The Annals of Statistics
1988, Vol. 16, No. 1, 113–119

TESTING THE (PARAMETRIC) NULL MODEL HYPOTHESIS IN (SEMIPARAMETRIC) PARTIAL AND GENERALIZED SPLINE MODELS[1]

BY DENNIS COX, EUNMEE KOH, GRACE WAHBA AND BRIAN S. YANDELL

University of Illinois, University of Wisconsin-Madison, University of Wisconsin-Madison and University of Wisconsin-Madison

Cox and Koh (1986) considered the model $y_i = f(x(i)) + \varepsilon_i$, ε_i i.i.d. $N(0, \sigma^2)$, with the (parametric) null hypothesis $f(x)$, $x \in [0,1]$, a polynomial of degree $m-1$ or less, versus the alternative f is "smooth," based on the Bayesian model for f which leads to polynomial smoothing spline estimates for f. They showed that there was no uniformly most powerful test and found the locally most powerful (LMP) test. We extend their result to the generalized smoothing spline models of Wahba (1985) and to the partial spline models proposed and studied by Engle, Granger, Rice and Weiss (1986), Shiller (1984), Green, Jennison and Seheult (1985), Wahba (1984), Heckman (1986) and others. We also show that the test statistic has an intimate relationship with the behavior of the generalized cross validation (GCV) function at $\lambda = \infty$. If the GCV function has a minimum at $\lambda = \infty$, then GCV has chosen the (parametric) model corresponding to the null hypothesis; we show that if the LMP test statistic is no larger than a certain multiple of the residual sum of squares after (parametric) regression, then the GCV function will have a (possibly local) minimum at $\lambda = \infty$.

1. Introduction. Cox and Koh (1986) considered the model

$$y_i = f(x(i)) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $x \in [0,1]$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \sim N(0, \sigma^2 I)$, σ^2 known, and $f(x)$, $x \in [0,1]$, is a Gaussian stochastic process independent of ε satisfying

$$f(x) \sim \sum_{\nu=1}^{m} \alpha_\nu \Phi_\nu(x) + b^{1/2} Z(x), \tag{1.2}$$

where $\Phi_1, \dots, \Phi_m$ span the polynomials of degree less than or equal to $m-1$ and Z is an $(m-1)$-fold integrated Weiner process [Shepp (1966)]. The parameter vector $\alpha = (\alpha_1, \dots, \alpha_m)'$ may be considered as a fixed vector of unknown parameters or as a random vector having the improper prior distribution $N(0, \xi I)$ with $\xi \to \infty$. This model is called the (special) spline model because the Bayes estimate of f is a polynomial spline. Cox and Koh were interested in the (parametric) null hypothesis H_0: $b = 0$ versus H_1: $b > 0$, equivalently, H_0: f is a polynomial of degree at most $m-1$ versus H_1: f is "smooth," i.e., (1.2) holds.

Received October 1986; revised March 1987.

[1]Research supported in part by NSF Grants MCS-82-02560, DMS-86-03083 and DMS-84-04970 and AFOSR Grant 87-0171.

AMS 1980 *subject classifications*. 62F03, 62G10, 62H15, 62C10.

Key words and phrases. Partial splines, hypothesis testing in spline models, parametric versus smooth models.

They showed that there is no uniformly most powerful (UMP) test, and they constructed the locally most powerful (LMP) test.

It is the purpose of this note to show that the Cox–Koh results extend easily to the univariate and multivariate partial spline models proposed by Engle, Granger, Rice and Weiss (1986), Shiller (1984), Green, Jennison and Seheult (1985), Wahba (1984), Heckman (1986) and others, and to the generalized spline models considered in Wahba (1985) and to note an intimate relationship between the LMP test statistic and the GCV estimate of $\lambda = \sigma^2/nb$.

2. Generalized and partial spline models. To make clear the relationship between the special spline model and the generalizations we will be interested in, we review a few facts. Let the set $\{x(1),\dots,x(n)\}$ contain at least m distinct points and let f_λ be the unique minimizer, in the Sobolev space $W_2^m\,[0,1]$, of

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x(i))\right)^2 + \lambda J_m(f). \tag{2.1}$$

Then it is known that

$$f_\lambda(x) = E(f(x)|y_1,\dots,y_n) \tag{2.2}$$

if $\lambda = \sigma^2/nb$. [See Kimeldorf and Wahba (1971) for α an unknown parameter and Wahba (1978) for α having an improper prior.]

