

## A SMOOTHING SPLINE BASED TEST OF MODEL ADEQUACY IN POLYNOMIAL REGRESSION

DENNIS COX<sup>1\*</sup> AND EUNMEE KOH<sup>2</sup>

<sup>1</sup>*Department of Statistics, University of Illinois, 101 Illini Hall, 725 S. Wright St.,  
Champaign, IL 61820, U.S.A.*

<sup>2</sup>*Department of Statistics, University of Wisconsin, 1210 West Dayton Street,  
Madison, WI 53706, U.S.A.*

(Received August 29, 1986; revised September 12, 1988)

**Abstract.** For the regression model  $y_i = f(t_i) + \varepsilon_i$  ( $\varepsilon_i$ 's iid  $N(0, \sigma^2)$ ), it is proposed to test the null hypothesis that  $f$  is a polynomial of degree less than some given value  $m$ . The alternative is that  $f$  is such a polynomial plus a scale factor  $b^{1/2}$  times an  $(m - 1)$ -fold integrated Wiener process. For this problem, it is shown that no uniformly (in  $b$ ) most powerful test exists, but a locally (at  $b = 0$ ) most powerful test does exist. Derivation and calculation of the test statistic is based on smoothing spline theory. Some approximations of the null distribution of the test statistic for the locally most powerful test are described. An example using real data is shown along with a computing algorithm.

*Key words and phrases:* Regression, model adequacy, smoothing splines.

### 1. Introduction

Consider the univariate regression model

$$(1.1) \quad y_i = f(t_i) + \varepsilon_i, \quad 1 \leq i \leq n,$$

where  $0 < t_1 < \dots < t_n < 1$  and  $\varepsilon_i \sim \text{iid } N(0, \sigma^2)$ . We assume  $\sigma^2$  is known. As formulated, there is no parametric form assumed for the regression function  $f(t)$ . Often, a polynomial model for  $f$  will be used, either because of theoretical considerations in the scientific application, or more often for the convenience of the statistical modeller. The simple linear regression  $f(t) = \beta_1 + \beta_2 t$  is, of course, one such model. We consider here the problem of testing the null hypothesis

---

\*This author's research was supported by the National Science Foundation under grants numbered DMS-8202560 and DMS-8603083.

$H_0$ :  $f$  is a polynomial of degree  $< m$ ,

where  $m$  is given (e.g.,  $m = 2$  in the simple linear regression context). Of course, the testing problem is not complete until we have specified the alternative. Our goal here is not to specify a parametric alternative, but to allow a fairly arbitrary departure from the null model. It is difficult to formalize such a general alternative, and we follow Wahba (1978) in modelling the departure from the polynomial null model by a stochastic process. Specifically, under  $H_1$  we assume

$$(1.2) \quad f(t) = \sum_{i=1}^m \alpha_i t^{i-1} + \sqrt{b} Z(t).$$

The summation on the r.h.s. represents the polynomial null model. Here,  $b \geq 0$  is a scale parameter, and  $Z(t)$  is the  $(m - 1)$ -fold integrated Wiener process. With these specifications, we can restate the hypothesis testing problem in terms of the single parameter  $b$ :

$$H_0: b = 0 \quad \text{vs.} \quad H_1: b > 0.$$

Even with the problem so simplified, it will be seen below that no UMP test exists. However, a locally most powerful (LMP) test does exist. We believe that the LMP test provides a very useful procedure for general purpose tests of adequacy of regression models.

There have been many studies on the model departure from a polynomial representation (see Smith (1973), Blight and Ott (1975), Steinberg (1983), Wecker and Ansley (1983) and Green *et al.* (1985)). In the first three papers, this problem is considered from a Bayesian point of view similar to the one used here.

In Subsection 2.1, we give a detailed description of the Bayesian model, i.e., the stochastic process prior for the possible departures from the polynomial model. In Subsection 2.2, a basis for the space of natural splines of degree  $(2m - 1)$  is described. This basis was discovered by Demmler and Reinsch (1975), and will prove useful for the theoretical analysis. In particular, simple representations are given for the relevant distributions in Subsection 2.3. For this setup, it is shown that no Uniformly Most Powerful test exists (Subsection 2.4), but a Locally Most Powerful test does exist (Subsection 2.5). We briefly describe the test. For a function  $f$  let  $\mathbf{f} = (f(t_1), \dots, f(t_n))'$  denote the vector of values at the observation points. Then, there is an  $n \times n$  symmetric, nonnegative definite matrix  $R$  such that for all natural spline functions  $f$  (see definition in Subsection 2.2),

$$J(f) = \int_0^1 [f^{(m)}(t)]^2 dt = \mathbf{f}' R \mathbf{f}.$$

Let  $R^-$  denote the Moore-Penrose generalized inverse of  $R$ , then our test is given by

$$(1.3) \quad \text{reject } H_0 \quad \text{if} \quad T = \frac{y'R^-y}{n^2\sigma^2} > C,$$

where, of course, the critical value  $C$  is chosen to achieve a desired significance level. Some alternative characterizations of  $R$ , useful for computing, are given in Subsection 2.5. Determination of  $C$  requires knowledge of the null distribution of  $T$ . Some approximations discussed in Subsection 2.6 are based on standard results about weighted sums of independent chi-squared variables. One of these approximations (the Monte Carlo method) is put to use in Section 3, where the methodology is applied to an example. In this example  $\sigma^2$  is unknown. We utilize  $T = y'R^-y/n^2\hat{\sigma}^2$  where  $\hat{\sigma}^2$  is a nonparametric estimate of  $\sigma^2$ .

## 2. Mathematical theory

### 2.1 Specification of the alternatives

The prior distribution on the space of alternatives is suggested by the following result.

