VARYING DEGREE POLYNOMIAL REGRESSION

II. ASYMPTOTIC NORMALITY*

(Abbreviated Title: VARYING DEGREE POLYNOMIAL REGRESSION)

by

Russell D. Wolfinger
SAS Institute Inc.

and

A. Ronald Gallant
North Carolina State University

November 1989**

*AMS 1980 subject classifications. Primary 62J05; secondary 62E20, 62G05, 62F12.
Key words and phrases. Regression, nonparametric regression, polynomial regression, asymptotic normality, strong approximation, M-estimator.

**This research was supported by National Science Foundation Grant SES-8808015, North Carolina Agricultural Experiment Station Projects NCO-5593, NCO-3879, and the NCSU PAMS Foundation.

SUMMARY

for

VARYING DEGREE POLYNOMIAL REGRESSION

II. ASYMPTOTIC NORMALITY

by

Russell D. Wolfinger and A. Ronald Gallant

For a regression model $y_t = g^o(x_t) + e_t$, we set a normal-type confidence interval around the k^{th} derivative of the unknown function g^o, evaluated at some preassigned point. A consistent point estimator is found by regressing on polynomial basis functions, allowing the degree to increase with sample size. Asymptotic normality of this estimator is then shown by using a strong approxmation result of Einmahl (1987). The estimator is viewed as the solution to some member of a wide class of optimization problems, including those associated with least squares, maximum likelihood, and M-estimators; an M-estimator is used to illustrate the main results.

1. Introduction. This paper is a continuation of Wolfinger and Gallant (1989), henceforth referred to as (RC). In (RC) we consider univariate data $\{y_t\}_{t=1}^{n}$ generated according to

$$y_t = g^o(x_t) + e_t \qquad 1 \le t \le n$$

where g^o is an unknown regression function possessing m derivatives, the x_t's are observed iid realizations from the beta(a,b) distribution (a and b are known), and the e_t's are unobserved realizations from some distribution $\mathcal{P}(e)$ and are independent of the x_t's. We consistently estimate g^o (in a norm including derivatives up to order $\ell < m$) by fitting the parametric model

$$y_t = \sum_{j=1}^{p_n} \theta_j^o \, \varphi_j(x) + e_t \qquad 1 \le t \le n$$

where $\varphi_j(\cdot)$ is a multiple of the j^{th} Jacobi polynomial, θ_j^o is the j^{th} parameter, and p_n is some increasing function of n satisfying $p_n \le n$. This follows Example 1 of Cox (1988) who calls this procedure varying degree polynomial regression. See (RC) for a motivation.

Our goal for this paper is to find a confidence interval for the evaluation functional

$$D_o^k(g^o) = (d^k/dx^k)g^o(x)\Big|_{x=x_o}$$

where $k \le \ell$ and x_o is some preassigned point in (0,1). This goal is essentially achieved for least squares estimators in Andrews (1988), who proves asymptotic normality for series estimators in semiparametric and his addditive interactive models. We extend his results to the class of least mean distance estimators, i.e. those that can be viewed as the solution to an optimization problem of the form

$$\underset{\theta \in \Theta_n}{\text{minimize:}} \quad (1/n) \sum_{t=1}^{n} s(y_t, x_t, \varphi_t' \theta)$$

where Θ_n is some subset of $\mathbb{R}^{p_n}$; φ_t is the p_n-vector with j^{th} element $\varphi_j(x_t)$, and $s(\cdot,\cdot,\cdot)$ is a suitable objective function. Examples of objective functions include those associated with least squares, maximum likelihood, and M-estimators. The theory for least mean distance estimators with p_n bounded can be found in Chapter 3 of Gallant (1987); we thus extend these results to the case where p_n is increasing with n. Our goal is to find constraints on p_n that yield asymptotic normality for this entire class.

We now describe the general framework and preliminary assumptions. We assume the true regression function, g^o, is m-times differentiable with $D^m g^o(x)$ absolutely continuous, where D^m is the m^{th} differentiation operator. We also assume that it has the expansion

$$g^o(x) = \sum_{j=1}^{\infty} \theta_j^o \, \varphi_j(x)$$

where θ_j^o is the generalized Fourier coefficients corresponding to the j^{th} basis function $\varphi_j(\cdot)$, as defined in (RC). Define a norm on g^o by

$$\|g^o\|_{2m}^2 = \sum_{j=1}^{\infty} j^{2m} \, (\theta_j^o)^2$$

which we assume is finite. This is the $\|\cdot\|_{2m}$ norm defined by Cox; it is one member of his scale of norms that vary with powers of j in the above expression. Cox identifies this norm as being close to a weighted Sobolev norm, and the consistency results in (RC) are proven in $\|\cdot\|_{2\ell}$. We also use the usual Euclidean norm $\|\cdot\|_0$, which we write as $\|\cdot\|$.

Define

$$\hat{\theta}_n = \underset{\theta \in \Theta_n}{\operatorname{argmin}} \; (1/n) \sum_{t=1}^{n} s(y_t, x_t, \varphi_t' \theta)$$

which we assume exists and is unique, and where

$$\Theta_n = \{\theta \in \mathbb{R}^{p_n} \mid \|\theta\|_{2\ell} \le B\}$$

$$\|\theta\|_{2\ell} = \sum_{j=1}^{p_n} j^{2\ell} \, \theta_j^2$$

and $B = \|g^o\|_{2\ell} + 1$. B is thus unknown, but in practice this is irrelevant because the choice of p_n from (RC) guarantees that $\hat{\theta}_n$ satisfies $\|\hat{\theta}_n\|_{2\ell} < B$ a.s. for n sufficiently large.

Our estimator of $D_o^k(g^o)$ is $\rho_o' \hat{\theta}_n$, where ρ_o is the p_n-vector with j^{th} element $D_o^k(\varphi_j)$. This estimator is consistent in a weighted $\mathcal{L}^2$ norm; see (RC). To show the asymptotic normality of $\rho_o' \hat{\theta}_n$, we use the *first order equations:*

$$\hat{\theta}_n = \theta_n^o - (1/n) \, \mathcal{J}_n^{-1} \sum_{t=1}^{n} \varphi_t \, X_{nt}$$

where θ_n^o is the p_n-vector with j^{th} element θ_j^o and

$$X_{nt} = (\partial/\partial g) s(y_t, x_t, g) \Big|_{g = \varphi_t' \theta_n^o}$$

Also

$$\mathcal{J}_n = (1/n) \sum_{t=1}^{n} \varphi_t \varphi_t' \, \tilde{w}_{nt}$$

is the *Hessian matrix* where

$$\tilde{w}_{nt} = (\partial^2/\partial g^2) s(y_t, x_t, g)\Big|_{g = \varphi_t' \tilde{\theta}_{nt}}$$

and

$$\tilde{\theta}_{nt} = \lambda_t \hat{\theta}_n + (1-\lambda_t)\theta_n^0$$

for some $\lambda_t \in [0,1]$. These equations are obtained from the first order conditions of the optimization problem (Luenberger, 1984), followed by a Taylor series expansion about θ_n^0. The first order conditions are free of Lagrange multipliers a.s. for n sufficiently large because of the consistency results of (RC). These equations correspond to the usual normal equations in the least squares case.

