Statistica Sinica **10**(2000), 819-837

F TESTS AND REGRESSION ANALYSIS OF VARIANCE BASED ON SMOOTHING SPLINE ESTIMATORS

L. A. Ramil Novo and W. González Manteiga

Universidade de Santiago de Compostela

Abstract: This paper examines the goodness-of-fit of a polynomial regression model. We derive the asymptotic distribution of two generalizations of the classical F test by means of spline estimators. Furthermore, we propose an analysis of variance technique that extends the classical one from linear to nonparametric regression, and we put forward a coefficient of determination for nonparametric models.

Key words and phrases: Analysis of variance, goodness of fit, hypothesis testing, multiple determination coefficient, nonparametric estimation.

1. Introduction

Consider the regression model

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where it is assumed that $m(x)$, the regression function, belongs to a space of smooth functions $\mathcal{M}$ and the ε_i are i.i.d. random variables with zero expectation and finite variance σ^2. It is also assumed that the design points are regularly distributed, with density f, in a finite interval, taken to be $[0, 1]$ without loss of generality. That is, $\int_0^{x_i} f(x)dx = (i - 0.5)/n$, $i = 1, \ldots, n$. We require f to be continuous, positive and bounded away from zero. Suppose we are interested in testing the goodness of fit of a pth order polynomial model:

$$\mathrm{H}_0 : \ m(x) \in \mathcal{M}_0 \quad \text{against} \quad \mathrm{H}_1 : \ m(x) \in \mathcal{M}\backslash\mathcal{M}_0, \tag{2}$$

where $\mathcal{M}_0$ represents the space formed by the polynomials of degree less than or equal to $p - 1$, $\mathcal{P}_p = \{\theta_1 + \theta_2 x + \cdots + \theta_p x^{p-1} : (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p\}$, and $\mathcal{M}$ is a larger space. If $\mathcal{M} = \mathcal{P}_q$ with $q > p$, we have the classical test of nullity of parameters $\theta_{p+1}, \ldots, \theta_q$. Our attention is focused on more general spaces of a nonparametric nature, such as the Sobolev spaces $\mathcal{W}_2^q[0, 1] = \{g$: the jth order derivative $g^{(j)}$ is absolutely continuous for $0 \leq j \leq q - 1$ and $\int_0^1 [g^{(q)}(x)]^2 dx$ is finite$\}$.

Despite this general setting, classical theory (see Seber (1977)) can be applied to motivate a test statistic, since both hypotheses in (2) allow us to consider linear estimators of m. It is possible to construct an $n \times n$ matrix M_i for each hypothesis so that the vector of estimates at the design points is $\hat{m}_i = M_i Y$. Under the null hypothesis, least squares estimation is available, but under the alternative there is a need to apply nonparametric estimation of the regression curve. Most of these estimators are local averages of the response data, $\hat{m}(x) = n^{-1} \sum_{j=1}^{n} W_{nj}(x) y_j$. Thus, possible choices for M_0 and M_1 are $P_0 = X_0(X_0^t X_0)^{-1} X_0^t$ with $X_0 = [x_i^{j-1}]_{i=1,\ldots,n, j=1,\ldots,p}$, and $H = [n^{-1} W_{nj}(x_i)]_{i,j=1,\ldots,n}$, respectively. In the classical setting with $\mathcal{M} = \mathcal{P}_q$, M_1 is given as $P_1 = X_1(X_1^t X_1)^{-1} X_1^t$ where $X_1 = [x_i^{j-1}]_{i=1,\ldots,n,j=1,\ldots,q}$.

For (2), the following three types of statistics are developed.

1. Residual sums of squares $RSS(M_1, M_0) = Y^t(M_1 - M_0)^t(M_1 - M_0)Y = \sum_{i=1}^{n}[\hat{m}_1(x_i) - \hat{m}_0(x_i)]^2$, or more generally $RSS_D = Y^t D Y$, where D is a positive semidefinite matrix.

2. χ^2-type statistics of the form RSS_D/σ^2. Under the null hypothesis with i.i.d. $N(0, \sigma^2)$ errors, $RSS(P_1, P_0)/\sigma^2$ follows a χ^2_ν distribution with $\nu = \mathrm{tr}(P_1 - P_0)$. When D is a function of the hat matrix H of a nonparametric estimator, a transformation of RSS_D such as $S = [2\mathrm{tr}(D^2)]^{-1/2}[\sigma^{-2} RSS_D - \mathrm{tr}(D)]$ has a distribution which can be approximated by a standardized χ^2_ν (see Ramil Novo and González Manteiga (1998)).

3. F statistics of the form $F = \hat{\sigma}_0^2/\hat{\sigma}^2$, where $\hat{\sigma}_0^2$ and $\hat{\sigma}^2$ are estimators of σ^2 generalizing those used in the classical F test statistic,

$$F_C = \frac{[RSS(I, P_0) - RSS(I, P_1)]/\mathrm{tr}(P_1 - P_0)}{RSS(I, P_1)/\mathrm{tr}(I - P_1)}. \tag{3}$$

Replacing P_1 by H, where H is the hat matrix of an appropriate smoothing technique, gives an F statistic like the one proposed by Cleveland and Devlin (1988) (see also Hastie and Tibshirani (1990, Chap. 3) and Azzalini and Bowman (1993)):

$$F_{CD} = \frac{[RSS(I, P_0) - RSS(I, H)]/\mathrm{tr}((I - P_0)^t(I - P_0) - (I - H)^t(I - H))}{RSS(I, H)/\mathrm{tr}((I - H)^t(I - H))}. \tag{4}$$

Section 2 examines the asymptotic distribution of the above statistic. We consider both convergence to normality and χ^2 approximations. For our purpose, an appropriate choice of H is the matrix associated with a spline estimator of order $2q$ with $q \geq p$ (considering $q = p$ is convenient to avoid strong smoothness assumptions about $m(x)$). Since $HX_0 = X_0$, this choice provides exact estimation when the relation is polynomial; a consequence is that $\hat{m}$ is unbiased under the null hypothesis. The same could be said of local polynomial regression of order $q \geq p$, although in this case H does not have the interesting property of symmetry. This type of estimator has already been considered by Cleveland and

Devlin (1988) and Azzalini and Bowman (1993), though they consider different ways of approximating the distribution of the F_{CD} statistic when the errors are normally distributed. In contrast to the above papers, we consider spline estimators. Furthermore, we study the asymptotic distribution of the F_{CD} statistic together with the generalization of related topics. Given that Cleveland and Devlin (1988) considered a multiple regression setting, we suggest that our theory can be generalized to deal with more general parametric models than polynomials ones.

Section 3 is dedicated to analysis of variance associated with the F_{CD} statistic and other decompositions of the variance related to classical analysis of variance techniques. Section 4 examines a new way to generalize the F test based on the projections of the response data vector onto the spaces $\mathcal{M}_0$ and $\mathcal{M}$. The same idea can be used to generalize the multiple determination coefficient in a natural way. Section 5 illustrates the theory with a simple example based on real data, as well as a simulation study. The Appendix provides the proofs.

2. Distribution of the Cleveland and Devlin F Statistic

Let H be the matrix of weights associated with the spline estimator of order $2p$, that is, $H = [n^{-1}W_{\lambda j}(x_i)]_{i,j=1,\ldots,n}$, where $\hat{m}(x) = n^{-1}\sum_{j=1}^{n} W_{\lambda j}(x)y_j$ is the function which minimizes

$$M_{n\lambda}(g) = n^{-1}\sum_{i=1}^{n}[g(x_i) - y_i]^2 + \lambda \int_0^1 [g^{(p)}(x)]^2 dx \tag{5}$$

