DATA SMOOTHING BY POLYNOMIAL-TRIGONOMETRIC REGRESSION

by

R. L. Eubank and Paul Speckman

December 1987

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

DATA SMOOTHING BY POLYNOMIAL–TRIGONOMETRIC REGRESSION

R. L. Eubank

Southern Methodist University

Paul Speckman

University of Missouri–Columbia

ABSTRACT

Let μ be a function defined on a finite interval [a,b], and suppose that $y_1,...,y_n$ are uncorrelated observations with $E(y_j) = \mu(t_j)$ and $var(y_j) = \sigma^2$, $j = 1,...,n$, where the t_j are fixed in [a,b]. We consider a method of estimating μ by regression on a combination of low order polynomial terms and trigonometric terms. Estimation based on trigonometric functions alone is known to suffer from bias problems at the boundaries due to the periodic nature of the fitted functions. We show that these boundary problems are alleviated by adding low order polynomial terms. The utility of the method is illustrated with examples, and asymptotic results show the estimators are competitive with other nonparametric procedures.

Keywords: Nonparametric regression; Trigonometric regression; Polynomial regression

1. INTRODUCTION

The traditional approach to fitting regression models of uncertain form is to regress on a suitable number of polynomial terms. Occasionally polynomial regression is replaced by regression on sine and cosine terms, and it has even been noted (see e.g., Graybill (1976)) that it can be useful to include a few polynomial functions when performing trigonometric regression analysis. In this paper we show that regression with both polynomial and trigonometric terms has certain practical and theoretical advantages over either method separately. Our results demonstrate that regression analysis using a combination of trigonometric functions and low order polynomials is a viable approach to estimating unknown regression curves and is competitive with a number of widely studied nonparametric regression techniques.

Consider the regression model with bivariate observations $(t_1, y_1),...,(t_n, y_n)$ satisfying

$$y_i = \mu(t_i) + \epsilon_i, i = 1,...,n, \tag{1.1}$$

where the ϵ_i are zero mean, uncorrelated random errors having common variance σ^2, and μ is an unknown regression function that we wish to estimate. The t_i are assumed to be design points falling in some finite interval [a,b]. There are many effective methods of estimating μ in (1.1) including kernel, nearest neighbor, spline and smoothing spline estimators. However, the methods invariably advocated in textbooks and undoubtedly most widely used in practice are regression on a polynomial or trigonometric function basis. There are several obvious reasons for this. First of all, least squares is simple to implement using available statistical software. In addition, statistics for inference, diagnostic analysis, model selection, etc. are readily available. In contrast, most other nonparametric regression methods generally require access to specialized code which may even require

modification to provide these capabilities.

Despite the popularity of trigonometric and polynomial regression estimators, the two methods have certain shortcomings which seem to be common knowledge. Polynomial regression is frequently subject to problems of multicollinearity. Although this difficulty can be avoided by using orthogonal polynomials, the simplicity of the method is then lost. Despite this problem, polynomial regression can be shown to attain the theoretical optimal rate of convergence for mean square error in certain settings (see e.g. Rafajlowicz (1987), Cox (1987)).

The problems associated with trigonometric regression (TR) estimators stem mainly from their behavior at the boundaries of [a,b]. Since trigonometric functions are periodic, the estimator will also share this property, regardless of whether or not the true regression curve is periodic. An unfortunate consequence of this fact is that a TR estimator of μ cannot be relied on to satisfactorily resolve the behavior of the regression curve near the boundaries. This characteristic is manifested in slower rates of convergence for trigonometric estimators than would be anticipated for competing estimators. For example, mean squared error convergence rates for trigonometric series estimators can be as slow as $n^{-1/2}$ (globally) or $n^{-2/3}$ (locally) for a twice differentiable, nonperiodic regression function (c.f. Hall (1981,1983) and Eubank, Hart, and Speckman (1987)), rather than the uniform optimal $n^{-4/5}$ rate attained by both kernel and smoothing spline estimators. In essence, the boundary behavior of a trigonometric series estimator dominates its mean squared error. When a data driven method based on a mean squared error estimate such as cross validation is used to choose the number of trigonometric functions in the regression, the result is often an estimator involving too many terms that undersmooths and exhibits anomalous wiggles.

4

In the next section we show that a simple solution to difficulties associated with TR estimators is to regress on a few low order polynomial terms as well as the trigonometric functions. This provides a natural boundary modification to the TR estimator that results in a method obtaining "proper" rates of convergence. Since the sine and cosine functions are orthogonal for equally spaced data, one can anticipate fewer collinearity difficulties than in polynomial regression. Two examples and a small simulation are also discussed which demonstrate the improvements that can be realized from the proposed method.

Similar results to those in Section 2 are given in Section 3 for estimators computed using either sine or cosine functions alone. We then summarize our findings in Section 4.

2. POLYNOMIAL–TRIGONOMETRIC REGRESSION

Assume that model (1.1) holds and without loss of generality set a=0 and b=2π. To estimate $\mu(t)$ we will use, for some integers d $\geq$ 0 and $\lambda \geq$ 0, the estimator

$$\mu_\lambda(t) = b_0 + \sum_{j=1}^{d} b_j t^j + \sum_{j=1}^{\lambda} (c_j \cos jt + s_j \sin jt), \qquad (2.1)$$

where the b_j, c_j and s_j are obtained by regression on the y data. More specifically, the coefficients in (2.1) are the minimizers of

$$\sum_{i=1}^{n} [y_i - B_0 - \sum_{j=1}^{d} B_j t_i^j - \sum_{j=1}^{\lambda} (C_j \cos jt_i + S_j \sin jt_i)]^2 \qquad (2.2)$$

with respect to the B_j, j=0,...,d, and (C_j, S_j), j=1,...,λ. When $\lambda = 0$ in (2.1), this is taken to mean that μ_λ is a polynomial estimator. We will hereafter refer to μ_λ as a polynomial trigonometric regression (PTR) estimator.