The general smoothing spline model [see Wahba (1985) for more details] begins with a reproducing kernel Hilbert space H_Q of real valued functions of x for x in some domain I ($= E^d$, for example), an M-dimensional subspace of H_Q spanned by $\Phi_1,\dots,\Phi_M$ and $L_1,\dots,L_n$, n bounded linear functionals on H_Q. The model is

$$y_i = L_i f + \varepsilon_i, \qquad i = 1,\dots,n, \tag{2.3}$$

where the ε_i's are as before. Let T be the $n \times M$ matrix with (i,ν)th entry $L_i\Phi_\nu$. We shall always assume that T is of full column rank, that is, the least squares regression of y onto span$\{\Phi_1,\dots,\Phi_M\}$ is unique. Then the generalized spline estimate f_λ of f is the unique minimizer in H_Q of

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - L_i f\right)^2 + \lambda\|P_1 f\|_Q^2, \tag{2.4}$$

where P_1 is the orthogonal projection of f onto H_{Q_1}, the orthocomplement in H_Q of span$\{\Phi_1,\dots,\Phi_M\}$. Equation (2.2) holds here, with the Bayes model

$$f(x) \sim \sum_{\nu=1}^{M}\alpha_\nu\Phi_\nu(x) + b^{1/2}Z(x), \tag{2.5}$$

where now $Z(x)$, $x \in I$, is a family of zero mean Gaussian random variables with the (prior) covariance

$$EZ(x)Z(x') = Q_1(x,x'), \tag{2.6}$$

where Q_1 is the reproducing kernel for H_{Q_1}. This prior differs from the one considered by Cox and Koh for the setting of Section 1, but only on the space of polynomials of degree less than or equal to $m - 1$. The test derived in the next section will be the same, and the prior covariance (2.6) is more convenient.

A popular example is the thin plate spline case, where $I = E^d$, the $\Phi_1, \dots, \Phi_M$ are the $M = \binom{d+m-1}{d}$ polynomials of total degree less than m in d variables $(x_1, \dots, x_d)$ with $2m - d > 0$ and $\|P_1 f\|^2 = J_m^d(f)$ is given by

$$J_m^d(f) = \sum_{\alpha_1 + \cdots + \alpha_d = m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right)^2 dx_1 \cdots dx_d. \tag{2.7}$$

In the thin plate spline case, the hypothesis that $b = 0$ can be viewed as the hypothesis that f is a polynomial in d variables of total degree less than m versus the alternative that f is a fairly arbitrary "smooth" function.

The partial spline model (which is really a special case of the general spline model) allows a response that is the sum of a "smooth" function of $x = (x_1, \dots, x_d)$ and a parametric function of x and some other concomittant variables s, that is,

$$y_i = L_i g(\cdot\,; s_i) + \varepsilon_i, \tag{2.8}$$

where

$$g(x; s) = f(x) + \sum_{j=1}^{p} \beta_j \Psi_j(x; s). \tag{2.9}$$

Here f is as in (2.5) and the Ψ_j's are p linearly independent functions such that $L_i \Psi_j(\cdot, s_i)$ is well defined for each i, j. The vector $\beta = (\beta_1, \dots, \beta_p)'$ may be considered as a nuisance parameter or as $N(0, \xi I_p)$ with $\xi \to \infty$, similar to α. Let T be as before, let S be the $n \times p$ matrix with ijth entry $L_i \Psi_j \equiv L_i \Psi_j(\cdot\,; s_i)$ and suppose $X = (T : S)$ has rank $(M + p)$. Then an estimate g_λ of g is obtained by finding $f_\lambda \in H_Q$ and $\beta_\lambda \in E^p$ as the unique minimizer of

$$\frac{1}{n} \sum_{i=1}^{n} \left(y_i - L_i f - \sum_{j=1}^{p} \beta_j L_i \Psi_j \right)^2 + \lambda \|P_1 f\|_Q^2. \tag{2.10}$$

Then

$$g_\lambda(x; s) = f_\lambda(x) + \sum_{j=1}^{p} \beta_{j,\lambda} \Psi_j(x; s) = E(g(x; s) | y_1, \dots, y_n), \tag{2.11}$$

with $\lambda = \sigma^2/nb$. We have in fact just relabeled the domain in the general spline model and adjoined $\mathrm{span}\{\Psi_1, \dots, \Psi_p\}$ to H_Q and called the generic element of this enlarged space g. Now, the hypothesis that $b = 0$ is the "null model" hypothesis that g is of the parametric form

$$g(x; s) = \sum_{\nu=1}^{M} \alpha_\nu \Phi_\nu(x) + \sum_{j=1}^{p} \beta_j \Psi_j(x; s), \tag{2.12}$$

versus the alternative that g is a "smooth" function of f plus a parametric function of x and s of the form (2.9).