over $g \in \mathcal{W}_2^p[0,1]$. It is well-known (see for example, Cox (1983)) that $\hat{m}$ belongs to the space of natural splines $\mathcal{S}_n^p = \{g : g \in \mathcal{C}^{2p-2}[0,1],\ g$ is a polynomial of degree $2p-1$ on $[x_i, x_{i+1}]$, $i = 1,\ldots,n-1$, and of degree $p-1$ on $[0,x_1]$, $[x_n,1]\}$. Demmler and Reinsch (1975) introduced a basis $\{\phi_{kn}\}_{k=1}^{n}$ for $\mathcal{S}_n^p$ satisfying $n^{-1}\sum_{i=1}^{n}\phi_{jn}(x_i)\phi_{kn}(x_i) = \delta_{jk}$ and $\int_0^1 \phi_{jn}^{(p)}(x)\phi_{kn}^{(p)}(x)dx = \gamma_{kn}\delta_{jk}$ $(j,k = 1,\ldots,n)$, with $0 = \gamma_{1n} = \cdots = \gamma_{pn} < \gamma_{(p+1)\,n} \le \cdots \le \gamma_{nn}$, such that $\{\phi_{kn}\}_{k=1}^{p}$ forms a basis for $\mathcal{P}_p$ (see Ragozin (1985)). With this basis, $\hat{m}$ can be expressed as $\hat{m}(x) = \sum_{j=1}^{p}\hat{\theta}_j x^{j-1} + \sum_{j=p+1}^{n}(1+\lambda\gamma_{jn})^{-1}\alpha_{jn}\phi_{jn}(x)$, where $\alpha_{jn} = n^{-1}\sum_{i=1}^{n}\phi_{jn}(x_i)y_i$ and $\hat{m}_0(x) = \sum_{j=1}^{p}\hat{\theta}_j x^{j-1}$ is the least squares estimator (see Eubank (1988, Section 5.3.3)). A detailed study of the relation between spline estimators and the least squares estimator can be seen in Eubank (1984). A remarkable property of the hat matrix H shown in his work is that $HX_0 = X_0$ (and thus $HP_0 = P_0$). Furthermore, writing H in terms of the Demmler and Reinsch basis, $H = \Phi\Gamma\Phi^t$, where $\Phi = [n^{-1/2}\phi_{jn}(x_i)]_{i,j=1,\ldots,n}$ and $\Gamma = \text{diag}[(1+\lambda\gamma_{jn})^{-1}]$, shows the symmetry of this matrix. All these properties make spline estimators especially attractive for testing polynomial models.

Writing $H = \Phi\Gamma\Phi^t$ is useful to the study of asymptotic properties of splines (see Eubank (1988, Section 6.3.2)). The traces of the first two powers of H are given by

$$n^{-1}\mathrm{tr}(H^r) \sim (2\pi n\lambda^{1/(2p)})^{-1}c(f)\int(1+x^{2p})^{-r}dx, \quad r=1,2, \tag{6}$$

where $c(f) = \int_0^1 f(x)^{1/(2p)}dx$ (see Speckman (1981) or Eubank (1988, Section 6.3.2)). It has been proved that the same representation holds for powers of greater order.

The properties of spline estimators have also been studied using the relationship between $\hat{m}$ and a function $\tilde{m}$ which minimizes a continuous version of (5), given by

$$M_\lambda(g) = \int_0^1[g(x)-m(x)]^2 f(x)dx + \lambda\int_0^1[g^{(p)}(x)]^2dx,$$

over $g \in \mathcal{W}_2^p[0,1]$. Cox (1984) proved that if $m \in \mathcal{L}_2[0,1]$, $\tilde{m}$ can be characterized as the unique solution of the differential equation

$$(-1)^p\lambda f^{-1}(x)\frac{d^{2p}g(x)}{dx^{2p}} + g(x) = m(x) \quad x \in [0,1],$$

with boundary conditions $g^{(j)}(0) = g^{(j)}(1) = 0$, $j = p,\ldots,2p-1$. This allows the representation $\tilde{m}(x) = \int G_\lambda(x,t)m(t)f(t)dt$, where $G_\lambda(x,t)$ is the Green's function for the differential operator $(-1)^p\lambda D^{2p} + f$ acting on $\mathcal{N}_p[0,1] = \{g \in \mathcal{C}^{2p}[0,1] : g^{(j)}(0) = g^{(j)}(1) = 0 \text{ for } p \le j \le 2p-1\}$ (see Cox (1993)) .

For each $t \in [0,1]$ let $A_{nt}(g) = n^{-1}\sum_{i=1}^n g^2(x_i) - 2g(t) + \lambda\int_0^1[g^{(p)}(x)]^2dx$, and let $W_\lambda(x,t)$ be the function which minimizes this functional. Given that $W_{\lambda j}(x) = W_\lambda(x,x_j)$, $W_\lambda(x,t)$ plays a role for $\hat{m}$ analogous to what $G_\lambda(x,t)$ does for $\tilde{m}$. Silverman (1984) proved that under appropriate conditions, $W_\lambda(x,t)$ can be approximated by a Priestley-Chao kernel,

$$W_\lambda(x,t) \simeq \frac{1}{f(t)}\frac{1}{h(t)}K(\frac{x-t}{h(t)}),$$

where $h(t) = [\lambda/f(t)]^{1/(2p)}$ and $K(u)$ is the kernel whose Fourier transform is $\varphi(x) = (1+x^{2p})^{-1}$, that is,

$$K(u) = (2\pi)^{-1}\int(1+x^{2p})^{-1}\exp(-ixu)dx. \tag{7}$$

Nychka (1995) proved that $W_\lambda(x,t)$ can be approximated uniformly by $G_\lambda(x,t)$. The kernel in (7) was also studied by Messer and Goldstein (1993).

We now present a result which is important for establishing a relationship between the F_{CD} statistic based on a spline of order $2p$ and the χ^2-type statistics S_{ES} (Eubank and Spiegelman (1990)) and S_C (Chen (1994)):

$$S_{ES} = [2\text{tr}(H - P_0)^4]^{-1/2}[\sigma^{-2}Y^t(H - P_0)^2Y - \text{tr}(H - P_0)^2] \tag{8}$$

and

$$S_C = [2\text{tr}(H - H^2)^2]^{-1/2}[\sigma^{-2}Y^t(H - H^2)Y - \text{tr}(H - H^2)]$$

(these statistics are motivated in the following section).

Lemma 1. *Let H be the matrix associated with a spline estimator of order $2p$, and A be an $n \times n$ matrix with all elements equal to n^{-1}. The following identities hold:*

a) $(I - P_0)^2 - (I - H)^2 = (H - P_0)^2 + 2(H - H^2)$,

b) $(I - A)^2 = (I - H)^2 + (H - P_0)^2 + 2(H - H^2) + (P_0 - A)^2$,

c) $(I - A)^2 = (I - H)^2 + (H - A)^2 + 2(H - H^2)$.

Theorem 1 below gives the relationship between the numerator in F_{CD} and the statistics S_{ES} and S_C, as well as the asymptotic distributions implied by this relationship. Let K denote the kernel (7) and, for simplicity, let $RSS(I, P_0)$ and $RSS(I, H)$ be written as RSS_0 and RSS_1, respectively. Our assumptions A, B and C will be used to show some of the results in the theorem. Assumption A is used in Chen (1994) and Jayasuriya (1996) to apply Nychka's (1995) approximations to the spline weight function. This assumption has been shown to hold with some additional assumptions about f and p. See Chen (1994), Messer and Goldstein (1993) and Nychka's (1995). Messer and Goldstein's (1993) results imply that equations (9) and (10) hold with constant ζ equal to 0 (see below), and f constant. Nychka (1995) showed that they also hold for $\zeta = 0$, $p = 1$ and f with a uniformly continuous derivative. Thus, assumption A could be replaced by any of these additional assumptions about f and p.

A. Let $G_\lambda(s, t)$ be Green's function for the differential operator $(-1)^p\lambda D^{2p} + f$ with domain $\mathcal{N}_p[0, 1]$, and $h = \lambda^{1/(2p)}$. There exist finite, positive constants α, ζ and k such that for all $s, t \in [0, 1]$,

$$|G_\lambda(s,t)| \le \frac{k}{h^{1+\zeta}} \exp(-\alpha|s - t|/h), \qquad \left|\frac{\partial}{\partial s}G_\lambda(s,t)\right| \le \frac{k}{h^{2+\zeta}} \exp(-\alpha|s - t|/h). \tag{9}$$

If $s \neq t$, then

$$\left|\frac{\partial^2}{\partial s \partial t}G_\lambda(s,t)\right| \le \frac{k}{h^{3+\zeta}} \exp(-\alpha|s - t|/h). \tag{10}$$

Either $(\partial^2/\partial s\partial t)G_\lambda(s,t)$ exists for $s=t$, in which case (10) holds, or for all continuous functions g on [0,1],

$$\left|\frac{\partial}{\partial s}\int_0^1[\frac{\partial}{\partial t}G_\lambda(s,t)]g(t)dt\right| \le \left(\frac{k}{h^{3+\zeta}}\right)\left[\int_0^1 \frac{1}{2}\exp(-\alpha|s-t|/h)|g(t)|dt + |g(s)|\right].$$

B. $0<\mu_4=E[\varepsilon_i^4]<\infty$.

C. The ε_i are normally distributed.

Assumptions A, B and C are used to form the following conditions.

