Trigonometric Series Regression Estimators with an Application to Partially Linear Models

R. L. Eubank*

Southern Methodist University

J. D. Hart†

Texas A & M University

AND

Paul Speckman

University of Missouri, Columbia

Communicated by the Editors

Let $\mu$ be a function defined on an interval $[a, b]$ of finite length. Suppose that $y_1, \ldots, y_n$ are uncorrelated observations satisfying $E(y_j) = \mu(t_j)$ and $\text{var}(y_j) = \sigma^2$, $j = 1, \ldots, n$, 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 $\mu$ are obtained. Our integrated squared bias approximations closely parallel those of Hall 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 whose nonparametric estimator was a partial spline.

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 trigono-
metric 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 [12], Greblicki and Pawlak
[5], and Rafajlowicz [10]. In contrast, series estimators have played a
prominent role in the estimation of probability densities. (See, e.g.,
[8, 6, 7].) 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.

Assume that observations \( y_1, \ldots, y_n \) are obtained following the model

\[
y_i = \mu(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}
\]

where the \( \varepsilon_i \) are zero mean uncorrelated errors with common variance \( \sigma^2 \),
\( \mu \) is an unknown regression function and the \( t_i \) are design points satisfying
\( a \leq t_1 < \cdots < t_n \leq b \) for finite constants \( a \) and \( b \). The objective is to estimate
\( \mu \) assuming only that it satisfies certain smoothness conditions.

In many cases it is possible to represent \( \mu \) in (1.1) as a Fourier series
involving sine and/or cosine functions \( \{ \phi_j \}_{j=1}^{\infty} \). More precisely, it is often
possible to write \( \mu = \sum_{j=1}^{\infty} \beta_j \phi_j \) for Fourier coefficients \( \beta_j \). Thus, if
estimators \( \hat{\beta}_j \) can be derived for the \( \beta_j \), the first \( \lambda \) terms in the series can
be estimated to produce an estimator \( \hat{\mu}_\lambda = \sum_{j=1}^{\lambda} \hat{\beta}_j \phi_j \) for \( \mu \). We will refer to
estimators of this type as trigonometric series (TS) estimators.

To assess the performance of an estimator \( \hat{\mu}_\lambda \) for \( \mu \), we use the integrated
mean squared error (IMSE)

\[
R(\hat{\mu}_\lambda) = \int_a^b E(\mu(t) - \hat{\mu}_\lambda(t))^2 \, dt. \tag{1.2}
\]

In the next section we study the asymptotic behavior of \( R(\hat{\mu}_\lambda) \) for TS
estimators constructed from both sine and cosine functions or either sine or
cosine functions alone. It is seen that, unless \( \mu \) satisfies certain boundary
conditions, neither the sine nor sine and cosine estimators perform up to
the level of competing estimators such as kernel or smoothing spline
estimators. 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 deter-
mining the large sample properties of \( R(\hat{\mu}_\lambda) \). 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 [7] and also analyze the partial integrated mean squared error (PIMSE)

\[
R_i(\mu_j) = \int_{a-\varepsilon}^{b-\varepsilon} E(\mu(t) - \mu_j(t))^2 dt,
\]

where \( \varepsilon \) is a positive constant satisfying \( \varepsilon < (b - a)/2 \). When viewed through this performance window, TS estimators fare somewhat better than before in terms of 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 \( R_i(\mu_j) \). 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 \([a, b]\).

In Section 3 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., [11]). Proofs of all results are provided in Section 4.

2. TRIGONOMETRIC SERIES ESTIMATORS

Assume that model (1.1) holds with \( \mu \) absolutely continuous on \([a, b]\). Also, for notational simplicity, assume that \([a, b] = [-\pi, \pi]\) or \([a, b] = [0, \pi]\). Let \( s_0 = a, s_j = (t_j + t_{j+1})/2, j = 1, ..., n - 1, s_n = b \) and define (for \( j = 0, 1, ... \))

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

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 \( \mu \) by

\[
\hat{\mu}_2(t) = (2\pi)^{-1} \left[ a_0 + 2 \sum_{j=1}^{\lambda} (a_j \cos jt + b_j \sin jt) \right].
\]

(2.2)
If \([a, b] = [0, \pi]\), two other possible estimators are

\[
\mu_{i2}(t) = (2/\pi) \sum_{j=1}^{\infty} b_j \sin jt
\]

(2.3)

and

\[
\mu_{i3}(t) = \pi^{-1} \left[ a_0 + 2 \sum_{j=1}^{\infty} a_j \cos jt \right].
\]

(2.4)

The estimators \(\mu_{i2}, i = 1, 2, 3\), 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 of kernel estimators proposed by Gasser and Müller [4].

