat{\sigma}^2/\sigma^2)$ for $k > 1$. We wish to establish that $P(\tilde{k} = 0) \to 1 - \alpha$.

Given $\delta > 0$, it follows that

$$P\left(\bigcap_{k=1}^{n-p} \left\{\sum_{j=1}^{k} Z_{jn} \le kc(1-\delta)\right\}\right) - P(C_n)$$
$$\le P(\tilde{k} = 0) \le P\left(\bigcap_{k=1}^{n-p} \left\{\sum_{j=1}^{k} Z_{jn} \le kc(1+\delta)\right\}\right) + P(C_n),$$

where $C_n = \{|\hat{\sigma}/\sigma - 1| \ge \delta\}$. By assumption, $P(C_n) \to 0$ as $n \to \infty$. The other two probabilities in the upper and lower bounds on $P(\tilde{k} = 0)$ can be handled as before for fixed δ as $n \to \infty$. Then let $\delta \to 0$ to obtain the desired result. □

PROOF OF THEOREM 4.1. Clearly

$$P(\hat{k} \ge 1) \ge P\left(\sum_{j=1}^{k} \frac{n(\tilde{a}_{jn} + \alpha_{jn})^2}{\hat{\sigma}^2} > kc\right),$$

where $\tilde{a}_{jn} = n^{-1}\sum_{r=1}^{n} \varepsilon_r u_{jn}(x_r)$ and we assume that k is the smallest integer such that either $\liminf_{n\to\infty} \alpha_{kn} > 0$ or $\limsup_{n\to\infty} \alpha_{kn} < 0$. Assume without loss of generality that $\liminf_{n\to\infty} \alpha_{kn} > 0$. We have

$$P\left(\sum_{j=1}^{k} \frac{n(\tilde{a}_{jn} + \alpha_{jn})^2}{\hat{\sigma}^2} > kc\right)$$
$$\ge P\left(\frac{(n\tilde{a}_{kn} + \alpha_{kn})^2}{\hat{\sigma}^2} > kc\right)$$
$$\ge P\left(\frac{\sqrt{n}\,\tilde{a}_{kn}}{\hat{\sigma}} > \sqrt{kc} - \frac{\sqrt{n}\,\alpha_{kn}}{\hat{\sigma}}\right)$$
$$\ge P\left(\left\{\frac{\sqrt{n}\,\tilde{a}_{kn}}{\sigma} > \sqrt{kc}\,(1+\delta) - \frac{\sqrt{n}\,\alpha_{kn}}{\sigma}\right\} \cap \left\{\frac{\hat{\sigma}}{\sigma} \le 1 + \delta\right\}\right),$$

where $\delta > 0$. Thus it suffices to show that $P(\sqrt{n}\,\tilde{\alpha}_{kn}/\sigma > \sqrt{kc}(1+\delta) - \sqrt{n}\,\alpha_{kn}/\sigma) \to 1$ as $n \to \infty$. Using the Lindeberg–Feller theorem and the condition $\sup_{1 \le i \le n}|u_{kn}(x_i)| = o(\sqrt{n})$, it is easy to check that $\sqrt{n}\,\tilde{\alpha}_{kn}/\sigma$ converges in distribution to a standard normal random variable. The result then follows upon applying Polya's theorem and the fact that $\liminf_{n\to\infty} \alpha_{kn} > 0$. □

PROOF OF THEOREM 4.2. We will assume in our proof that σ is known; the extension to the case where σ is replaced by a consistent estimator is handled as in the proof of Theorem 3.1. We proceed as in the proof of Theorem 3.1 after noting that

$$\sum_{j=1}^{k} \frac{na_{jn}^2}{\sigma^2} - kc = \sum_{j=1}^{k}\left\{\frac{na_{jn}^2}{\sigma^2} - 1 - \alpha_{jn}^2\right\} - k(c-1) + \sum_{j=1}^{k}\alpha_{jn}^2.$$