Condition 1. Assumptions A and B, $p\ge 2$, $n\to\infty$ and $\lambda\to 0$ such that $n\lambda^{(4p+1)/(4p)}\to\infty$ and $n\lambda^{(3+4\zeta)/(4p)}\to\infty$.

Condition 2. Assumption C, $p\ge 2$, $n\to\infty$ and $\lambda\to 0$ such that $n\lambda\to\infty$.

For notational simplicity f is the uniform density function, only trivial changes are needed for a different f.

Let $S_{CD}=[2\mathrm{tr}((I-P_0)^2-(I-H)^2)^2]^{-1/2}[\sigma^{-2}(RSS_0-RSS_1)-\mathrm{tr}((I-P_0)^2-(I-H)^2)]$.

Theorem 1. *Under the assumptions of Model* (1)*,*

a) $[\mathrm{tr}((I-P_0)^2-(I-H)^2)]^{-1}[\sigma^{-2}(RSS_0-RSS_1)]-1=c_1(H)S_{ES}+c_2(H)S_C$, *where* $c_1(H)=[2tr(H-P_0)^4]^{1/2}/tr((H-P_0)^2+2(H-H^2))$ *and* $c_2(H)=2[2tr(H-H^2)^2]^{1/2}/tr((H-P_0)^2+2(H-H^2))$.

b) *If* H_0 *is true and Condition* 2 *is satisfied, then* $\sup_x|\Pr\{S_{CD}\le x\}-\Pr\{(2\nu)^{-1/2}(\chi^2_\nu-\nu)\le x\}|\le c\lambda^{1/(2p)}$*, where* c *is a constant that does not depend on* n *or* λ*, and* $\nu=(tr(2H-H^2)^3-p)^{-2}(tr(2H-H^2)^2-p)^3$.

c) $S_{CD}=k_1(H)S_{ES}+k_2(H)S_C$*, where each of* $k_1(H)$ *and* $k_2(H)$ *converges to a constant. If Condition* 1 *or* 2 *is satisfied, then*

  c.1) $S_{CD}\to N(0,1)$ as $n\to\infty$ under H_0, and

  c.2) $S_{CD}=S^0_{CD}+[2\sigma^4c(H)]^{-1/2}\|g\|^2+o_p(1)$, where S^0_{CD} has the same distribution as S_{CD} under H_0 and $c(H)=\lim_{n\to\infty}\lambda^{1/(2p)}\mathrm{tr}((I-P_0)^2-(I-H)^2)^2=\int(2K-K^{*2})^2$, if $m(x)=m_0(x)+c(n)g(x)$, where $m_0(x)\in\mathcal{P}_p$, $c(n)=n^{-1/2}(\lambda^{1/(2p)})^{-1/4}$, $g(x)\in\mathcal{W}^p_2[0,1]$ such that $\int_0^1 x^{j-1}g(x)dx=0$ for $j=1,\dots,p$, and K^{*2} is the convolution of K with itself.

Comments on Theorem 1.

1. Part a) of Theorem 1 states that the numerator of F_{CD} is an unbiased estimator of the error variance only under H_0.
2. Part b) provides χ^2 approximations for S_{CD} analogous to those studied by Ramil Novo and González Manteiga (1998) for S_{ES}, S_C and other related statistics. It is an extension of the classical result that under H_0, $\sigma^{-2}[RSS(I,P_0)-RSS(I,P_1)]\sim\chi^2_\nu$, $\nu=\mathrm{tr}(P_1-P_0)$. Part c) extends the convergence to normality that holds in the classical context with $\nu\to\infty$.

3. Replacing σ^2 by an estimator $\hat{\sigma}^2$ in S_{CD} gives a test statistic $\hat{S}_{CD}$ such that

$$S_{CD}-\hat{S}_{CD}=\hat{\sigma}^{-2}(\hat{\sigma}^2-\sigma^2)\{S_{CD}+[2\mathrm{tr}((I-P_0)^2-(I-H)^2)^2]^{-1/2}\mathrm{tr}((I-P_0)^2-(I-H)^2)\}. \tag{11}$$

Hence parts b) and c) are still valid for $\hat{S}_{CD}$ if $\hat{\sigma}^2 = \sigma^2 + o_p((\lambda^{1/(2p)})^{1/2})$. Hall and Marron (1990) and Buckley, Eagleson and Silverman (1988) proved that $\hat{\sigma}^2_{CD} = RSS_1/\mathrm{tr}(I-H)^2$ satisfies this property under the conditions in Theorem 1. Because the $\hat{S}_{CD}$ statistic corresponding to this estimator is $[2\mathrm{tr}((I-P_0)^2-(I-H)^2)^2]^{-1/2}\mathrm{tr}((I-P_0)^2-(I-H)^2)(F_{CD}-1)$, the limiting distribution of F_{CD} is also determined by Theorem 1. Hall and Marron also show that optimal estimation of σ^2 demands less smoothing than optimal estimation of the regression curve. This suggests that it might be convenient to use a different smoothing parameter in the numerator and denominator of the F_{CD} statistic. Choice of the smoothing parameter for the numerator should be made to optimally estimate the discrepancy with respect to the null hypothesis, while the smoothing parameter in the denominator should be chosen to optimally estimate the error variance.

3. Analysis of Variance

The classical analysis of variance allows for a natural extension to the new context with a nonparametric alternative, providing a way of expressing the quantities required for the F_{CD} test in an ANOVA table. Part a) of Lemma 1 gives the following decomposition,

$$Y^t(I-P_0)^2Y = Y^t(H-P_0)^2Y + 2Y^t(H-H^2)Y + Y^t(I-H)^2Y,$$

that is,

$$\sum_{i=1}^{n}[y_i-\hat{m}_0(x_i)]^2 = \sum_{i=1}^{n}[\hat{m}(x_i)-\hat{m}_0(x_i)]^2 + 2n\lambda\int_0^1[\hat{m}^{(p)}(x)]^2dx + \sum_{i=1}^{n}[y_i-\hat{m}(x_i)]^2. \tag{12}$$

This decomposition, which generalizes the classical one for linear models, allows for the following interpretation. The term on the left hand side, $SS_{NE} = \sum_{i=1}^n[y_i-\hat{m}_0(x_i)]^2$, is the *variability not explained by the model* of the null hypothesis. The first term on the right hand side, $SS_D = \sum_{i=1}^n[\hat{m}(x_i)-\hat{m}_0(x_i)]^2$, represents the *variability component due to the discrepancy with respect to the model* of the null hypothesis, and the second term, $SS_B = 2n\lambda\int_0^1[\hat{m}^{(p)}(x)]^2dx$, can be considered as the *variability component due to the bias of nonparametric estimation of the discrepancy with respect to the model* (note that $\int_0^1[\hat{m}^{(p)}(x)]^2dx = \int_0^1[d^p/dx^p(\hat{m}-\hat{m}_0)(x)]^2dx$). Finally, $SS_R = \sum_{i=1}^n[y_i-\hat{m}(x_i)]^2$ is the *residual variability.*

The above decomposition helps to compare χ^2 type statistics. The S_{ES} statistic in Eubank and Spiegelman (1990) and Jayasuriya (1996) uses SS_D, Chen's (1988) S_C uses SS_B, and S_{CD} is based on $SS_D + SS_B$, combining both measures of discrepancy with respect to the model.