**THEOREM 2.1.** (Wahba (1978)) *Let  $f(t)$ ,  $t \in [0, 1]$  have the prior distribution which is the distribution of the stochastic process  $X_\xi(t)$ ,  $t \in [0, 1]$ ,*

$$(2.1) \quad X_\xi(t) = \sum_{i=1}^m \theta_i \phi_i(t) + \sqrt{b} Z(t),$$

where  $\phi_i(t) = t^{i-1}/(i-1)!$ ,  $i = 1, 2, \dots, m$ ,  $b$  is fixed,  $\theta = (\theta_1, \dots, \theta_m)^t \sim N(\mathbf{0}, \xi \mathbf{I}_{m \times m})$ , and  $Z(t)$  is the  $(m-1)$ -fold integrated Wiener process (see Shepp (1966), pp. 321-324),

$$Z(t) = \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} dW(u).$$

Assume  $f$  is independent of the  $\varepsilon$ 's in (1.1). Then the polynomial spline  $f_{n,\lambda}$  which is the minimizer of

$$\frac{1}{n} \sum_{i=1}^n [y_i - f(t_i)]^2 + \lambda \int_0^1 [f^{(m)}(t)]^2 dt$$

has the property

$$f_{n,\lambda}(t) = \lim_{\xi \rightarrow \infty} E_{\xi}(f(t) | Y = y) ,$$

with  $\lambda = \sigma^2/nb$ , where  $E_{\xi}$  is the expectation over the posterior distribution of  $f(t)$  with prior (2.1).

In our problem we consider  $\theta$  to be deterministic. In this setup, testing  $H_0: f \in \mathbf{H}_0$  vs.  $H_1: f \notin \mathbf{H}_0$  is the same as testing  $H_0: b = 0$  vs.  $H_1: b > 0$ , where  $\mathbf{H}_0$  is the null space of  $J(f)$  and  $b$  is given in (2.1).

2.2 Conversion of the original problem

Let  $S_n^m$  denote the space of natural splines, where

$$S_n^m = \{s | s \in C^{2m-2}[0, 1], s \text{ is a polynomial of degree } 2m - 1 \text{ on } (t_i, t_{i+1}), i = 1, \dots, n - 1, \text{ and of degree } m - 1 \text{ on } [0, t_1), (t_n, 1]\} .$$

We consider a basis for  $S_n^m$  introduced by Demmler and Reinsch (1975) consisting of eigenfunctions  $\{\phi_{kn}\}_{k=1}^n$  along with eigenvalues  $\{\rho_{kn}\}_{k=1}^n$  satisfying

$$(2.2) \quad \frac{1}{n} \sum_{i=1}^n \phi_{jn}(t_i)\phi_{kn}(t_i) = \delta_{jk} ,$$

$$(2.3) \quad \int_0^1 \phi_{jn}^{(m)}(t)\phi_{kn}^{(m)}(t) dt = \rho_{kn}\delta_{jk} ,$$

for  $j, k = 1, \dots, n$  with  $0 = \rho_{1n} = \dots = \rho_{mn} < \rho_{m+1,n} \leq \dots \leq \rho_{nn}$ . Here,  $\delta_{jk}$  is Kronecker's delta. Note that  $\{\phi_{1n}, \dots, \phi_{mn}\}$  span the space of polynomials of degree  $\leq m - 1$ .

As long as we can find  $n$  basis functions, call them  $\{\beta_1, \dots, \beta_n\}$ , for  $S_n^m$ , we can build the Demmler-Reinsch basis from that basis. One such basis that is popular is based on the fact that  $W_2^m$  is a reproducing kernel Hilbert space (Aronszajn (1950)). For any  $t$  in  $[0, 1]$  there is an  $\eta(t) \in W_2^m$  such that  $g(t) = \langle g, \eta(t) \rangle_{W_2^m}$ , where  $\langle \cdot, \cdot \rangle_{W_2^m}$  is any valid inner product under which  $W_2^m$  is a Hilbert space. Then  $\eta(t_1), \dots, \eta(t_n)$  form a basis for  $S_n^m$ . Another popular basis discussed in Subsection 2.5 is obtained from  $B$ -splines (Lyche and Schumaker (1973)).

Since the Demmler-Reinsch basis functions have nice properties such as (2.2) and (2.3), we will use them to transform our problem. Define projections by

$$\tilde{y}_v = \frac{1}{n} \sum_{i=1}^n y_i \phi_{vn}(t_i), \quad v = 1, 2, \dots, n ,$$

and

$$\tilde{f}_v = \frac{1}{n} \sum_{i=1}^n f(t_i) \phi_{vn}(t_i), \quad v = 1, 2, \dots, n.$$

These are the generalized Fourier coefficients of the data vector and the sampled function. Our original null hypothesis now can be restated as  $H_0$ :  $\tilde{f}_v = 0$  for  $v = m + 1, \dots, n$ .

### 2.3 Calculation of distributions

In this subsection, we derive several distributional results that will be needed.

LEMMA 2.1. *Assume the errors  $\varepsilon_i$ 's in (1.1) are iid  $N(0, \sigma^2)$ . Then conditioned on  $f$ ,  $\tilde{y}_1, \dots, \tilde{y}_n$  are independent normal such that*

$$E[\tilde{y}_j | f] = \frac{1}{n} \sum_{i=1}^n f(t_i) \phi_{jn}(t_i) \equiv \tilde{f}_j,$$

$$\text{Var}[\tilde{y}_j | f] = \frac{\sigma^2}{n}.$$

PROOF.  $\tilde{y}$  can be expressed as  $\Phi' y$  where  $\Phi_{ij} = \phi_{jn}(t_i)/n$  and  $y = (y_1, \dots, y_n)'$ , while  $\tilde{f} = \Phi' f$  where  $f = (f(t_1), \dots, f(t_n))'$ . We know

$$\text{Law}[y | f] = N(f, \sigma^2 I),$$

then

$$\begin{aligned} \text{Law}[\tilde{y} | f] &= \text{Law}[\Phi' y | f] \\ &= N(\Phi' f, \sigma^2 \Phi' \Phi) \\ &= N\left(\Phi' f, \frac{\sigma^2}{n} I\right). \end{aligned}$$

This completes our proof.

