TRIGONOMETRIC SERIES REGRESSION ESTIMATORS WITH AN APPLICATION TO PARTIALLY LINEAR MODELS

by

R. L. Eubank, J. D. Hart, Paul Speckman

Technical Report No. SMU/DS/TR/210
Department of Statistical Science

December 1987

Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

Trigonometric Series Regression Estimators With an Application to Partially Linear Models

R. L. Eubank

Southern Methodist University

J. D. Hart

Texas A&M University

Paul Speckman

University of Missouri, Columbia

Abstract

Let μ be a function defined on an interval [a,b] of finite length. Suppose further that $y_1, \ldots, y_n$ are uncorrelated observations satisfying $E(y_j) = \mu(t_j)$ and $\mathrm{var}(y_j) = \sigma^2$, $j=1,\ldots,n$, where the t_j's are fixed design points. Asymptotic (as $n \to \infty$) approximations of the integrated mean squared error and the partial integrated mean squared error of trigonometric series type estimators of μ are obtained. Our integrated squared bias approximations closely parallel those of Hall (1981, 1983) in the setting of density estimation. Estimators that utilize only cosines are shown to be competitive with the so-called cut-and-normalized kernel estimators.

Our results for the cosine series estimator are applied to the problem of estimating the linear part of a partially linear model. An efficient estimator of the regression coefficient in this model is derived without undersmoothing the estimate of the nonparametric component. This differs from the result of Rice (1986) whose nonparametric estimator was a partial spline.

Keywords: Nonparametric regression; Fourier series; Rates of convergence

1. Introduction

There are currently a number of nonparametric regression estimators that have been studied extensively in the literature. Many of these, such as smoothing splines and kernel estimators, are closely related to trigonometric series estimators. It is thus surprising that asymptotic theory for the latter estimators is not as well developed as it is for other regression estimators. Apparently, the only published work on trigonometric series regression estimators is that of Rutkowski (1982), Greblicki and Pawlak (1985) and Rafajlowicz (1987). In contrast, series estimators have played a prominent role in the estimation of probability densities. (See, e.g., Kronmal and Tarter 1986 and Hall 1981, 1983.) In this paper we fill in one of the gaps in knowledge about the large sample behavior of trigonometric series regression estimators by giving characterizations of their asymptotic integrated mean squared error. We then apply these results to the problem of estimating the linear part of a partially linear model.

Assume that observations $y_1,\ldots,y_n$ are obtained following the nonparametric regression model

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

where the ϵ_i are zero mean uncorrelated errors with common variance σ^2, μ is an unknown regression function and the t_i are design points satisfying $a \le t_1 < \ldots < t_n \le b$ for finite constants a and b. The objective is to estimate μ assuming only that it satisfies certain smoothness conditions.

In many cases it is possible to represent μ in (1.1) as a Fourier series involving sine and/or cosine functions $\{\varphi_j\}_{j=1}^{\infty}$. More precisely, it is often possible to write $\mu \equiv \Sigma_{j=1}^{\infty}\beta_j\varphi_j$ for Fourier coefficients β_j. Thus, if estimators $\hat{\beta}_j$ can be derived for the β_j, the first λ terms in the series can be estimated to produce an estimator $\mu_\lambda = \Sigma_{j=1}^{\lambda}\hat{\beta}_j\varphi_j$ for μ. We will refer to estimators of this type as <u>trigonometric</u> <u>series</u> (TS) <u>estimators</u>. The large sample properties of TS estimators are of interest for a number of reasons apart from the previously noted connection between TS and other nonparametric estimators. For example, TS estimators are often used in practice by linear regression analysts when the true mean value function is unknown. Thus, it is important to obtain analytic results that allow us to see what is lost (or gained) by this practice over the use of other nonparametric estimators. It is also noteworthy that the smoothing parameter, λ, for TS estimators is integer valued. This is in contrast to kernel and smoothing spline estimators whose smoothing parameters take on a continuum of values. Since the level of smoothing must be chosen by the user through (possibly automated) examination of a number of values for the smoothing parameter, there are some definite computational advantages associated with the use of series type estimators. Asymptotic performance comparisons will therefore be useful in determining whether these computational savings can be realized without significant losses in estimation efficiency.

To assess the performance of an estimator μ_λ for μ, we use the integrated mean squared error (IMSE)

$$R(\mu_\lambda) = \int_a^b E(\mu(t)-\mu_\lambda(t))^2dt. \tag{1.2}$$

In the next section we study the asymptotic behavior of $R(\mu_\lambda)$ for TS estimators constructed from both sine and cosine functions or either sine or cosine functions alone. It is seen that, unless μ satisfies certain boundary conditions, neither the sine nor sine and cosine estimators perform up to the level of their competition. However, the estimator based on cosines alone is competitive with kernel estimators that have been cut and renormalized at the boundaries of [a,b].

The boundary behavior of TS estimators is a dominant factor in determining the large sample properties of $R(\mu_\lambda)$. Thus, the asymptotic behavior of the IMSE over [a,b] will generally not give an accurate picture of the estimator's performance in the interior of the interval. For this reason, we follow Hall (1983) and also analyze the partial integrated mean squared error (PIMSE)

$$R_\epsilon(\mu_\lambda) = \int_{a+\epsilon}^{b-\epsilon} E(\mu(t)-\mu_\lambda(t))^2 dt, \tag{1.3}$$

where ϵ is a positive constant satisfying $\epsilon < \frac{(b-a)}{2}$. When viewed through this performance window, TS estimators fare somewhat better than before in terms of asymptotic convergence rates. However, estimators based on sine or sine and cosine functions are still found to be deficient relative to kernel or smoothing spline estimators. In contrast, the TS estimator constructed only from the cosine functions can attain an $n^{-4/5}$ rate of decay for $R_\epsilon(\mu_\lambda)$. Thus, it seems to merit serious consideration as a competitor to second order kernel or cubic smoothing spline estimators which can also attain this same rate in the interior of [a,b].

Also of interest are the Cesàro means of TS estimators. We thus consider in Section 3 estimators of the form $\mu_\lambda^* = (\lambda+1)^{-1} \sum_{j=o}^{\lambda} \mu_j$, where μ_j is one of the TS estimators based on j terms. When the periodic extension of μ is sufficiently smooth, the Cesàro means are seen to have asymptotically larger IMSE and PIMSE than do the corresponding TS estimators.

In Section 4 we give an application of our work to the problem of parameter estimation in partially linear models. For a simple linear regression model with a covariate entering the model nonparametrically, we derive an estimator of the regression coefficient whose variance is of order n^{-1} and whose squared bias is $o(n^{-1})$. Thus, using our estimator, inference about the regression coefficient becomes feasible without the necessity of bias adjustments. This is in contrast to similar estimators derived from a smoothing spline viewpoint (see, e.g., Rice, 1986). Proofs of all results are provided in Section 5.

