JOURNAL OF MULTIVARIATE ANALYSIS 50, 30–40 (1994)

Series Estimation of Semilinear Models

STEPHEN G. DONALD

Boston University

AND

WHITNEY K. NEWEY

Massachusetts Institute of Technology

This paper discusses estimation of the semilinear model $E[y \mid x, z] = x'\beta + g(z)$ using series approximations to the unknown function $g(z)$ under much weaker conditions than heretofore given in the literature. In particular, we allow for z being multidimensional and to have a discrete distribution, features often present in applications. In addition, the smoothness conditions are quite weak: it will suffice for $\sqrt{n}$ consistency of $\hat{\beta}$ that the modulus of continuity of $g(z)$ and $E[x \mid z]$ be higher than one-fourth the dimension of z and that the number of terms be chosen appropriately.

1. INTRODUCTION

A regression model that is useful for estimating treatment effects while controlling for covariates in a very flexible way is a semilinear model

$$E[y \mid x, z] = x'\beta + g(z), \tag{1}$$

where x and β are $q \times 1$ vectors of regressors and parameters respectively, and $g(z)$ is an unknown function of an r-dimensional vector z. In this model β is the partial effect of treatment x given covariates z, where the covariate effect has unknown functional form but is restricted to enter additively. In this case β is the main object of interest. Alternatively, one might use this model to allow the regression to depend nonparametrically on a variable of interest z, while controlling for covariates x in a parsimonious way, in which case $g(z)$ is the main object of interest.

Received April 28, 1992; revised September 7, 1993.

AMS 1980 subject classifications: primary 62J05, 62J10; secondary 62G99.

Key words and phrases: partly linear model, power series approximation, spline approximation, semiparametric model.

0047-259X/94 $6.00

The purpose of this paper is to discuss estimation of this model using series approximations to $g(z)$, under much weaker conditions than heretofore given in the literature. In particular we allow for z to be multidimensional and to have a discrete distribution, features often present in applications. Also, the smoothness conditions we impose are quite weak: it will suffice for $\sqrt{n}$ consistency of $\hat{\beta}$ that the modulus of continuity of $g(z)$ and $h(z) = E(x \mid z)$ be higher than one-fourth the dimension of z and that the number of terms be chosen appropriately. Bickel and Ritov (1988) showed that for the case of a random vector of dimension one, a continuity modulus of 1/4 is a minimal requirement for $\sqrt{n}$ consistent estimation of the squared density integral. This suggests that in the similar semilinear model, where β can also be thought of as a functional of a nonparametric distribution, our conditions may be nearly minimal. Furthermore, the derivation of the convergence rate is straightforward, using simple results on projection matrices and expectations that are familiar from analysis of ordinary linear regression.

Previous work on semilinear regression models includes papers by Engle *et al.* (1984), Heckman (1986), Rice (1986), Robinson (1988), Speckman (1988), Chen (1988), Eubank and Speckman (1990), Eubank *et al.* (1990), Andrews (1991), and Chamberlain (1986). This paper is different from this literature in the features described above, although Robinson (1988) and Speckman (1988) have given some results for multivariate, continuously distributed z. In addition, we allow for heteroskedasticity and several types of series approximations, including splines, polynomials, and Fourier series.

2. The Estimators

The model can be written as

$$y_i = x_i'\beta + g(z_i) + \varepsilon_i, \qquad (i = 1, ..., n)$$

where x_i is of dimension q, z_i is of dimension r and $E[\varepsilon_i \mid x_i, z_i] = 0$.

The estimator is obtained by regressing y on x and functions that can approximate g. Let

$$p_K(z) = (p_{1K}(z), ..., p_{KK}(z))'$$

be a vector of approximating functions, such as power series or splines. Also, let p_K be the $n \times K$ matrix with ith row $p_{Ki} = p_K(z_i)$ and let $Q = p_K(p_K' p_K)^- p_K'$ be the projection matrix formed with the approximating functions, where $(p_K' p_K)^-$ denotes any generalized inverse. The estimator of β are the coefficients of x in the (possibly) singular linear regression of

y on x and $p_K(z)$. By the usual "profile" or "partialing out" formula, the estimator may be written as

$$\hat{\beta} = (x'(I-Q)\,x)^{-1}\,x'(I-Q)\,y$$

when $x'(I-Q)\,x$ is nonsingular, and the estimator is taken to be zero otherwise. Under the conditions that follow this matrix will be nonsingular with probability approaching 1, and $\hat{\beta}$ invariant to the g-inverse $(p_K' p_K)^-$. This is a series estimator like that considered in Chen (1988), Andrews (1991), Chamberlain (1986), and Newey (1991). However, unlike Chen (1988) and Andrews (1991), we allow for $p_K' p_K$ being asymptotically singular. This is a key feature of our analysis, because it allows for us to be agnostic about the distribution of z_i, e.g., z_i may be discrete.