The Gasser and Müller [4] kernel regression estimator is the convolution of a kernel with the piecewise constant function \(y_n(t) = \sum_{r=1}^{s_r} y_r I_r(t)\), where \(I_r\) is the indicator on \((s_{r-1}, s_r]\). Their estimator can thus be viewed as a natural extension of kernel density estimators to the case of mean value function estimation (cf. [9, 15]). Similarly the \(\mu_{i2}\) can be viewed as extensions of the Kronmal and Tarter [8] Fourier series density estimators. A referee has pointed out that one could also regard the regression estimators of Greblicki and Pawlak [5] as generalizations of the Kronmal–Tarter scheme.

Estimators of the form (2.2)–(2.4) have also been studied by Rutkowski [12], who shows certain pointwise and global consistency properties of the estimators. Rafajlowicz [10] and Greblicki and Pawlak [5] obtain upper bounds for \(L_2\) convergence rates for Fourier series estimators when, respectively, the regression function is periodic and the design is random. 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

\[
\int_{a}^{t_j} p(t) \, dt = (j - 1)/n, \quad j = 1, \ldots, n.
\]

(2.5)
Assume also that \( \mu' \) is continuous on \([a, b]\). If \( n, \lambda \to \infty \) in such a way that \( \lambda^2/n = O(1) \), the following results hold:

\[
R(\mu_{A1}) = \sigma^2 \lambda(n\pi)^{-1} \int_{-\pi}^{\pi} [p(t)]^{-1} dt + [\mu(\pi) - \mu(-\pi)]^2(\pi\lambda)^{-1} + o(\lambda/n + \lambda^{-1})
\]

\[
R(\mu_{A2}) = \sigma^2 \lambda(n\pi)^{-1} \int_{0}^{\pi} [p(t)]^{-1} dt + 2[\mu(\pi)^2 + \mu(0)^2](\pi\lambda)^{-1} + o(\lambda/n + \lambda^{-1}).
\]

If in addition \( \mu'' \) is absolutely integrable,

\[
R(\mu_{A3}) = \sigma^2 \lambda(n\pi)^{-1} \int_{0}^{\pi} [p(t)]^{-1} dt + 2[\mu'\pi)^2 + \mu'(0)^2] + (3\pi\lambda^2)^{-1} + o(\lambda/n + \lambda^{-3}).
\]

Under the conditions of Theorem 1, we see that the best rate of convergence for \( R(\mu_{ji}) \), \( i = 1, 2 \), is \( n^{-1/2} \). This is obtained by taking \( \lambda = c_0 n^{1/2} \). The cosine estimator \( \mu_{A3} \) performs considerably better with \( R(\mu_{A3}) = O(n^{-3/4}) \) when \( \lambda = c_1 n^{1/4} \). Thus, from the standpoint of IMSE, the cosine series estimator is to be preferred over either \( \mu_{A1} \) or \( \mu_{A2} \). 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, \( \mu_{A3} \) is comparable to a kernel estimator of \( \mu \) in this sense. Of course it is possible to utilize boundary kernels (see, e.g., [4]) to obtain the better rate of \( n^{-4/5} \) for the IMSE of a kernel estimator. Similar modifications are undoubtedly possible for \( \mu_{A3} \), although we will not pursue that topic here.