LEMMA 2.2. *Assume the  $\varepsilon_i$ 's are iid  $N(0, \sigma^2)$  and that  $f$  is given by (1.2) where  $Z$  is an  $(m - 1)$ -fold integrated Wiener process independent of the  $\varepsilon_i$ 's. Then  $\tilde{f}_{m+1}, \dots, \tilde{f}_n$  are independent normal with mean 0 and  $\text{Var}(\tilde{f}_j) = b\rho_{jn}^{-1}$ ,  $m < j \leq n$ .*

PROOF. Let us put

$$Z_1(t) = \sum_{j=1}^n z_j \phi_{jn}(t),$$

where  $\{\phi_{1n}, \dots, \phi_{mn}\}$  is the Demmler-Reinsch basis and

$$z_j = \frac{1}{n} \sum_{i=1}^n Z(t_i) \phi_{jn}(t_i),$$

where  $Z$  is the  $(m-1)$ -fold integrated Wiener process. Here  $\tilde{f}_j = \sqrt{b} z_j$ . Note that  $Z_1$  is the natural spline of degree  $(2m-1)$  that interpolates  $Z$  at  $t_1, \dots, t_n$ .

Now we will show the following:

$$(2.4) \quad z_j = \frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m)}(t) dW(t), \quad j > m.$$

Theorem 4.5.6 of Arnold ((1974), p. 74) states that  $\int_0^1 \phi_n^{(m)}(t) dW$  is normally distributed with distribution  $N(\mathbf{0}, \mathbf{S})$ , where  $\phi_n^{(m)}(t) = (\phi_{m+1,n}^{(m)}(t), \dots, \phi_{mn}^{(m)}(t))^t$  and  $\mathbf{S}$  is the  $(n-m) \times (n-m)$  matrix whose  $(i, j)$  element is  $\int_0^1 \phi_{m+i,n}^{(m)}(t) \cdot \phi_{m+j,n}^{(m)}(t) dt = \rho_{m+j,n} \delta_{ij}$ . Thus, the proof is done once we establish (2.4).

Let us first consider  $h \in W_2^m[0, 1]$  and let

$$\tilde{h}_j = \frac{1}{n} \sum_{i=1}^n h(t_i) \phi_{jn}(t_i),$$

$$g_1(t) = \sum_{j=1}^n \tilde{h}_j \phi_{jn}(t),$$

$$g_2(t) = h(t) - g_1(t).$$

Now

$$\begin{aligned} \int_0^1 \phi_{jn}^{(m)}(t) h^{(m)}(t) dt &= \int_0^1 \phi_{jn}^{(m)}(t) g_1^{(m)}(t) dt + \int_0^1 \phi_{jn}^{(m)}(t) g_2^{(m)}(t) dt \\ &= I_1 + I_2, \quad \text{say.} \end{aligned}$$

By the definition of  $g_1$  and (2.3)

$$I_1 = \int_0^1 \phi_{jn}^{(m)}(t) g_1^{(m)}(t) dt = \rho_{jn} \tilde{h}_j.$$

So for  $j > m$  (which implies  $\rho_{jn} > 0$ )

$$(2.5) \quad \tilde{h}_j = \frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m)}(t) g_1^{(m)}(t) dt .$$

Thus, if  $Z$  had  $m$  derivatives, (2.4) would follow from (2.5).

Let us denote the  $n$ -fold integrated Wiener process by  $Z_{(n)}$ , so  $Z = Z_{(m-1)}$ . Then we have

$$(2.6) \quad Z_{(0)}(t) = W(t) ,$$

$$(2.7) \quad Z_{(n)}(t) = \int_0^t Z_{(n-1)}(u) du ,$$

(see Shepp (1966), p. 327). If we apply the fundamental theorem of calculus  $(m - 1)$  times to (2.7), then we get

$$Z_{(m-1)}^{(m-1)}(t) = W(t) .$$

Since  $W$  is a.s. continuous on  $[0, 1]$ ,  $Z_{(m-1)}$  is a.s. in  $C^{m-1}[0, 1]$ , the space of  $(m - 1)$  times continuously differentiable functions, and hence  $Z_{(m-1)} \in W_2^{m-1}$ , but  $Z_{(m-1)} \notin W_2^m$ . Thus, it will be necessary to modify (2.5) to apply to functions in  $W_2^{m-1}$ .

Now  $\phi_{jn}$ 's are natural splines of degree  $(2m - 1)$  with knots  $t_1, \dots, t_n$ , so by Lemma 3.1 of Lyche and Schumaker (1973), we have

$$\int_0^1 \phi_{jn}^{(m)}(t) g_2^{(m)}(t) dt = \sum_{i=1}^n a_i g_2(t_i) ,$$

for some constants  $a_1, \dots, a_n$ . Since  $g_2(t_i) = 0, 1 \leq i \leq n$  (because  $g_1$  interpolates  $h$  at  $t_1, \dots, t_n$ ), we have  $I_2 = 0$ . So we have

$$(2.8) \quad \tilde{h}_j = \frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m)}(t) h^{(m)}(t) dt .$$

Assuming  $m \geq 2$  and selecting  $j > m$ , by applying an integration by parts to (2.8), we obtain

$$(2.9) \quad \frac{1}{n} \sum_{i=1}^n h(t_i) \phi_{jn}(t_i) = - \frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m+1)}(t) h^{(m-1)}(t) dt ,$$

since  $\phi_{jn}^{(m+1)}(0) = \phi_{jn}^{(m+1)}(1) = 0$  by the definition of natural splines. (Note that  $\phi_{jn}^{(m+1)} \in L_2[0, 1]$  when  $m \geq 2$ .) Even though we have only shown (2.9) is valid for  $h \in W_2^m[0, 1]$ , the right-hand side of (2.9) is defined whenever  $h \in W_2^{m-1}[0, 1]$ . Now we show it holds for  $h \in W_2^{m-1}[0, 1]$ .