There are two parameters in (2.1), d and λ. The parameter d will be fixed so that a low order polynomial, say of degree two or three, is included in the regression, while λ will be allowed to vary. We will show that the polynomial terms give the PTR estimator satisfactory boundary behavior, and the parameter λ, which defines the number of sine and cosine terms, can be manipulated to give a desired amount of smoothing to the data. Smaller values of λ correspond to smoother estimators, and larger values will provide rougher estimates. In practice, one must select a suitable λ to construct an estimator. One approach is to use trial and error through examination of various fits. We will also discuss the use of automated selection methods.

To see the motivation for μ_λ, consider the case d = 2, and suppose that μ in (1.1) admits an absolutely continuous first derivative with a square integrable second derivative. This is the standard situation in which $O(n^{-4/5})$ is the best possible uniform mean squared error convergence rate. If $\mu(0) = \mu(2\pi)$ and $\mu'(0) = \mu'(2\pi)$, then a trigonometric regression estimator alone achieves the "correct" rate. However, μ typically does not satisfy such boundary conditions, and in this case the trigonometric regression with no boundary adjustment is far from optimal (see Eubank, Hart, and Speckman (1987)). However, suppose q(t) is a quadratic polynomial such that $q(2\pi) - q(0) = \mu(2\pi) - \mu(0)$ and $q'(2\pi) - q'(0) = \mu'(2\pi) - \mu'(0)$, and let

$$\mu(t) = \mu_0(t) + q(t).$$

Clearly μ_0 is twice differentiable with $\mu_0(0) - \mu_0(2\pi) = \mu_0'(0) - \mu_0'(2\pi) = 0$. Thus μ can be written as a quadratic polynomial plus a function μ_0 that can be well approximated by sine and cosine functions. The form of the proposed estimator is a direct consequence of this fact. Heuristically, the trigonometric functions are expected to model μ_0 while the polynomial part of the estimator will account for the boundary behavior of μ. We now establish that this is what actually occurs.

2.1 ASYMPTOTIC PROPERTIES

We are interested in the global behavior of μ_λ as an estimator of μ and therefore confine our study to the (total) mean squared error or risk,

$$R_n(\lambda) = n^{-1} \sum_{i=1}^{n} E(\mu(t_i) - \mu_\lambda(t_i))^2 \tag{2.3}$$

We assume that the t_i are distributed approximately as a sample from a distribution with cdf W and continuous positive density w on $[a,b]$. If $a \le t_1 < t_2 < ... < t_n \le b$, let W_n be the corresponding empirical cdf

$$W_n(t) = k/n \text{ for } t_k \le t < t_{k+1},$$

where $t_0 = a$ and $t_{n+1} = b$, and let δ_n be the Kolmogorov–Smirnov distance between W_n and W,

$$\delta_n = \sup_t |W(t) - W_n(t)|.$$

Concerning $R_n(\lambda)$ in (2.3) we are able to establish the following result.

Theorem 1. Assume that μ has $(d-1)$ absolutely continuous derivatives with $\mu^{(d)}$ square integrable. Then

$$R_n(\lambda) = O(\lambda^{-2d}) + \sigma^2(2\lambda + d + 1)/n + O(\delta_n \lambda^{-2d+1}). \tag{2.4}$$

In the applications we have in mind, either the t_i are the order statistics from a sample of size n, or they are generated by the relation

$$\int_a^{t_i} w(t)dt = i/n, \quad i = 1,...,n.$$

In the first case, $\delta_n = O(n^{-1/2}\log\log n)$ a.s. (see e.g. Csörgő and Révész (1981)), while $\delta_n = O(n^{-1})$ in the second. An immediate consequence is that we obtain the rate $O(n^{-2d/(2d+1)})$ for $R_n(\lambda)$ by taking λ proportional to $n^{1/(2d+1)}$ in either case. Stone (1982) and Speckman (1985) have shown that $n^{-2d/(2d+1)}$ is the best uniform rate for linear estimators over functions with the same smoothness properties as μ. This rate is to be compared to a rate of $n^{-1/2}$ for unmodified trigonometric series estimators (c.f. Eubank, Hart, and Speckman (1987)).

The proof of Theorem 1 is instructive, and a sketch of the main steps will be given here. Many of the details are relegated to the Appendix.

Proof. Let $\mu = (\mu(t_1),...,\mu(t_n))'$, $y = (y_1,...,y_n)'$, and let $S_{\lambda n}$ denote the hat matrix for regression of y on $1, t,...,t^d$ and $\{\cos jt, \sin jt; j=1,...,\lambda\}$. Then

$$R_n(\lambda) = B_n^2(\lambda) + \sigma^2(2\lambda + d + 1)/n$$

and

$$B_n^2(\lambda) = n^{-1}\|(I - S_{\lambda n})\mu\|_n^2,$$

where $\|\cdot\|_n^2$ denotes the Euclidean norm.