3. Results. We are now in a position to reduce the problem of testing that g is of the special parametric form (2.12) in the model (2.9) to the Cox–Koh setup. We always assume that the "parametric design matrix" X defined following (2.9) is of full column rank. Let Σ be the $n \times n$ matrix with ijth entry

$$EL_iZL_jZ = L_{i(x)}L_{j(x')}Q_1(x, x'), \tag{3.1}$$

where $L_{i(x)}$ means the linear functional applied to what follows considered as a function of x. Let $\theta = (\alpha' : \beta')'$. Then if we look at θ as a fixed unknown parameter, we have

$$y \sim N(X\theta, b\Sigma + \sigma^2 I) \tag{3.2}$$

and if we adopt the improper prior $\theta \sim N(0, \xi I)$ with $\xi \to \infty$, we have

$$y \sim N(0, \xi XX' + b\Sigma + \sigma^2 I). \tag{3.3}$$

Now, let R be any $n - (M + p) \times n$ matrix with $RR' = I$ and $RX = 0$, and let

$$u = Ry. \tag{3.4}$$

Then

$$u \sim N(0, bR\Sigma R' + \sigma^2 I) \tag{3.5}$$

for either model (3.2) or (3.3). Let $\Gamma D \Gamma'$ be the eigenvalue–eigenvector decomposition of $R\Sigma R'$, with λ_ν, $\nu = 1, \dots, n - (M + p)$, the diagonal elements of D. Let

$$\tilde{y} = \Gamma' u. \tag{3.6}$$

Then the $\tilde{y}_\nu$ are independent with

$$\tilde{y}_\nu \sim N(0, b\lambda_\nu + \sigma^2), \qquad \nu = 1, \dots, \tilde{n}, \tag{3.7}$$

where $\tilde{n} = n - (M + p)$.

THEOREM 1. *Let $y_1, \dots, y_n$ be given by* (2.8) *where g is given by* (2.9). *Consider the problem of testing $H_0 : g$ given by* (2.12) *versus $H_1 : g$ given by* (2.9) *with $b > 0$ in* (2.5). *Let $\lambda_1, \dots, \lambda_{\tilde{n}}$ be the preceding eigenvalues. Let Y denote the family of tests invariant under translations by vectors in span$\{(L_1\Phi_\nu, \dots, L_n\Phi_\nu)$: $1 \le \nu \le M \cup (L_1\Psi_j(\cdot; s_1), \dots, L_n\Psi_j(\cdot; s_n))$: $1 \le j \le p\}$.*

(a) *If there are at least two distinct eigenvalues $\lambda_i \ne \lambda_k$, then no UMP test exists in Y.*

(b) *There exists an LMP test in Y (at $b = 0$). It rejects when*

$$\tilde{T}(\tilde{y}) = \sum_{\nu=1}^{\tilde{n}} \lambda_\nu \tilde{y}_\nu^2 \tag{3.8}$$

is too large.

PROOF. Observe that $\tilde{y}$ is a maximal invariant under the group of translations, so tests in Y are functions of $\tilde{y}$. The rest of the proof follows as in Cox and Koh (1986) using the distributional results (3.7). □

Substituting (3.4) and (3.6) into (3.5) gives our main result.

THEOREM 2.

$$\tilde{T}(\tilde{y}) = T(y) = y'R'R\Sigma R'Ry \tag{3.9}$$

is the LMP test statistic for H_0: g has the parametric form (2.12) *versus the alternative, g has the* (*semiparametric*) *form* (2.9).

We remark that if $R\Sigma R'$ is not of full rank, then the $\tilde{y}_\nu$'s which correspond to $\lambda_\nu = 0$ do not appear in (3.8), nevertheless, the right-hand side of (3.9) may be used to compute $T(y)$.