Let us define the linear functional  $\lambda_1: W_2^{m-1} \rightarrow R$  by

$$\lambda_1(h) = -\frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m+1)}(t) h^{(m-1)}(t) dt,$$

where  $R$  is the set of real numbers. Then by the Cauchy-Schwarz inequality

$$\begin{aligned} |\lambda_1(h)| &\leq \frac{1}{\rho_{jn}} \|\phi_{jn}^{(m+1)}\|_{L_2[0,1]} \|h^{(m-1)}\|_{L_2[0,1]} \\ &\leq \frac{1}{\rho_{jn}} \|\phi_{jn}^{(m+1)}\|_{L_2[0,1]} \|h\|_{W_2^{m-1}[0,1]}. \end{aligned}$$

So  $\lambda_1$  is a bounded, hence continuous, linear functional on  $W_2^{m-1}$  (Theorem 3.3.2 of Ash (1972)). Now define another linear functional  $\lambda_2: W_2^{m-1} \rightarrow R$  by

$$\lambda_2(h) = \frac{1}{n} \sum_{i=1}^n h(t_i) \phi_{jn}(t_i).$$

Thus  $\lambda_2$  is continuous since it is a linear combination of evaluation functionals at  $t_1, \dots, t_n$ . Further, by (2.9) the linear functionals  $\lambda_1, \lambda_2$  agree on  $W_2^m$ , which is dense in  $W_2^{m-1}$ . Two continuous linear functionals agreeing on a dense set must be identical, so  $\lambda_1 = \lambda_2$ . Thus, (2.9) holds for  $h \in W_2^{m-1}$ .

Since  $Z \in W_2^{m-1}$ , we have

$$\begin{aligned} z_j &= -\frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m+1)}(t) Z^{(m-1)}(t) dt \\ &= -\frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m+1)}(t) W(t) dt. \end{aligned}$$

This last integral equals

$$= \frac{1}{\rho_{jn}} \int_0^1 \phi_{jn}^{(m)}(t) dW(t),$$

by applying integration by parts. Thus, we have shown (2.4) for  $m \geq 2$ . If  $m = 1$ , then  $\phi_{jn}^{(1)}$  is a step function, and the right-hand side of (2.8) equals

$$\frac{1}{\rho_{jn}} \sum_{i=1}^n \phi_{jn}^{(1)}(t_i) [h(t_{i+1}) - h(t_i)].$$

Now  $W_2^1$  is dense in  $C[0, 1]$ , and this last expression extends to a bounded linear functional on  $C[0, 1]$ . Since  $Z_{(1)} = W$  is in  $C[0, 1]$ , the argument goes



through as before.

2.4 *Nonexistence of a UMP test*

We have the following setup;

$$\begin{aligned} \tilde{y}_v &= \frac{1}{n} \sum_{i=1}^n y_i \phi_{vn}(t_i) \\ &= \frac{1}{n} \sum_{i=1}^n f(t_i) \phi_{vn}(t_i) + \frac{1}{n} \sum_{i=1}^n \varepsilon(t_i) \phi_{vn}(t_i) \\ &= \tilde{f}_v + \tilde{\varepsilon}_v, \end{aligned}$$

where  $\tilde{\varepsilon}_v$  denotes  $(1/n) \sum_{i=1}^n \varepsilon(t_i) \phi_{vn}(t_i)$ . As we observed,  $\tilde{\varepsilon} \sim N(0, \sigma^2 I/n)$  and Law  $(\tilde{y}_v | \tilde{f}_v) = N(\tilde{f}_v, \sigma^2/n)$ . For  $v > m$ , since  $\tilde{f}_v$  is random and independent of  $\tilde{\varepsilon}_v$ , we have by Lemma 2.2

$$\begin{aligned} \text{Law}(\tilde{y}_v) &= \text{Law}(\tilde{f}_v + \tilde{\varepsilon}_v) \\ &= N\left(0, b\rho_{vn}^{-1} + \frac{\sigma^2}{n}\right). \end{aligned}$$

**THEOREM 2.2.** *For all  $n$  sufficiently large, there exists no UMP test for  $H_0: b = 0$  vs.  $H_1: b > 0$ .*

**PROOF.** Computing the likelihood ratio at an alternative  $b$ ,

$$\begin{aligned} \frac{f(\tilde{y}|b)}{f(\tilde{y}|0)} &= \prod_{i=m+1}^n \left( \frac{\frac{\sigma^2}{n}}{b\rho_{jn}^{-1} + \frac{\sigma^2}{n}} \right)^{1/2} \exp \left[ \frac{1}{2} \sum \frac{n\tilde{y}_j^2}{\sigma^2} - \frac{1}{2} \sum \frac{\tilde{y}_j^2}{b\rho_{jn}^{-1} + \frac{\sigma^2}{n}} \right] \\ &= \prod_{i=m+1}^n \left( 1 + \frac{nb}{\sigma^2} \rho_{jn}^{-1} \right)^{-1/2} \exp \left[ \frac{1}{2} \sum \frac{n\tilde{y}_j^2}{\frac{\sigma^4}{nb} \left( \frac{nb}{\sigma^2} + \rho_{jn} \right)} \right]. \end{aligned}$$

For testing  $H_0: b = 0$  vs.  $H_1: b = b_0, b_0 > 0$ , there exists a Most Powerful test  $\phi_0$  such that

(2.10a)  $E_0 \phi_0 = \alpha$  and

(2.10b)  $\phi_0 = \begin{cases} 1 & \text{if } T_0 > c_0, \\ 0 & \text{if } T_0 < c_0, \end{cases}$

where  $T_0 = \sum [\tilde{y}_j^2 / (nb_0/\sigma^2 + \rho_{jn})]$ . Given another alternative  $b_1 > 0$ ,  $b_1 > b_0$ , for testing  $H_0: b = 0$  vs.  $H_1: b = b_1$ , there exists a Most Powerful test  $\phi_1$  such that

$$(2.11a) \quad E_0 \phi_1 = \alpha \quad \text{and}$$

$$(2.11b) \quad \phi_1 = \begin{cases} 1 & \text{if } T_1 > c_1, \\ 0 & \text{if } T_1 < c_1, \end{cases}$$

where  $T_1 = \sum [\tilde{y}_j^2 / (nb_1/\sigma^2 + \rho_{jn})]$ . These two Most Powerful tests,  $\phi_0$  and  $\phi_1$ , are uniquely determined by (2.10a), (2.10b) and (2.11a), (2.11b), respectively, except on a set which has Lebesgue measure 0 (see Lehmann (1959), Theorem 1, p. 65).