Now define $p(t)$ to be the polynomial of degree d uniquely determined by the conditions

$$\int_0^{2\pi} p(t)dt = 0,$$

$$p^{(j)}(2\pi) - p^{(j)}(0) = \mu^{(j)}(2\pi) - \mu^{(j)}(0), \quad j = 1,...,d,$$

and set

8

$$\mu_0(t) = \mu(t) - p(t).$$

Then μ_0 satisfies the periodic boundary conditions

$$\mu_0^{(j)}(0) = \mu_0^{(j)}(2\pi), \ j = 1,...,d, \tag{2.5}$$

and since $\mathbf{p} = (p(t_1),...,p(t_n))'$ is in the column span of $\mathbf{S}_{\lambda n}$, it follows that

$$B_n^2(\lambda) = n^{-1}\|(I - \mathbf{S}_{\lambda n})\mu_0\|_n^2,$$

where $\mu_0 = (\mu_0(t_1),...,\mu_0(t_n))'$. Further, if $\mathbf{T}_{\lambda n}$ is the hat matrix for regression on $\{1,$ $\cos jt, \sin jt, j=1,...,\lambda\}$, we have $\mathrm{span}(\mathbf{T}_{\lambda n}) \subset \mathrm{span}(\mathbf{S}_{\lambda n})$, hence

$$B_n^2(\lambda) \le n^{-1}\|(I - \mathbf{T}_{\lambda n})\mu_0\|_n^2.$$

Thus we need only establish that $n^{-1}\|(I - \mathbf{T}_{\lambda n})\mu_0\|_n^2$ is $O(\lambda^{-2d})$ to complete the proof.

As a result of Lemmas 3 and 1 in the Appendix,

$$n^{-1}\|(I - \mathbf{T}_n)\mu_0\|_n^2 \le \beta \, B^2(\lambda) + O(\delta_n \lambda^{-2d+1}),$$

where $\beta < \infty$ is a constant,

$$B^2(\lambda) = \int_0^{2\pi} [\mu_0(t) - (T_\lambda \mu_0)(t)]^2 dt,$$

and $T_\lambda \mu_0$ is the $L_2[0,1]$ projection of μ_0 onto the linear span of 1 and $\{\sin jt, \cos jt: j = 1,...,\lambda\}$.

It will be notationally convenient to formulate the remainder of the proof in terms

of complex exponentials rather than sines and cosines. In doing so we implicitly make use of the well known fact that $T_\lambda \mu_0$ is equivalently the projection of μ_0 onto $\{e^{ikt}; -\lambda \leq k \leq \lambda\}$ if μ_0 is real.

An arbitrary function μ in L_2 has the Fourier series representation

$$\mu(t) = \sum_{k=-\infty}^{\infty} a_k e^{ikt}, \qquad (2.6a)$$

where

$$a_k = a_k(\mu) = (2\pi)^{-1} \int_0^{2\pi} \mu(t) e^{-ikt} dt. \qquad (2.6b)$$

It follows that

$$(T_\lambda \mu)(t) = \sum_{k=-\lambda}^{\lambda} a_k e^{ikt}. \qquad (2.7)$$

If μ has d derivatives in L_2, repeated integration by parts on (2.6b) shows that

$$a_k(\mu) = (2\pi)^{-1} \sum_{j=0}^{d-1} [\mu^{(j)}(0) - \mu^{(j)}(2\pi)](ik)^{-j-1} + (ik)^{-d} a_k(\mu^{(d)}). \qquad (2.8)$$

In particular, by (2.5)

$$a_k(\mu_0) = (ik)^{-d} a_k(\mu_0^{(d)}).$$

Using (2.7) with this result and Parseval's equality,

$$B^2(\lambda) = \sum_{|k|>\lambda} |a_k(\mu_0)|^2$$
$$\leq \lambda^{-2d} \sum_{k=-\infty}^{\infty} |a_k(\mu_0^{(d)})|^2 \qquad (2.9)$$

$$= \lambda^{-2d} \int_0^{2\pi} (\mu_0^{(d)}(t))^2 dt,$$

and (2.4) is obtained. □

The proof of Theorem 1 makes it clear that the key to improving convergence of trigonometric regression estimators is to obtain the right boundary behavior for μ_0. This translates into faster convergence of the Fourier coefficients assuming μ is sufficiently smooth.

The conclusions of Theorem 1 are under the assumption that d is fixed and known, while in practice a value for d must be chosen. It is, of course, difficult to tell by visual inspection of data what boundary modifications are needed for a particular regression curve. However, one of the nice properties of this estimator is that decisions of this nature can be guided by examining the t–statistics for the polynomial coefficients. If a particular polynomial coefficient is not statistically significant, this provides an indication that the corresponding boundary adjustment is not necessary.

2.2 SELECTION OF λ

In this section we discuss the use of data driven methods for selecting λ, the number of sine and cosine terms to be included in the estimator. Attention will be focused on two methods for estimating the value of λ that minimizes the loss

$$L_n(\lambda) = n^{-1} \sum_{i=1}^{n} (\mu(t_i) - \mu_\lambda(t_i))^2. \tag{2.10}$$

A number of procedures are currently available for estimating the minimizer of (2.10); see, e.g., Rice (1984) and Li (1985,1987). These methods include cross–validation (or PRESS), generalized cross–validation (GCV), and unbiased risk estimation (or, equivalently, Mallow's C_p). The latter two techniques are particularly simple to utilize

with μ_λ.

Let

$$\text{RSS}(\lambda) = \sum_{i=1}^{n} (y_i - \mu_\lambda(t_i))^2.$$

Then the GCV and unbiased risk criteria for selecting λ are

$$\text{GCV}(\lambda) = n\text{RSS}(\lambda)/(n - 2\lambda - d - 1)^2 \tag{2.11}$$

and

$$R(\lambda) = n^{-1}\text{RSS}(\lambda) + \sigma^2(2\lambda + d + 1)/n, \tag{2.12}$$

respectively. Note that, apart from estimating σ^2 in (2.12), both criterion functions require only the additional computation of the residual sum of squares. This quantity will generally be available as output from any regression software package. The parameter σ^2 in (2.12) can be estimated using, for example, the error variance estimator proposed by Gasser, Sroka and Steinmetz (1986).