If \( \mu \) is smoothly periodic one can parallel the work of Hall [6] and establish improved rates of convergence for the three TS estimators. For example, if \( \mu(0) = \mu(\pi) \), the IMSE for \( \mu_{A1} \) can be made to decay at a rate of \( n^{-3/4} \) by choosing \( \lambda = c_2 n^{1/4} \). Similar results hold for \( \mu_{A2} \) and \( \mu_{A3} \). 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 \( \mu_{A1} \) and \( \mu_{A2} \). However, this phenomenon has a simple explanation. The cosine series expansion of \( \mu \) with support on \([0, \pi]\) is identical to the Fourier series (i.e., sine and cosine) expansion of a function \( \mu^* \) on \([-\pi, \pi]\) obtained by reflecting \( \mu \) about zero. Thus, the bias for \( \mu_{A3} \) will be the same as that for \( \mu_{A1} \) in estimating \( \mu^* \). Since \( \mu^*(-\pi) = \mu^*(\pi) \), this translates into an \( n^{-3/4} \) rate as noted above.
While $\mu_{\lambda 3}$ appears to be the preferred estimator for general $\mu$, there are cases where the use of $\mu_{\lambda 2}$ is advisable. To see when this occurs, observe that the sine series $2\pi^{-1} \sum_{j=-\infty}^{\infty} \beta_j \sin jt$ of the function $\mu$ (defined on $[0, \pi]$) is the Fourier series of the odd function $\mu_0(t) = \text{sgn}(t) \mu(|t|), \ -\pi \leq t \leq \pi$ (where $\text{sgn}(0) = 1$). Now, suppose that $\mu'(0+)$ and $\mu'(\pi-)$ exist and that $\mu(0) = \mu(\pi) = 0$. Then $\mu_0$ is differentiable at 0 and satisfies $\mu_0(-\pi) = \mu_0(\pi)$ and $\mu'_0(-\pi+) = \mu'_0(\pi-)$. 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 not smaller than $c\lambda^{-3}$, whereas for $\mu_{\lambda 2}$ it can be as small as $O(\lambda^{-5})$ (see [6]). Hence, if one knows that $\mu$ vanishes at 0 and $\pi$ 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 $\mu'$ 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 [6] and Lemmas 2 and 3 of Section 4). Thus, by taking $\lambda = c_3 n^{1/3},$ $R(\mu_{\lambda 1}(t) - \mu_{\lambda 1}(t))^2$ 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 $\mu'$ is continuous on $[a, b]$, then for any $0 < \varepsilon < \pi$,

$$R_\varepsilon(\mu_{\lambda 1}) = \sigma^2 \lambda (n\pi)^{-1} \int_{-\pi + \varepsilon}^{\pi - \varepsilon} \left[ p(t) \right]^{-1} dt + \left[ \mu(\pi) - \mu(-\pi) \right]^2 [2(\pi \lambda)^2]^{-1}$$

$$\times \int_{\varepsilon}^{\pi} (1 - \cos t)^{-1} dt + o(\lambda/n + \lambda^{-2}), \quad (2.9)$$

while for any $0 < \varepsilon < \pi/2$,

$$R_\varepsilon(\mu_{\lambda 2}) = \sigma^2 \lambda (n\pi)^{-1} \int_{\varepsilon}^{\pi - \varepsilon} \left[ p(t) \right]^{-1} dt + \left[ \mu(\pi)^2 + \mu(0)^2 \right] [\pi \lambda]^{-2}$$

$$\times \int_{\varepsilon}^{\pi - \varepsilon} (1 - \cos t)^{-1} dt + o(\lambda/n + \lambda^{-2}). \quad (2.10)$$
If, in addition, \( \mu'' \) is of bounded variation,

\[
R_c(\mu_{i3}) = \sigma^2 \lambda (n\pi)^{-1} \int_{-\pi}^{\pi} [p(t)]^{-1} dt + [\mu'(\pi)^2 + \mu'(0)^2] \pi^{-2} \lambda^{-4} \\
\times \int_{-\pi}^{\pi} (1 - \cos t)^{-1} dt + o(\lambda/n + \lambda^{-4}).
\] (2.11)

By choosing \( \lambda = c_{11} n^{1/3} \) we see that PIMSE of both \( \mu_{i1} \) and \( \mu_{i2} \) can be made to decay at a rate of \( n^{-2/3} \). If we take \( \lambda = c_{15} n^{1/5} \) then \( R_k(\mu_{i3}) = O(n^{-4/5}) \). 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.

It is also possible to proceed as in Hall [7] and obtain parallels of Theorems 1 and 2 for Cesaro means of the \( \mu_{ii} \), \( i = 1, 2, 3 \). Unfortunately one finds that these estimators do not provide improvements, asymptotically, over the \( \mu_{ii} \) and, in fact, can perform much worse. The problem with Cesaro mean estimators is that, like kernel estimators, their bias does not continue to decrease as the smoothness of \( \mu \) increases. For a more detailed discussion of these issues, see Eubank, Hart, and Speckman [3].