In the homoskedastic case where $E[\varepsilon_i^2 \mid x_i, z_i] = \sigma^2$ we will be able to show that variance estimation is possible. The usual estimator of the variance covariance matrix in that case is

$$\hat{\sigma}^2(x'(I-Q)\,x/n)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n-q-K}\,(y - x\hat{\beta})'\,(I-Q)(y - x\hat{\beta}).$$

Although in the context of the partially linear model one is often interested in estimating β while g is treated as a nuisance parameter, there are cases where an estimate of g is needed. If one is interested in estimating the function g, then an obvious estimator is

$$\hat{g}(z) = p_K(z)'\,(p_K' p_K)^-\,p_K'(y - x\hat{\beta}),$$

which is a nonparametric projection estimator, where one projects the residual $y - x\hat{\beta}$ on basis functions of z. The estimator will not be invariant to the choice of g-inverse, although $\hat{g}(z_i)$ will be. Of course if β was known exactly then we would be in the situation of a fully nonparametric regression since

$$y - x\beta = g(z) + \varepsilon.$$

We will provide results for two types of basic functions. The first are power series,

$$p_{kK}(z) = \prod_{j=1}^{r} (z_j)^{\lambda_j(k)},$$

where $(\lambda_1(k), ..., \lambda_r(k))$ is a vector of nonnegative integers, distinct for different k, equal to any vector of nonnegative integers for some k and

$\sum_{j=1}^{r} \lambda_j(k)$ increasing in k. The second are interaction splines where we assume that the bounds on the support of z are known, which without loss of generality is assumed to be the cube $\prod_{j=1}^{r} [-1, 1]$. We will also assume $L+1$ fixed evenly spaced "knots," denoted $\bar{z}_{j,l}$ for $l = 1, 2, ..., L+1$. If the knots are evenly spaced then

$$\bar{z}_{j,l} = -1 + \frac{(l-1)}{L+1}.$$

An mth degree spline sequence can be defined by

$$p_k^j(z_j) = \begin{cases} (z_j)^{k-1}, & \text{for} \quad 1 \leqslant k \leqslant m, \\ \{(z_j - \bar{z}_{j,k-m})_+\}^m, & \text{for} \quad k \geqslant m+1, \end{cases}$$

where $(\cdot)_+ = 1(\cdot > 0)(\cdot)$. To obtain multivariate splines one mulitplies the univariate splines

$$p_{kK}(z) = \prod_{j=1}^{r} p_{\lambda_j(k)}^j(z_j), \qquad \text{for} \quad k = 1, 2, ...$$

3. Convergence Rates

The convergence rates we derive will depend on approximation error bounds for g and h. We specify that there are functions $e_g(K)$ and $e_h(K)$ satisfying magnitude restrictions specified below, such that there are π and η with

$$\sup_{n \geqslant 1} \left\{ \sum_{i=1}^{n} E[\{g(z_i) - p_K(z_i)' \pi\}^2]/n \right\}^{1/2} \leqslant e_g(K)$$

$$\max_j \sup_{n \geqslant 1} \left\{ \sum_{i=1}^{n} E[\{h_j(z_i) - p_K(z_i)' \eta\}^2]/n \right\}^{1/2} \leqslant e_h(K),$$

where $h_j(z_i) = E(x_{ji} \mid z_i)$ with x_{ji} being the ith observation on the jth regressor. We will assume throughout that $h_j(z)$ is the same for different observations.