To estimate the minimizer of (2.10) one can use the minimizer of either (2.11) or (2.12). The optimality of this type of selection method is addressed in work by Li (1987). Using his results we can conclude that if the ϵ_i satisfy certain mild restrictions and if $\inf_\lambda nR_n(\lambda) \to \infty$, then

$$L_n(\hat{\lambda})/\inf_\lambda L_n(\lambda) \overset{p}{\to} 1,$$

where $\hat{\lambda}$ is a minimizer of either (2.11) or (2.12). We have been able to demonstrate this result under certain conditions on the Fourier coefficients of μ, but the proof is outside the scope of this paper.

2.3 EXAMPLES

In this section we illustrate a number of points from previous discussions through two examples and a small scale simulation. To begin, we will examine the voltage drop data from Montgomery and Peck (1982). These data consist of 41 readings on the battery voltage drop in a guided missile motor at equally spaced time points during its flight. The data are shown in Figure 1. The t variable represents time in seconds rescaled so that all points fall in the interval $[0, 2\pi]$.

A number of TR and quadratic (i.e. with $d = 2$) PTR estimators were fit to the voltage drop data. The corresponding values of the GCV and unbiased risk criteria are given in Table 1. The unbiased risk estimates were computed using the Gasser–Sroka–Steinmetz estimator of σ^2. Notice that both criteria seem to point to the use of 12 trigonometric functions (6 sines and 6 cosines) for TR while indicating that only 4 trigonometric functions (2 sines and 2 cosines) are needed for the PTR estimator. Actually, there is very little difference between the loss estimates for $\lambda = 1$ or 2 for the PTR estimator. Since the t–statistics were also very small for the coefficients of the last two sine and cosine terms, we elected to use $\lambda = 1$ rather than 2 in this case. The two resulting fits using TR and PTR are shown in Figure 1.

Notice from Figure 1 the marked difference in the boundary behavior of the PTR and TR estimators. The TR estimator also exhibits some anomalous wiggles near the peak in the data which are not reflected in the PTR estimator. These wiggles disappear if fewer terms are used in the TR estimator, thereby indicating that undersmoothing has occurred and that our selection criteria have pointed toward the use of too many terms. As noted in the introduction, this difficulty can be attributed to properties of the loss for a TR estimator of a nonperiodic regression curve.

Also shown in Figure 1 is a cross–validated cubic smoothing spline (SS) estimator

of the underlying regression function. Note the similarity between the SS and PTR estimators. This provides a further indication that the PTR estimator has adjusted correctly at the boundaries.

We were also interested in the performance of the PTR estimator relative to polynomial regression. We therefore fit polynomials of order one through eight to the data. GCV indicated that a quartic polynomial provided the best fit. The resulting estimator is shown in Figure 2 and seems to be quite comparable to the other fits to this data. Our primary reason for comparing these estimators, however, was to demonstrate the difference in the conditioning of the design matrices for the two cases.

A commonly used index of the conditioning of a matrix is the ratio of its smallest and largest singular values. We computed these indices for a number of comparable cases for the quadratic PTR and polynomial regression estimators. The ratios of these index numbers are given in Table 2. Thus, for example, the second row of Table 2 corresponds to the quartic polynomial estimator and the quadratic PTR estimator using the first sine and cosine functions. The ratio of index numbers for this case is 2.73 indicating that collinearity is roughly three times worse for polynomial regression than it is for PTR. The ratios grow large quite rapidly supporting our belief that collinearity has been significantly reduced through the use of PTR.

To further investigate the PTR estimator we conducted a small simulation patterned after the voltage drop example. For this purpose we used the model fit to the voltage drop data by Montgomery and Peck (1982). They found the data to be well modeled by a cubic spline with knots at 6.5 and 13 seconds; the knots correspond roughly to course changes in the missile. We therefore simulated from model (1.1) with n = 41 and μ taken to be the Montgomery and Peck cubic spline fit to the voltage drop data. The random errors were simulated from a normal distribution with variance $\sigma^2 = .07$. This value for σ^2 coincides with the estimate of the error variance for the Montgomery and Peck fit.

The basic simulation experiment was repeated a total of 100 times. For each of the 100 data sets, TR, quadratic PTR and cubic SS estimators were computed with optimized smoothing parameters for each estimator derived from GCV. To assess the performance of the three estimators we computed the squared error loss, as defined in (2.10), for each estimator over the 100 data sets. The results are summarized in Table 3. The PTR estimator has the smallest average loss followed closely by the SS estimator.

We analyzed the data from our simulation as a randomized block design with 100 blocks and three treatment levels corresponding to the three estimators used in the study. The Friedman test indicated a highly significant treatment effect. Follow up treatment comparisons using the Wilcoxon signed rank statistic revealed all three treatments to be significantly different. The largest P-value was .0046 for the comparison involving smoothing splines and PTR estimators.

While the improvement of the PTR and SS estimators over the TR estimator is no surprise, we were somewhat surprised at the PTR estimator's performance relative to smoothing splines. Subsequent simulations using other functions have given similar results, however, so this should not be regarded as a fluke. Our limited experience with the PTR estimator suggests that it performs similarly to a smoothing spline estimator provided the function does not change too rapidly. Sharp peaks in a regression function that are reflected by the data will tend to cause a ringing or Gibb's type phenomenon away from the peak that will not be present with smoothing splines. We should also note that all our investigations have been in the context of an equally spaced design. The variable bandwidth nature of smoothing splines (Silverman (1984)) may pay performance dividends for unequally spaced data.

Our second example extends the application of PTR curve fitting to an analysis of covariance problem. The data are from a Rothamsted mildew control experiment and are found in Draper and Guttman (1980). In the experiment, the effect on yield of four mildew control spray applications were tested: none, early spring, late spring and repeated. The

experiment was arranged as a single column of 38 plots in 9 blocks of 4 plots each with an extra plot on each end. When analyzed as a conventional block design, there is strong evidence of block effect. Using nonparametric techniques, Green, Jennison, and Seheult (1985) and Speckman (1986) analyzed these data assuming an additive model of the form

$$y_i = x_i'\beta + \mu(i) + \epsilon_i. \tag{2.13}$$

Here $x_i'\beta$ represents the treatment effect, and $\mu(t)$ for $t = i = 0,...,37$ is the plot effect.