The rejection region for  $\phi_0$  is the exterior of an ellipsoid given by

$$\left\{ (\tilde{y}_{m+1}, \dots, \tilde{y}_n): \sum_{i=m+1}^n \frac{\tilde{y}_i^2}{\left(\frac{nb_0}{\sigma^2} + \rho_{in}\right)} > c_0 \right\},$$

and for  $\phi_1$  is

$$\left\{ (\tilde{y}_{m+1}, \dots, \tilde{y}_n): \sum_{i=m+1}^n \frac{\tilde{y}_i^2}{\left(\frac{nb_1}{\sigma^2} + \rho_{in}\right)} > c_1 \right\}.$$

If  $\phi_0$  is UMP, then it must be unique, and  $P(\phi_1 = \phi_0) = 1$ . Thus the two tests are essentially the same; then all of the axes are the same, which implies

$$c_0(nb_0 + \sigma^2\rho_{in}) = c_1(nb_1 + \sigma^2\rho_{in}), \quad m < i \leq n.$$

Thus  $c_1 < c_0$  since  $b_1 > b_0$ , but subtracting the  $j$ -th equation from the  $i$ -th equation gives

$$(2.12) \quad (\rho_{in} - \rho_{jn})c_0 = (\rho_{in} - \rho_{jn})c_1 \quad \text{for } i, j > m.$$

Here we need only show there exist  $i, j$  such that  $\rho_i \neq \rho_j$ . We know  $\rho_{in}$ 's have a property that says there exist  $k_1, k_2$  such that

$$k_1 i^{2m} \leq \rho_{i+m,n} \leq k_2 i^{2m}, \quad i = 1, 2, \dots, n - m,$$

see Utreras (1983). Thus, if we have a large enough  $n$ , then there exist  $i_1, i_2, i_1 < i_2$ , such that  $k_2 i_1^{2m} < k_1 i_2^{2m}$ , so there exist two distinct nonzero

eigenvalues (we conjecture that if  $n \geq m + 2$ , then there are two distinct nonzero eigenvalues). From (2.12), we have  $c_0 = c_1$ , which contradicts  $c_0 > c_1$ . So the two most powerful tests are different. This completes our proof.

2.5 LMP test

THEOREM 2.3. *There exists a LMP test for  $H_0: b = 0$  vs.  $H_1: b > 0$ .*

PROOF. Let us put  $\tilde{y}_{>m} = (\tilde{y}_{m+1}, \dots, \tilde{y}_n)^t$ . The probability density function of  $\tilde{y}_{>m}$  given  $b$  is

$$l(\tilde{y}_{>m}|b) = \frac{1}{(2\pi)^{(n-m)/2}} \prod_{j=m+1}^n \left( b\rho_{jn}^{-1} + \frac{\sigma^2}{n} \right)^{-1/2} \exp \left[ -\frac{1}{2} \sum_{j=m+1}^n \frac{\tilde{y}_j^2}{\left( b\rho_{jn}^{-1} + \frac{\sigma^2}{n} \right)} \right],$$

and the derivative w.r.t.  $b$  is

$$l'(\tilde{y}_{>m}|b) = l(\tilde{y}_{>m}|b) \sum_{j=m+1}^n \left( \frac{\rho_{jn}^{-1}}{\left( b\rho_{jn}^{-1} + \frac{\sigma^2}{n} \right)^2} \tilde{y}_j^2 - \frac{\frac{1}{2} \rho_{jn}^{-1}}{b\rho_{jn}^{-1} + \frac{\sigma^2}{n}} \right).$$

The LMP test will reject when  $l'(\tilde{y}|0) > kl(\tilde{y}|0)$  (Ferguson (1967), p. 235) since we want to maximize  $(d/db) \int \phi_d(\tilde{y})l(\tilde{y}|b)du|_{b=0}$  subject to  $\int \phi_d(\tilde{y}) \cdot l(\tilde{y}|0)du = \alpha$ , where  $\phi_d$  is the critical function. Thus, when

$$\sum_{j=m+1}^n \left( \frac{\rho_{jn}^{-1}}{\left( b\rho_{jn}^{-1} + \frac{\sigma^2}{n} \right)^2} \tilde{y}_j^2 - \frac{\frac{1}{2} \rho_{jn}^{-1}}{b\rho_{jn}^{-1} + \frac{\sigma^2}{n}} \right) \Bigg|_{b=0} > c_1,$$

we reject the null hypothesis. Then we can use  $T = \sum_{j=m+1}^n (\rho_{jn}^{-1} \tilde{y}_j^2 / \sigma^2)$  as our test statistic, so we have the LMP test given by

$$(2.13) \quad \text{reject } H_0 \quad \text{if} \quad \sum_{j>m} \frac{\rho_{jn}^{-1} \tilde{y}_j^2}{\sigma^2} \geq c.$$

Here we derive an alternative formula for  $T$  which does not use the Demmler-Reinsch basis and is useful for computations. Let  $\Phi$  denote the  $n \times n$  matrix with entry  $\Phi_{ij} = \{\phi_{jn}(t_i)/n\}$ . Then  $T = y^t \Phi \rho_M^{-1} \Phi^t y$  where  $\rho_M$  is an  $n \times n$  diagonal matrix with diagonal entries  $\rho_{in}$ , and  $\rho_M^{-1}$  is the Moore-