We assume that the largest derivative of interest, ℓ, satisfies

$$\ell \geq h + 1/2$$

where h = max(a,b,1/2), and recall a and b are the parameters of the beta distribution of the x_t's. Also, define the matrix Φ as being the $n \times p_n$ matrix with rows φ_t', and let

$$G_n = (1/n)\, \Phi'\Phi.$$

Note that G_n is nonsingular wp1, and we assume the existence of positive constants L and U such that

$$L \leq \lambda_{min}(G_n) \leq \lambda_{max}(G_n) \leq U$$

for all n, where $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ denote minimum and maximum eigenvalues, respectively.

Finally, we assume that the objective function s is real valued and has the form $s[Y(e,x),x,g(x)]$, where $Y(e,x) = g^o(x) + e$ and g is some function of x. To be reasonable, s should be some measure of distance between $g_0(x)$ and $g(x)$, and to avoid measurability problems we assume that it is continuous in all three of its arguments.

We now illustrate our main results with the example from (RC).

EXAMPLE (M-estimator). Use the objective function

$$s[Y(e,x),x,\varphi(x)'\theta] = \rho[g^o(x) + e - \varphi(x)'\theta]$$

where

$$\rho(u) = \log\cosh(u/2).$$

Define also

$$\Psi(u) = (d/du)\,\rho(u) = 1/2 \tanh(u/2).$$

Assume that the errors possess finite r^{th} moments, where $r > 2 + 1/\ell$, and that $\mathcal{E}\,\rho(e) = \mathcal{E}\,\Psi(e) = 0$. This would be satisfied if the error density is symmetric about zero. We set

$$\alpha = \min\left(\frac{\ell - (1+\gamma)(1+2\ell)/r}{h + 2\ell(1+2\ell)}, \frac{1}{4i + 4}\right) - \delta$$

for a sufficiently small $\gamma > 0$ and $\delta > 0$, where $i = \max(a+1,b+1,1/2)$. Assume that m is large enough to satisfy

$$\alpha \geq 1/(2m-2q-3) > 0$$

where $q = \max(2\ell+a-1, 2\ell+b-1, \ell-1/2)$.

Let $p_n \approx n^\alpha$, where the symbol $\approx$ means that the r.h.s. can be bounded above and below by constant multiples of the l.h.s. Then conditionally on

$\{x_t\}$,

$$\frac{\rho_o'\hat{\theta}_n - D_o^k(g^o)}{\sigma_n} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

where σ_n is the scale factor defined in Section 3. Our approximate confidence interval takes the form

$$\rho_o'\hat{\theta}_n \pm z_{\alpha/2}\,\hat{\sigma}_n$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}_n$ is an estimator of σ_n satisfying

$$|\hat{\sigma}_n - \sigma_n| \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

The remainder of the paper is organized as follows. In Section 2 we list our primary assumptions, and in Section 3 we state our main results. These are then applied to the example in Section 4, proving the claims made above. We give the proofs of the main theorems are given in Section 5, and then list some possible extensions of our work in Section 6. The Appendix contains the proofs of three lemmas stated in Section 3.

2. The assumptions. Assumptions S1-S6 and P1-P5 from (RC) are in force throughout the paper. We make the following additional assumptions, also grouped into two categories: the objective function (Assumptions S7-S10), and rate constraints (Assumptions P6-P8).

ASSUMPTION S7. The second derivative of the objective function (with respect to its third argument) evaluated at the true regression function can be factorized as

$$(\partial^2/\partial g^2)s[g^o(x) + e,x,g]\Big|_{g = g^o(x)} = v_1(x)\, w_1(e)$$

where $v_1(\cdot)$ and $w_1(\cdot)$ satisfy the following conditions:

There exist positive constants L_1 and U_1 such that

$$L_1 \le \inf_{x\in\mathcal{X}} v_1(x) \le \sup_{x\in\mathcal{X}} v_1(x) \le U_1$$

and if we define

$$d = \mathcal{E}_{\mathcal{P}}[w_1(e)]$$

then d exits and is finite and positive. The integral $\mathcal{E}_{\mathcal{P}}[w_1^4(e)]$ also exists and is finite, where $\mathcal{E}_{\mathcal{P}}$ represents integration with respect to $\mathcal{P}(e)$ only.

Also, we assume that the third derivative of the objective function (with respect to its third argument) exists and is bounded, i.e. there exists a constant C_3 such that

$$\sup_{0<n<\infty}\ \sup_{\theta\in\Theta_n}\ \sup_{e\in\mathcal{E}}\ \sup_{x\in\mathcal{X}} |(\partial^3/\partial g^3)s[g^o(x) + e,x,g]|\Big|_{g = \varphi(x)'\theta} \le C_3$$

where $\mathcal{E}$ is the support of $\mathcal{P}(e)$ and $\mathcal{X} = (0,1)$.

For the next assumption, define

$$g_n^o(x) = \sum_{j=1}^{p_n} \theta_j^o \varphi_j(x)$$

as the truncated counterpart of g^o.

ASSUMPTION S8. Define the following function:

$$X_n(e,x) = (\partial/\partial g) s[g^o(x) + e,x,g]\Big|_{g = g_n^o(x)}.$$

Then

$$\sup_{0<n<\infty} \sup_{x\in\mathcal{X}} \mathcal{E}_{\mathcal{P}}[|X_n(e,x)|^8] < \infty.$$

ASSUMPTION S9. The first derivative of the objective function (with respect to its third argument) evaluated at the true regression function can be factorized as

$$(\partial/\partial g) s[g^o(x) + e,x,g]\Big|_{g = g^o(x)} = v_0(x)\, w_0(e)$$

where $v_0(\cdot)$ and $w_0(\cdot)$ satisfy the following conditions:

There exist positive constants L_0 and U_0 such that

$$L_0 \le \inf_{x\in\mathcal{X}} v_0(x) \le \sup_{x\in\mathcal{X}} v_0(x) \le U_0,$$

$$\mathcal{E}_{\mathcal{P}}[w_0(e)] = 0,$$

and if we define

$$\sigma_0^2 = \mathcal{E}_{\mathcal{P}}[w_0^2(e)]$$

then σ_0^2 exits and is finite and positive.

ASSUMPTION S10. The functions $w_0(\cdot)$ and $w_1(\cdot)$ as defined in Assumptions S9 and S7 respectively possess bounded first derivatives.

REMARK 2.1. Assumption S7 is used to invert the Hessian matrix. Assumption S8 allows us to invoke the strong approximation result of Einmahl (1987) with the best possible rate. Assumptions S9 and S10 help us to estimate the scale factor σ_n.