The analysis of Green et. al. (1985) and their graphs strongly suggest that even blocks of only four plots each are too large to assume homogeneous error variances. Their analysis results in a decomposition of variability into treatment effects and a spatial effect corresponding to $\mu(t)$ for which the residuals are satisfactory. It is also clear from their analysis that no simple low order polynomial model (e.g. linear or quadratic) for spatial effect will be adequate.

We analyzed these data with the PTR approach of modeling $\mu(t)$ as a function of $1, t, t^2$, and $\{\cos jt, \sin jt: j = 1,...,8\}$. In the spirit of the original design, the two end plots were omitted from the analysis, and we took $t_i = 2\pi(i-1)/36$ to scale the plot position to the interval $[0,2\pi]$. The choice $\lambda = 8$ here was the minimizer of GCV on the complete model. With 14 degrees of freedom for error, MSE $= .0081$ and the test for treatment effect had F $= 71.0$ on 3 and 14 degrees of freedom, results close to those obtained by Green et. al. (1985) and Speckman (1986). In contrast, the standard analysis for a block design yielded MSE $= .0363$ on 24 degrees of freedom and F $= 28.8$ for treatment effect. Although both analyses show a significant treatment effect, the large reduction in MSE from the PTR fit could be important.

One notable feature of PTR is that conventional F–statistics are obtained. This is in contrast to other nonparametric methods such as in Green et. al. (1985) and Speckman (1986) which are not based on projections. The second feature is that this analysis is very

16

simple to perform using a statistical package such as SAS. Both of the other approaches cited above require specialized software for their implementation.

One could of course account for the additive effect of plot position by modeling $\mu(t)$ as a polynomial of unusually high degree. For these data, it appears that GCV chooses a polynomial of degree 14. Without constructing orthogonal polynomials, this approach can have difficulties. Using MINITAB as configured for an IBM 4381 computer, for example, we could not fit a polynomial of degree larger than eight. SAS apparently works with polynomials of degree up to 15, although there are unusual statistics reported (e.g. 0 degrees of freedom for some parameter estimates). In this situation, however, the trigonometric functions are orthogonal, so neither program has any problem with fitting models with even as many as 30 terms.

An analysis requiring such high order fits is admittedly unusual, and model (2.13) may be somewhat suspect if the fitted error term is too small. In our view, however, this kind of application can be worthwhile if only for its diagnostic value. In the context of field trials, one could try such a technique when for some reason the blocks failed to be homogeneous. Moreover, F—tests can be used to compare models under this approach.

3. OTHER ESTIMATORS

In this section we examine the asymptotic behavior of two alternatives to the PTR estimator based on either the sine or cosine functions alone. Throughout this section the t_i's will be assumed to lie in $[0,\pi]$.

The standard motivation for the TR estimator of Section 2 is the $L_2[0,2\pi]$ convergence of the Fourier series (2.5). Similar results could be obtained based on other series expansions for μ (see e.g. Rafajlowicz (1987)). If we consider functions $\mu \in L_2[0,\pi]$, either sines or cosines alone form an orthonormal basis, and the following expansions hold in L_2:

$$\mu(t) = \sum_{j=0}^{\infty} C_j \cos jt = \sum_{j=1}^{\infty} S_j \sin jt,$$

where

$$C_0 = C_0(\mu) = \pi^{-1} \int_0^{\pi} \mu(t)dt,$$

$$C_j = C_j(\mu) = (2/\pi) \int_0^{\pi} \mu(t) \cos jt\, dt, \ j \geq 1,$$

and

$$S_j = S_j(\mu) = (2/\pi) \int_0^{\pi} \mu(t) \sin jt\, dt, \ j \geq 1.$$

If μ is differentiable, integration by parts yields

$$C_j(\mu) = -j^{-1} S_j(\mu')$$

and

$$S_j(\mu) = j^{-1} \mathscr{B}_j(\mu) - j^{-1} C_j(\mu'),$$

$$(3.1)$$

with

$$\mathscr{B}_j(\mu) = (2/\pi)[\mu(0) - (-1)^j \mu(\pi)].$$

As seen in the proof of Theorem 1, rates of decay of the Fourier coefficients for μ are directly related to convergence rates for trigonometric estimators. Using (3.1) repeatedly, we see that if μ is sufficiently differentiable,

$$
\begin{aligned}
C_j(\mu) &= -j^{-2} \mathscr{B}_j(\mu') + j^{-2} C_j(\mu'') \\
&= -j^{-2} \mathscr{B}_j(\mu') - j^{-3} S_j(\mu''') \\
&= -j^{-2} \mathscr{B}_j(\mu') - j^{-4} \mathscr{B}_j(\mu''') + j^{-4} C_j(\mu^{(iv)})
\end{aligned}
$$

$$(3.2)$$

etc.

and

$$S_j(\mu) = j^{-1} \mathscr{B}_j(\mu) + j^{-2} S_j(\mu'') \tag{3.3}$$

$$= j^{-1} \mathscr{B}_j(\mu) + j^{-3} \mathscr{B}_j(\mu'') - j^{-3} C_j(\mu''')$$

etc.

Then by analyzing expressions similar to (2.9) and using (3.2)–(3.3), one can show that the sine series has $O(\lambda^{-1})$ bias if μ' exists but $\mathscr{B}_j(\mu) \neq 0$, and the cosine series has $O(\lambda^{-3})$ bias if μ'' exists but $\mathscr{B}_j(\mu') \neq 0$. This leads to convergence rates of only $O(n^{-1/2})$ and $O(n^{-3/4})$ in the two cases (c.f. Hall (1981,1983) and Eubank, Hart, and Speckman (1987)). However, if appropriate boundary modifications can be made so that the $\mathscr{B}_j$ terms vanish, higher rates of convergence are possible.