Penrose  $g$ -inverse of  $\rho_M$ . Here the polynomial spline  $f_n(t) \equiv \sum_{v=1}^n \tilde{f}_v \phi_{vn}(t)$  is the element of  $S_n^m$  that interpolates  $f$ . Note that

$$\begin{aligned} \int_0^1 (f_n^{(m)}(t))^2 dt &= \sum_{v=1}^n \tilde{f}_v^2 \rho_v \\ &= \mathbf{f}' \Phi \rho_M \Phi' \mathbf{f} \\ &\equiv \mathbf{f}' R \mathbf{f}. \end{aligned}$$

Thus,  $T = \mathbf{y}' R^- \mathbf{y} / n^2 \sigma^2$ , with  $R^- = n^2 \Phi \rho_M^- \Phi'$  (since  $\Phi' \Phi = 1/n$ ), the Moore-Penrose generalized inverse of  $R$  (see Rao (1973), p. 26). So we can summarize our test as follows:

$$\text{reject } H_0: b = 0 \text{ vs. } H_1: b > 0 \quad \text{if} \quad T = \frac{\mathbf{y}' R^- \mathbf{y}}{n^2 \sigma^2} \text{ is too large.}$$

We derive another expression for  $T$  in terms of  $B$ -splines (Lyche and Schumaker (1973) and de Boor (1978)). The formulae are valid for any basis for the natural spline functions, although  $B$ -splines are recommended for their local support properties. Let  $B_1(t), \dots, B_n(t)$  be the basis for  $S_n^m$ , then if  $f \in S_n^m$ ,

$$f(t) = \sum_{i=1}^n c_i B_i(t),$$

and  $\mathbf{f}' = (f(t_1), \dots, f(t_n)) = \mathbf{B}' \mathbf{c}$ , where  $B_{ij} = B_i(t_j)$ ,  $i, j = 1, \dots, n$ . Then

$$\begin{aligned} J(f) &= \sum_i \sum_j c_i \left[ \int B_i^{(m)}(t) B_j^{(m)}(t) dt \right] c_j \\ &\equiv \mathbf{c}' Q \mathbf{c} \\ &= \mathbf{f}' B^{-1} Q (B')^{-1} \mathbf{f} \\ &= \mathbf{f}' R \mathbf{f}, \\ R &= B^{-1} Q (B')^{-1}, \end{aligned}$$

where  $Q_{ij} = \int B_i^{(m)} B_j^{(m)}$ , the matrix of  $L_2$  inner product of  $m$ -th derivatives of  $B$ -splines. Note that  $Q$  is a  $(2m-1)$  banded matrix and  $B$  is an  $(m-1)$  banded matrix by the local support properties for  $B$ -splines.

## 2.6 Distribution of the test statistic

To get the critical value  $c$  we have to know the null distribution of  $T$ . There is no exact closed form for the distribution of  $T$ . We will consider

some approximations. Assume  $X_1, \dots, X_k$  are iid  $N(0, 1)$ . We want to approximate the distribution of  $Q = \sum_{i=1}^k \alpha_i X_i^2$ , where  $\alpha_i$ 's are positive constants. For our setting,  $k = n - m$  and  $\alpha_i = \rho_{i+m,n}^{-1}/n$ . There are several approximations using different methods (Kotz *et al.* (1970)). Their cumulative distribution functions,  $G_k(\alpha, q)$ , are the following:

(1) By power series expansion

$$G_k(\alpha, q) = \left(\frac{q}{2}\right)^{k/2} \sum_{i=0}^{\infty} \frac{c_i^p (-1)^i \left(\frac{q}{2}\right)^i}{\Gamma\left(\frac{k}{2} + i + 1\right)},$$

where  $c_0 = \prod_{j=1}^k \alpha_j^{-1/2}$ ,  $c_i^p = (1/i) \sum_{r=0}^{i-1} d_{i-r}^p c_r^p$ , with  $d_i^p = (1/2) \sum_{j=1}^k \alpha_j^{-i}$ . This method will probably not work well for the tail region of interest.

(2) By Laguerre expansion

$$G_k(\alpha, q) = G\left(k, \frac{q}{\beta}\right) + \sum_{i=1}^{\infty} c_i^l \left(\frac{(i-1)!}{\Gamma\left(\frac{k}{2} + i\right)}\right) \left(\frac{q}{2\beta}\right)^{k/2} e^{-q/2} L_{i-1}^{k/2}\left(\frac{y}{2\beta}\right),$$

where  $G(k, a)$  denotes the cumulative  $\chi^2$  distribution function with  $k$  degrees of freedom and  $c_0 = 1$ ,  $c_i^l = (1/i) \sum_{r=0}^{i-1} d_{i-1-r}^l c_r^l$ , with  $d_i^l = (1/2) \sum_{j=1}^k (1 - \alpha_j/\beta)^i$ , and  $L_i^{(a)}(x) = (1/i!) e^x x^{-a} (d^i/dx^i)(e^{-x} x^{i+a})$ , and  $\beta$  can be chosen by user.

(3) By  $\chi^2$ -expansion

$$G_k(\alpha, q) = \sum c_i^c G\left(k + 2i, \frac{q}{\beta}\right),$$

where  $c_0^c = \prod_{j=1}^k (\beta/\alpha_j)^{1/2}$ ,  $c_i^c = (1/i) \sum_{r=0}^{i-1} d_{i-r}^c c_r^c$ , with  $d_i^c = (1/2) \sum_{j=1}^k (1 - \beta/\alpha_j)^i$ . Here in the case of (2) and (3) Ruben (1962) suggested that the best choice of  $\beta$  for computational purposes is  $\beta = 2\alpha_1 \alpha_2 (\alpha_1 + \alpha_n)^{-1}$ .