Defining $S_{rn} = \sum_{j=1}^{r}\{na_{jn}^2/\sigma^2 - 1 - \alpha_{jn}^2\}$, we have, for all m, n sufficiently large,

$$\bigcap_{r=m}^{n-p}\left\{\frac{S_{rn}}{r} < \frac{(c-1)}{2}\right\} \subset \bigcap_{r=m}^{n-p}\left\{\frac{S_{rn}}{r} < (c-1) - \sum_{j=1}^{r}\frac{\alpha_{jn}^2}{r}\right\}, \tag{5.4}$$

due to the fact that $c > 1$ and $\sum_{j=1}^{r}\alpha_{jn}^2/r \to 0$ as $r \to \infty$. Using (5.4) and condition (iv) in Theorem 4.2, one can argue as in the proof of Theorem 3.1 that

$$P(\hat{k} = 0) - P\left(\bigcap_{r=1}^{m}\left\{\sum_{j=1}^{r}\frac{na_{jn}^2}{\sigma^2} < rc\right\}\right) \to 0$$

as $n \to \infty$, where $\{m\}$ is any unbounded sequence of positive integers. Defining $p_n = P(\bigcap_{r=1}^{m_n}\{\sum_{j=1}^{r}(Z_j + \alpha_{jn})^2 < rc\})$, we can use Theorem 13.3 of Bhattacharya and Ranga Rao (1976) to verify that there exists an unbounded sequence $\{m_n\}$ such that

$$P\left(\bigcap_{r=1}^{m_n}\left\{\sum_{j=1}^{r}\frac{na_{jn}^2}{\sigma^2} < rc\right\}\right) - p_n \to 0$$

as $n \to \infty$.

The next step is to argue that $p_n - p_n^* \to 0$ as $n \to \infty$, where $p_n^* = P(\bigcap_{r=1}^{m_n}\{\sum_{j=1}^{r}(Z_j + a_j)^2 < rc\})$. The proof of this result relies on conditions (i) and (ii), but is very tedious and hence is omitted. Details are available from the authors.

The final step is to show that $p_n^* \to p \equiv P(\bigcap_{r=1}^{\infty}\{\sum_{j=1}^{r}(Z_j + a_j)^2 < rc\})$. To do this, it is enough to verify that $\sum_{r=m_n}^{\infty}P(\sum_{j=1}^{r}(Z_j + a_j)^2 \ge rc) \to 0$. This quantity is bounded by

$$\sum_{r=m_n}^{\infty} e^{-rct}E\left[\exp\left\{t\sum_{j=1}^{r}(Z_j + a_j)^2\right\}\right], \tag{5.5}$$

where $0 < t < (1 - c^{-1})/4$. The expectation in (5.5) is bounded by

$$\exp\left\{(t + 4t^2) \sum_{j=1}^{r} a_j^2\right\}(1 - 4t)^{-r/4}.$$

Using condition (iii) and the fact that $0 < t < (1 - c^{-1})/4$, it is now easily checked that for all m_n bigger than some m^* the summand in (5.5) is bounded by $\exp(-dr)$ for a positive constant d, and the proof is complete. □

Acknowledgment. The authors are grateful to two referees whose comments helped to clarify several points in the original manuscript.

REFERENCES

AKAIKE, H, (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21** 243–247.

BHANSALI, R. J. and DOWNHAM, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. *Biometrika* **64** 547–551.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

DURBIN, J. and KNOTT, M. (1972). Components of Cramér–von Mises statistics. *J. Roy. Statist. Soc. Ser. B* **34** 290–307.

EUBANK, R. L. and LARICCIA, V. N. (1990). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. Unpublished manuscript.

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

GASSER, TH., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

HÄRDLE, W. and MAMMEN, E. (1990). Comparing nonparametric versus parametric fits. Unpublished manuscript.

HART, J. D. (1985). On the choice of a truncation point in Fourier series density estimation. *J. Statist. Comput. Simulation* **21** 95–116.

JAYASURIYA, B. (1990). Testing for polynomial regression using nonparametric regression techniques. Ph.D. dissertation, Dept. Statistics, Texas A & M Univ.

RAZ, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: A randomization approach. *J. Amer. Statist. Assoc.* **85** 132–138.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

SERFLING, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist.* **41** 1227–1234.

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126.

SPITZER, F. (1956). A combinatorial lemma and its applications to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143