2. Trigonometric Series Estimators

Assume that model (1.1) holds with μ absolutely continuous on [a,b]. Also, for notational simplicity, assume that $[a,b] = [-\pi,\pi]$ or $[a,b] = [0,\pi]$. Let $s_o = a$, $s_j = (t_j+t_{j+1})/2$, $j=1,\ldots,n-1$, $s_n = b$ and define (for $j=0,1,\ldots$)

$$a_j = \sum_{r=1}^{n} y_r \int_{s_{r-1}}^{s_r} \cos jt \, dt$$

and

$$b_j = \sum_{r=1}^{n} y_r \int_{s_{r-1}}^{s_r} \sin jt \, dt.$$

Then, if $[a,b] = [-\pi,\pi]$ we can estimate μ by

$$\mu_{\lambda 1}(t) = (2\pi)^{-1}[a_o+2\sum_{j=1}^{\lambda}(a_j\cos jt + b_j\sin jt)]. \tag{2.2}$$

If $[a,b] = [0,\pi]$, two other possible estimators are

$$\mu_{\lambda 2}(t) = (2/\pi)\sum_{j=1}^{\lambda} b_j \sin jt \tag{2.3}$$

and

$$\mu_{\lambda 3}(t) = \pi^{-1}[a_o+2\sum_{j=1}^{\lambda} a_j\cos jt]. \tag{2.4}$$

The estimators $\mu_{\lambda i}$, i=1,2,3, are motivated by the fact that the sine and cosine functions form an orthonormal basis for $L_2[-\pi,\pi]$, whereas either the sine or cosine functions provide an orthonormal basis for $L_2[0,\pi]$. In defining the estimates of the Fourier coefficients in (2.1), we have used integrals of the trigonometric functions rather than evaluations at the t_j. This is similar to modifications to kernel estimators proposed by Gasser and Müller (1979) which improve estimator bias and automatically adjust the estimator to take account of nonuniform designs. Similar gains are realized here. (See Lemma 2 in Section 5 and comments later in this section.) Estimators of the form (2.2) - (2.4) have also been studied by Rutkowski (1982), who shows certain pointwise and global consistency properties of the estimators. Rafajlowicz (1986) and Greblicki and Pawlak (1985) obtain upper bounds for L_2 convergence rates for estimators similar to the ones we are considering in the case of a periodic regression function and random designs, respectively. A characterization of the asymptotic IMSE of the estimators (2.2) - (2.4) is provided by the following theorem. In the sequel, when h is a function defined only on [a,b], we say that h is continuous on [a,b] if it is continuous on (a,b), right continuous at a, and left continuous at b.

Theorem 1. Assume that the t_j are generated by a positive continuously differentiable density p on [a,b] through the relation

$$\int_a^{t_j} p(t)dt = (j-1)/n,\ j=1,\ldots,n. \tag{2.5}$$

Assume also that μ' is continuous on [a,b]. If $n,\lambda \to \infty$ in such a way that $\lambda^2/n = 0(1)$, the following results hold:

$$R(\mu_{\lambda 1}) = \sigma^2\lambda(n\pi)^{-1}\int_{-\pi}^{\pi} [p(t)]^{-1}dt + [\mu(\pi)-\mu(-\pi)]^2(\pi\lambda)^{-1} + o(\lambda/n+\lambda^{-1}) \tag{2.6}$$

$$R(\mu_{\lambda 2}) = \sigma^2\lambda(n\pi)^{-1}\int_0^{\pi}[p(t)]^{-1}dt + 2[\mu(\pi)^2+\mu(0)^2](\pi\lambda)^{-1} + o(\lambda/n+\lambda^{-1}). \tag{2.7}$$

If in addition μ'' is absolutely integrable,

$$R(\mu_{\lambda 3}) = \sigma^2\lambda(n\pi)^{-1}\int_0^{\pi} [p(t)]^{-1}dt + 2[\mu'(\pi)^2+\mu'(0)^2](3\pi\lambda^3)^{-1}+o(\lambda/n+\lambda^{-3}). \tag{2.8}$$

Under the conditions of Theorem 1, we see that the best rate of convergence for $R(\mu_{\lambda i})$, i = 1,2, is $n^{-1/2}$. This is obtained by taking $\lambda = c_o n^{1/2}$. The cosine estimator $\mu_{\lambda 3}$ performs considerably better with $R(\mu_{\lambda 3}) = 0(n^{-3/4})$ when $\lambda = c_1 n^{1/4}$. Thus, from the standpoint of IMSE, the cosine series estimator is to be preferred over either $\mu_{\lambda 1}$ or $\mu_{\lambda 2}$. It is worth mentioning that the best rate of convergence for the IMSE of a kernel estimator (of order two) which has been renormalized at the boundary (so that the observation weights sum to one) is also $n^{-3/4}$; so, $\mu_{\lambda 3}$ is comparable to a kernel estimator of μ in this sense. Of course it is possible to utilize boundary kernels (see, e.g., Gasser and Müller, 1979) to obtain the better rate of $n^{-4/5}$ for the IMSE of a kernel estimator. Similar modifications are undoubtedly possible for $\mu_{\lambda 3}$, although we will not pursue that topic here.

If μ satisfies certain boundary conditions, improved rates of convergence result for the three TS estimators. For example, if for some $r \geq 1$ $\mu^{(j)}(-\pi) = \mu^{(j)}(\pi)$, $j=0,1,\ldots,r-1$, the second term in (2.6) can be replaced by

$$[\mu^{(r)}(\pi) - \mu^{(r)}(-\pi)]^2[(2r+1)\pi\lambda^{2r+1}]^{-1} \tag{2.9}$$

and the remainder term becomes $o(\lambda/n) + o(\lambda^{-(2r+1)})$. Thus, if the periodic extension of μ is continuous (i.e., $\mu(\pi) = \mu(-\pi)$), the IMSE for $\mu_{\lambda 1}$ can be made to decay at a rate of $n^{-3/4}$ by choosing $\lambda = c_2 n^{1/4}$. Similar results hold for $\mu_{\lambda 2}$ and $\mu_{\lambda 3}$. Unfortunately, most regression functions will not be smoothly periodic; so one cannot routinely expect such improved performance in practice.

It may at first seem surprising that the cosine series estimator performs better than its counterparts $\mu_{\lambda 1}$ and $\mu_{\lambda 2}$. However, this phenomenon has a simple explanation. The cosine series expansion of μ with support on $[0,\pi]$ is identical to the Fourier series (i.e., sine and cosine) expansion of a function μ^* on $[-\pi, \pi]$ obtained by reflecting μ about zero. Thus $\mu^*(-\pi) = \mu^*(\pi)$ and we can use (2.6) along with (2.9) and r=1 to obtain (2.8).