3. 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 non-parametric 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

\[
y_i = \beta x_i + f(t_i) + \varepsilon_i, \quad i = 1, \ldots, n,
\] (3.1a)

where the \( \varepsilon_i \) are independent, zero mean random variables with common variance \( \sigma^2 \), \( f \) is some unknown function of the covariate \( t \), and \( \beta \) is an unknown regression coefficient. The \( t_i \) satisfy \( 0 \leq t_1 < \cdots < t_n \leq \pi \) and, following Rice [11] and Speckman [13], the \( x_i \) are assumed to admit a regression model in \( t \). Specifically, we assume the \( x_i \) follow the model

\[
x_i = g(t_i) + \eta_i,
\] (3.1b)
where \( g \) is an unknown function, the \( \eta_i \) are independent, zero mean random variables with some common variance \( \theta^2 \), and the \( \varepsilon_i \)‘s and \( \eta_i \)‘s are independent of each other.

For any set of numbers \( z_1, \ldots, z_n \) define

\[
z_{it} = a_0(z)/\pi + (2/\pi) \sum_{j=1}^{\lambda} a_j(z) \cos j t_i,
\]

(3.2)

with

\[
a_j(z) = \sum_{r=1}^{n} z_r \int_{t_{r-1}}^{t_r} \cos j t \, dt.
\]

Then our proposed estimator of \( \beta \) is

\[
\hat{\beta} = \sum_{i=1}^{n} (x_i - x_{it})(y_i - y_{it}) \sum_{r=1}^{n} (x_r - x_{rt})^2.
\]

(3.3)

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 \( \beta \). The definition of \( \hat{\beta} \) in (3.3) is a similar type of adjustment.

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

**Theorem 3.** Let \( x = (x_1, \ldots, x_n)' \) and assume that \( f \) and \( g \) both have continuous derivatives and second derivatives of bounded variation on \([0, \pi]\).

Let

\[
e(\lambda) = \lambda/n + 2(3\pi\lambda^3)^{-1} \max \{g'(\pi)^2 + g'(0)^2, f'(\pi)^2 + f'(0)^2\} + R_{n, \lambda},
\]

(3.4)

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

\[
\beta - E[\hat{\beta} \mid x] = o_p(e(\lambda)).
\]

(3.5)

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

\[
\text{Var}(\hat{\beta} \mid x) = \sigma^2 n^{-1} \theta^{-2} + o_p(n^{-1}).
\]

(3.6)

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 \( \beta \) can be conducted using \( \hat{\beta} \) without the necessity of adjusting for the bias from the non-parametric 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 \( \beta \) (see [11]). The fact that \( \hat{\beta} \) is asymptotically normal with mean \( \beta \) implies that confidence intervals and tests for \( \beta \) can be conducted using standard parametric methods.

Theorem 3 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 \( R(\mu_{i,j}) = o(1/\sqrt{n}) \), \( j = 1, 2 \).

4. PROOFS OF THEOREMS

To prove the results in Section 2 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 \( s_0 = a, s_j = (t_j + t_{j+1})/2, j = 1, \ldots, n - 1 \), and \( s_n = b \), then \( \max_j |s_j - s_{j-1}| = O(n^{-1}) \).

**Lemma 2.** Assume that \( \mu \) has a continuous derivative on \( [a, b] \). Let

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

and

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

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

**Lemma 3.** Consider a quantity of the form

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

where the \( s_j \) are defined as in Lemma 1 with \( p \) continuously differentiable and where \( K_\lambda \) is a continuously differentiable function. Then,

\[
C_\lambda(t) - n^{-1} \int_a^b \left[ K_\lambda^2(t-s)/p(s) \right] \, ds + O(n^{-2}) \left\{ \int_a^b |K_\lambda'(t-s) K_\lambda(t-s)| \, ds + \int_a^b K_\lambda^2(t-s) \, ds \right\}.
\]
Proof. Using the mean value theorem for integrals and the uniform differentiability of $p$, we have, for points $s_{r-1, r}$ with $s_{r-1} \leq \xi_r, \theta_r \leq s_r$,