Let $SS_T = \sum_{i=1}^n (y_i - \overline{y})^2$ be the *total variability* of the response variable and $SS_M = \sum_{i=1}^n (\hat{m}_0(x_i) - \overline{y})^2$ be the *variability explained by the model* of the null hypothesis. Then part b) of Lemma 1 leads to the following ANOVA table:

Variability	Sum of squares	Degrees of freedom	Mean squares
Explained by the model	SS_M	$p-1$	$SS_M/(p-1)$
Discrepancy with the model	SS_D	$\mathrm{tr}(H^2)-p$	$SS_D/(\mathrm{tr}(H^2)-p)$
Bias	SS_B	$2\mathrm{tr}(H)-2\mathrm{tr}(H^2)$	$SS_B/(2\mathrm{tr}(H)-2\mathrm{tr}(H^2))$
Residual	SS_R	$n+\mathrm{tr}(H^2)-2\mathrm{tr}(H)$	$SS_R/(n+\mathrm{tr}(H^2)-2\mathrm{tr}(H))$
Total	SS_T	$n-1$	

This table suggests several tests. For example, assuming that $m \in \mathcal{P}_p$, we could test the null hypothesis that the predictor variable has no effect using the classical F test

$$F = \frac{SS_M}{SS_T - SS_M} \cdot \frac{n-p}{p-1} = \frac{SS_T - SS_{NE}}{SS_{NE}} \cdot \frac{n-p}{p-1}. \tag{13}$$

Avoiding the assumption of a polynomial model, we could have a test statistic to study the significance of the predictor variable, given by

$$F = \frac{SS_T - SS_R}{SS_R} \cdot \frac{n + \mathrm{tr}(H^2) - 2\mathrm{tr}(H)}{2\mathrm{tr}(H) - \mathrm{tr}(H^2) - 1}. \tag{14}$$

Replacing σ^2 by its estimator $\hat{\sigma}_R^2$ in (8) leads to an expression of Eubank and Spiegelman's statistic by mean squares: $F = [2(\mathrm{tr}(H^4) - p)]^{-1/2}(\mathrm{tr}(H^2) - p)(\hat{\sigma}_D^2/\hat{\sigma}_R^2 - 1)$. Likewise, Chen's statistic can be expressed as $F = [2\mathrm{tr}(H - H^2)^2]^{-1/2}\mathrm{tr}(H - H^2)(\hat{\sigma}_B^2/\hat{\sigma}_R^2 - 1)$.

The total variability also admits a simpler decomposition. By part c) of Lemma 1:

$$\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n (\hat{m}(x_i) - \overline{y})^2 + 2n\lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx + \sum_{i=1}^n [y_i - \hat{m}(x_i)]^2, \tag{15}$$

where $SS_E = \sum_{i=1}^n (\hat{m}(x_i) - \overline{y})^2$ measures the *variability explained* by the regression. This new decomposition is, in fact, an abbreviated version of the previous one given that $SS_E = SS_M + SS_D$. Equation (15) provides an alternative test statistic to that given by (14):

$$F = \frac{SS_E}{SS_R} \cdot \frac{n + \mathrm{tr}(H^2) - 2\mathrm{tr}(H)}{\mathrm{tr}(H^2) - 1} = \frac{SS_T - (SS_R + SS_B)}{SS_R} \cdot \frac{n + \mathrm{tr}(H^2) - 2\mathrm{tr}(H)}{\mathrm{tr}(H^2) - 1}.$$

Two special cases of this statistic were studied in Eubank and Hart (1993) and Eubank and LaRiccia (1993), among others.

We have shown that decompositions of the variance generalize the classical ones. In the classical context, the sums of squares are associated with an orthogonal decomposition. In the following section we examine orthogonality in the nonparametric context. As we will see, the classical inner product is not used since it is not the most convenient one for our purpose.

4. Natural Generalization of the F Test

The spline estimator can be seen as a projection of the response variable onto $\mathcal{W}_2^p \equiv \mathcal{W}_2^p[0,1]$ in the following way (see Cox (1984) or Speckman (1981)).

Let $X = \mathbb{R}^n \times \mathcal{L}_2[0,1]$, where $\mathbb{R}^n$ is the n-dimensional Euclidean space. Let us consider for $\lambda > 0$ the norm in X given by $\|(Y,g)\|_X^2 = \|Y\|_n^2 + \lambda\|g\|^2$, where $\|Y\|_n^2 = n^{-1}\sum_{i=1}^n y_i^2$ and $\|g\|^2 = \int_0^1 [g(x)]^2 dx$ are the ordinary norms in $\mathbb{R}^n$ and $\mathcal{L}_2[0,1]$ respectively. If $n \geq p$, $g^\star = ((g(x_1),\ldots,g(x_n))^t, g^{(p)})$ is an injective bounded linear transformation from $\mathcal{W}_2^p$ into X. Hence, by the Closed Range and Open Mapping Theorems, $\mathcal{W}_2^{p\star}$, the image of $\mathcal{W}_2^p$, is closed, and $\mathcal{W}_2^p$ is complete under the induced norm

$$\|g\|_{n,\lambda}^2 = \|g^\star\|_X^2 = n^{-1}\sum_{i=1}^n g^2(x_i) + \lambda\int_0^1 [g^{(p)}(x)]^2 dx.$$

Letting $Y^\star = ((y_1,\ldots,y_n)^t, 0) \in X$; the spline estimator $\hat{m}(x)$ of order $2p$ is the unique element in $\mathcal{W}_2^p$ that solves the minimization problem, $\min_{g\in\mathcal{W}_2^p} \|Y^\star - g^\star\|_X^2$, and thus, $\hat{m}^\star = ((\hat{m}(x_1),\ldots,\hat{m}(x_n))^t, \hat{m}^{(p)})$ is the projection of $Y^\star$ onto $\mathcal{W}_2^{p\star}$. Considering $\hat{m}_0^\star = ((\hat{m}_0(x_1),\ldots,\hat{m}_0(x_n))^t, 0) = (P_0Y, 0) \in X$,

$$\|Y^\star - \hat{m}_0^\star\|_X^2 = \|Y^\star - \hat{m}^\star\|_X^2 + \|\hat{m}^\star - \hat{m}_0^\star\|_X^2, \qquad (16)$$

since $\hat{m}^\star$ is the perpendicular projection of $Y^\star$ onto the closed subspace $\mathcal{W}_2^{p\star}$ in the norm $\|\ \|_X$. However if we write $\hat{m} = (\hat{m}(x_1),\ldots,\hat{m}(x_n))^t$ and $\hat{m}_0 = (\hat{m}_0(x_1),\ldots,\hat{m}_0(x_n))^t$, then the summands in (16) are $\|Y^\star - \hat{m}_0^\star\|_X^2 = \|Y - \hat{m}_0\|_n^2 = n^{-1}SS_{NE}$, $\|Y^\star - \hat{m}^\star\|_X^2 = \|Y - \hat{m}\|_n^2 + \lambda\|\hat{m}^{(p)}\|^2 = n^{-1}SS_R + (2n)^{-1}SS_B$ and $\|\hat{m}^\star - \hat{m}_0^\star\|_X^2 = \|\hat{m} - \hat{m}_0\|_n^2 + \lambda\|\hat{m}^{(p)}\|^2 = n^{-1}SS_D + (2n)^{-1}SS_B$. In this way we arrive at the same decomposition of variance as in (12) but seen from a different angle. Equation (16) generalizes the classical decomposition for $\mathcal{M} = \mathcal{P}_q$

(with $q > p$) used in the numerator of the classical F test in (3). In view of this, we can consider a new generalization of the F statistic given by:

$$F_N = \frac{\|\hat{m}^\star - \hat{m}_0^\star\|_X^2 \ / \ \mathrm{tr}(H - P_0)}{\|Y^\star - \hat{m}^\star\|_X^2 \ / \ \mathrm{tr}(I - H)} = \frac{\|Y^\star - \hat{m}_0^\star\|_X^2 - \|Y^\star - \hat{m}^\star\|_X^2}{\|Y^\star - \hat{m}^\star\|_X^2} \cdot \frac{\mathrm{tr}(I - H)}{\mathrm{tr}(H - P_0)}. \tag{17}$$

By analogy with the classical F statistic we use the following notation: $RSS_0 = \sum_i [y_i - \hat{m}_0(x_i)]^2 = Y^t(I - P_0)Y$ and $RSS_1 = \sum_i [y_i - \hat{m}(x_i)]^2 + n\lambda \int_0^1 [\hat{m}^{(p)}(x)]^2 dx = Y^t(I - H)Y$.

Let $S_N = [2(\mathrm{tr}(H^2) - p)]^{-1/2}[\sigma^{-2}(RSS_0 - RSS_1) - (\mathrm{tr}(H) - p)]$.