(4) By  $\beta\chi_v^2$  approximation

Satterthwaite (1946) suggested the approximation of  $Q$  by  $R = \beta\chi_v^2$ , where  $\beta$  and  $v$  are chosen to make the first two moments agree with those of  $G_k(\alpha, q)$ . Thus, after some calculation

$$v = \frac{\left(\sum_{j=1}^k \alpha_j\right)^2}{\sum_{j=1}^k \alpha_j^2} \quad \text{and} \quad \beta = \frac{\sum_{j=1}^k \alpha_j^2}{\sum_{j=1}^k \alpha_j}.$$

## (5) By Monte Carlo

One very simple method for obtaining the null distribution of  $T$  is Monte Carlo. We describe the computation of a  $p$ -value for a given observed value  $t$  of  $T$ . For each Monte Carlo trial, generate iid  $N(0, 1)$  random variables  $X_1, \dots, X_{n-m}$  and compute a value of  $T = (1/n) \cdot \sum_{i=1}^{n-m} \rho_{i+m,n}^{-1} X_i^2$ . Then the Monte Carlo estimate of the  $p$ -value is the proportion of Monte Carlo trials wherein  $T \geq t$ . About 8000 Monte Carlo trials are required to estimate a  $p$ -value near .05 with  $\pm .005$  accuracy (95% confidence).

## 3. Application to real data

We use the algorithm of Lyche and Schumaker (1973) to compute  $B$ -splines and the inner products of their  $m$ -th derivatives. As mentioned in Subsection 2.5, our test statistic  $T = \left( \sum_{j>m} \rho_j^{-1} \tilde{y}_j^2 \right) / \sigma^2$  is the same as  $y^t (B^{-1} Q (B^t)^{-1})^{-1} y / n^2 \sigma^2$ , where  $B$  is the  $B$ -spline evaluation matrix whose  $(i, j)$  element is  $B_i(t_j)$  and  $Q$  is the inner products of  $m$ -th derivatives of  $B$ -splines with  $Q_{ij} = \int B_i^{(m)}(t) B_j^{(m)}(t) dt$ . For  $R = B^{-1} Q (B^t)^{-1}$ , we can decompose as  $R = U A U^t$  where  $U$  is an orthogonal matrix and  $A$  is a diagonal matrix with  $n - m$  nonzero  $\lambda_j$ 's and  $m$  zeros as its diagonal elements. Here we note that  $\lambda_j = \rho_j / n$  because  $R = \Phi \rho_M \Phi^t = U (\rho_M / n) U^t = U A U^t$ . Then its Moore-Penrose  $g$ -inverse  $R^-$  will be  $U A^- U^t$  where  $A^-$  has  $n - m$  nonzero element as  $1/\lambda_j$  and  $m$  zeros.

Thus our test statistic  $T$  will be expressed as

$$T = \frac{y^t (U A^- U^t) y}{n^2 \sigma^2} = \sum_{j=1}^{n-m} \frac{\lambda_j^{-1}}{n^2 \sigma^2} (U_j^t y)^2,$$

where  $U_j$  is the  $j$ -th column vector of  $U$  corresponding to  $\lambda_j$ .

We compute the  $p$ -value by Monte Carlo as described in the previous section. Our test statistic  $T$  can be compared with 8000  $t$ 's where  $t = \sum_{j=1}^{n-m} (\lambda_j^{-1} / n^2) Z_j^2$  and the  $Z_j$ 's are iid  $N(0, 1)$ . The computation was done in double precision using the IMSL routine GGNML to generate the  $Z_j$ 's.

When  $\sigma^2$  is not known, then we can use the estimate of  $\sigma^2$  proposed by Rice (1984) based on successive differences  $y_{i+1} - y_i$ :

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2,$$

where it is assumed that the  $t_i$ 's are sorted in ascending order.

A numerical example follows.

It is generally thought that the percentage of fruits attacked by codling moth larvae is greater on apple trees bearing a small crop. The data shown in Table 1 are adapted from the results of an experiment in Hansberry and Richardson (1935) containing evidence about this phenomenon. A straight line was fitted to this data in Snedecor and Cochran (1980) (see Fig. 1). We apply our test to this data and obtain a  $p$ -value of .13, which suggests that the simple linear model is adequate.

A residual plot (or other diagnostic) can provide some hint of model inadequacy, but it is difficult to assess statistical significance from such a plot. In the example, one may be tempted to conclude that the linear fit is

Table 1. Percentage of wormy fruits on size of apple crop.

Tree	Size of crops (hundreds)	Wormy fruits (percentage)
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
12	40	27

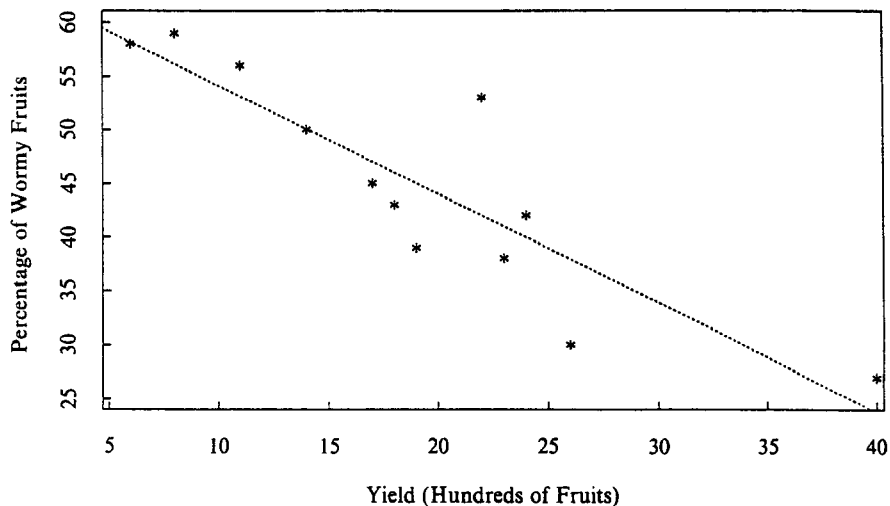


Fig. 1. Codling moth data.

inadequate (see Fig. 2), but the results presented here indicate that the apparent lack of fit could be the result of chance alone, although the moderately small  $p$ -value is still suggestive of model inadequacy that may become more apparent with more data.