For our first two rate assumptions, we define

$$B(p) = \sum_{j=1}^{p} \sup_{x\in\mathcal{X}} \varphi_j^2(x)$$

$$C(p) = \sum_{j=1}^{p} [\sup |\psi_j(x)|]^2$$

where $\psi_j(x) = (d/dx)\varphi_j(x)$.

ASSUMPTION P6. The truncation point p_n satisfies

$$\lim_{n\to\infty} \frac{p_n^4 B(p_n)}{n} = 0.$$

ASSUMPTION P7. The truncation point p_n satisfies $C(p_n) = O(n^{1/2-\gamma})$ for some $\gamma > 0$.

ASSUMPTION P8. m is large enough to satisfy

$$\alpha \geq 1/(2m-2q-3) > 0$$

where $p_n \approx n^{\alpha}$ and $q = \max(2\ell+a-1, 2\ell+b-1, \ell-1/2)$.

REMARK 2.2. By results from Szego (1975) on Jacobi polynomials and from our choice of normalizing constants [see (RC)]

$$\sup_{x\in\mathcal{X}} |\varphi_j(x)| = O(j^{h-1/2})$$

$$\sup_{x\in\mathcal{X}} |(d/dx)\varphi_j(x)| = O(j^{i+1/2}).$$

These can be used to show that

$$B(p) = O(p^{2h}) \quad \text{as} \quad p \to \infty$$

$$C(p) = O[p^{2(i+1)}] \quad \text{as} \quad p \to \infty.$$

An alternative procedure would be to use basis functions that have $B(\cdot)$ and $C(\cdot)$ bounded; however, one would then have to manage declining eigenvalues in the regression matrix. See Gallant (1989) for such an approach. Nonetheless, in terms of α, Assumptions P6 and P7 are respectively equivalent to

$$\alpha < 1/(2h+4)$$

$$\alpha < 1/(4i+4).$$

Since $h \leq i$, Assumption P7 thus implies Assumption P6. Assumption P8 implies Assumption P4 of (RC) because $h \leq q$.

3. Statement of the main results. We define some more notation. Recall the Hessian matrix

$$\mathcal{J}_n = (1/n) \sum_{t=1}^{n} \varphi_t \varphi_t' \tilde{w}_{nt}$$

and by Assumption S7 we can write

$$\tilde{w}_{nt} = v_1(x_t)\, w_1(e_t) + [g^o(x_t) - \varphi_t'\vartheta_{nt}](\partial^3/\partial g^3) s(y_t, x_t, g)\Big|_{g = \tilde{g}_n(x_t)}$$

where $\tilde{g}_n(x_t)$ is between $g^o(x_t)$ and $\varphi_t'\vartheta_{nt}$. Let V_1 be the n x n diagonal matrix with elements $\{v_1(x_t)\}$, let W_1 be the n x n diagonal matrix with elements $\{w_1(e_t)\}$, and let W_2 be the n x n diagonal matrix with elements

$$\{[g^o(x_t) - \varphi_t'\vartheta_n](\partial^3/\partial g^3) s(y_t, x_t, g)\Big|_{g = \tilde{g}_n(x_t)}\}.$$

We can thus write

$$\begin{aligned}\mathcal{J}_n &= (1/n)\, [d\Phi' V_1 \Phi + \Phi'(V_1 W_1 - dV_1)\Phi + \Phi' W_2 \Phi] \\ &= (1/n)\, (d\Phi' V_1 \Phi)^{1/2}\, [I - T_n]\, (d\Phi' V_1 \Phi)^{1/2}\end{aligned}$$

where $(d\Phi' V_1 \Phi)^{1/2}$ is the Cholesky factor of $d\Phi' V_1 \Phi$, I is the p_n x p_n identity matrix,

$$T_n = H'(dV_1 - V_1 W_1)H - H' W_2 H$$

$$H = \Phi(d\Phi' V_1 \Phi)^{-1/2}.$$

LEMMA 3.1 *Under Assumptions S1-S7 and P1-P6,*

$$|\lambda_{max}(T_n)| \;=\; o_p[\frac{p_n^3\ B(p_n)}{n}]^{1/2}.$$

PROOF. The proof follows from Lemma 1 of Severini and Wong (1987). See the Appendix. ∎

COROLLARY. *Under the assumptions of Lemma 3.1, define*

$$S_n \;=\; \sum_{\ell=1}^{\infty} T_n^{\ell}.$$

Then

$$|\lambda_{max}(S_n)| \;=\; o_p[\frac{p_n^3\ B(p_n)}{n}]^{1/2}.$$

PROOF. The proof follows from the dominating first term of S_n. ∎

REMARK 3.1. If T is a real symmetric matrix with $|\lambda_{max}(T)| < 1$, then we have expansion

$$(I - T)^{-1} \;=\; \sum_{\ell=0}^{\infty} T^{\ell}$$

where convergence is with respect to the operator norm (Kreyzig, 1978). Lemma 3.1 thus allows us to invert the Hessian matrix and write

$$\begin{aligned}\mathcal{J}_n^{-1} &= n\ (d\Phi' V_1 \Phi)^{-1/2}\ (\sum_{\ell=0}^{\infty} T_n^{\ell})\ (d\Phi' V_1 \Phi)^{-1/2} \\ &= n\ (d\Phi' V_1 \Phi)^{-1} \;+\; n\ R_n\end{aligned}$$

where

$$R_n = (d\Phi' V_1 \Phi)^{-1/2} S_n (d\Phi' V_1 \Phi)^{-1/2}.$$

LEMMA 3.2 *Under Assumptions S1-S7 and P1-P6, let ρ be an arbitrary vector of length p_n. Then*

$$|\rho' R_n \sum_{t=1}^{n} \varphi_t X_{nt}| = \|\rho' (d\Phi' V_1 \Phi)^{-1/2}\| \; o_p[\frac{p_n^2 \sqrt{B(p_n)}}{\sqrt{n}}].$$

PROOF. Define X_n to be the n-vector $\{X_{nt}\}_{t=1}^{n}$. Then the first order equations are

$$\begin{aligned} \Phi' X_n &= n \mathcal{J}_n (\hat{\theta}_n - \theta_n^o) \\ &= (d\Phi' V_1 \Phi)^{1/2} (I - T_n) (d\Phi' V_1 \Phi)^{1/2} (\hat{\theta}_n - \theta_n^o). \end{aligned}$$