We thank a referee for the following remark: Let $P = R'R\Sigma R'R$, then $T = y'Py$ and it can be shown that $(n\lambda)^2 J(f_\lambda) = (y - f_\lambda)'P(y - f_\lambda)$. Thus $(n\lambda)^2 J(f_\lambda)$ and $T(y)$ estimate the same thing, namely $\varepsilon' P \varepsilon$, under the null model, and are identical when $\lambda = \infty$. Another referee has observed that Example 4.16 of Cox and Hinkley (1974) has discussed the test of (3.8) in the context of components of variance.

4. A connection between the LMP test and the GCV estimate of λ. There is an interesting relationship between the test statistic $\tilde{T}(\tilde{y})$ of (3.8), and the GCV estimate $\hat{\lambda}$ of λ. If $\hat{\lambda}$ is infinity, then GCV has chosen the null model. $\hat{\lambda}$ is the minimizer of $V(\lambda)$ given [see Wahba (1985)] by

$$V(\lambda) = \frac{\sum_{\nu=1}^{\tilde{n}} (n\lambda/[n\lambda + \lambda_\nu])^2 \tilde{y}_\nu^2}{\left(\sum_{\nu=1}^{\tilde{n}} (n\lambda/[n\lambda + \lambda_\nu])\right)^2}. \tag{4.1}$$

THEOREM 3. *$V(\lambda)$ has a (possibly local) minimum at $\lambda = \infty$ whenever*

$$T(y) = \sum_{\nu=1}^{\tilde{n}} \lambda_\nu \tilde{y}_\nu^2 \le \frac{1}{\tilde{n}} \left(\sum_{\nu=1}^{\tilde{n}} \lambda_\nu \right) \left(\sum_{\nu=1}^{\tilde{n}} \tilde{y}_\nu^2 \right). \tag{4.2}$$

Note that $\sum_{\nu=1}^{\tilde{n}} \tilde{y}_\nu^2$ is the residual sum of squares after least squares regression on the (parametric) null model.

PROOF. Let $\gamma = 1/n\lambda$ and let

$$\tilde{V}(\gamma) = \frac{\sum_{\nu=1}^{\tilde{n}} (1/[1 + \gamma\lambda_\nu])^2 \tilde{y}_\nu^2}{\left(\sum_{\nu=1}^{\tilde{n}} (1/[1 + \gamma\lambda_\nu])\right)^2} = \frac{N(\gamma)}{D(\gamma)}, \quad \text{say.} \tag{4.3}$$

Then $V(\lambda)$ has a minimum at $\lambda = \infty$ if and only if $\tilde{V}(\gamma)$ has a minimum at $\gamma = 0$, and this occurs when $\tilde{V}'(0) \ge 0$, equivalently, when

$$D(0)N'(0) \ge N(0)D'(0), \tag{4.4}$$

or

$$(\tilde{n})^2 \left(-2 \sum_{\nu=1}^{\tilde{n}} \lambda_\nu \tilde{y}_\nu^2 \right) \ge \left(\sum_{\nu=1}^{\tilde{n}} \tilde{y}_\nu^2 \right) \left(-2\tilde{n} \sum_{\nu=1}^{\tilde{n}} \lambda_\nu \right). \qquad \square$$

Note that under the null hypothesis the right and left sides of (4.2) have the same expectation. However, since the distribution of T may not be very symmetric about its mean, we did an example simulation to estimate the probability that GCV will pick the null model when it is true. We selected the example of the cubic smoothing spline ($J_m(f) = \int_0^1 f^{(2)}(x)\,dx$) with the null model a straight line, and $x(i) = i/(n+1)$, with $n = 100$. We computed the $\tilde{n} = 98$ eigenvalues λ_ν using GCVPACK as described in Section 5. We computed 1000 independent samples of

$$t = \frac{\tilde{n}\left(\sum_{\nu=1}^{\tilde{n}} \lambda_\nu \tilde{y}_\nu^2\right)}{\left(\sum_{\nu=1}^{\tilde{n}} \lambda_\nu\right)\left(\sum_{\nu=1}^{\tilde{n}} \tilde{y}_\nu\right)}$$

with the $\tilde{y}_\nu$'s standard normal pseudorandom numbers from CMLIB. We found that 673 of the t's were less than 1. Hence, in this case there is about a 67% ($\pm 3.6\%$) chance that GCV will pick the null model when it is true, whereas the LMP test will pick the null model 95% of the time in such circumstances, assuming the usual 0.05 level of significance. Thus the test of hypothesis is more conservative in rejection of the null hypothesis, as one would expect. Nonetheless, we conjecture that in general the model selected by GCV will probably be not far from the null model when it is true.