For specific approximating functions, the magnitude of these approximation errors will be determined by the degree of smoothness of the functions being approximated. Define a function $f(z)$ to be smooth of degree d_f if it is continuously differentiable of order equal to the largest integer $[d_f]$, strictly less than d_f and there is a constant C such that for each partial derivative $\partial f(z)$ of order $[d_f]$,

$$|\partial f(z) - \partial f(\bar{z})| \leqslant C\, |\bar{z} - z|^{d_f - [d_f]}.$$

The following assumption is made regarding the unknown functions g and h and the support.

Assumption 1. The support of z is compact and there are extensions of $g(z)$ and $h(z)$ to a compact cube, containing the support such that the extensions are smooth of degree d_g and d_h, respectively. Also, $p_{kK}(z)$ are either power series or splines.

Under Assumption 1, it follows from Lorentz (1986, p. 90) for power series, and from Schumaker (1981, Thm. 12.8) for splines, that

$$e_f(K) = O(K^{-d_f/r}).$$

Note that one advantage of the power series is that one does not need to know the support of z. Splines, however, may be more robust to extreme points in the z data. Similar approximation rates for trigonometric polynomial approximations also follow from Lorentz (1986, p. 90), but we do not discuss such approximations here.

Let

$$u_i = x_i - h(z_i)$$

and let u_{ji} be the jth element of u_i. We make the following assumption regarding the data generating process. In the statement of the assumption δ and Δ will be used to denote generic small and large positive constants, respectively.

Assumption 2. The data are independent observations and

(i) $\text{Var}(y_i \mid x_i, z_i) \leqslant \Delta < \infty$ for all i;

(ii) $(1/n)\sum_{i=1}^n E(u_i u_i') = \bar{A}_n$ is uniformly positive definite;

(iii) $(1/n)\sum_{i=1}^n E(\varepsilon_i^2 u_i u_i') = \bar{B}_n$ is uniformly positive definite;

(iv) $\text{Var}(x_i \mid z_i) \leqslant \Delta < \infty$ is bounded for all i;

(v) $(1/n)\sum_{i=1}^n E(|u_{ji}\varepsilon_i|^{2+\delta}) \leqslant \Delta < \infty$ for each $j = 1, ..., q$ for some $\delta > 0$;

(vi) $(1/n)\sum_{i=1}^n E(|u_{ji}|^{2+\delta}) \leqslant \Delta < \infty$ for each j for some $\delta > 0$.

Conditions (i) and (iv) guarantee that the unknown functions g and h can be consistently estimated using the series (spline) approximations and allow for heteroskedasticity in the residuals ε_i and u_i. Conditions (ii), (iii), (v), and (vi) will allow one to apply the Central Limit Theorem of Liapunov to the estimator one obtains when h is known, given by

$$\tilde{\beta} = (u'u)^{-1} u'(u\beta + \varepsilon),$$

where $u = (u_1, ..., u_n)'$. Condition (ii) is required for identification of β. As is usually the case, x should not include a constant, otherwise this condition will fail. The results will be stated both in terms of the order of approximation for the unknown functions and in terms of restrictions on the way that K is allowed to increase and on the degree of smoothness of the unknown functions. In stating the second set of conditions it is understood from the above discussion that the condition relates to both spline and power series approximations. The first result concerns the convergence rate for $\hat{\beta}$.

THEOREM 1. *Given Assumption* 2, *and assuming that* $K = K(n) \to \infty$ *in such a way that* $K(n)/n \to 0$, *and either*

(i) $e_g(K(n)) \to 0$ *and* $e_h(K(n)) \to 0$, *or*

(ii) *Assumption* 1 *is satisfied with* $d_g > 0$ *and* $d_h > 0$, *then*

$$\hat{\beta} - \beta = O_p(n^{-1/2}) + O_p(e_g(K)\, e_h(K)) + O_p(e_g(K)\, n^{-1/2}) + O_p(e_h(K)\, n^{-1/2}) + O_p(K^{1/2} n^{-1}).$$

The hypotheses on K here are imposed to give consistency of $x'(I-Q)\,x/n$. Given the expansion implied by Theorem 1 it is then easy to prove the following result regarding the limiting distribution of $\hat{\beta}$.