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

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

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 $P_n(s) = r/n$, $\theta_r \leq s < \theta_{r+1}$ for $1 \leq r < n$, $P_n(s) = 0$ for $s < \theta_1$, and $P_n(s) = 1$ for $s \geq \theta_n$. Using the previous equation we can then express $C_\lambda$ as

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

It is easy to show that $\sup_{a \leq s \leq b} |P_n(s) - P(s)| \leq 2/n$. Use integration by parts (cf. Billingsley [1, Theorem 18.4]) to obtain

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

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

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

and the lemma follows.

Proof of Theorems 1 and 2. We indicate only how to prove (2.6) and (2.9) as the proof of the other results follows a similar pattern. To obtain (2.6) we begin by noting that

$$E(\mu(t) - \mu_{\lambda_1}(t))^2 = \text{Var } \mu_{\lambda_1}(t) + (\mu(t) - E\mu_{\lambda_1}(t))^2.$$

Now observe that $\mu_{\lambda_1}$ can be written as $\sum_{j=1}^n y_j \int_{-\pi}^{\pi} K_\lambda(t-s) \, ds$ with $K_\lambda$ the Dirichlet kernel, i.e., $K_\lambda(u) = (2\pi)^{-1} \sum_{|j| \leq \lambda} e^{i j u}$. Thus, an application of Lemma 3 gives

$$\text{var } \mu_{\lambda_1}(t) = n^{-1} \int_{-\pi}^{\pi} [K_\lambda^2(t-s)/p(s)] \, ds + O \left( \frac{\lambda^2 \log \lambda}{n^2} \right) + O(\lambda/n^2) \quad (4.1)$$

uniformly in $t$. In applying the lemma we have used the facts that $\max_{-\pi \leq s \leq \pi} |K_\lambda(s)| = O(\lambda^2)$, $\int_{-\pi}^{\pi} K_\lambda^2(s) \, ds = 2\lambda + 1$ and $\int_{-\pi}^{\pi} |K_\lambda(s)| \, ds = O(\log \lambda)$. The last bound is the Lebesgue constant (cf. [2, Proposition 1.2.3]).
To finish the proof of (2.6) it remains to deal with the squared bias term

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

An application of Parseval's relation along with Lemma 2 and arguments similar to those in Hall [6] reveal that

$$B^2 = [\mu(\pi) - \mu(-\pi)]^2 (2\pi)^{-1} + \sigma(\pi)^{-1} + O(\sqrt{\lambda/n}) + O((\lambda/n)^3). \quad (4.2)$$

Equation (2.6) follows immediately from (4.1) and (4.2).

The proof of (2.9) is similar to that of (2.6) but relies on work in Hall [7] rather than [6]. One uses Lemma 3 to provide an expression for the estimator's variance and then applies results in Hall [7] to characterize the asymptotic behavior of $$\int_{-\pi}^{\pi} [K^2(t-s)/p(s)] \, ds.$$ The squared bias is handled using Lemma 2 which allows us to separate the bias into a sum involving the unestimated Fourier coefficients of $$\mu$$ 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 [7], while, using Lemma 2,

$$|A_0 - E\theta_0|/(2\pi) + (1/\pi) \left| \sum_{j=1}^{j} [(A_j - E\theta_j) \cos jt + (B_j - E\theta_j) \sin jt] \right| = O(\lambda/n),$$

uniformly in $$t$$. Upon combining all these results one obtains (2.9).

The proofs for (2.7)–(2.8) and (2.10)–(2.11) can be obtained by analogous arguments to those required for (2.6) and (2.9). The only new difficulty that arises is in obtaining an approximation to the variances of $$\mu_{i,2}$$ and $$\mu_{i,3}$$. Using an extension of Lemma 3 one can show that the integrated variance of $$\mu_{i,2}, i = 2, 3, \text{ is well approximated by}$$

$$n^{-1} \{ \int_{-\pi}^{\pi} [K_{i}^2(t-s)/p(s)] \, ds + 2(-1)^{i+1} \sum_{j=0}^{n+1} \int_{0}^{\pi} [K_{i}(t-s)K_{i}(t+s)/p(s)] \, ds \, dt \}.$$ 

One now uses the fact that $$K_{i}(-t) = K_{i}(t)$$ and that $$\int_{0}^{\pi} |K_{i}(t)| \, dt = O(\log \lambda)$$ to justify the approximation that was employed.