Using these equations, the Cauchy-Schwartz inequality, and the definition of λ_{max} we have

$$\begin{aligned} &|\rho' R_n \sum_{t=1}^{n} \varphi_t X_{nt}| \\ = \;& |\rho' (d\Phi' V_1 \Phi)^{-1/2} S_n (d\Phi' V_1 \Phi)^{-1/2} \Phi' X_n| \\ \leq \;& \|\rho' (\Phi' V_1 \Phi)^{-1/2}\| \; |\lambda_{max}[S_n (I - T_n)]| \; |\lambda_{max}^{1/2}(V_1)| \; \|\Phi(\hat{\theta}_n - \theta_n^o)\| \\ = \;& \|\rho' (\Phi' V_1 \Phi)^{-1/2}\| \; o_p[p_n^{3/2} \sqrt{B(p_n)}/\sqrt{n}] \; O(1) \; O_p(\sqrt{p_n}) \end{aligned}$$

by the Corollary to Lemma 3.1 and Theorem 3.7 of (RC). ∎

We now give a strong approximation result. First, note that if we ignore R_n above, then the first order equations reduce to

$$\hat{\theta}_n = \theta_n^o - (d\Phi' V_1 \Phi)^{-1} \sum_{t=1}^{n} \varphi_t X_{nt}$$

where recall that

$$X_{nt} = (\partial/\partial g) s(y_t, x_t, g)\Big|_{g = g_n^o(x_t)}$$

Subtracting the expectation of this expression with respect to $\mathscr{P}(e)$ and premultiplying by ρ' yields

$$\rho'\hat{\theta}_n - \mu_n = \rho'(d\Phi' V_1 \Phi)^{-1} \sum_{t=1}^{n} \varphi_t \xi_{nt}$$

where

$$\xi_{nt} = X_{nt} - \mathscr{E}_{\mathscr{P}}(X_{nt})$$

$$\mu_n = \rho'\theta_n^o - \rho'(d\Phi' V_1 \Phi)^{-1} \sum_{t=1}^{n} \varphi_t \mathscr{E}_{\mathscr{P}}(X_{nt}).$$

Note that $\{\xi_{nt}: t = 1,..,n;\ n = 1,...\}$ is a triangular array of rowwise independent random variables, with each ξ_{nt} having a zero mean and a variance s_{nt}^2, say, that depends upon $g_n^o(x_t)$. Using common probability space methods, we can construct a version of the array $\{\xi_{nt}\}$, call it $\{Y_{nt}\}$, on a new probability space. By a version we mean that $\{Y_{nt}\}$ is also rowwise independent and that Y_{nt} has exactly the same distribution as ξ_{nt} for every n and t. On this new space we can also construct a triangular array of $N(0, s_{nt}^2)$ random variables, call it $\{Z_{nt}\}$, that is close to $\{Y_{nt}\}$ in the following sense.

LEMMA 3.3 Let $\{\xi_{nt}\}$, $\{Y_{nt}\}$, and $\{Z_{nt}\}$ be as above and define

$$S_n = \Sigma_{t=1}^n Y_{nt} \qquad T_n = \Sigma_{t=1}^n Z_{nt}.$$

Assume that

$$\sup_{0<n<\infty} \max_{1\le t\le n} \mathscr{E}_{\mathscr{P}}[|\xi_{nt}|^s] < \infty$$

for some $2 < s \le 8$. Then a construction is possible such that

$$\max_{1\le t\le n} |S_t - T_t| = o(n^{2/s + \gamma}) \quad \text{a.s.}$$

where γ is any positive real number, and the a.s. is with respect to the new probability measure.

PROOF. The proof follows from Proposition 1 of Einmahl (1987). See the Appendix. ∎

LEMMA 3.4. *Under Assumptions S7 and S9, there exists a sequence $\{r_n\}$ of real numbers that converge to zero and satisfy*

$$\mathrm{Var}_{\mathscr{P}} X_n(e,x) = v_0^2(x)\,\sigma_0^2 + r_n^2$$

for all n, where $X_n(e,x)$ is defined in Assumption S8.

PROOF. By Taylor's theorem and Assumptions S7 and S9

$$\begin{aligned} X_n(e,x) &= (\partial/\partial g)s[g^0(x) + e,x,g]\Big|_{g = g_n^0(x)} \\ &= v_0(x)\,w_0(e) + v_1(x)\,w_1(e)\,[g^0(x) - g_n^0(x)] \\ &\quad + [g^0(x) - g_n^0(x)]^2\,(\partial^3/\partial g^3)s[g^0(x) + e,x,g]\Big|_{g = \bar{g}_n(x)} \end{aligned}$$

where $\bar{g}_n(x)$ is between $g^0(x)$ and $g_n^0(x)$. The lemma follows from

Assumptions S7 and S9 and the fact that

$$\sup_{x\in\mathcal{X}} [g^{o}(x) - g_n^{o}(x)] = o(1)$$

[see Lemma 3.5 of (RC)]. ∎

Now define

$$\sigma_n^2 = \mathrm{Var}_{\mathcal{P}}\{ \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t \; Z_{nt}\}$$

$$= \rho'(d\Phi'V_1\Phi)^{-1} [\sum_{t=1}^{n} \varphi_t \varphi_t' s_{nt}^2] (d\Phi'V_1\Phi)^{-1}\rho.$$

REMARK 3.2. By Lemma 3.4,

$$\sigma_n^2 = (\sigma_0^2/d^2) \rho'(\Phi'V_1\Phi)^{-1} (\Phi'V_0\Phi) (\Phi'V_1\Phi)^{-1}\rho + r_n^2 \rho'(\Phi'V_1\Phi)^{-1} (\Phi'\Phi) (\Phi'V_1\Phi)^{-1}\rho$$

$$= (\sigma_0^2/d^2) \rho'(\Phi'V_1\Phi)^{-1} (\Phi'V_0\Phi) (\Phi'V_1\Phi)^{-1}\rho \, [1 + o(1)].$$

where V_0 is the $n \times n$ diagonal matrix with elements $\{v_0^2(x_t)\}$. Note that σ_n^2 should therefore decline at the same rate as $(1/n)\rho'(G_n)^{-1}\rho$ because of Assumptions S7 and S9. Also note that to obtain a consistent estimate of σ_n^2, one only needs consistent estimates of σ_0 and d; these are found in Theorem 3.8 below.

LEMMA 3.5. *Under Assumptions S7-S9 and P7,*

$$(1/\sigma_n) \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t (Y_{nt} - Z_{nt}) \longrightarrow 0 \quad \text{a.s. as} \quad n \to \infty.$$

PROOF. See the Appendix. ∎

THEOREM 3.6. *Under Assumptions S1-S9 and P1-P7,*

$$\frac{\rho'\hat{\theta}_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty$$

where ρ is an arbitrary vector of length $\mathbb{R}^{p_n}$ and

$$\mu_n = \rho'\theta_n^o - \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t \, \mathcal{E}_{\mathcal{P}}(X_{nt}).$$

THEOREM 3.7. *Under Assumptions S1-S10 and P1-P8*

$$\frac{\mu_n - D_o^k(g^o)}{\sigma_n} \longrightarrow 0 \qquad \text{as } n \to \infty.$$

where μ_n and σ_n are defined as above with $\rho = \rho_o$.