Theorem 2. *Under the assumptions of the Model* (1),

a) $[tr(H - P_0)]^{-1}[\sigma^{-2}(RSS_0 - RSS_1)] - 1 = c_1(H)S_{ES} + c_2(H)S_C$, *where* $c_1(H) = [2tr(H - P_0)^4]^{1/2}/tr(H - P_0)$ *and* $c_2(H) = [2tr(H - H^2)^2]^{1/2}/tr(H - P_0)$.

b) *If* H_0 *is true and Condition* 2 *is satisfied, then* $\sup_x |\Pr\{S_N \le x\} - \Pr\{(2\nu)^{-1/2} (\chi^2_\nu - \nu) \le x\}| \le c\lambda^{1/(2p)}$, *where* c *is a constant which does not depend on* n *or* λ *and* $\nu = (tr(H^3) - p)^{-2}(tr(H^2) - p)^3$.

c) $S_N = k_1(H)S_{ES} + k_2(H)S_C$, *where each of* $k_1(H)$ *and* $k_2(H)$ *converges to a constant, and, if one of the Conditions* 1 *or* 2 *is satisfied,*

  c.1) $S_N \to N(0,1)$ *as* $n \to \infty$ *under* H_0, *and*

  c.2) $S_N = S_N^0 + [2\sigma^4 c(H)]^{-1/2} \|g\|^2 + o_p(1)$, *where* S_N^0 *has the same distribution as* S_N *under* H_0 *and* $c(H) = \lim_{n\to\infty} \lambda^{1/(2p)} tr(H^2) = \int K^2$, *if* $m(x) = m_0(x) + c(n)g(x)$, *where* $m_0(x) \in \mathcal{P}_p$, $c(n) = n^{-1/2}(\lambda^{1/(2p)})^{-1/4}$ *and* $g(x) \in \mathcal{W}_2^p[0,1]$ *such that* $\int_0^1 x^{j-1} g(x)dx = 0$ *for* $j = 1, \ldots, p$.

A decomposition analogous to that in equations (15) and (16) allows us to define a coefficient of determination for regression models that generalizes the classical one in a natural way. Let $m \in \mathcal{M} = P_q$ or $\mathcal{W}_2^q$ $(q \ge 1)$ and $\hat{m}(x)$ be the function which minimizes $\|Y^\star - g^\star\|_X^2$ over $g \in \mathcal{M}$. We can define the model's $\mathcal{M}$ coefficient of determination as

$$R^2_{\mathcal{M}} = \frac{\|\hat{m}^\star - \bar{Y}^\star\|_X^2}{\|Y^\star - \bar{Y}^\star\|_X^2} = \frac{\text{explained variability}}{\text{total variability}}, \tag{18}$$

where $\bar{Y}^\star = ((\bar{y}, \ldots, \bar{y})^t, 0) \in X$. With this definition, when the departure model is $\mathcal{M} = \mathcal{P}_q$, $R^2_{\mathcal{M}}$ is the usual multiple coefficient of determination.

The natural F statistic in (17), can now be expressed as

$$F_N = \frac{R^2 - R_0^2}{1 - R^2} \cdot \frac{\mathrm{tr}(I - H)}{\mathrm{tr}(H - P_0)},$$

where R^2 and R_0^2 are the coefficients of determination for $\mathcal{M} = \mathcal{W}_2^p$ and $\mathcal{M}_0 = \mathcal{P}_p$ respectively, generalizing the usual expression for linear models.

Final remarks

Both S_{CD} and S_N can be expressed as $S = [2\text{tr}(D^2)]^{-1/2}[\sigma^{-2}Y^tDY - \text{tr}(D)]$, where $D = (2H - H^2) - P_0$ in S_{CD} and $D = H - P_0$ in S_N. This permits us to show the consistency of the corresponding tests for fixed alternatives with an argument like the one in Eubank and Spiegelman (1990). Consider $m(x) = m_0(x) + g(x)$, with $m_0(x) \in \mathcal{P}_p$ and $g(x) \in \mathcal{W}_2^p[0,1]$, such that $\int_0^1 x^{j-1}g(x)dx = 0$ for $j = 1,\ldots,p$. It holds that $n^{-1}[2\text{tr}(D^2)]^{1/2}S = \sigma^{-2}n^{-1}g^tDg + n^{-1}[\sigma^{-2}\varepsilon^tD\varepsilon - \text{tr}(D)] + 2\sigma^{-2}n^{-1}g^tD\varepsilon$. Under either Condition 1 or 2, the last two summands in this equation are $o_p(1)$, $n^{-1}g^tDg \to \|g\|^2$ and $n^{-1}[2\,\text{tr}(D^2)]^{1/2} \to 0$. Thus consistency is guaranteed.

5. Examples and Simulation Results

We begin this section by analyzing a mortality table from Green and Silverman (1994, Chap. 5, p.101). Of interest is the relationship between the age x and the natural logarithm of the estimated mortality rate, y. The mortality rate for each age group is the ratio between the annual number of deaths and the number of individuals in the group. A plot of y versus x shows that the relationship is approximately linear between the ages 65 and 92.

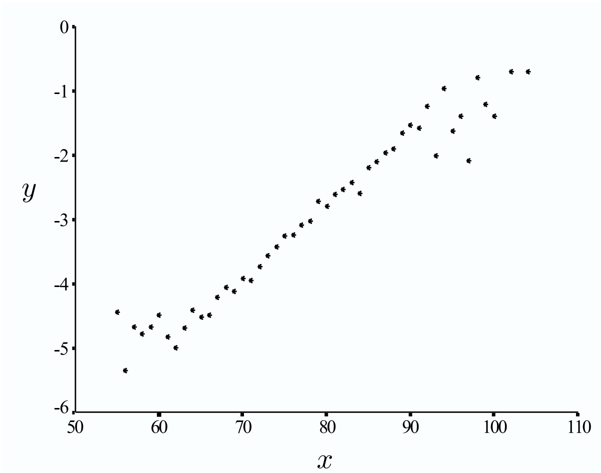

Figure 1. Logarithm of the annual mortality rate versus age.

To test this hypothesis we considered a transformation of x by recentering and rescaling such that $x_i = (i - 0.5)/28$. We applied three different tests: the classical F test for linear models assuming that $m(x) \in \mathcal{P}_4$ and tests based on F_{CD} and F_N - the latter two defined in terms of a cubic spline estimator. We chose $\mathcal{P}_4$ as the alternative model for the classical F test since a polynomial of degree three provides a good fit in the whole region (see Figure 1). For the same reason we use a cubic spline for nonparametric F tests. Results of these three tests are presented in the ANOVA tables below. For the F_{CD} ANOVA table, SS_{NE}, SS_D, SS_B and SS_R are taken from equation (12), and in the F_N ANOVA table, $SS_D = n\|\hat{m}^\star - \hat{m}_0^\star\|_X^2$ and $SS_R = n\|Y^\star - \hat{m}^\star\|_X^2$. Evaluating the general cross validation function in a wide range of bandwidths (see Craven and Wahba (1979)) for the data corresponding to ages between 65 and 92 one observes that, although decreasing, the function becomes nearly constant after $h = \lambda^{1/4} = 0.0925$. Therefore, we consider $h = 0.0925$ an adequate value to show the F tables based on splines. Nonetheless, how to select h in practice to get the most appropriate ANOVA table needs to be studied.

CLASSICAL F TEST - ANOVA TABLE

Source	SS	D.F.	MS
D	.027903	2	.013951
R	.191895	24	.007996
NE	.219798	26	

F_C= 1.744863 p value=.196106

F_{CD} TEST - ANOVA TABLE

Source	SS	D.F.	MS
D	.010137	1.563893	.006482
B	.019528	1.572836	.012416
R	.190133	22.863272	.008316
NE	.219798	26	

F_{CD}= 1.13724 p value=.323964

F_N TEST - ANOVA TABLE

Source	SS	D.F.	MS
D	.019901	2.35031	.008467
R	.199897	23.64969	.008452
NE	.219798	26	

F_N=1.00178 p value=.379358

The coefficients of determination in (18) are .991204, .992321 and .992001 for a straight line fit, a cubic polynomial fit and a cubic spline fit, respectively. They all have similar values and none of the tests is significant. However, nonparametric tests show stronger evidence that a polynomial of degree one is enough to get a satisfactory fit. Including in the analysis all pairs of data for which there were a reasonable number of individuals to estimate the mortality rate (ages between 55 and 97) gives the following p values : $p < .0001$ for the classical F test assuming $m(x) \in \mathcal{P}_4$, $p = .6371$ assuming only that $m(x) \in \mathcal{P}_3$, and $p < .0001$ for both the F_{CD} and F_N tests (based on a bandwidth selected by the generalized cross validation procedure with $h = .0523$). Thus, nonparametric tests again give satisfactory p values.