Our strategy in making boundary adjustments is the same here as in Section 2. We let p be a polynomial of degree d such that $\mu_0(t) = \mu(t) - p(t)$ satisfies suitable boundary conditions. In particular, for the sine series, let d be odd and choose $p(t)$ so that $\mu_0(0) = \mu_0(\pi) = \mu_0''(0) = \mu_0''(\pi) = ... = \mu_0^{(d-1)}(0) = \mu_0^{(d-1)}(\pi) = 0$. Then $\mathscr{B}_j(\mu_0) = \mathscr{B}_j(\mu_0'') = ... = \mathscr{B}_j(\mu_0^{(d-1)}) = 0$. With the cosine series, d is even and p is chosen to satisfy $\mu_0'(0) = \mu_0'(\pi) = ... = \mu_0^{(d-1)}(0) = \mu_0^{(d-1)}(\pi) = 0$ so that $\mathscr{B}_j(\mu_0') = ... = \mathscr{B}_j(\mu_0^{(d-1)}) = 0$. The proof of Theorem 1 can then be adapted to obtain the following result.

Theorem 2. Suppose the t_i are generated as in Theorem 1 for a positive density w on $[0,\pi]$, and assume that μ has (m−1) absolutely continuous derivatives with $\mu^{(m)} \in L_2[0,\pi]$.

(i) Let d be odd and define $\mu_{\lambda s}$ to be the estimator of μ obtained by regression on 1, t,..., t^d and sinjt, $j = 1,...,\lambda$. Then

$$R_{ns}(\lambda) = n^{-1} \sum_{i=1}^{n} E[\mu(t_i) - \mu_{\lambda s}(t_i)]^2$$
$$= O(\lambda^{-2m}) + \sigma^2(\lambda + d + 1)/n + O(\delta_n \lambda^{-2m+1})$$

for m = d or d + 1.

(ii) Let d be even and define $\mu_{\lambda c}$ to be the estimator of μ obtained by regression on 1, t,..., t^d and cosjt, $j = 1,...,\lambda$. Then

$$R_{nc}(\lambda) = n^{-1} \sum_{i=1}^{n} E[\mu(t_i) - \mu_{\lambda c}(t_i)]^2$$

$$= O(\lambda^{-2m}) + \sigma^2(\lambda + d + 1)/n + O(\delta_n \lambda^{-2m+1})$$

for $m = d$ or $d + 1$.

In practice, one would probably want to use either $\mu_{\lambda s}$ with a linear polynomial (d = 1) or $\mu_{\lambda c}$ with a quadratic polynomial (d = 2). As a result of Theorem 2, $\mu_{\lambda s}$ with d = 1 obtains the uniform optimal rates $O(n^{-2/3})$ and $O(n^{-4/5})$ for μ' or μ'' in L_2 by taking $\lambda = O(n^{1/3})$ or $\lambda = O(n^{1/5})$ respectively. Similarly, if μ'' or μ''' is in L_2, $\mu_{\lambda c}$ with d = 2 obtains the rates $O(n^{-4/5})$ and $O(n^{-6/7})$ by taking $\lambda = O(n^{1/5})$ and $\lambda = O(n^{1/7})$ respectively. Thus the sine series linear PTR might be more suitable for curves thought to be relatively "rough", and the cosine series quadratic PTR might be better for curves thought to be relatively "smooth".

We also included the cosine regression (CR) and quadratic polynomial–cosine regression (PCR) estimators in our simulation example of Section 2.3. The results for the loss are reported in Table 3. Wilcoxon signed rank tests indicate that CR is not significantly better than TR (P = .48) but is significantly worse than SS, PCR or PTR (the largest P–value being less than 10^{-4}). On the other hand, PCR is not found to be significantly different, at the .05 level, from either PTR or SS.

We have not as yet had sufficient experience with μ_λ, $\mu_{\lambda c}$ and $\mu_{\lambda s}$ to be able to recommend one over the other. For unmodified trigonometric series estimators, Eubank, Hart, and Speckman (1987) found that cosines had preferable asymptotics. However all three PTR type estimators achieve the same rates of convergence, so convergence rates do not provide a discriminating factor in this instance.

4. SUMMARY AND CONCLUSIONS

In this paper we have examined the properties of a simple alternative to trigonometric or polynomial regression obtained by essentially combining the two methods. The PTR procedure retains the implementational simplicity of the two methods while

appearing to modulate some of their associated difficulties. In particular, convergence rates have been derived for the estimator which show it performs comparably to other nonparametric methods such as kernel or smoothing spline estimators and outperforms what would be expected from trigonometric regression alone. Simulation results indicate these conclusions carry over to small samples.

The PTR estimator has a number of advantageous qualities deriving from its connection with ordinary linear regression methodology. Foremost among these is that it can be computed using standard statistical software without the requirement of special code. Consequently the estimator seems ideal for use in situations where a simple, easily computed smoother is required. One illustration of this was provided by the analysis of covariance example in Section 2.3. As another potential application area we mention additive nonparametric regression methods such as those discussed by Friedman and Stuetzle (1981), Breiman and Friedman (1985), and Stone (1985).

Another benefit of the linear regression nature of PTR is that the choice of the correct degree of smoothing is really a variable selection problem concerning the proper number of polynomial, sine and cosine terms to be employed. Thus such determinations can be made using any of a number of familiar methods (c.f. the discussion in Section 2.2). Tools such as t–statistics for the coefficient estimates can also be used to aid in this process.

Cleveland (1979) has noted the usefulness of robust methods for scatterplot smoothing. We note in passing that PTR can be easily adapted for this purpose by interfacing the estimator with any of a number of possible robust regression routines.