THEOREM 3.8. *Under Assumptions S1-S10 and P1-P5, define*

$$\hat{\sigma}_{0n}^2 = (1/n) \sum_{t=1}^{n} w_0^2(\hat{e}_{nt}) \qquad\qquad \hat{d}_n = (1/n) \sum_{t=1}^{n} w_1(\hat{e}_{nt})$$

where $\hat{e}_{nt} = y_t - \varphi_t'\hat{\theta}_n$. Then as $n \to \infty$,

$$\hat{\sigma}_{0n}^2 \xrightarrow{p} \sigma_0^2 \qquad\qquad \hat{d}_n \xrightarrow{p} d.$$

REMARK 3.3. Our final confidence interval takes the form

$$\rho'\hat{\theta}_n \pm z_{\alpha/2} \, (\hat{\sigma}_0/\hat{d}) \, \surd\{\rho'(\Phi'V_1\Phi)^{-1} \, (\Phi'V_0\Phi) \, (\Phi'V_1\Phi)^{-1}\rho\}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. An alternative to this type of interval is the regression percentile method used by Efron (1989).

4. Application to the example. We first verify Assumptions S7-S10 for the M-estimator. For Assumption S7

$$(\partial^2/\partial g^2)s[g^o(x) + e,x,g]\Big|_{g = g^o(x)} = 1/4\ \mathrm{sech}^2(e/2)$$

so we choose $v_1(x) = 1/4$ for all $x \in \mathcal{X}$ and $w_1(e) = \mathrm{sech}^2(e/2)$. Note that

$$d = \mathcal{E}_{\mathcal{P}}\mathrm{sech}^2(e/2)$$

exits and is finite by nature $\mathrm{sech}^2(\cdot)$ and we assume that the error distribution is such that d is also positive. The integral $\mathcal{E}_{\mathcal{P}}[w_1^4(e)]$ also exists and is finite. Finally

$$\begin{aligned}(\partial^3/\partial g^3)s[g^o(x) + e,x,g] &= -1/8\ \mathrm{sech}^2\{[e+g^o(x)-g]/2\}\\ &\qquad \times \tanh\{[e+g^o(x)-g]/2\}\end{aligned}$$

and C_3 can be taken to be 1/8.

For Assumption S8

$$X_n(e,x) = -\Psi[g^o(x) - g_n^o(x) + e]$$

which is bounded for all n and x because $|\tanh(\cdot)|$ is bounded above by 1.

For Assumption S9

$$\begin{aligned}(\partial/\partial g)s[g^o(x) + e,x,g]\Big|_{g = g^o(x)} &= -\Psi(e)\\ &= -1/2\ \tanh(e/2)\end{aligned}$$

So $v_0(x)$ can be taken to be 1/2 for all x and $w_0(e) = -\Psi(e)$, which we assume has mean zero. We also have

$$\sigma_0^2 = \mathcal{E}_{\mathcal{P}} \tanh^2(e/2)$$

which we assume is positive.

For Assumption S10

$$w_0(e) = -\tanh(e/2) \qquad w_1(e) = \operatorname{sech}^2(e/2),$$

both of which also have bounded first derivatives.

REMARK 4.1. Note that in Assumptions S7 and S9, $v_1(x)$ and $v_0(x)$ do not depend on x. This is also the case with the least squares objective function (which also satisfies Assumptions S7-S10). We include this generality to make for an easier extension to more complex scenarios.

For the rate constraints, note that our choice of α and m satisfy Assumptions P1-P5 of (RC), and thus the consistency results of that paper apply. Since we make Assumption P8 outright, the only new constraint is Assumption P7 (see Remark 2.2), which is satisfied by using the $\min(\cdot,\cdot)$.

5. Proof of main theorems. In this section are given the proofs of the theorems stated in Section 3. To summarize, Theorem 3.6 is our asymptotic normality result, conditional on $\{x_t\}$. It utilizes Lemmas 3.1-3.2, which invert the Hessian matrix, and Lemmas 3.3-3.5, which borrow from a strong approximation result of Einmahl (1987). Theorem 3.7 shows that our standardized bias, $[\mu_n - D_o^k(g^o)]/\sigma_n$, converges to zero; it makes use of the lower bound on the growth rate of p_n imposed by Assumption P8. Theorem 3.8 constructs consistent estimates of d and σ by using an empirical integration of residuals. The factorizations in Assumptions S7 and S9 make the above analyses tractable.

PROOF OF THEOREM 3.6. By the first order equations we have

$$\frac{\rho'\hat{\theta}_n - \mu_n}{\sigma_n} = (1/\sigma_n)\ \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t\ \xi_{nt}$$

$$+\ (1/\sigma_n)\ \rho' R_n \sum_{t=1}^{n} \varphi_t\ X_{nt}.$$

Using the common probability space methods of Lemma 3.3, the first term on the r.h.s. has the same distribution as

$$(1/\sigma_n)\ \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t\ Y_{nt} = N(0,1) + \rho'(d\Phi'V_1\Phi)^{-1} \sum_{t=1}^{n} \varphi_t\ (Y_{nt} - Z_{nt}).$$

By Lemma 3.5 the last term in this expression is o(1) a.s. For the second term on the r.h.s., recall that in the proof of Lemma 3.5 we derived the following lower bound for σ_n using Assumptions S7 and S9:

$$\sigma_n \geq (\sigma_0/d)\ [1 + o(1)]\ \|\rho'(\Phi'V_1\Phi)^{-1/2}\| \times \lambda_{min}[(\Phi'V_1\Phi)^{-1/2}\ (\Phi'V_0\Phi)\ (\Phi'V_1\Phi)^{-1/2}]$$

$$\geq (\sigma_0/d)\ [1 + o(1)]\ \|\rho'(\Phi'V_1\Phi)^{-1/2}\|\ \ (U\ U_1)/(L\ L_0^2)$$

Using this along with Lemma 3.2 we have

$$|(1/\sigma_n)\ \rho' R_n \sum_{t=1}^{n} \varphi_t\ X_{nt}| \leq (d/\sigma_0)\ \frac{\|\rho'(\Phi'V_1\Phi)^{-1/2}\|}{\|\rho'(\Phi'V_1\Phi)^{-1/2}\|}\ o_p[\frac{p_n^2\ \sqrt{B(p_n)}}{\sqrt{n}}]$$

which is $o_p(1)$ by Assumption P6. Slutsky's theorem (see Serfling, 1980) yields the desired result. ∎

REMARK 5.1. Note that the N(0,1) distribution does not depend on the sequence $\{x_t\}$ upon which we have been conditioning. This implies that

$$\frac{\rho'\hat{\theta}_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty$$

unconditionally as well.

REMARK 5.2. Our proof strategy for asymptotic normality is quite different from that used in say Portnoy (1985) or Eubank (1988). These works make direct use of iid-type central limit theorems, while we use the independent-non-identically distributed result of Einmahl. Einmahl's work is an extension of iid strong approximation results such as those found in Csorgo and Revesz (1981).