Simulation results

In order to study the performance of nonparametric F tests with small sample sizes, carried out a small simulation study to examine the level and power of these tests. We generated 1,000 samples of size $n = 50$ from each of the following models:

model 0: $y_i = 1 + 2x_i + \varepsilon_i$ and model 1: $y_i = 1 + 2x_i + x_i^2 + \varepsilon_i$,
where $x_i = (i - 0.5)/50$, $i = 1, \ldots, 50$, and the errors ε_i are i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.25$. We used each sample to test the polynomial model of degree one ($\mathrm{H}_0 : m(x) \in \mathcal{P}_2$) with the following tests: the classical F test with alternative hypothesis $m(x) \in \mathcal{P}_3$, tests based on S_{CD}, S_N, S_{ES} and S_C (assuming that σ^2 is known), tests based on F_{CD} and F_N (since σ^2 is not usually known), and finally, two bootstrap versions based on $B = 500$ resamples from each original data set. Resamples were generated by a naive bootstrap from the residuals (see González Manteiga and Cao Abad (1993)) and were used to calculate the bootstrap version of the corresponding $RSS_0 - RSS_1$ in the numerator of both Cleveland-Devlin and natural F statistics. Nonparametric tests are based on a cubic spline estimator. All the tests were carried out at a significance level $\alpha = 0.05$. The four line charts below show the percentage of rejections for values of $h = \lambda^{1/4}$ (the equivalent bandwidth) varying from 0.02 to 0.26 with a step of 0.01. Figure 2 represents the percentage of rejections with test statistics requiring σ^2 to be known (S_{CD}, S_N, S_{ES} and S_C). Critical values were calculated using both normal and χ^2 approximations. With respect to χ^2 approximations we used Theorems 1 and 2, and Theorems 2.2 and 2.4 of Ramil Novo and González Manteiga (1998). We only show results for χ^2 approximations as the normal approximations lead to a higher than expected level under the null hypothesis for all the S statistics considered. This was already observed by Eubank and Spiegelman (1990) and Chen (1994) for their S statistics. Charts on Figure 3 represent the percentage of rejections for tests applicable when the error variance is not known (such as those based on F_{CD} and F_N) and the corresponding bootstrap tests. For the same reason as with the S tests, critical values for the F statistics were obtained through the χ^2 approximations.

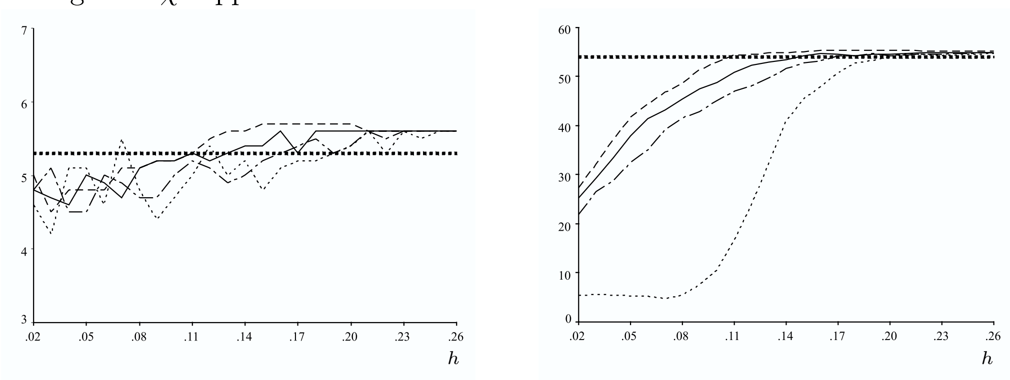

Figure 2. Percentage of rejections assuming σ^2 known. Left: Model 0. Right: Model 1. Test statistics: —— natural S_N, – - – Cleveland and Devlin's S_{CD}, – – – Eubank and Spiegelman's S_{ES}, and, - - - - Chen's S_C, all based on χ^2 critical values. Horizontal reference line: classical F test.

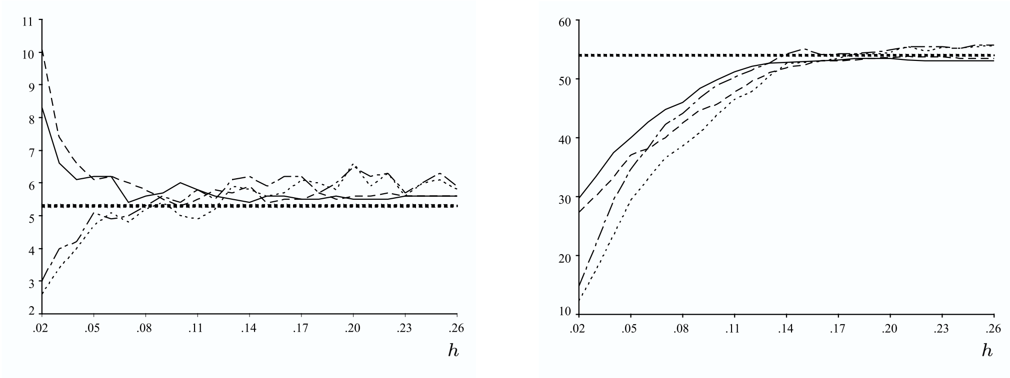

Figure 3. Percentage of rejections assuming σ^2 unknown. Left: Model 0. Right: Model 1. F tests: —— natural F_N, – – – Cleveland and Devlin's F_{CD}. Bootstrap tests: – - – natural, - - - - Cleveland and Devlin. Horizontal reference line: classical F test.

Results of our simulation study reveal several interesting facts.

1. Considering the performance of S_N, S_{CD}, F_N and F_{CD} statistics under the null hypothesis (see charts on the left), one sees that χ^2 approximations give a rate of rejections very similar to that of the classical F test and, therefore, very close to the α level. However, a comparison between Figure 2-left and Figure 3-left charts shows that when σ^2 is known the problem of the bandwidth selection is not relevant since the S tests are not very sensitive to the selection of the bandwidth.

2. Under the alternative (Model 1, charts on the right), S_N and S_{CD} can be decomposed in three terms, namely $S = S^0 + S^1 + S^2$ (see the final remarks of Section 4 and Appendix for details). The component S^0 has the same distribution as S under H_0 and S^2 has zero expectation, thus $\Pr\{S \geq c_\alpha\} = \Pr\{S^0 \geq c_\alpha - S^1 - S^2\}$ depends on h through S^1.

3. A comparison among all the S statistics (see Figure 2) shows that the power of S_{ES} is bigger than the others. However this should be interpreted with care, because S_{ES} tends to have a larger type I error. The power of the natural S or F tests is higher than that of the corresponding Cleveland and Devlin tests, but they have a similar behavior to the classical F test, for a wide range of bandwidths. Chen's S statistic has poor power for small bandwidths. There seems to be a common range of optimal bandwidths for all S statistics. The optimal selection of h for these S and F statistics is still an open problem that needs research.

4. Both bootstrap tests (Cleveland and Devlin's and the natural) show good performance, but selecting a good bandwidth is important. Moreover, they appeared to be less stable than the corresponding S or F tests.

Our final conclusion is that if one selects an adequate bandwidth, there is hardly anything to lose using nonparametric F tests when classical methods are applicable. Using nonparametric estimators, we can estimate a much wider class of smooth functions, and thus avoid strong smoothness and shape assumptions about the regression curve.

Acknowledgements

This work was partially supported by the Spanish Ministry of Education and Culture with the DGES Grant PB95-0826 and by the Xunta de Galicia (Spain) under Research Project XUGA 20701B96.

The authors wish to express their thanks to an anonymous referee, the Chair-Editor and the Associate Editor for a number of valuable comments toward improvement of this paper. The authors are also grateful to Rosalía Vázquez Álvarez for her careful supervision of the writing.

Appendix

Proof of Lemma 1.

Part a). Write $I = (I - H)^2 + H^2 + 2(H - H^2)$. Now, observing that $(I - P_0) = (I - P_0)^2$ and that $H^2 - P_0 = (H - P_0)^2$ (since H and P_0 are symmetric matrices, $HP_0 = P_0$ and P_0 is idempotent), we obtain

$$(I - P_0)^2 = (I - H)^2 + (H - P_0)^2 + 2(H - H^2).$$

Part b). It is an immediate consequence of Part a) as $(I - P_0)^2 = (I - A)^2 - (P_0 - A)^2$.