We remark that, intuitively, if Q_1 behaves like a Green's function, then the $\tilde{y}_\nu$'s that correspond to large λ_ν's generally are measures of the "low frequency" components of y (perpendicular to the null model) whereas the $\tilde{y}_\nu$'s corresponding to small λ_ν's are measuring the "high frequency" components. We conjecture that roughly similar (approximate, asymptotic) results as these can be obtained for the penalized likelihood estimates with GCV of O'Sullivan (1983), O'Sullivan, Yandell and Raynor (1986), Green (1985) and O'Sullivan and Wahba (1985).

5. Some remarks concerning the computation of T. Let $F:G$ be the QR decomposition of X,

$$X = F:G = (F_1 : F_2)\begin{pmatrix} G_1 \\ 0 \end{pmatrix}, \tag{5.1}$$

where F is orthogonal and G_1 is lower triangular. Then R can be taken as F_2'. Let the Cholesky factorization of $F_2'\Sigma F_2$ be LL'. Then

$$T(y) = \|L'F_2'y\|^2. \tag{5.2}$$

Cox and Koh discuss some approximations to the distribution of $T(y)$, which depends on the nonzero values of $\lambda_\nu \sigma^2$, $\nu = 1, \dots, \tilde{n}$. The λ_ν's can be computed as the squares of the singular values of $L'F_2'$ by using the singular value decomposition in LINPACK [Dongarra, Bunch, Moler and Stewart (1979)]. The subroutine library GCVPACK [Bates, Lindstrom, Wahba and Yandell (1987)] can be used to compute f_λ, g_λ and the GCV estimate of λ and with slight modifications can return $T(y)$, both in the partial thin plate spline case with evaluation data and in general.

In the thin plate spline (TPS) case a reproducing kernel is known [see, for example, Wahba and Wendelberger (1980)], but it is much easier to work with the so-called "semikernel" $E_m(x, x')$ defined in that paper (also called a "variogram" in the kriging literature), which gives the covariances of generalized divided differences. It can be shown, using the reproducing kernel Q_1 in Wahba and Wendelberger (1980), that the matrix K with i, jth entry $L_{i(x)}L_{j(x')}E_m(x, x')$ satisfies $K = \Sigma + B$, where $F_2'BF_2 = 0$ so that K may be used instead of Σ in (3.9) and elsewhere.

REFERENCES

BATES, D. M., LINDSTROM, M. J., WAHBA, G. and YANDELL, B. (1987). GCVPACK-routines for generalized cross validation. *Comm. Statist. B—Simulation Comput.* **16** 263–297.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

COX, D. and KOH, E. (1986). A smoothing spline based test of model adequacy in polynomial regression. Technical Report 787, Dept. Statistics, Univ. Wisconsin-Madison.

DONGARRA, J., BUNCH, J., MOLER, C. and STEWART, G. (1979). *Linpack Users' Guide*. SIAM, Philadelphia.

ENGLE, R. F., GRANGER, C. W. J., RICE, J. and WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

GREEN, P. J. (1985). Penalized likelihood for general semi-parametric regression models. Technical Report 2819, Mathematics Research Center, Univ. Wisconsin-Madison.

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.

HECKMAN, N. E. (1986). Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B* **48** 244–248.

KIMELDORF, G. and WAHBA, G. (1971). Some results on tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95.

O'SULLIVAN, F. (1983). The analysis of some penalized likelihood estimation schemes. Technical Report 726, Dept. Statistics, Univ. Wisconsin-Madison.

O'SULLIVAN, F. and WAHBA, G. (1985). A cross validated Bayesian retrieval algorithm for non-linear remote sensing experiments. *J. Comput. Phys.* **59** 441–455.

O'SULLIVAN, F., YANDELL, B. and RAYNOR, W. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–103.

SHEPP, L. A. (1966). Radon–Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

SHILLER, R. J. (1984). Smoothness priors and nonlinear regression. *J. Amer. Statist. Assoc.* **79** 609–615.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

WAHBA, G. (1984). Cross validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H. A. David and H. T. David, eds.) 205–235. Iowa State Univ. Press, Ames, Ia.

WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.* **13** 1378–1402.

WAHBA, G. and WENDELBERGER, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Rev.* **108** 1122–1145.

DENNIS COX
DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801

EUNMEE KOH
GRACE WAHBA
BRIAN S. YANDELL
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706-1693