While $\mu_{\lambda 3}$ appears to be the preferred estimator for general μ, there are cases where the use of $\mu_{\lambda 2}$ is advisable. To see when this occurs, observe that the sine series $2\pi^{-1}\sum_{j=1}^{\infty}\beta_j \sin jt$ of the function μ (defined on $[0,\pi]$) is the Fourier series of the odd function

$$\mu_o(t) = \operatorname{sgn}(t)\mu(|t|), \ -\pi \le t \le \pi$$

(where $\operatorname{sgn}(0) = 1$). Now, suppose that $\mu'(0+)$ and $\mu'(\pi-)$ exist and that $\mu(0) = \mu(\pi) = 0$. Then μ_o is differentiable at 0 and satisfies $\mu_o(-\pi) = \mu_o(\pi)$ and $\mu_o'(-\pi+) = \mu_o'(\pi-)$. Generally speaking, then, $\mu_{\lambda 2}$ is preferable to $\mu_{\lambda 3}$ and (the appropriate version of) $\mu_{\lambda 1}$ when $\mu(0) = \mu(\pi) = 0$ and $\mu'(0+) \neq \mu'(\pi-)$. Under these conditions, the integrated squared bias for $\mu_{\lambda 1}$ and $\mu_{\lambda 3}$ is no smaller than $c\lambda^{-3}$, whereas for $\mu_{\lambda 2}$ it can be as small as $O(\lambda^{-5})$ (see Hall, 1981). Hence, if one knows that μ vanishes at 0 and π but has no other information about the function, then $\mu_{\lambda 2}$ appears to be the right choice among the $\mu_{\lambda i}$.

The slow rates of convergence noted for the $R(\mu_{\lambda i})$, i=1,2,3, are primarily due to the boundary behavior of the estimators. To see this, we observe, for example, that if μ' is absolutely integrable, then for any fixed $t \in (-\pi,\pi)$ the bias of $\mu_{\lambda 1}(t)$ is $O(\lambda^{-1})$ and its variance is $O(\lambda/n)$ (see Hall, 1981 and Lemmas 2 and 3 of Section 5). Thus, by taking $\lambda = c_3 n^{1/3}$, $E(\mu(t) - \mu_{\lambda 1}(t))^2$ can be made to decay at a rate of $n^{-2/3}$ rather than the $n^{-1/2}$ obtained from Theorem 1. The conclusion to be drawn from this is that IMSE does not give an accurate picture of how TS estimators perform over the majority of $[a,b]$. A more appropriate measure for this purpose is the PIMSE defined in (1.3). The next theorem provides a summary of the asymptotic PIMSE behavior of TS estimators.

Theorem 2. Assume that the t_j are as defined in Theorem 1 and that n, $\lambda \to \infty$ in such a way that $\lambda^2/n = O(1)$. If μ' is continuous on [a,b], then for any $0 < \epsilon < \pi$

$$R_\epsilon(\mu_{\lambda 1}) = \sigma^2\lambda(n\pi)^{-1}\int_{-\pi+\epsilon}^{\pi-\epsilon}[p(t)]^{-1}dt + [\mu(\pi)-\mu(-\pi)]^2[2(\pi\lambda)^2]^{-1}\int_{\epsilon}^{\pi}(1-\cos t)^{-1}dt$$
$$+ o(\lambda/n+\lambda^{-2}), \qquad (2.10)$$

while for any $0 < \epsilon < \pi/2$

$$R_\epsilon(\mu_{\lambda 2}) = \sigma^2\lambda(n\pi)^{-1}\int_{\epsilon}^{\pi-\epsilon}[p(t)]^{-1}dt + [\mu(\pi)^2+\mu(0)^2](\pi\lambda)^{-2}\int_{\epsilon}^{\pi-\epsilon}(1-\cos t)^{-1}dt$$
$$+ o(\lambda/n+\lambda^{-2}). \qquad (2.11)$$

If, in addition, μ'' is of bounded variation

$$R_\epsilon(\mu_{\lambda 3}) = \sigma^2\lambda(n\pi)^{-1}\int_{\epsilon}^{\pi-\epsilon}[p(t)]^{-1}dt + [\mu'(\pi)^2+\mu'(0)^2]\pi^{-2}\lambda^{-4}\int_{\epsilon}^{\pi-\epsilon}(1-\cos t)^{-1}dt$$
$$+ o(\lambda/n+\lambda^{-4}). \qquad (2.12)$$

By choosing $\lambda = c_4 n^{1/3}$ we see that PIMSE of both $\mu_{\lambda 1}$ and $\mu_{\lambda 2}$ can be made to decay at a rate of $n^{-2/3}$. If we take $\lambda = c_5 n^{1/5}$ then $R_\epsilon(\mu_{\lambda 3}) = O(n^{-4/5})$. This is the same type of behavior one would expect from a kernel estimator (of order two) or a cubic smoothing spline. Thus, the cosine series estimator compares favorably to other popular nonparametric estimators in the interior of the interval of estimation.

The estimators given in (2.2) - (2.4) are derived from a particularly simple method of estimating the Fourier coefficients for μ. However, these estimators are not the standard choice. The usual approach to estimating Fourier coefficients is through least squares. Thus, to conclude this section we point out some implications of our results for Fourier series estimation by least squares.

For simplicity, let us restrict attention to an estimator of μ based on both sines and cosines. A common approach in this case is to estimate μ by $m_\lambda = \hat{\alpha}_o + \Sigma_{j=1}^{\lambda}[\hat{\alpha}_j \cos jt + \hat{\beta}_j \sin jt]$, where the $\hat{\alpha}_j$ and $\hat{\beta}_j$ are obtained by minimizing

$$\sum_{r=1}^{n}(y_r - \alpha_o - \sum_{j=1}^{\lambda}[\alpha_j \cos jt + \beta_j \sin jt])^2$$

with respect to the α_j and β_j. Note that the construction of such estimators will generally require matrix inversion, thereby indicating that $\mu_{\lambda 1}$ is more computationally expedient.

Define an alternative norm for the set of all square integrable functions on [a,b] by

$$||f||_p^2 = \int_a^b f(t)^2 p(t)dt,$$

where $f \in L_2[a,b]$ and p is our design density. The estimator m_λ provides an estimate of the projection $P_\lambda \mu$ of μ onto the linear span of the functions 1, sinjt, cosjt, ..., $\sin\lambda t$, $\cos\lambda t$ under $||\cdot||_p$. We can use this fact to obtain rates of convergence for m_λ, modulo quadrature error.

Let $(\tilde{P}_\lambda \mu)$ denote the $L_2[0,1]$ projection of μ onto the sine and cosine functions of period at most λ. Since p is positive and bounded there are positive constants B and C such that

$$B\int_a^b (\mu(t) - (\tilde{P}_\lambda \mu)(t))^2 dt \le \int_a^b (\mu(t) - (P_\lambda \mu)(t))^2 p(t)dt \le C\int_a^b (\mu(t) - (\tilde{P}_\lambda \mu)(t))^2 dt.$$

It follows from results in Section 5 that the integrated squared bias for $\mu_{\lambda 1}$ is well approximated by $\int_a^b (\mu(t) - (\tilde{P}_\lambda \mu)(t))^2 dt$. Thus, if we can assume that

$$\int_a^b (\mu(t) - Em_\lambda(t))^2 p(t)dt \sim \int_a^b (\mu(t) - (P_\lambda \mu)(t))^2 p(t)dt$$

then, as a result of Theorem 1,

$$\int_a^b E(\mu(t) - m_\lambda(t))^2 p(t)dt \sim \int_a^b (\mu(t)-(P_\lambda \mu)(t))^2 p(t)dt + \sigma^2(2\lambda+1)/n$$

$$= O(\lambda^{-1}) + O(\lambda/n).$$

Consequently, one would anticipate the same type of convergence rate for m_λ and for $\mu_{\lambda 1}$. Similar inferences can be made concerning rates of convergence for the least squares parallels of $\mu_{\lambda 2}$ and $\mu_{\lambda 3}$.