Part c). Proof of a) is valid for any matrix H associated with a spline of order $2q$ with $q \geq p$, thus c) is a particular case of a) as $P_0 = A$ when $X_0 = (1, \ldots, 1)^t$.

Let the r-times ($r \geq 2$ an integer) convolution product of the kernel (7) be written as K^{*r}, so that $K^{*2} = K * K$, etc. For $r = 0, 1$, let K^{*r} be K^r. The following lemma shows that asymptotic expressions of $n^{-1}\mathrm{tr}(H^r)$ in terms of the powers of φ can be expressed in terms of Silverman's kernel. As a consequence, integrals of linear combinations of the powers of φ can be expressed in terms of K.

Lemma 2. *Let* $K(u) = (2\pi)^{-1} \int (1 + x^{2p})^{-1} \exp(-ixu)dx$. *Then*

$$(2\pi)^{-1} \int (1 + x^{2p})^{-r} dx = \int K(x) K^{*(r-1)}(x) dx = K^{*r}(0) \quad \text{for } r = 1, 2, \ldots$$

In particular, when $r = 2s$, $(2\pi)^{-1} \int (1 + x^{2p})^{-r} dx = \int [K^{*s}(x)]^2 dx$ *for* $s = 1, 2, \ldots$.

Proof. Given $f \in \mathcal{L}_1(\mathbb{R})$, denote by $\hat{f}(t) = \int f(x)\exp(ixt)dx$ its Fourier transform. The Fourier Inverse Theorem gives $\varphi(x) = \hat{K}(x)$ and $\hat{\varphi} = 2\pi K$. Since $(f * g)\hat{} = \hat{f} \cdot \hat{g}$, by Fubini's Theorem it holds that

$$\begin{aligned}
&(2\pi)^{-1}\int \varphi(x)^r dx = (2\pi)^{-1}\int \varphi(x)[\hat{K}(x)]^{r-1}dx \\
&= (2\pi)^{-1}\int \varphi(x)(K^{*(r-1)})\hat{}\,(x)dx = (2\pi)^{-1}\int K^{*(r-1)}(t)[\int \varphi(x)\exp(ixt)dx]dt \\
&= (2\pi)^{-1}\int K^{*(r-1)}(t)\hat{\varphi}(t)dt = \int K(t)K^{*(r-1)}(t)dt = K^{*r}(0).
\end{aligned}$$

See Bhattacharya and Rao (1986, Chap. 2) for properties of Fourier transforms.

Proofs of Theorems.

First note that all the S statistics are of the type $S = [2\mathrm{tr}(D^2)]^{-1/2}[\sigma^{-2}Y^tDY - \mathrm{tr}(D)]$, where D is $(I-P_0)^2 - (I-H)^2 = 2H - H^2 - P_0$ in the case of S_{CD}, is $H - P_0$ for S_N, is $(H-P_0)^2$ for S_{ES}, and is $H - H^2$ for S_C.

Part a) of both theorems concerns the terms $[Y^tDY/\mathrm{tr}(D)]/\sigma^2 - 1$. Setting $RSS_0 = RSS(I, P_0)$ and $RSS_1 = RSS(I, H)$, $\sigma^{-2}(RSS_0 - RSS_1) = \sigma^{-2}Y^tDY$ where $D = (I-P_0)^2 - (I-H)^2$. By Lemma 1, $D = (H-P_0)^2 + 2(H - H^2)$. Thus,

$$\begin{aligned}
\sigma^{-2}Y^tDY - \mathrm{tr}(D) = \sigma^{-2}(Y^t(H-P_0)^2Y - \mathrm{tr}(H-P_0)^2) + 2(\sigma^{-2}Y^t(H-H^2)Y \\
-\mathrm{tr}(H-H^2)). \qquad (19)
\end{aligned}$$

Analogously, defining $RSS_0 = Y^t(I-P_0)Y$ and $RSS_1 = Y^t(I-H)Y$, we obtain $\sigma^{-2}(RSS_0 - RSS_1) = \sigma^{-2}Y^tDY$ with $D = H - P_0 = (H-P_0)^2 + (H-H^2)$ (as $(H-P_0)^2 = H^2 - P_0$). Thus,

$$\begin{aligned}
\sigma^{-2}Y^tDY - \mathrm{tr}(D) = \sigma^{-2}(Y^t(H-P_0)^2Y - \mathrm{tr}(H-P_0)^2) + (\sigma^{-2}Y^t(H-H^2)Y \\
-\mathrm{tr}(H-H^2)). \qquad (20)
\end{aligned}$$

Dividing (19) and (20) by the corresponding $\mathrm{tr}(D)$ term, the equality stated in a) of each of the theorems is attained:

$$[Y^tDY/\mathrm{tr}(D)]/\sigma^2 - 1 = c_1(H)S_{ES} + c_2(H)S_C, \qquad (21)$$

where $c_1(H) = [2\mathrm{tr}(H-P_0)^4]^{1/2}/\mathrm{tr}(D)$ and $c_2(H) = 2[2\mathrm{tr}(H-H^2)^2]^{1/2}/\mathrm{tr}(D)$ if $D = (I-P_0)^2 - (I-H)^2$, $c_2(H) = [2\mathrm{tr}(H-H^2)^2]^{1/2}/\mathrm{tr}(D)$ if $D = H - P_0$.

To prove b) and c), consider a decomposition of each S into three terms, $S = S^0 + S^1 + S^2$, where $S^0 = [2\mathrm{tr}(D^2)]^{-1/2}[\sigma^{-2}\varepsilon^tD\varepsilon - \mathrm{tr}(D)]$, $S^1 = \sigma^{-2}nc^2(n)[2\,\mathrm{tr}(D^2)]^{-1/2}\,[n^{-1}g^tDg]$ and $S^2 = 2\sigma^{-2}c(n)\,[2\mathrm{tr}(D^2)]^{-1/2}[g^tD\varepsilon]$. The summands S^i corresponding to Eubank and Spiegelman's statistic (with $D = H^2 - P_0$) and

Chen's statistic (with $D = H - H^2$) will be denoted as S^i_{ES} and S^i_C, $i = 0$, 1 and 2.

When H_0 is true, $S = S^0$. This permits us to obtain the χ^2 approximations found in b) of each theorem as an application of the uniform bound of Equation 3 in Buckley and Eagleson (1988, Section 2). We use the following notation for constants involved in that bound: $\beta = \max_{1\le i\le n} \lambda_i^2/\sum_{k=1}^n \lambda_k^2$, $\eta = \sum_{k=1}^n \lambda_k^4/(\sum_{k=1}^n \lambda_k^2)^2$ and $\nu = (\sum_{k=1}^n \lambda_k^2)^3/(\sum_{k=1}^n \lambda_k^3)^2$, where the λ_i are the eigenvalues of D, and thus $\mathrm{tr}(D^r) = \sum_{k=1}^n \lambda_k^r$. Since $\max_{1\le i\le n} \lambda_i^4/(\sum_{k=1}^n \lambda_k^2)^2 = \beta^2 \le (\sum_{k=1}^n \lambda_k^4)/(\sum_{k=1}^n \lambda_k^2)^2 = \eta$, it is possible to work only with traces of powers of D to obtain an asymptotic expression of Buckley and Eagleson's bound.