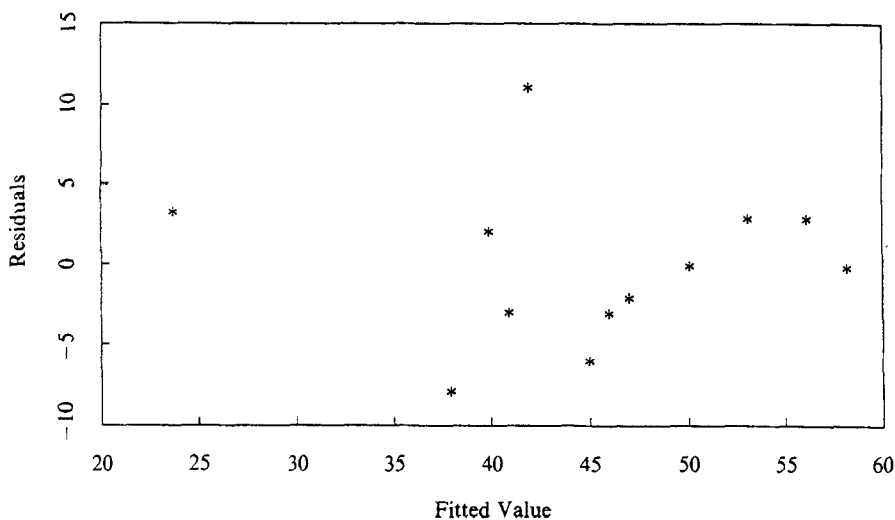


Fig. 2. Residuals from codling moth data.

#### 4. Discussion

The proposed test maximizes the derivative of the *average* power at  $b = 0$ , where the average is with respect to the “prior” distribution on the departure from a polynomial. Thus, the power may be quite low at a specific alternative. Nonetheless, we expect the test will perform well against smooth alternatives, as will now be explained.

The Demmler-Reinsch basis functions are analogous to sines and cosines, and the test statistic is a weighted sum of the squares of the Fourier coefficient analogous  $\tilde{y}_j$  of the data (see equation (2.13)). The weights are larger for the lower “frequencies”. As smoother functions have “energy” concentrated at lower frequencies, the test will have better power at smoother alternatives.

It is possible to generalize the methodology presented in Section 1 to any case where the null hypothesis is that the regression function belongs to a linear space  $H_0$  which is the null space of a quadratic functional  $J(\cdot)$  which can be associated with the inner product of the reproducing kernel Hilbert space for a Gaussian measure. This is done in Cox *et al.* (1988).

The enquiring reader may be interested in the selection of  $m$ , the order of the polynomial. We anticipate that  $m = 2$  will be most frequently used in practice, corresponding to simple linear regression. In general, we do not



advocate the procedure proposed here as a means of selecting  $m$ , but instead recommend the use of a genuine model selection criterion for that purpose. See for example Shibata (1981). The test here will be useful when a polynomial regression model of a given order has already been posited and one wishes to check its adequacy.

## REFERENCES

- Arnold, L. (1974). *Stochastic Differential Equations—Theory and Applications*, Wiley, New York.
- Aronszajn, N. (1950). Theory of reproducing kernels, *Trans. Amer. Math. Soc.*, **58**, 337–404.
- Ash, R. (1972). *Real Analysis and Probability*, Academic Press, New York.
- Blight, B. and Ott, L. (1975). A Bayesian approach to model inadequacy for polynomial regression, *Biometrika*, **62**, 79–88.
- Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, *Ann. Statist.*, **16**, 113–119.
- de Boor, C. (1978). *A Practical Guide to Splines*, Springer, New York.
- Demmler, A. and Reinsch, C. (1975). Oscillation matrices with spline smoothing, *Numer. Math.*, **24**, 375–382.
- Ferguson, T. (1967). *Mathematical Statistics*, Academic Press, New York.
- Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing, *J. Roy. Statist. Soc. Ser. B*, **47**, 299–315.
- Hansberry, T. R. and Richardson, C. H. (1935). A design for testing technique in codling moth spray experiments, *Iowa State Coll. J. Sci.*, **10**, 27–36.
- Kotz, S., Johnson, N. and Boyd, D. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2, Wiley, New York.
- Lehmann, E. (1959). *Testing Statistical Hypotheses*, Wiley, New York.
- Lyche, T. and Schumaker, L. (1973). Computation of smoothing and interpolating natural splines via local bases, *SIAM J. Numer. Anal.*, **10**, 1027–1038.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.
- Rice, J. (1984). Bandwidth choice for nonparametric regression, *Ann. Math. Statist.*, **12**, 1215–1230.
- Rubens, H. (1962). Probability content of regions under spherical normal distributions IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables, *Ann. Math. Statist.*, **33**, 542–570.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components, *Biometrika*, **2**, 110–114.
- Shepp, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures, *Ann. Math. Statist.*, **37**, 321–354.
- Shibata, R. (1981). An optimal selection of regression variables, *Biometrika*, **68**, 45–54.
- Smith, A. F. M. (1973). Bayes estimates in one-way and two-way models, *Biometrika*, **60**, 319–329.
- Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*, The Iowa State University Press, Iowa.
- Steinberg, D. (1983). Bayesian model for response surface and their implications for experimental design, Thesis, Department of Statistics, University of Wisconsin-Madison, Wisconsin.

- Utreras, F. (1983). Natural spline functions, their associated eigenvalue problem, *Numer. Math.*, **42**, 107–117.
- Wahba, G. (1978). Improper priors, spline smoothing, and the problem of guarding against model errors in regression, *J. Roy. Statist. Soc. Ser. B*, **49**, 364–372.
- Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to non-linear regression and smoothing, *J. Amer. Statist. Assoc.*, **78**, 81–89.