In summary, the PTR estimator provides a simple method of conducting nonparametric regression that in many cases should perform comparably to more sophisticated procedures. The method and its variants therefore appear to provide useful additions to the arsenal of nonparametric regression estimators.

APPENDIX

We use the following notation in proving Theorem 1. Let $0 < \alpha < \beta < \infty$ be constants such that $\alpha \leq w(t) \leq \beta$, $t \in [0,2\pi]$. Denote the ordinary $L_2[0,2\pi]$ and the weighted $L_2(w)$ norms by

$$\|f\|^2 = \int_0^{2\pi} f(t)^2 dt$$

and

$$\|f\|_w^2 = \int_0^{2\pi} f(t)^2 w(t) dt$$

respectively.

Lemma 1. Let V be an arbitrary subspace of $L_2[0,2\pi]$, and let P and P_w denote projection onto V with respect to $\|\cdot\|$ and $\|\cdot\|_w$ respectively. If $f \in L_2[0,2\pi]$, then

(i) $\|(I - P_w)f\|_w^2 \leq \beta \|(I - P)f\|^2$.

(ii) $\|(P - P_w)f\|^2 \leq \gamma^2 \|(I - P_w)f\|^2$, where $\gamma^2 = (\beta/\alpha) - 1$.

Proof. The first assertion follows immediately from the inequalities

$$\alpha \|(I - P)f\|^2 \leq \alpha \|(I - P_w)f\|^2 \leq \|(I - P_w)f\|_w^2 \leq \|(I - P)f\|_w^2 \leq \beta \|(I - P)f\|^2.$$

Moreover, these inequalities imply that

$$\|(I - P_w)f\|^2 \leq (\beta/\alpha) \|(I - P)f\|^2. \tag{A.1}$$

Because $P_w f \in V$,

$$\|(I - P_w)f\|^2 = \|(I - P)f\|^2 + \|(P - P_w)f\|^2,$$

and so (ii) follows from (A.1).

Lemma 2. Let $T_{\lambda w}$ and T_{λ} denote, respectively, the $L_2(w)$ and $L_2[0,2\pi]$ projection operators for the linear span of 1 and $\{\sin jt, \cos jt: j=1,\ldots,\lambda\}$. Then if f is absolutely continuous, $f(0) = f(2\pi)$, and $f' \in L_2$,

$$\|f' - (T_{\lambda w}f)'\| \le (1 + \gamma)\|f' - (T_{\lambda}f)'\|.$$

Proof. Let

$$(T_{\lambda w}f)(t) = \sum_{k=-\lambda}^{\lambda} b_{k\lambda}e^{ikt}$$

and

$$(T_{\lambda}f)(t) = \sum_{k=-\lambda}^{\lambda} a_k e^{ikt}.$$

We then have

$$\|(T_{\lambda}f)' - (T_{\lambda w}f)'\|^2 = \sum_{k=-\lambda}^{\lambda} k^2 |a_k - b_{k\lambda}|^2$$

$$\le \lambda^2 \sum_{k=\lambda}^{\lambda} |a_k - b_{k\lambda}|^2$$

$$= \lambda^2 \|(T_{\lambda} - T_{\lambda w})f\|^2$$

$$\le \lambda^2 \gamma^2 \|(I - T_{\lambda})f\|^2,$$

where the last inequality is by (ii) of Lemma 1. Under the assumptions on f,

$$\|f'\|^2 = \sum_{k=-\infty}^{\infty} k^2 |a_k|^2 < \infty.$$

Thus

$$\lambda^2 \|(I - T_\lambda)f\|^2 \le \sum_{|k| > \lambda} k^2 |a_k|^2 = \|f' - (T_\lambda f)'\|^2,$$

and an application of the triangle inequality completes the proof.

Lemma 3.

$$n^{-1}\|(I - T_{n\lambda})\mu_0\|^2 \le \int_0^{2\pi} [\mu_0(t) - (T_{\lambda w})\mu_0(t)]^2 w(t) + O(\delta_n \lambda^{-2d+1}).$$

Proof. To begin, note that

$$n^{-1}\|(I - T_{n\lambda})\mu_0\|^2 \le n^{-1} \sum_{i=1}^{n} [\mu_0(t_i) - (T_{\lambda w}\mu_0)(t_i)]^2$$

$$= \int_0^{2\pi} [\mu_0(t) - (T_{\lambda w}\mu_0)(t)]^2 dW_n(t). \tag{A.2}$$

For any function f with a continuous derivative, integration by parts (cf. Billingsley (1986), Theorem 18.4) and the Cauchy–Schwarz inequality give the bound

$$\left| \int_0^{2\pi} f^2(t)dW_n(t) - \int_0^{2\pi} f^2(t)dW(t) \right|$$

$$= \left| 2\int_0^{2\pi} f(t)f'(t)[W_n(t) - W(t)]dt \right|$$

$$\le 2\delta_n \int_0^{2\pi} |f(t)f'(t)|\, dt$$

$$\le 2\delta_n \|f\|\|f'\|.$$

To prove the lemma, we apply this inequality to (A.2) with $f(t) = \mu_0(t) - (T_{\lambda w}\mu_0)(t)$ and note that it suffices to show that $\|f\| = O(\lambda^{-d})$ and $\|f'\| = O(\lambda^{-d+1})$.

$$\|(I - T_{\lambda w})\mu_0\| = O(\lambda^{-d}).$$

An application of Lemma 2 yields

$$\|\tfrac{d}{dt}(I - T_{\lambda w})\mu_0\|^2 \leq (1 + \gamma)\|\tfrac{d}{dt}(I - T_\lambda)\mu_0\|^2$$

$$= (1 + \gamma) \sum_{|k| > \lambda} k^2 |a_k(\mu_0')|^2$$

$$= O(\lambda^{-2d+2}),$$

where the last bound is computed as in (2.9) again, and the proof is complete.