PROOF OF THEOREM 3.7. We can write $[\mu_n - D_o^k(g^o)]/\sigma_n$ as

$$(1/\sigma_n) \sum_{j=p_n+1}^{\infty} \theta_j^o D_o^k(\varphi_j) \quad + \quad (1/\sigma_n)\, \rho'(d\Phi' V_1 \Phi)^{-1} \sum_{t=1}^{n} \varphi_t\, \mathcal{E}_{\mathcal{P}}(X_{nt}).$$

Using the lower bound for σ_n given in the proof of Theorem 3.6. and Cauchy-Schwartz, the absolute value of the second term is less than or equal to

$$C\, \|(d\Phi' V_1 \Phi)^{-1/2} \sum_{t=1}^{n} \varphi_t\, \mathcal{E}_{\mathcal{P}}(X_{nt})\|$$

where C is a constant independent of n. By the argument used in the last part of the proof of Lemma 3.6 of (RC), this term is o(1). As for the first term above, use the same lower bound for σ_n to conclude that

$$(1/\sigma_n) \quad = \quad O(\sqrt{n})$$

provided that $\rho'\rho$ does not converge to zero (it should not converge to zero because the differentiation operator invokes multiplication by powers of j). In fact, $\rho'\rho$ may go to infinity with n, but no faster than p_n^{2q+2} where $q = \max(2\ell+a-1, 2\ell+b-1, \ell-1/2)$, because

$$\sup_{x\in\mathcal{X}} |D^k\varphi_j(x)| \quad = \quad O(j^{q+1/2}).$$

This result along with

$$\theta_j^o \quad = \quad o(j^{-m})$$

shows that the first term satisfies

$$|(1/\sigma_n)\sum_{j=p_n+1}^{\infty} \theta_j^o\, D_o^k(\varphi_j)| \;=\; o(\surd n\; p_n^{-m+q+3/2})$$

which is o(1) by Assumption P8. ▌

REMARK 5.3. Theorem 3.7 shows that our standardized bias converges to zero with the only new assumption being the lower bound on p_n given in Assumption P8. This conflicts with many of the spline and kernel results in which a different but similar standardized bias converges to some positive constant such as 1/4. If one really wanted to match such results, one could fix m and then choose p_n so that the first term in the proof of Theorem 5 above converges to some constant. This approach seems to deviate from our goal of setting a confidence interval on $D_0^k(g^o)$. We thus make Assumption P8 and proceed with a straightforward normal-type confidence interval.

PROOF OF THEOREM 3.8. By Taylor's theorem and Assumption S10

$$w_0(\hat{e}_{nt}) \;=\; w_0(e_t) + [g^o(x_t) - \hat{g}_n(x_t)]\; c_{nt}$$

where

$$c_{nt} \;=\; (d/de)w_0(e)\Big|_{e\,=\,\tilde{e}_{nt}}$$

and $\tilde{e}_{nt}$ is some number between e_t and $[e_t + g^o(x_t) - \hat{g}_n(x_t)]$. Squaring this expression and summing over t gives

$$\hat{\sigma}_{0n}^2 \;=\; (1/n)\sum_{t=1}^{n} w_0^2(e_t) \;+\; (1/n)\sum_{t=1}^{n}[g^o(x_t) - \hat{g}_n(x_t)]^2\, c_{nt}^2$$

$$+\; (2/n)\sum_{t=1}^{n} w_0(e_t)\;[g^o(x_t) - \hat{g}_n(x_t)]\; c_{nt}.$$

The first term on the r.h.s. converges in probability to σ_0^2 by the weak law of large numbers. The second term is $O_p(p_n/n)$ by Theorem 3.7 of (RC) and the boundedness of c_{nt}. The final term is $O_p(\sqrt{p_n}/\sqrt{n})$ by Cauchy-Schwartz. A similar argument yields the consistency of $\hat{d}_n$. ∎

6. Possible extensions. We conclude by discussing some possibilities for further research. First, our assumption that the x_t's follow a beta distribution could be generalized; see Remark 1.1 of (RC). Next, we assume that p_n is a deterministic function of n. The empirical evidence in Eastwood (1987) suggests that it may be more appropriate to let p_n depend upon the data. He develops techniques to make the extension from deterministic to adaptive truncation rules, and they appear to be applicable to our setting. The fact that we let p_n increase like a power of n is in contrast to Gallant (1989), in which p_n may be required to increase only at a logarithmic rate. In that case, he also must require g^o to be infinitely differentiable in order for the bias terms to converge to zero. If one is willing to make this assumption though, our techniques should be extendable to the Fourier flexible form basis functions (which consist of constant, linear, and quadratic terms, and then additional terms of the form sin(kx) and cos(kx), k = 1,2,3,...).

Another extension of interest would be the inclusion of a nuisance parameter in the objective function, as is done in Chapter 3 of Gallant (1988). This would allow one to employ such methods as iteratively rescaled M-estimation and two and three stage least squares. The generalization to the case of multivariate dynamic models in Chapter 7 of Gallant would also be useful.

Though we only consider the linear functional $D_0^k(\cdot)$, one might be interested in setting a confidence interval around $h(g_0)$, where $h(\cdot)$ is some nonlinear functional. This could probably be most easily accomplished by another application of Taylor's theorem and the imposition of regularity condidtions on $h(\cdot)$. Our method of strong approximation seems suited to such a generalization. The common probability space approach is also useful in obtaining a Gaussian approximation to a relevant stochastic process (see Cox, 1984).

Finally, a general theory of selection procedures that encompasses our problem and the associated kernel and spline problems would be desirable.

APPENDIX

In this Appendix we give the proofs of Lemmas 3.1, 3.3, and 3.5.

PROOF OF LEMMA 3.1. Note that

$$|\lambda_{max}(T_n)| \le \frac{(1/n)\ |\lambda_{max}[\Phi'(dV_1 - V_1W_1)\Phi - \Phi'W_2\Phi]|}{(1/n)\ \lambda_{min}(d\Phi'V_1\Phi)}$$

$$\le \frac{(1/n)\ |\lambda_{max}[\Phi'(dV_1 - V_1W_1)\Phi - \Phi'W_2\Phi]|}{d\ U\ U_1}.$$

by Assumption S7. By Cauchy-Schwartz

$$(1/n)\ \lambda_{max}[\Phi'W_2\Phi] = \sup_{\|\xi\|=1} (1/n) \sum_{t=1}^{n} (\varphi_t'\xi)^2\ w_2(e_t, x_t, \vartheta_n)$$

$$\le \sup_{\|\xi\|=1} (1/n)\ [\sum_{t=1}^{n} (\varphi_t'\xi)^4]^{1/2}\ [\sum_{t=1}^{n} w_2^2(e_t, x_t, \vartheta_n)]^{1/2}$$