3. Cesàro Means of TS Estimators

We now consider the Cesàro means of the estimators $\mu_{\lambda 1}$ and $\mu_{\lambda 3}$. (For the sake of simplicity we do not examine $\mu_{\lambda 2}$; as indicated before this estimator is to be preferred only under a special set of conditions.) One reason for being interested in Cesàro means estimators is the well-known result that continuous periodic functions μ may be uniformly arbitrarily well approximated by the Cesàro means of Fourier series. In contrast, the Fourier series of a continuous function may diverge (see, e.g., Butzer and Nessel 1971). Furthermore,

$$\mu^*_{\lambda i} = (\lambda+1)^{-1}\sum_{j=0}^{\lambda} \mu_{\lambda i}, \; i=1,3,$$

will be positive whenever the data $y_1, \ldots, y_n$ are positive. This, of course, is a desirable property if μ is known to be positive. It will be seen, though, that if μ is sufficiently smooth, then the IMSE and PIMSE for the Cesàro means $\mu^*_{\lambda i}$ converge at a slower rate than they do for $\mu_{\lambda i}$. This must be balanced against any qualitative improvements obtained with $\mu^*_{\lambda i}$.

To establish the following IMSE and PIMSE results it is sufficient to use Lemmas 2 and 3 of Section 5 and the bias expansions of Hall (1981, 1983). Hence, the proofs of Theorems 3 and 4 will merely be outlined in Section 5.

<u>Theorem 3</u>. Let the design points t_j satisfy the conditions of Theorem 1 and suppose that $n,\lambda \to \infty$ in such a way that $\lambda^2/n = O(1)$. We can then establish the following results. If μ' is continuous on $[-\pi,\pi]$, then

$$R(\mu^*_{\lambda 1}) = \sigma^2\lambda(3\pi n)^{-1}\int_{-\pi}^{\pi} [p(t)]^{-1}dt + 2[\mu(\pi)-\mu(-\pi)]^2(\lambda\pi)^{-1} + o(\lambda/n+\lambda^{-1}). \tag{3.1}$$

If $\mu(\pi) = \mu(-\pi)$, and if μ' satisfies a uniform Lipschitz condition on $[-\pi,\pi]$ of order $\alpha > 1/2$, then

$$R(\mu^*_{\lambda 1}) = \sigma^2\lambda(3\pi n)^{-1}\int_{-\pi}^{\pi} [p(t)]^{-1}dt + \lambda^{-2}\int_{-\pi}^{\pi} |\mu'(t)|^2dt + o(\lambda/n+\lambda^{-2}). \tag{3.2}$$

If μ' is continuous on $[0,\pi]$ and μ'' is absolutely integrable on $(0,\pi)$, then

$$R(\mu^*_{\lambda 3}) = \sigma^2\lambda(3\pi n)^{-1}\int_0^{\pi} [p(t)]^{-1}dt + \lambda^{-2}\int_0^{\pi} |\mu'(t)|^2 dt + o(\lambda/n+\lambda^{-2}). \tag{3.3}$$

We see that the optimal value of $R(\mu^*_{\lambda 1})$ converges at the same rate, $n^{-1/2}$, as does that of $R(\mu_{\lambda 1})$ when $\mu(\pi) \neq \mu(-\pi)$. However, when $\mu(\pi) = \mu(-\pi)$, the integrated squared bias of $\mu^*_{\lambda 1}$ converges more slowly than does that of $\mu_{\lambda 1}$. Similarly, the integrated squared bias of $\mu^*_{\lambda 3}$ is larger than that of $\mu_{\lambda 3}$. The problem with the Cesàro means estimators is that, like kernel estimators, their bias does not continue to decrease as the smoothness of μ increases. Whenever $\sum_{j=\lambda+1}^{\infty} |A_j|^2 + |B_j|^2 = o(\lambda^{-2})$ (where the A_j's and B_j's are the Fourier coefficients of μ), the integrated squared bias of $\mu^*_{\lambda i}$ is dominated by

$$c\lambda^{-2}\sum_{j=1}^{\lambda} j^2[|A_j|^2 + |B_j|^2].$$

Hence, unlike $\mu_{\lambda i}$, $\mu^*_{\lambda i}$ cannot take advantage of the fact that $\sum_{j=\lambda+1}^{\infty} (A_j^2+B_j^2)$ may be small.

In the next theorem we give PIMSE results for $\mu^*_{\lambda 1}$ and $\mu^*_{\lambda 3}$. The quantity $\tilde{\mu}$ appearing below is the Hilbert transform of μ, i.e.,

$$\tilde{\mu}(t) = (2\pi)^{-1}\int_0^{\pi} [\mu(t-v)-\mu(t+v)]\cot(v/2)dv,$$

where μ is extended from $(-\pi,\pi]$ to $(-\infty,\infty)$ by periodicity. (If μ is defined on $[0,\pi]$, it is first extended to $(-\pi,\pi]$ by $\mu(-t) = \mu(t)$, $0 < t < \pi$.)

<u>Theorem 4</u>. Let the design points t_j satisfy the conditions of Theorem 1, and

suppose that μ' is continuous on [a,b]. If $n,\lambda \to \infty$ and $\lambda^2/n = O(1)$, the following results hold.

i) Suppose that $\tilde{\mu}$ and $(\tilde{\mu})'$ exist and are bounded on $[-\pi+\epsilon,\pi-\epsilon]$ for each $\epsilon > 0$. If

$$\sup_{|t|\leq\pi-\epsilon} \int_0^\pi v^{-2}[\tilde{\mu}(t+v)-\tilde{\mu}(t-v)-2v(\tilde{\mu})'(t)]dv < \infty$$

for each $\epsilon > 0$, then, for $0 < \epsilon < \pi$,

$$R_\epsilon(\mu^*_{\lambda 1}) = \sigma^2\lambda(3n\pi)^{-1}\int_{-\pi+\epsilon}^{\pi-\epsilon} [p(t)]^{-1}dt + \lambda^{-2}\int_{-\pi+\epsilon}^{\pi-\epsilon} |(\tilde{\mu})'(t)|^2 dt + o(\lambda/n+\lambda^{-2}), \quad (3.4)$$

provided the second of the two integrals does not vanish.

ii) Suppose that $|\mu''|$ is integrable on $(0,\pi)$ and that $(\tilde{\mu})'$ is continuous and of bounded variation on each interval $[\epsilon,\pi-\epsilon]$, $0 < \epsilon < \pi$. Then, for $0 < \epsilon < \pi$,

$$R_\epsilon(\mu^*_{\lambda 3}) = \sigma^2\lambda(3n\pi)^{-1}\int_{\epsilon}^{\pi-\epsilon} [p(t)]^{-1}dt + \lambda^{-2}\int_{\epsilon}^{\pi-\epsilon} |(\tilde{\mu})'(t)|^2 dt + o(\lambda/n+\lambda^{-2}). \quad (3.5)$$

As with IMSE, we see that PIMSE of Cesàro mean estimators will be asymptotically larger than that for the TS estimators when μ or its periodic extension are sufficiently smooth. To summarize the PIMSE results of Sections 2 and 3, we note that $\mu_{\lambda 3}$ is the only estimator considered whose PIMSE can attain the convergence rate $n^{-4/5}$ without imposing boundary conditions on μ. Other Fourier series estimates one could consider are the singular integral estimates of Hall (1983). He shows that, in density estimation, the PIMSE of these estimators can be of order $n^{-4/5}$ with no boundary conditions on μ.