REFERENCES

Billingsley, P. (1986), "*Probability and Measure, (Second Edition)*, Wiley: New York.

Breiman, L. and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *J. Amer. Statistic Assoc. 78*, 131–136.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *J. Amer. Statist. Assoc. 74*, 829–836.

Cox, D. (1987), "Approximation of Linear Regression on Nested Subspaces," manuscript.

Csörgő, M. and Révész, P., (1981), *Strong Approximations in Probability and Statistics*, New York: Academic Press.

Draper and Guttman (1980), "Incorporating Overlap Effects From Neighboring Units Into Response Surface Models," *Appl. Statist.*, 29, 128–134.

Eubank, R.L., Hart, J.D., and Speckman, P. (1987), "Trigonometric Series Regression Estimators With an Application to Partially Linear Models," Manuscript.

Friedman, J. H. and Stuetzle, W. (1981), "Projection Pursuit Regression," *J. Amer. Statist. Assoc. 76*, 817–823.

Gasser, T., Sroka, L., and Jennen–Steinmetz, C. (1986), "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, 73, 625–633.

Graybill, F. (1976), *Theory and Application of the Linear Model*, Duxbury: North Scituate.

Green, P., Jennison, C., and Seheult, A. (1985), "Analysis of Field Experiments by Least Squares Smoothing", *J.R. Statist. Soc., Ser. B.*, 47, 299–315.

Hall, P. (1981), "On Trigonometric Series Estimates of Densities", *Ann. Statist.,9*, 683–685.

Hall, P. (1983), "Measuring the Efficiency of Trigonometric Series Estimates of a Density," *J. Multivariate Anal. 13*, 234–256.

Li, K. C. (1985), "From Stein's Unbiased Risk Estimates to the Method of Generalized Cross Validation," *Ann. Statist.*, 13, 1352–1377.

Li, K. C. (1987), "Asymptotic Optimality for C_P, C_L, cross–validation and generalized cross–validation: Discrete index set," *Ann. Statist.* 15, 958–975.

Montgomery, D.C. and Peck, E.A. (1982), *Introduction to Linear Regression Analysis*, New York: John Wiley.

Rafajlowicz, E. (1987), "Nonparametric Orthogonal Series Estimators of Regression: A Class Attaining the Optimal Convergence Rate in L_2," *Statist. and Probab. Letters* 5, 219–224.

Rice, J. (1984) "Bandwidth Choice for Nonparametric Regression," *Ann. Statist.*, 12, 1215–1230.

Silverman, B.W. (1984), "Spline Smoothing: The Equivalent Variable Kernel Method," *Ann. Statist.* 12, 898–916.

Speckman, P. (1985), "Spline Smoothing and Optimal Rates of Convergence in Nonparametric Regression Models," *Ann. Statist.*, 13, 970–983.

Speckman, P. (1986), "Kernel Smoothing in Partial Linear Models", Manuscript.

Stone, C. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression", *Ann. Statist.*, 10, 1040–1053.

Stone, C. (1985), "Additive regression and other nonparametric models," *Ann. Statist.* 13, 689–705.

Table 1. GCV and Unbiased Risk Estimates for TR and PTR Estimators

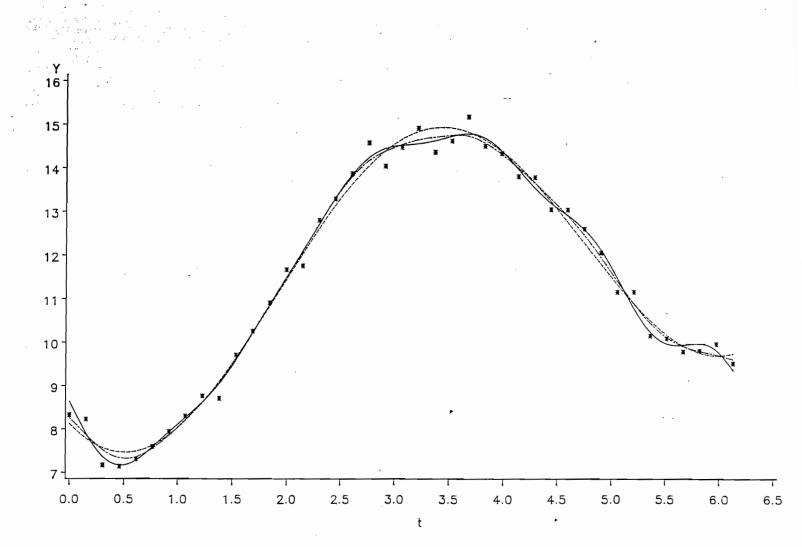| λ	TR		PTR	
	GCV	R	GCV	R
0	6.731	6.410	1.251	1.084
1	.235	.212	.086	.083
2	.149	.132	.084	.071
3	.139	.119	.092	.086
4	.122	.105	.010	.091
5	.103	.092	.102	.092
6	.090	.086	.102	.091
7	.097	.090	.101	.093
8	.114	.097	.117	.098
9	.127	.101	.138	.104
10	.154	.108	.158	.109

Table 2. Collinearity Comparison of PTR and Polynomial Regression

Number of Terms in the Estimator	Collinearity Index Ratio
3	1
5	2.73
7	8.06
9	11.32

Table 3. Summary Statistics for the Loss

Method	Mean	Standard Deviation
TR	.0325	.0095
SS	.0210	.0092
PTR	.0197	.0127
CR	.0322	.0116
PCR	.0204	.0129

FIGURE 1. TR, PTR, AND SS FITS TO THE
VOLTAGE DROP DATA

FIGURE 2. PTR AND POLYNOMIAL FITS TO
THE VOLTAGE DROP DATA