THEOREM 2. *Given the conditions of Theorem* 1 *and either*

(i) $\sqrt{n}\, e_g(K(n))\, e_h(K(n)) \to 0$ *or*

(ii) *Assumption* 1 *is satisfied and* $\sqrt{n}\, K(n)^{-(d_g + d_h)/r} \to 0$, *then* $\sqrt{n}(\hat{\beta} - \beta) = O_p(1)$ *and*

$$(\bar{A}_n^{-1} \bar{B}_n \bar{A}_n^{-1})^{-1/2} \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I_q).$$

In addition, if $E[\varepsilon_i^2 \mid x_i, z_i] = \sigma^2$, *then*

$$\frac{1}{\hat{\sigma}} (x'(I-Q)\,x)^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} N(0, I_q).$$

This result requires weaker restrictions on the distribution of the regressor and the rate of increase of K than the result of Eubank *et al.* (1990). In particular it is only required that $K/n \to 0$ which is somewhat weaker than the requirement that $K^2/n \to 0$, used in Eubank *et al.* (1990), and the requirement that $K^\gamma/n \to 0$ for some $\gamma > 1$, used in Chen (1988). In addition, all the results except the last one allow for heteroskedasticity in the residuals, and z is not required to be continuously distributed or even have a continuous component.

Under Assumption 1, $K(n)$ satisfying the hypotheses of Theorem 2 will exist if $d_g + d_h > r/2$. For example, if z is univariate (so $r = 1$), then if g and h are both Lipschitz of order 1/4 and either one is Lipschitz of slightly higher order then this condition is satisfied. Also, for the case considered by Eubank *et al.* (1990), the functions g and h were assumed to be continuously differentiable with a second derivative of bounded variation. In our notation this corresponds to a smoothness index of 2. Then requirement (ii) of Theorem 2 is $K^8/n \to \infty$, which is weaker than the condition $K^6/n \to \infty$ that appears in that paper.

Next we consider convergence rates for the estimator of g defined in the previous section. The difference between g and $\hat{g}$ will be measured by the square root of the sample MSE defined as

$$\frac{1}{\sqrt{n}} \|\hat{g} - g\| = \left\{\frac{1}{n} \sum_{i=1}^{n} (\hat{g}(z_i) - g(z_i))^2\right\}^{1/2}.$$

The convergence rate result is contained in

THEOREM 3. *Under the conditions of Theorem* 1,

$$\frac{1}{\sqrt{n}} \|\hat{g} - g\| = O_p(K^{1/2} n^{-1/2}) + O_p(e_g(K)),$$

and hence if Assumption 1 *and* $K = cn^{r/(2d_g + r)}$, *then*

$$\frac{1}{\sqrt{n}} \|\hat{g} - g\| = O_p(n^{-d_g/(2d_g + r)}).$$

Under Assumption 1 the optimal rate of convergence is that given in the second conclusion. By comparing Theorems 2 and 3 we can see that the optimal rate of growth of K for estimation of g will result in $\sqrt{n}$-consistency of $\hat{\beta}$ if $d_h > r/2$, a condition similar to that of Chen (1988). However, it is interesting to note that the optimal rate of growth of K for estimation of g may not be the same one that minimizes the remainder terms in $\hat{\beta} - \beta$. From Theorem 1 it is apparent that the optimal rate for the $\hat{\beta}$ remainder also depends on d_h. Also, even if $d_h = d_g$, the growth rate for K that will minimize the remainder terms in Theorem 1 is $n^{2r/(4d_g + r)}$, which is faster than the optimal rate for estimation of g. Thus, for this estimator, although $\sqrt{n}$-consistency can be obtained when K grows at the optimal rate for estimation of g, a faster rate for K can reduce the remainder. Since a larger K tends to reduce bias (analogously to a smaller band-width in a kernel estimator), the larger K for minimizing the remainder corresponds to undersmoothing. This under-smoothing is not required for $\sqrt{n}$-consistency, but may be optimal.

4. Proofs of Results

The proofs proceed by using the Cauchy–Schwarz and Markov inequalities to bound remainders. This approach works well because of the regression form of the estimator. It avoids the use of the complicated central limit theorem used by Chen (1988), while producing sharper results.

Proof of Theorem 1. Throughout the proof the notation C and Δ will be used for generic positive constants. Define the $n \times q$ matrix,

$$h = (h(z_1), ..., h(z_n))'.$$

We use a subscript j to indicate an n vector of all observations on the jth component. Thus, for example, x_j will refer to the n vector of all observations of the jth regressor. The notation M is used for $I - Q$. A fact used throughout the proof is that $Mp = 0$, where we suppress the K subscript for notational convenience.