Let $D = 2H - H^2 - P_0$. Since $HP_0 = P_0$, $D^r = (2H - H^2)^r - P_0$ $r = 1, 2, \ldots$. Let $t_r = (2\pi)^{-1}\int(1 + x^{2p})^{-r}dx$. Given that $n^{-1}\mathrm{tr}(H^r) \sim (n\lambda^{1/(2p)})^{-1}t_r$, then $n^{-1}\mathrm{tr}(D^r) \sim (n\lambda^{1/(2p)})^{-1}d_r$, where $d_r = \sum_{k=0}^r r!/[k!(r-k)!]2^k(-1)^{r-k}t_{2r-k}$. Thus $\eta = \{[n^{-2}\mathrm{tr}(D^4)]/\ [n^{-1}\mathrm{tr}(D^2)]^2\} \sim d_0\lambda^{1/(2p)}$, where $d_0 = d_4/d_2^2$. Analogously, $\nu^{-1} = \{n^{-1}[n^{-1}\mathrm{tr}(D^3)]^2/\ [n^{-1}\mathrm{tr}(D^2)]^3\} \sim d_0'\lambda^{1/(2p)}$, where $d_0' = d_3^2/d_2^3$. Then, letting $c(x) = (10 + 3(1 - 8/x)^{-2})/2\pi$, as $\lambda \to 0$, both $c(\beta^{-1})$ and $c(\nu)$ converge to a constant c_0. Constant c in Theorem 1 can be obtained by observing that $c(\beta^{-1})\eta + c(\nu)\nu^{-1} \sim c_0(d_0\lambda^{1/(2p)} + d_0'\lambda^{1/(2p)})$. Constant c in b) of Theorem 2 can be obtained in a similar way by taking $D = H - P_0$, bearing in mind that $(H - P_0)^r = H^r - P_0$ and that $n^{-1}\mathrm{tr}(H^r) \sim (n\lambda^{1/(2p)})^{-1}t_r$.

c). Equation (21) gives the expressions of S_{CD} and S_N statistics as linear combinations of S_{ES} and S_C since

$$S = [2\mathrm{tr}(D^2)]^{-1/2}[\sigma^{-2}Y^tDY - \mathrm{tr}(D)] = [2\mathrm{tr}(D^2)]^{-1/2}\mathrm{tr}(D)\{[Y^tDY/\mathrm{tr}(D)]/\sigma^2 - 1\},$$

and thus $S = k_1(H)S_{ES} + k_2(H)S_C$, where $k_i(H) = [2\mathrm{tr}(D^2)]^{-1/2}\mathrm{tr}(D)c_i(H)$.

Consider first $D = H - P_0$, that is $S = S_N$. In this case, $k_1(H) = [\mathrm{tr}(H - P_0)^4/\mathrm{tr}(D^2)]^{1/2} \sim k_1 = [\int(K^{*2})^2/\int K^2]^{1/2}$ and $k_2(H) = [\mathrm{tr}(H - H^2)^2/\mathrm{tr}(D^2)]^{1/2} \sim k_2 = [\int(K - K^{*2})^2/\int K^2]^{1/2}$, by Lemma 2.

c.1). When H_0 is true, $S_N = S_N^0 = S_{ES}^0 k_1(H) + S_C^0 k_2(H)$. Thus if H_0 and Condition 1 hold, convergence to a standard normal is a straightforward consequence of Theorem 2.1 in Jayasuriya (1996) and Theorem 1 in Chen (1994) since these show, respectively, that S_{ES}^0 and S_C^0 are asymptotically $N(0, 1)$.

If H_0 and Condition 2 hold, asymptotic normality can be deduced from the Lemma in Eubank and Spiegelman (1990). Certainly, if $\lambda_1 \le \cdots \le \lambda_n$ are the eigenvalues of $D = H - P_0$, it holds that $0 \le \beta^2 = (\max_i \lambda_i^2/\sum_{k=1}^n \lambda_k^2)^2 \le \eta = [n^{-2}\mathrm{tr}(D^4)]/[n^{-1}\mathrm{tr}(D^2)]^2 \sim a\lambda^{1/(2p)}$, where $a = t_4/t_2^2$. Consequently $\beta = \max_i \lambda_i^2/\sum_{k=1}^n \lambda_k^2 \to 0$.

c.2). Taking into account that $c(n) = n^{-1/2}(\lambda^{1/(2p)})^{-1/4}$ along with the assumptions about g, it holds that $S_{ES}^1 k_1(H) = [2\sigma^4\lambda^{1/(2p)}\mathrm{tr}(D^2)]^{-1/2}[n^{-1}g^t(H^2 -$

$P_0)g] \sim [2\sigma^4 c(H)]^{-1/2}\ \|g\|^2$, and that $S_C^1 k_2(H) = [2\sigma^4 \lambda^{1/(2p)} \mathrm{tr}(D^2)]^{-1/2}[n^{-1}g^t(H - H^2)g] \sim \lambda[2\sigma^4 c(H)]^{-1/2}\|g_\lambda^{(p)}\|^2$, where $c(H) = \lim_{n\to\infty} \lambda^{1/(2p)}\mathrm{tr}(D^2) = \int K^2$. Here g_λ denotes the spline function of order $2p$ corresponding to $\{x_i, g(x_i)\}_{i=1}^n$. Thus as $\lambda \to 0$, it follows that $S_N^1 = S_{ES}^1 k_1(H) + S_C^1 k_2(H) = [2\sigma^4 \lambda^{1/(2p)} \mathrm{tr}(D^2)]^{-1/2}$ $[n^{-1}g^t Dg] \sim [2\sigma^4 c(H)]^{-1/2}\|g\|^2$. Similar arguments allow us to conclude that $S_N^2 = S_{ES}^2 k_1(H) + S_C^2 k_2(H)$ is negligible, since both $S_{ES}^2 k_1(H)$ and $S_C^2 k_2(H)$ are negligible.

Consider now $D = 2H - H^2 - P_0$, that is, $S = S_{CD}$. In this case, $k_1(H) = [\mathrm{tr}(H - P_0)^4/\mathrm{tr}(D^2)]^{1/2} \sim k_1 = [\int (K^{*2})^2 / \int (2K - K^{*2})^2]^{1/2}$ and $k_2(H) = 2[\mathrm{tr}(H - H^2)^2/\mathrm{tr}(D^2)]^{1/2} \sim k_2 = 2[\int (K - K^{*2})^2 / \int (2K - K^{*2})^2]^{1/2}$. The rest of the proof follows from the same arguments as those given for $D = H - P_0$.

References

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55**, 549-557.

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions.* Robert E. Krieger Publishing Company, Florida.

Buckley, M. J. and Eagleson, G. K. (1988). An approximation to the distribution of quadratic forms in normal random variables. *Austral J. Statist.* **30**A, 150-159.

Buckley, M. J., Eagleson, G. K. and Silverman, B. W. (1988). The estimation of the residual variance in nonparametric regression. *Biometrika* **75**, 189-199.

Chen, J. C. (1994). Testing goodness of fit of polynomial models via spline smoothing techniques. *Statist. Probab. Lett.* **19**, 65-76.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83**, 596-610.

Cox, D. D. (1983). Asymptotics for *M*-type smoothing splines. *Ann. Statist.* **11**, 530-551.

Cox, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21**, 789-813.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31**, 377-403.

Demmler, A. and Reinsch, C. (1975). Oscillation matrices with spline functions. *Numer. Math.* **24**, 375-382.

Eubank, R. L. (1984). The hat matrix for smoothing splines. *Statist. Probab. Lett.* **2**, 9-14.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

Eubank, R. L. and Hart, J. D. (1993). Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* **80**, 89-98.

Eubank, R. L. and LaRiccia (1993). Testing for no effect in nonparametric regression. *J. Statist. Plann. Inference* **36**, 1-14.

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.

González-Manteiga, W. and Cao-Abad, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test* **2**, 161-188.

Green P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall, London.

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

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

Jayasuriya, B. R. (1996). Testing for polynomial regression using nonparametric regression techniques. *J. Amer. Statist. Assoc.* **91**, 1626-1631.

Messer, K. and Goldstein, L.(1993). A new class of kernels for nonparametric curve estimation. *Ann. Statist.* **21**, 179-195.

Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.* **23**, 1175-1197.

Ragozin, D. L. (1985). The discrete k-functional and spline smoothing of noisy data. *SIAM J. Numer. Anal.* **22**, 1243-1254.

Ramil Novo, L. A. and González Manteiga, W. (1998). χ^2 goodness of fit tests for polynomial regression. *Comm. Statist.-Simula.* **27**, 229-258.

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

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

Speckman, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. Technical Report, Dept. of Mathematics, University of Oregon, Eugene, OR, U.S.A.

Departamento de Estatística e Investigación Operativa, Universidade de Santiago de Compostela, Facultade de Matemáticas, 15706 Santiago de Compostela (A Coruña), Spain.

E-mail: wences@zmat.usc.es

Departamento de Estatística e Investigación Operativa, Universidade de Santiago de Compostela, Facultade de Ciencias, 27002 Lugo, Spain.

E-mail: lramil@lugo.usc.es

(Received February 1998; accepted September 1999)