4. An Application to Partially Linear Models

There has been much interest recently in semi-parametric statistical models. One variety of semi-parametric model is the partly linear model which contains both a linear parametric term and an additive nonparametric term involving one or more covariates. The interest is usually in obtaining efficient estimates of the linear parameters in the model. In this section we present an application of our work in Section 2 to the problem of estimating the regression coefficient in a simple partially linear model.

For simplicity, we confine attention to the case of only one linear predictor and one covariate. It will be assumed that

$$y_i = \beta x_i + f(t_i) + \epsilon_i, \quad i=1,\dots,n, \tag{4.1a}$$

where the ϵ_i are independent, zero mean random variables with common variance σ^2, f is some unknown function of the covariate t, and β is an unknown regression coefficient. The t_i satisfy $0 \le t_1 < \dots < t_n \le \pi$ and, following Rice (1986) and Speckman (1986), the x_i are assumed to admit a regression model in t. Specifically, we assume the x_i follow the model

$$x_i = g(t_i) + \eta_i, \tag{4.1b}$$

where g is an unknown function, the η_i are independent, zero mean, random variables with some common variance θ^2, and the ϵ_i's and η_i's are independent of each other.

For any set of numbers $z_1, \ldots, z_n$ define

$$z_{\lambda i} = a_o(z)/\pi + (2/\pi)\sum_{j=1}^{\lambda} a_j(z)\cos jt_i\,, \tag{4.2}$$

with $a_j(z) = \sum_{r=1}^{n} z_r \int_{s_{r-1}}^{s_r} \cos jt\,dt$. Then our proposed estimator of β is

$$\hat{\beta} = \sum_{i=1}^{n} (x_i - x_{\lambda i})(y_i - y_{\lambda i}) / \sum_{r=1}^{n} (x_r - x_{\lambda r})^2\,. \tag{4.3}$$

The motivation for this estimator stems from analysis of covariance. In that setting both y and x are adjusted for the covariate t and then residuals are regressed on residuals to estimate β. The definition of $\hat{\beta}$ in (4.3) is a similar type of adjustment.

Concerning $\hat{\beta}$ we are able to establish the following result.

<u>Theorem 5</u>. Let $\underline{x} = (x_1, \ldots, x_n)$ and assume that f and g both have continuous derivatives and second derivatives of bounded variation on $[0,\pi]$. Let

$$e(\lambda) = \lambda/n + 2(3\pi\lambda^3)^{-1}\max\{g'(\pi)^2+g'(0)^2,\ f'(\pi)^2+f'(0)^2\} + o(1)\,, \tag{4.4}$$

where $o(1) = o(\lambda^3/n^2) + o(\lambda/n) + o(\lambda^{-3})$, and assume that $t_j = (j-1)\pi/n$, $j=1,\ldots,n$. Then, if $\lambda, n \to \infty$ in such a way that $\lambda^2/n = 0(1)$,

$$\beta - E[\hat{\beta}|\underline{x}] = O_p(e(\lambda)). \tag{4.5}$$

If $\lambda^2/n \to 0$ and $\lambda^6/n \to \infty$,

$$\mathrm{Var}(\hat{\beta}|\underline{x}) = \sigma^2 n^{-1} \theta^{-2} + o_p(n^{-1}). \tag{4.6}$$

In addition, if $E|\eta_i|^{2+\delta}$ is uniformly bounded for some $\delta>0$, $\sqrt{n}\ (\hat{\beta}-\beta)$ converges in distribution to a $N(0,\sigma^2\theta^{-2})$ random variable.

If $\lambda^2/n \to 0$ and $\lambda^6/n \to \infty$, then, as a result of (2.8), $\sqrt{n}\ e(\lambda)=o(1)$. Thus, for n sufficiently large, the bias of $\hat{\beta}$ is negligible relative to its standard deviation. This has the consequence that inference for β can be conducted using $\hat{\beta}$ without the necessity of adjusting for the bias from the nonparametric part of the model. This is quite different from what transpires in the smoothing spline setting where the squared bias may dominate the mean squared error of the analogous estimator of β (see Rice, 1986). The fact that $\hat{\beta}$ is asymptotically normal with mean β has the implication that confidence intervals and tests for β can be conducted using standard parametric methods.

Theorem 5 can be easily extended to include estimation of more than one regression coefficient and nonuniform designs in t. Apparently results of this nature do not extend to estimators based on the sine or sine and cosine series without undersmoothing to ensure that $R(\mu_{\lambda j}) = o(1/\sqrt{n})$, j=1,2.

5. Proofs of Theorems

To prove the results in Sections 2 and 3 we require three lemmas. The proofs of Lemmas 1 and 2 are elementary and therefore omitted.

Lemma 1. Assume that the t_j are generated by a positive, continuous density on $[a,b]$ through relation (2.5). If $s_o=a$, $s_j=(t_j+t_{j+1})/2$, $j=1,\ldots,n-1$, and $s_n=b$, then $\max_j|s_j-s_{j-1}|=0(n^{-1})$.

Lemma 2. Assume that μ has a continuous derivative on $[a,b]$. Let

$$A_j = \int_a^b \mu(t)\cos jt\,dt,\ j=0,1,2,\ldots,$$

and

$$B_j = \int_a^b \mu(t)\sin jt\,dt,\ j=1,2,\ldots\ .$$

Then $A_j-E(a_j)$ and $B_j-E(b_j)$ are $0(n^{-1})$ uniformly in j.