To prove Theorem 3 we require two further lemmas. First, however, we introduce some additional notation.

Let $$K$$ be the $$n \times n$$ matrix whose $$(ij)$$th element is

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

Note that $$K$$ transforms a vector of constants to the vector of “fitted values” under the TS cosine estimator. Also define the vectors $$\eta = (\eta_1, ..., \eta_n)'$$, $$\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$$, $$f = (f(t_1), ..., f(t_n))'$$, and $$g = (g(t_1), ..., g(t_n))'$$ and, for any vector $$z = (z_1, ..., z_n)'$$, let $$\tilde{z} = (I - K)z$$ and $$\|z\|^2 = \sum_{j=1}^{n} z_j^2.$$
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 $n^{-1} \| \hat{f} \|^2$ and $n^{-1} \| \hat{g} \|^2$ are both $O(e(\lambda))$, where $e(\lambda)$ is defined in (3.4).

LEMMA 5. Under model (3.1)

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

(ii) $\text{tr } K'K = O(\lambda)$

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

and

$$\eta \hat{f} = O_p(\| \hat{f} \|).$$

The proof of Lemma 4 rests on the following result. If, for example, $z_i = f(t_i) + \alpha_i$ with the $\alpha_i$ 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 $n^{-1} \| \hat{f} \|^2 = \int_0^\pi (f(t) - Ef_{\lambda,3}(t))^2 \, dt + O(n^{-1})$. This can be established as follows. Let $P(t) = t$ 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,

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

$$\leq \left| \int_0^\pi (f(t) - Ef_{\lambda,3}(t))^2 \, d(P_n - P)(t) \right|$$

$$\leq O(n^{-1}) \int_0^\pi |f(t) - Ef_{\lambda,3}(t)| \, |f'(t) - Ef'_{\lambda,3}(t)| \, dt,$$

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 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

$$\hat{\beta} = \hat{\beta}'\hat{y}/\| \hat{y} \|^2 = \beta + (\hat{\beta}'\hat{y} + \hat{\beta}'\hat{e})/\| \hat{y} \|^2. \quad (4.3)$$

Thus to establish (3.5) it suffices to show that, under the conditions of Theorem 3,

$$n^{-1} \| \hat{\beta} \|^2 = \theta^2 + o_p(1) \quad (4.4)$$

and

$$n^{-1} \hat{\beta} \hat{f} = O_p(\epsilon(\lambda)). \quad (4.5)$$
We can write \( n^{-1} \| \hat{x} \|^2 = n^{-1} \| \hat{g} \|^2 + 2n^{-1} \hat{g} \cdot \hat{n} + n^{-1} \| \hat{n} \|^2 \). By Lemma 4, \( n^{-1} \| \hat{g} \|^2 = O(e(\lambda)) \). Observe that \( \| \hat{n} \|^2 = n \| \eta - K\eta - n^{-1} K' \eta \|^2 = n\theta^2 + O_p(\sqrt{n}) + O_p(\sqrt{n}) \) by Lemma 5. Thus \( n^{-1} \hat{g} \cdot \hat{n} \leq O_p(\sqrt{e(\lambda)}) = o_p(1) \). Collecting these estimates proves (4.4).

To verify (4.5), write \( n^{-1} \hat{x}' \hat{f} = n^{-1} \hat{g}' \hat{f} + n^{-1} \hat{f}' \hat{n} - n^{-1} \eta' \kappa \). Using Lemma 4, \( n^{-1} \hat{g}' \hat{f} \) is found to be \( O(e(n)) \). Lemmas 4 and 5 then show that \( n^{-1} \hat{f}' \hat{n} \) and \( n^{-1} \eta' \kappa \) are \( O(e(n)) \). Thus (4.5) has been shown.