Note that (ii) implies (i) so that it is sufficient to prove the result under (i). First,

$$\hat{\beta} = \beta + (x'Mx/n)^{-1} x'M(g + \varepsilon)/n.$$

Thus it suffices to prove $(x'Mx/n)^{-1} = O_p(1)$ and

$$x'M(g+\varepsilon)/n = O_p(n^{-1/2}) + O_p(e_g(K)\, e_h(K)) + O_p(e_g(K)\, n^{-1/2}) + O_p(e_h(K)\, n^{-1/2}) + O_p(K^{1/2}n^{-1}).$$

The first of these follows by

$$x'Mx/n - u'u/n = o_p(1) \tag{2}$$

and

$$u'u/n - \bar{A}_n = o_p(1), \tag{3}$$

since $\bar{A}_n$ is uniformly positive definite and finite given Assumptions 2(ii) and 2(vi). Note that (3) follows from the law of large numbers given Assumption 2(vi). To show (2), note that since $x = h + u$,

$$\frac{1}{n} x'Mx - \frac{1}{n} u'u = \frac{1}{n} h'Mh + \frac{1}{n} h'Mu + \frac{1}{n} u'Mh + \frac{1}{n} u'Qu.$$

We show that each of the terms on the right hand side is $o_p(1)$ and then (2) will follow. By the Cauchy–Schwarz inequality it is sufficient to show that the elements of the diagonals of these four matrices is $o_p(1)$. To show that the result holds for the first term, by the Markov inequality

$$\frac{1}{n} h_j' M h_j = \frac{1}{n} (h_j - p\eta_j)' M (h_j - p\eta_j)$$

$$\leqslant \frac{1}{n} (h_j - p\eta_j)' (h_j - p\eta_j) = O_p(e_h(K(n))^2) = o_p(1). \tag{4}$$

For the second term the result holds by Markov's inequality since

$$\frac{1}{n} h_j' M u_j = \frac{1}{n} (h_j - p\eta_j)' M u_j$$

and

$$E\left(\frac{1}{n^2} (h_j - p\eta_j)' M E[u_j u_j' \mid z_1, ..., z_n] M (h_j - p\eta_j)\right) \leqslant C \frac{e_h(K(n))^2}{n} \to 0 \tag{5}$$

using independence of the data and Assumption 2(iv). The same applies to the third term. For the fourth term, the result follows from Markov's inequality since

$$\frac{1}{n} E(u_j' Q u_j) = \frac{1}{n} E(\operatorname{tr}(Q E[u_j u_j' \mid z_1, ..., z_n] Q)) \leqslant C \frac{K}{n} \to 0.$$

Next, we can write

$$x' M(g + \varepsilon)/n = u'\varepsilon/n + h' M g/n + h' M \varepsilon/n + u' M g/n - u' Q \varepsilon/n.$$

We need only consider each element of each term separately. For the first term, we have by Chebyshev's inequality that

$$P(|u_j'\varepsilon/\sqrt{n}| > \Delta) \leqslant \frac{\sum_{i=1}^{n} E(\varepsilon_i^2 u_{ji}^2)}{n\Delta^2},$$

hence $u'\varepsilon/\sqrt{n} = O_p(1)$, so that $u'\varepsilon/n = O_p(n^{-1/2})$. The second term can be shown to be $O_p(e_h(K(n)))\, O_p(e_g(K(n)))$ by using a method similar to (4). Using a method similar to (5) and Chebyshev's inequality, one can show the third term to be $O_p(e_h(K)\, n^{-1/2})$. The fourth term is $O_p(e_g(K)\, n^{-1/2})$ by using a similar argument. The final term $u'Q\varepsilon/n$ has conditional expectation zero, so by Chebyshev's inequality

$$E\left(\frac{1}{n^2} u_j' Q E(\varepsilon\varepsilon' \mid x_1, ..., x_n, z_1, ..., z_n) Q' u_j\right) \leqslant CE\left(\frac{1}{n^2} u_j' Q u_j\right) \leqslant C \frac{K}{n^2}$$

so that the last term is $O_p(K^{1/2} n^{-1})$. Q.E.D.