Lemma 3. Consider a quantity of the form

$$C_\lambda(t) = \sum_{j=1}^{n} [\int_{s_{j-1}}^{s_j} K_\lambda(t-s)ds]^2,$$

where the s_j are defined as in Lemma 1 with p continuously differentiable, and where K_λ is a continuously differentiable function. Then,

$$C_\lambda(t) = n^{-1}\int_a^b [K_\lambda^2(t-s)/p(s)]ds + 0(n^{-2})\{\int_a^b |K_\lambda'(t-s)K_\lambda(t-s)|ds + \int_a^b K_\lambda^2(t-s)ds\}.$$

Proof. Using the mean value theorem for integrals and the uniform differentiability of p, we have, for points ξ_r, θ_r with $s_{r-1} \le \xi_r$, $\theta_r \le s_r$,

$$\int_{s_{r-1}}^{s_r} K_\lambda(t-s)ds = (s_r-s_{r-1})K_\lambda(t-\theta_r) = K_\lambda(t-\theta_r)/(np(\xi_r))$$

$$= [K_\lambda(t-\theta_r)/(np(\theta_r))](1+0(n^{-1})),$$

where the $O(n^{-1})$ term does not depend on r. Let P(s) be the cdf on [a,b] with density p(s), and define $P_n(s) = r/n$, $\theta_r \le s < \theta_{r+1}$ for $1 \le r < n$, $P_n(s) = 0$ for $s < \theta_1$, and $P_n(s) = 1$ for $s \ge \theta_n$. Using the previous equation we can then express C_λ as

$$C_\lambda(t) = n^{-1}\int_a^b [K_\lambda(t-s)/p(s)]^2\, dP_n(s)(1+O(n^{-1})).$$

It is easy to show that $\sup_{a\le s\le b}|P_n(s)-P(s)|\le 2/n$. Use integration by parts (cf. Billingsley (1986), Theorem 18.4) to obtain

$$\left|n^{-1}\int_a^b (K_\lambda(t-s)/p(s))^2 dP_n(s) - n^{-1}\int_a^b (K_\lambda(t-s)/p(s))^2 dP(s)\right|$$

$$\le n^{-1}\int_a^b |P_n(s)-P(s)|\ \left|\tfrac{\partial}{\partial s}(K_\lambda(t-s)/p(s))^2\right|ds$$

$$= O(n^{-2})\int_a^b |(K_\lambda'(t-s)p(s)+K_\lambda(t-s)p'(s))K_\lambda(t-s)/p(s)^3|ds,$$

and the lemma follows.

<u>Proof of Theorems 1 and 2</u>. We indicate only how to prove (2.6) and (2.10) as the proof of the other results follows a similar pattern. To obtain (2.6) we begin by noting that

$$E(\mu(t)-\mu_{\lambda 1}(t))^2 = \mathrm{Var}\ \mu_{\lambda 1}(t) + (\mu(t)-E\mu_{\lambda 1}(t))^2.$$

Now observe that $\mu_{\lambda 1}$ can be written as $\Sigma_{j=1}^n y_j \int_{s_{j-1}}^{s_j} K_\lambda(t-s)ds$ with K_λ the Dirichlet kernel, i.e., $K_\lambda(u) = (2\pi)^{-1}\Sigma_{|j|\le\lambda} e^{iju}$. Thus, an application of Lemma 3 gives

$$\mathrm{var}\ \mu_{\lambda 1}(t) = n^{-1}\int_{-\pi}^{\pi} [K_\lambda^2(t-s)/p(s)]ds + O(\tfrac{\lambda^2\log\lambda}{n^2}) + O(\lambda/n^2) \tag{5.1}$$

uniformly in t. In applying the lemma we have used the facts that $\max_{-\pi \le s \le \pi} |K_\lambda'(s)| = O(\lambda^2)$,

$$\int_{-\pi}^{\pi} K_\lambda^2(s)ds = 2\lambda+1$$

and

$$\int_{-\pi}^{\pi} |K_\lambda(s)|ds = O(\log \lambda).$$

The last bound is the Lebesgue constant (cf. Butzer and Nessel (1971), Prop. 1.2.3).

To finish the proof of (2.6) it remains to deal with the squared bias term

$$B^2 = \int_{-\pi}^{\pi} (\mu(t)-E\mu_{\lambda 1}(t))^2 dt.$$

An application of Parseval's relation along with Lemma 2 and arguments similar to those in Hall (1981) reveals that

$$B^2 = [\mu(\pi)-\mu(-\pi)]^2(\pi\lambda)^{-1} + o(\lambda^{-1}) + O(\sqrt{\lambda}/n) + O((\lambda/n)^2). \tag{5.2}$$

Equation (2.6) follows immediately from (5.1) and (5.2).

The proof of (2.10) is similar to that of (2.6) but relies on work in Hall (1983) rather than Hall (1981). One uses Lemma 3 to provide an expression for the estimator's variance and then applies results in Hall (1983) to characterize the asymptotic behavior of $\int_{-\pi+\epsilon}^{\pi-\epsilon} \int_{-\pi}^{\pi} [K_\lambda^2(t-s)/p(s)]ds$. The squared bias is handled using Lemma 2 which allows us to separate the bias into a sum involving the unestimated Fourier coefficients of μ and a sum depending on the estimation biases for the $2\lambda+1$ estimated Fourier coefficients. The properties of the first sum follow from results in Hall (1983), while, using Lemma 2,

$$|A_o - Ea_o|/(2\pi) + (1/\pi)\left|\sum_{j=1}^{\lambda}[(A_j - Ea_j)\cos jt + (B_j - Eb_j)\sin jt]\right| = O(\lambda/n),$$

uniformly in t. Upon combining all these results one obtains (2.10).

The proofs for (2.7) - (2.8) and (2.11) - (2.12) can be obtained by analogous arguments to those required for (2.6) and (2.10). The only new difficulty that arises is in obtaining an approximation to the variances of $\mu_{\lambda 2}$ and $\mu_{\lambda 3}$. Using an extension of Lemma 3 one can show that the integrated variance of $\mu_{\lambda i}$, i=2,3, is well approximated by $n^{-1}\{\int_{-\pi}^{\pi}[K_\lambda^2(t-s)/p(s)]ds+2(-1)^{i+1}\int_0^\pi\int_0^\pi[K_\lambda(t-s)K_\lambda(t+s)/p(s)]ds\,dt\}$. One now uses the fact that $K_\lambda(-t) = K_\lambda(t)$ and that $\int_{-\pi}^{\pi}|K_\lambda(t)|dt = O(\log\lambda)$ to justify the approximation that was employed.

<u>Proofs of Theorems 3 and 4</u>. The bias of, e.g., $\mu^*_{\lambda 3}$ is

$$E[\mu^*_{\lambda 3}(t)] - \mu(t) = -(2/\pi)[(\lambda+1)^{-1}\sum_{j=1}^{\lambda} jA_j\cos jt + \sum_{j=\lambda+1}^{\infty} A_j\cos jt] + O(\lambda/n)$$

$$= b_\lambda(t)+O(\lambda/n),$$

where $O(\lambda/n)$ holds uniformly in t. This follows from Lemma 2. Now, if (a^*,b^*) is any subinterval of $[0,\pi]$,

$$\int_{a^*}^{b^*}\{E[\mu^*_{\lambda 3}(t)]-\mu(t)\}^2dt = \int_{a^*}^{b^*}[b_\lambda(t)]^2dt+o(\lambda/n),$$

since the latter integral tends to 0 and $\lambda/n \to 0$. The integrated squared bias terms in (3.3) and (3.5) follow immediately using the results of Hall (1981, 1983).