$$\le [\lambda_{max}(G_n)\ B(p_n)/n]^{1/2}\ C_3\ [\sum_{t=1}^{n} (g^o(x_t) - \varphi_t'\vartheta_n)^2]^{1/2}$$

by the final part of Assumption S7. The sum in this expression equals

$$\sum_{t=1}^{n} [g^o(x_t) - g_n^o(x_t)]^2 + \lambda^2 \sum_{t=1}^{n} [g_n^o(x_t) - \hat{g}_n(x_t)]^2$$

$$- 2\lambda \sum_{t=1}^{n} [g^o(x_t) - g_n^o(x_t)]\ [g_n^o(x_t) - \hat{g}_n(x_t)]$$

where λ is some number between 0 and 1. By Lemma 4 and Assumption P4 the first term in this expression is $O(p_n)$ and the second term is $O_p(p_n)$ by Theorem 3 of Section 3.4. The cross-product term is $O_p(p_n)$ by Cauchy-Schwartz. Therefore

$$(1/n)\ \lambda_{max}[\Phi' W_2 \Phi] = O_p\{\surd[p_n\ B(p_n)/n]\}$$
$$= o_p[\frac{p_n^3\ B(p_n)}{n}]^{1/2}.$$

We now show that

$$|(1/n)\ \lambda_{max}[\Phi'(dV_1 - V_1 W_1)\Phi]| = o_p[\frac{p_n^3\ B(p_n)}{n}]^{1/2}$$

using Lemma 1 of Severini and Wong (1987), which we state now as our Lemma A.1.

LEMMA A.1. Let $(f, \{\Xi_n\}, d)$ satisfy condition (L) with constants $\{M_n\}$ and let $(f, \{\Xi_n\})$ satisfy condition (W). If for some sequence of constants $\{\epsilon_n\}$

$$A_n H(\epsilon\epsilon_n / 2M_n, \Xi_n, d) = o(n\epsilon\epsilon_n^2)$$

for every $\epsilon > 0$, then

$$\sup_{\xi\in\Xi_n} |\mathscr{E}_n f(e,x;\xi) - \mathscr{E}\ f(e,x;\xi)| = o_p(\epsilon_n). \quad \blacksquare$$

Conditions (L) and (W) are given below when we verify them, and recall that $H(\cdot,\cdot,\cdot)$ is the metric entropy function as defined in (RC). In order to make use of this result, note that

$$(1/n)\ \lambda_{max}[\Phi'(dV_1 - V_1 W_1)\Phi] = \sup_{\|\xi\|=1} (1/n) \sum_{t=1}^{n} (\varphi_t'\xi)^2\ v_1(x_t)\ [d - w_1(e_t)]$$

$$= \sup_{\xi\in\Xi_n} \mathscr{E}_n f(e,x;\xi)$$

where we have defined

$$f(e,x;\xi) = [\varphi(x)'\xi]^2 \, v_1(x) \, [d - w_1(e)]$$

and

$$\Xi_n = \{\xi \in \mathbb{R}^{p_n} \mid \xi'\xi = 1\}.$$

Note further that

$$|(1/n)\, \lambda_{max}[\Phi'(dV_1 - V_1W_1)\Phi]| \le \sup_{\xi\in\Xi_n} |\mathscr{E}_n f(e,x;\xi) - \mathscr{E} f(e,x;\xi)|$$

because $\mathscr{E} f(e,x;\xi) = 0$ and we have interchanged the absolute value and the supremum. So the desired result will follow provided we can verify conditions (L) and (W) and can choose

$$\epsilon_n = [\frac{p_n^3 \, B(p_n)}{n}]^{1/2}.$$

We now verify the two conditions.

Condition (L). There exists a sequence of real-valued Borel-measurable functions $\{m_n(\cdot,\cdot)\}$ such that

$$|f(e,x;\xi_1) - f(e,x;\xi_2)| \le m_n(e,x)\, d(\xi_1,\xi_2)$$

for all e, x, and for all $\xi_1,\xi_2 \in \Xi_n$, where

(i) $\mathscr{E}\, m_n(e,x) \le M_n < \infty$ for all n

(ii) $\mathrm{Var}\; m_n(e,x) = o(nM_n^2)$ as $n \to \infty$. ∎

Note that for our case (using Cauchy-Schwartz)

$$|f(e,x;\xi_1) - f(e,x;\xi_2)| \le |v_1(x)|\, |d - w_1(e)|\, \|\varphi(x)\|^2 \, (\|\xi_1\| + \|\xi_2\|) \times \|\xi_1 - \xi_2\|,$$

so we can choose $M_n = O[B(p_n)]$ for part (i) and part (ii) obtains easily.

Condition (W). There exists a sequence of constants $\{A_n\}$ such that

(i) $A_n = \mathcal{E} \sup_{\xi \in \Xi_n} f^2(e,x;\xi) < \infty$ for all n

(ii) $\mathrm{Var} \sup_{\xi \in \Xi_n} f^2(e,x;\xi) = o(nA_n^2)$ as $n \to \infty$. ∎

For our case (again using Cauchy-Schwartz)

$$A_n = \mathcal{E} \sup_{\|\xi\|=1} [\varphi(x)'\xi]^4 \, v_1^2(x) \, [d - w_1(e)]^2$$

$$\leq U_1^2 \, \mathrm{Var}_{\mathcal{P}}[w_1(e)] \;\; \mathcal{E} \, [\sum_{j=1}^{p_n} \varphi_j^2(x)]^2$$

$$\leq U_1^2 \, \mathrm{Var}_{\mathcal{P}}[w_1(e)] \, B(p_n) \, p_n$$

where we have used Assumption S7 and the orthonormal property of the basis functions. By a similar argument

$$\mathrm{Var} \sup_{\|\xi\|=1} [\varphi(x)'\xi]^4 \, v_1^2(x) \, [d - w_1(e)]^2$$

$$\leq \mathcal{E} \, \{ \sup_{\|\xi\|=1} [\varphi(x)'\xi]^4 \, v_1^2(x) \, [d - w_1(e)]^2 \}^2$$

$$\leq U_1^4 \, \mathcal{E}_{\mathcal{P}}[w_1^4(e)] \, B^3(p_n) \, p_n$$

It is sufficient to verify condition (ii) for an upper bound for A_n rather than for A_n itself. Using the upper bound derived above, this reduces to requiring that $B(p_n)/np_n \to 0$ as $n \to \infty$, which is true by Assumption P6. We thus have verified conditions (L) and (W), and we now proceed to showing that

$$A_n H(\epsilon\epsilon_n/2M_n, \Xi_n, \|\cdot\|) = o(n\epsilon\epsilon_n^2)$$

for every $\epsilon > 0$, where

$$\epsilon_n = \left[\frac{p_n^3 \; B(p_n)}{n}\right]^{1/2}.$$

By Gallant (1989)