Proof of Theorem 2. It follows from Theorem 1 and the Liapunov Central Limit Theorem (which applies given (iii) and (v) of Assumption 2) that

$$(\bar{A}_n^{-1}\bar{B}_n\bar{A}_n^{-1})^{-1/2}\sqrt{n}(\hat{\beta}-\beta)=\bar{B}_n^{-1/2}u'\varepsilon/\sqrt{n}+o_p(1)$$
$$\xrightarrow{d} N(0, I_q).$$

To show the final conclusion it suffices to show that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ since $\bar{B}_n = \sigma^2\bar{A}_n$, hence

$$(\bar{A}_n^{-1}\bar{B}_n\bar{A}_n^{-1})^{-1/2}=\sigma^{-1}\bar{A}_n^{1/2}$$

and

$$\frac{\hat{\sigma}}{\sigma}(\bar{A}_n^{-1/2}(x'Mx/n)^{1/2}) \xrightarrow{p} I$$

by $\sigma^2>0$ and the proof of Theorem 1. Note that $n/(n-q-K) \to 1$, so that it suffices to show the result with n replacing $n-q-K$ in the denominator of $\hat{\sigma}^2$. Using the proof of Theorem 1 and the law of large numbers applied to $\varepsilon'\varepsilon/n$, it follows that

$$(y-x\beta)'\, M(y-x\beta)/n \xrightarrow{p} \sigma^2.$$

Thus it suffices to show that

$$(y-x\hat{\beta})'\, M(y-x\hat{\beta})/n-(y-x\beta)'\, M(y-x\beta)/n=o_p(1).$$

But this follows by using the result that

$$|\,\|a\|^2-\|b\|^2| \leqslant \|a-b\|^2+2\,\|a-b\|\,\|b\|,$$

where $\|\cdot\|$ is the Euclidean norm with distance matrix M, and the fact that

$$(\hat{\beta}-\beta)'\, x'Mx(\hat{\beta}-\beta)/n=o_p(1),$$

which follows from the proof of Theorem 1. Q.E.D.

Proof of Theorem 3. Note that

$$\hat{g}-g=Qh(\beta-\hat{\beta})+Qu(\beta-\hat{\beta})-Mg+Q\varepsilon.$$

The result follows from the triangle inquality and the following results. First,

$$\begin{aligned}&\frac{1}{n}(\beta-\hat{\beta})'\, hQh(\beta-\hat{\beta})\\ &\quad=O_p(n^{-1}+e_g(K)^2\, e_h(K)^2+e_g(K)^2\, n^{-1}+e_h(K)\, n^{-1}+K/n^2)\\ &\quad=O_p(K/n+e_g(K)^2)\end{aligned}$$

by $h'Qh/n$ bounded in probability and Theorem 1. Similarly,

$$\frac{1}{n}(\beta-\hat{\beta})' u'Qu(\beta-\hat{\beta}) = O_p(K/n + e_g(K)^2).$$

For the third term,

$$\frac{1}{n} g'Mg \leqslant e_g(K(n))^2,$$

and for the last term,

$$\frac{1}{n} \varepsilon' Q \varepsilon = O_p(K/n),$$

and the first result follows. The second result then follows trivially. Q.E.D.

References

[1] Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica* **59** 307–345.

[2] Bickel, P. J., and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya* **50** 381–393.

[3] Chamberlain, G. (1986). Notes on semiparametric regression. Mimeo, Department of Economics, Harvard University.

[4] Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16** 136–146.

[5] Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, A. (1984). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

[6] Eubank, R. L., Hart, J. D., and Speckman, P. (1990). Trigonometric series regression estimators with an application to partially linear models. *J. Multivariate Anal.* **32** 70–83.

[7] Eubank, R. L., and Speckman, P. (1990). Curve fitting by polynomial trigonometric regression. *Biometrika* **77** 1–9.

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

[9] Lorentz, G. G. (1986). *Approximation of Functions*. Chelsea, New York.

[10] Newey, W. K. (1991). *Series Estimation of Regression Functionals*. Working paper, Department of Economics, MIT.

[11] Rice, J. (1986). Convergence rates for partially splined estimates. *Statist. Probab. Lett.* **4** 203–208.

[12] Robinson, P. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954.

[13] Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.

[14] Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

[14a] Speckman, P. (1988). Kernel smoothing in partial linear models, *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

[15] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.