To obtain the variance of $\mu^*_{\lambda 3}(t)$, we first note that

$$\mu^*_{\lambda 3}(t) = \sum_{j=1}^{n} y_j \int_{s_{j-1}}^{s_j} [F_\lambda(t-u)+F_\lambda(t+u)]du,$$

where F_λ is the Fejer Kernel. Using Lemma 3 and a simple modification of it, we have

$$\mathrm{var}(\mu^*_{\lambda 3}(t)) = \sigma^2 n^{-1} \int_0^\pi [F_\lambda(t-u)+F_\lambda(t+u)]^2 [p(u)]^{-1} du + O(\lambda^2/n^2+\lambda^4/n^3),$$

where the second term holds uniformly in t. We have

$$\int_0^\pi \mathrm{var}(\mu^*_{\lambda 3}(t))dt = \sigma^2 n^{-1} \int_0^\pi [p(u)]^{-1} \{\int_{-\pi}^\pi F_\lambda^2(t)dt$$

$$+2\int_0^\pi F_\lambda(t-u)F_\lambda(t+u)dt\}du + O(\lambda^2/n^2 + \lambda^4/n^3).$$

Since F_λ is nonnegative,

$$\int_0^\pi [p(u)]^{-1} \int_0^\pi F_\lambda(t-u)F_\lambda(t+u)dt\ du \le \sup[p(u)]^{-1} \int_{-\pi}^\pi \int_{-\pi}^\pi F_\lambda(t-u)F_\lambda(t+u)dt\ du =$$

$$\sup[p(u)]^{-1} \int_{-\pi}^\pi (2\pi)^{-1} \sum_{j=-\lambda}^{\lambda} (1-|j|/(\lambda+1))^2 e^{2iju}\ du = \sup[p(u)]^{-1}.$$

The integrated variance term in (3.3) now follows from the above argument. The variance term in (3.5) is derived using the variance approximation above and an argument as on pp. 243-244 of Hall (1983). Establishing (3.1), (3.2), and (3.4) is done in an analogous manner.

To prove Theorem 5 we require two further lemmas. First, however, we introduce some additional notation.

Let K be the n×n matrix whose $(ij)^{th}$ element is

$$\pi^{-1}(s_j - s_{j-1}) + (2/\pi)\sum_{r=1}^{\lambda}\cos(rt_i)\int_{s_{j-1}}^{s_j}\cos(rs)ds.$$

Note that K transforms a vector of constants to the vector of "fitted values" under the TS cosine estimator. Also define the vectors $\underline{\eta} = (\eta_1, \ldots, \eta_n)'$, $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$, $\underline{f} = (f(t_1), \ldots, f(t_n))'$ and $\underline{g} = (g(t_1), \ldots, g(t_n))'$ and, for any vector $\underline{z} = (z_1, \ldots, z_n)'$, let $\tilde{\underline{z}} = (I-K)\underline{z}$ and $||\underline{z}||^2 = \Sigma_{j=1}^{n} z_i^2$.

Lemma 4. Let f" and g" be of bounded variation on $[0,\pi]$, $t_i = \pi(i-1)/n$, $i=1,\ldots,n$ and assume that $n, \lambda \to \infty$ in such a way that $\lambda/n \to 0$. Then $n^{-1}||\tilde{\underline{f}}||^2$ and $n^{-1}||\tilde{\underline{g}}||^2$ are both $O(e(\lambda))$, where $e(\lambda)$ is defined in (4.4).

Lemma 5. Under model (4.1)

i) $||\underline{\eta}||^2 = O_p(n)$

ii) $trK'K = O(\lambda)$

iii) $||K\underline{\eta}||^2 = O_p(\lambda) = ||K'\underline{\eta}||^2$

and

$$\underline{\eta}'\tilde{\underline{f}} = O_p(||\tilde{\underline{f}}||).$$

The proof of Lemma 4 rests on the following result. If, for example, $z_i = f(t_i) + \alpha_i$ with the α_i zero mean uncorrelated variables with some common variance, and $f_{\lambda 3}$ is the TS cosine series estimator of f computed under this model, then $n^{-1}||\tilde{\underline{f}}||^2 = \int_0^\pi (f(t) - Ef_{\lambda 3}(t))^2 dt + O(n^{-1})$. This can be established as follows. Let $P(t) = t/n$ and let P_n be the distribution function that places point mass n^{-1} at each of the points $t_1, \ldots, t_n$.

Then, $|n^{-1}||\tilde{f}||^2 - \int_0^\pi (f(t) - Ef_{\lambda 3}(t))^2 dt|$

$= |\int_0^\pi (f(t) - Ef_{\lambda 3}(t))^2 \; d(P_n - P)(t)| \le O(n^{-1}) \int_0^\pi |f(t) - Ef_{\lambda 3}(t)||f'(t) - Ef'_{\lambda 3}(t)| dt,$

through integration by parts. By the Cauchy-Schwarz inequality, the latter integral is bounded by the product of the $L_2(0,1)$ norms of $f - Ef_{\lambda 3}$ and $f' - (Ef_{\lambda 3})'$. Now use Theorem 1 and an extension of Theorem 1 to see that this product is $O(\lambda^{-3})$. The proof of Lemma 5 is elementary and therefore omitted.

Using the notation introduced above we have

$$\hat{\beta} = \underline{\tilde{x}}'\underline{\tilde{y}}/||\underline{\tilde{x}}||^2 = \beta + (\underline{\tilde{x}}'\underline{\tilde{f}} + \underline{\tilde{x}}'\underline{\tilde{\epsilon}})/||\underline{\tilde{x}}||^2. \tag{5.3}$$

Thus to establish (4.5) it suffices to show that, under the conditions of Theorem 5,

$$n^{-1}||\underline{\tilde{x}}||^2 = \theta^2 + o_p(1) \tag{5.4}$$

and

$$n^{-1}\underline{\tilde{x}}'\underline{\tilde{f}} = O_p(e(\lambda)). \tag{5.5}$$

We can write $n^{-1}||\underline{\tilde{x}}||^2 = n^{-1}||\underline{\tilde{g}}||^2 + 2n^{-1}\underline{\tilde{g}}'\underline{\tilde{\eta}} + n^{-1}||\underline{\tilde{\eta}}||^2$. By Lemma 4, $n^{-1}||\underline{\tilde{g}}||^2 = O(e(\lambda))$. Observe that $||\underline{\tilde{\eta}}||^2 = \underline{\eta}'\underline{\eta} - \underline{\eta}'K\underline{\eta} - \underline{\eta}'K'\underline{\eta} + ||K\underline{\eta}||^2 = n\theta^2 + o_p(n) + O_p(\sqrt{n\lambda}) + O_p(\lambda)$ by Lemma 5. Thus $|n^{-1}\underline{\tilde{g}}'\underline{\tilde{\eta}}| \le O_p(\sqrt{e(\lambda)}) = o_p(1)$. Collecting these estimates proves (5.4).