For the proof of (3.6) first observe that

\[
\text{Var}(\hat{\beta} \mid x) - \sigma^2 \| \hat{x} \|^2 - \| \hat{x} \|^4 (\| K' \hat{x} \|^2 - \hat{x}'K\hat{x} - \hat{x}'K'\hat{x}).
\]

Now \( \| K' \hat{x} \| \leq \| K' \eta \| + \| K' \| \| g - Kx \| = O_p(\sqrt{\lambda}) + O(\sqrt{\lambda}) O_p(\sqrt{ne(\lambda)}) = O_p(ne(\lambda)) \) and, as a result of (4.4), we know that \( \| \hat{x} \| = O(\sqrt{n}) \). Thus \( \| \hat{x}'K\hat{x} \| \) and \( \| \hat{x}'K'\hat{x}\| \) are both \( O(n^{3/2}e(\lambda)) \). Combining these estimates with (4.4) and the fact that, under the conditions of the theorem, \( \sqrt{n} e(\lambda) = o(1) \) completes the proof of (3.6).

To establish asymptotic normality for \( \hat{\beta} \), first write \( \hat{\beta} = \mathbf{c}'_n y_n / \| \hat{x}_n \|^2 \), where \( \mathbf{c}'_n = \mathbf{x}'_n (I - K') (I - K) \). Here we explicitly display the dependence on \( n \), and we will write \( \mathbf{c}'_n = (c_1, ..., c_m) \), \( \mathbf{g}'_n = (g(t_1), ..., g(t_m)) \), etc. If \( \lambda^2/n \to 0 \) and \( \lambda^6/n \to \infty \), by (3.5), (3.6), and (4.4) it suffices to prove \( n^{-1/2} c'_n \varepsilon_n \to N(0, \sigma^2) \). This will follow from the Lindeberg condition by showing that

\[
\max_{1 \leq i \leq n} n^{-1/2} |c_{in}| \xrightarrow{P} 0. \tag{4.6}
\]

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 (4.6) implies that

\[
E[\exp(it(n^{-1/2} c'_n \varepsilon_n) \mid x_n] \xrightarrow{P} \exp(-t^2 \sigma^2/2).
\]

The proof of (4.6) requires an estimate. First recall that the sup norm of \( c_n \) is \( \| c_n \|_\infty = \max_{1 \leq i \leq n} |c_{in}| \) and the sup norm of the matrix \( K = [K_{ij}] \) is \( \| K \|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |K_{ij}| \) (cf. [14]). Hence \( \| c_n \|_\infty \leq (1 + \| K \|_\infty)(1 + \| K \|_\infty) \| x_n \|_\infty \). Thus, since \( K(s) = \int_{y-1}^y \frac{1}{2} (K_1(s - t) + K_2(s + t)) \) \( ds \), where \( K_1 \) is the Dirichlet kernel, and \( \sum_{j} |K_{ij}| \leq \frac{1}{2} \int_{-\infty}^\infty (|K_1(s - t)| + |K_2(s + t)|) \) \( ds < \int_{-\infty}^\infty |K_2(u)| \) \( du = O(\log \lambda) \) uniformly in \( i \), we have \( \| K \|_\infty = O(\log \lambda) \). To estimate \( \| K' \|_\infty \), use the integral mean value theorem to obtain \( \sum_{j} |K_{ij}| = O(n^{-1}) \sum_{j} (K_1(t - \theta) + K_2(t + \theta)) \) for \( s_j, \theta_j, \leq \theta_j \leq s_j \). 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

\[
\| c_n \|_\infty = O((\log \lambda)^2) \| x_n \|_\infty \leq O((\log \lambda)^2)(\| g_n \|_\infty + \| n \|_\infty).
\]
Clearly \( \| g_n \|_\infty \) is uniformly bounded. To estimate \( \| \eta_n \|_\infty \), note that by the Markov inequality we have for any constant \( m_n \)

\[
P( \max_{1 \leq i \leq n} |\eta_{in}| > m_n) \leq \sum_{i=1}^{n} P(|\eta_{in}| \geq m_n) = O(nm_n^{-(2+\delta)}),
\]

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 \( \| \eta_n \|_\infty = o_p(n^p) \) and, hence, \( n^{-1/2} \| c_n \|_\infty = O(n^{-1/2}(\log \lambda)^{p}) \) \( \| \eta_n \|_\infty = o_p(n^{p-1/2}(\log \lambda)^{p}) \). The last term is \( o_p(1) \) by the conditions on \( \lambda \), and the proof is complete.

REFERENCES