To verify (5.5), write $n^{-1}\underline{\tilde{x}}'\underline{\tilde{f}} = n^{-1}\underline{\tilde{g}}'\underline{\tilde{f}} + n^{-1}\underline{\eta}'\underline{\tilde{f}} - n^{-1}\underline{\eta}'K\underline{\tilde{f}}$. Using Lemma 4, $n^{-1}\underline{\tilde{g}}'\underline{\tilde{f}}$ is found to be $O(e(\lambda))$. Lemmas 4 and 5 then show that $n^{-1}\underline{\eta}'\underline{\tilde{f}}$ and $n^{-1}\underline{\eta}'K\underline{\tilde{f}}$ are $O_p(\sqrt{e(\lambda)/n})$ and $O_p(\sqrt{\lambda/n})O(\sqrt{e(\lambda)}) = O_p(e(\lambda))$, respectively. Thus, (5.5) has been shown.

For the proof of (4.6) first observe that

For the proof of (4.6) first observe that

$$\mathrm{Var}\ (\hat{\beta}|\underline{x}) - \sigma^2||\underline{\tilde{x}}||^{-2} = ||\underline{\tilde{x}}||^{-4}(||K'\underline{\tilde{x}}||^2 - \underline{\tilde{x}}'K\underline{\tilde{x}} - \underline{\tilde{x}}'K'\underline{\tilde{x}}).$$

Now $||K'\underline{\tilde{x}}|| \le ||K'\underline{\eta}|| + ||K'||\ ||\underline{g}-K\underline{x}|| = O_p(\sqrt{\lambda}) + O(\sqrt{\lambda})O_p(\sqrt{ne(\lambda)}) = O_p(ne(\lambda))$ and, as a result of (5.4), we know that $||\underline{\tilde{x}}|| = O(\sqrt{n})$. Thus $|\underline{\tilde{x}}'K\underline{\tilde{x}}|$ and $|\underline{\tilde{x}}'K'\underline{\tilde{x}}|$ are both $O(n^{3/2}e(\lambda))$. Combining these estimates with (5.4) and the fact that, under the conditions of the theorem, $\sqrt{n}\ e(\lambda) = o(1)$ completes the proof of (4.6).

Proof of asymptotic normality for $\hat{\beta}$

To establish asymptotic normality for $\hat{\beta}$, first write $\hat{\beta} = \underline{c}_n'\underline{y}_n/||\underline{\tilde{x}}_n||^2$, where $c_n' = \underline{x}_n'(I-K')(I-K)$. Here we explicitly display the dependence on n, and we will write $\underline{c}_n' = (c_{1n},\ldots,c_{nn})$, $\underline{g}_n' = (g(t_{1n}),\ldots,g(t_{nn}))$, etc. If $\lambda^2/n \to \infty$ and $\lambda^6/n \to \infty$, by (4.5), (4.6) and (5.4) it suffices to prove

$$n^{-1/2}\underline{c}_n'\ \underline{\epsilon}_n \Rightarrow N(0,\sigma^2).$$

This will follow from the Lindeberg condition by showing that

$$\max_{1\le i\le n} n^{-1/2}|c_{in}| \xrightarrow{P} 0. \tag{5.6}$$

Note that the coefficients c_{in} are random rather than constant as in the usual statement of the Lindeberg condition. However, the usual case extends to the present situation because (5.6) implies that

$$E[\exp(it(n^{-1/2}\underline{c}_n'\underline{\epsilon}_n)|\underline{x}_n] \xrightarrow{P} \exp(t^2\sigma^2/2).$$

The proof of (5.6) requires an estimate. Recall that the sup norm of $\underline{c}_n$ is $||\underline{c}_n||_\infty = \max_{1\le i\le n} |c_{in}|$ and the sup norm of the matrix $K=[K_{ij}]$ is $||K||_\infty = \max_{1\le i\le n} \sum_{j=1}^{n}|K_{ij}|$ (cf. Stewart (1973)). Hence

$$||c_n||_\infty \le (1+||K'||_\infty)(1+||K||_\infty)||\underline{x}_n||_\infty.$$

Since $K_{ij}= \int_{s_{j-1}}^{s_j} \frac{1}{2}(K_\lambda(t_i-s) + K_\lambda(t_i+s))ds$, where K_λ is the Dirichlet kernel, and

$$\sum_j|K_{ij}| \le (1/2)\int_0^\pi (|K_\lambda(t_i-s)|+|K_\lambda(t_i+s)|)ds < \int_{-\pi}^{\pi} |K_\lambda(u)|du = O(\log \lambda)$$

uniformly in i, we have $||K||_\infty = O(\log\lambda)$. To estimate $||K'||_\infty$, use the integral mean value theorem to obtain

$$\sum_j |K_{ij}|= O(n^{-1}) \sum_i |K_\lambda(t_i-\theta_j) + K_\lambda(t_i+\theta_j)|$$

for $s_{j-1}\le\theta_j\le s_j$. A quadrature argument similar to the one for Lemma 3 then yields $||K'||_\infty = O(\log\lambda + \lambda^2/n)$. Thus with the assumption $\lambda^2/n = O(1)$, we obtain

$$||c_n||_\infty= O((\log\lambda)^2)||\underline{x}_n||_\infty \le O((\log\lambda)^2) (||\underline{g}_n||_\infty+||\underline{\eta}_n||_\infty).$$

Clearly $||\underline{g}_n||_\infty$ is uniformly bounded. To estimate $||\underline{\eta}_n||_\infty$, note that by the Markov inequality we have for any constant m_n

$$P(\max_{1\le i\le n} |\eta_{in}|> m_n) \le \sum_{i=1}^{n} P(|\eta_{in}|\ge m_n) = O(nm_n^{-(2+\delta)}),$$

since $E|\eta_{in}|^{2+\delta}$ is assumed bounded. Thus with $m_n = n^p$ for some p satisfying $1/(2+\delta)<p<1/2$, say $p = (2+\delta)/(4+2\delta)$, we have

$$||\underline{\eta}_n||_\infty = o_p(n^p)$$

and

$$n^{-1/2}||\underline{c}_n||_\infty = O(n^{-1/2}(\log\lambda)^2)||\underline{\eta}_n||_\infty = o_p(n^{p-1/2}(\log\lambda)^2).$$

The last term is $o_p(1)$ by the conditions on λ, and the proof of the theorem is complete.

References

Billingsley, P. (1986). Probability and Measure, John Wiley and Sons, New York.

Butzer, P. and Nessel, R. (1971). Fourier Analysis and Approximation, Academic Press, New York.

Gasser, Th. and Müller, H. G. (1979). Kernel estimation of regression functions, in Smoothing Techniques for Curve Estimation, eds. Th. Gasser and M. Rosenblatt, Heidelberg: Springer-Verlag, pp. 23-68.

Greblicki, W. and Pawlak, M. (1985). Fourier and Hermite series estimates of regression functions. Ann. Inst. Statist. Math A 37, 443-454.

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.

Kronmal, R. and Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. J. Amer. Statist. Assoc. 63, 925-952.

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. (1986). Convergence rates for partially splined models. Statist. and Probab. Letters 4, 203-208.

Rutkowski, L. (1982). On system identification by nonparametric function fitting. IEEE Trans. Automatic Control AC-27, 225-227.

Speckman, P. (1986). Kernel smoothing in partial linear models. Manuscript.

Stewart, G. (1973). Introduction to Matrix Computations. Academic Press, New York.