$$H(\epsilon, \Xi_n, \|\cdot\|) = \log(2p_n) + (p_n - 1)\log(2/\epsilon + 1).$$

Thus by choosing $A_n = O[B(p_n)\; p_n]$ and $M_n = O[B(p_n)]$ we have for some constant c_1

$$A_n H(\epsilon_n/2M_n, \Xi_n, \|\cdot\|) \leq c_1 \; B(p_n) \; p_n \; \{\log(2p_n) + p_n \log[4B(p_n)/\epsilon_n + 1]\}$$

$$\leq c_2 \; B(p_n) \; p_n \; \{\log(p_n) + p_n \log[B(p_n)/\epsilon_n]\}$$

for another constant c_2 and for n sufficiently large. Now substituting $p_n \approx n^{\alpha}$, $B(p_n) = O(n^{2h\alpha})$, and $\epsilon_n \approx n^{-\beta}$, the bound becomes

$$c_2 \; \alpha \; n^{2h\alpha + \alpha} \log(n) \; + \; c_2 \; (2h\alpha + \beta) \; n^{2h\alpha + 2\alpha} \log(n)$$

$$= O[n^{2h\alpha + 2\alpha} \log(n)].$$

Now $n\epsilon_n^2 \approx n^{1-2\beta}$, and so the desired result holds provided

$$2h\alpha + 2\alpha < 1 - 2\beta$$

or

$$\beta < 1/2 - \alpha - h\alpha.$$

Finally,

$$\epsilon_n = [\frac{p_n^3 \; B(p_n)}{n}]^{1/2}$$

corresponds to $\beta = 1/2 - 3\alpha/2 - h\alpha$, and since ϵ_n can be replaced by $\epsilon\epsilon_n$ for any $\epsilon > 0$ in the above argument, the result is proved. ∎

PROOF OF LEMMA 3.3. By Proposition 1 of Einmahl (1987), a construction is possible such that for every n

$$P_0(\max_{1\le t\le n} |S_t - T_t| > c_1\delta_n) \;\le\; c_2[\delta_n^{-s}K_{sn}(\delta_n) + \delta_n^{-2}L_n(\delta_n)]$$

where P_0 is the new probability measure, $\{\delta_n\}$ is a sequence of positive real numbers satisfying

$$n^{1/4+\epsilon} = O(\delta_n)$$

for some $\epsilon > 0$, c_1 and c_2 are positive constants independent of n and δ_n,

$$K_{sn}(\delta) = \sum_{t=1}^{n} \mathcal{E}_{\mathcal{P}}[|\xi_{nt}|^s \, 1\{|\xi_{nt}| \le \delta\}],$$

and

$$L_n(\delta) = \sum_{t=1}^{n} \mathcal{E}_{\mathcal{P}}[|\xi_{nt}|^2 \, 1\{|\xi_{nt}| > \delta\}].$$

Note that both $\delta^{-s}K_{sn}(\delta)$ and $\delta^{-2}L_n(\delta)$ are less than or equal to

$$\delta^{-s} \sum_{t=1}^{n} \mathcal{E}_{\mathcal{P}}[|\xi_{nt}|^s]$$

which is $\delta^{-s}O(n)$ by hypothesis. Choosing

$$\delta_n = n^{2/s} + \gamma$$

and applying the Borel-Cantelli lemma yield the desired result. ∎

PROOF OF LEMMA 3.5. By Remark 3.2 and Cauchy-Schwartz we have

$$\frac{|\rho'(d\Phi'V_1\Phi)^{-1}\ \Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})|}{\sigma_n}$$

$$= \frac{|\rho'(d\Phi'V_1\Phi)^{-1}\ \Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})|}{\surd\{(\sigma_0^2/d^2)\ \rho'(\Phi'V_1\Phi)^{-1}\ (\Phi'V_0\Phi)\ (\Phi'V_1\Phi)^{-1}\rho\ [1 + o(1)]\}}$$

$$\leq \frac{\|(G_n)^{-1/2}\ \Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})\|}{\surd n\ \ [1 + o(1)]}\ \ \frac{d\ U\ U_1^2}{\sigma_0\ L\ L_0^2}$$

by Assumptions S7 and S9. Note that

$$\|G_n^{-1/2}\ \Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})\|\ \leq\ (1/\surd L)\ \|\Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})\|.$$

We now use integration by parts as in Cox (1984) to show that

$$\|\Sigma_{t=1}^n\ \varphi_t\ (Y_{nt} - Z_{nt})\|\ \leq\ \max_{1\leq t\leq n} |S_t - T_t|\ \ [C(p_n)]^{1/2}$$

For a real number $u \in [0,1]$ and for every n, define the partial sum process

$$\bar{V}_n(u)\ =\ \sum_{t=1}^{[nu]} Y_{nt}$$

where [·] denotes the greatest integer function. Let μ_n be the empirical distribution function of the x_t's, i.e.

$$\mu_n(x) = \sum_{t=1}^{n} I(x_t \le x)$$

where $I(\cdot)$ is the indicator function, and on (0,1) define

$$V_n(x) = \bar{V}_n[\mu_n(x)].$$

Define $\bar{W}_n(\cdot)$ and $W_n(\cdot)$ analogously with Y_{nt} replaced by Z_{nt}. Then we can write

$$\sum_{t=1}^{n} \varphi_t \, (Y_{nt} - Z_{nt}) = \int_0^1 \varphi(x)\, d(V_n - W_n)$$

where the integral is Lebesgue-Stieltjes and $\varphi(x)$ is the p_n-vector with elements j^{th} element $\{\varphi_j(x)\}$. Now integrating by parts

$$\int_0^1 \varphi(x)\, d(V_n - W_n) = [V_n(1) - W_n(1)]\, \varphi(1) - \int_0^1 [V_n(x) - W_n(x)]\, \psi(x)\, dx$$

where $\psi(x)$ is the p_n-vector with j^{th} element $\{(d/dx)\varphi_j(x)\}$. Note that

$$V_n(1) - W_n(1) = S_n - T_n$$

and that

$$\sup_{x\in\mathcal{X}} [V_n(x) - W_n(x)] = \max_{1\le t\le n} |S_t - T_t|.$$

So by definition of C(p) and the fact that $B(p) \le C(p)$,

$$\|\Sigma_{t=1}^{n} \varphi_t \, (Y_{nt} - Z_{nt})\| \;=\; \| \int_0^1 \varphi(x)\, d(V_n - W_n)\|$$

$$\le \max_{1 \le t \le n} |S_t - T_t| \;\; [C(p_n)]^{1/2}$$

which is the desired inequality.

Now combining our inequalities we have

$$\frac{|\rho'(d\Phi' V_1 \Phi)^{-1} \, \Sigma_{t=1}^{n} \varphi_t \, (Y_{nt} - Z_{nt})|}{\sigma_n} \le \frac{\max_{1 \le t \le n} |S_t - T_t| \; [C(p_n)]^{1/2}}{\sqrt{n}\; (\sigma_0/d)\; [1 + o(1)]} \; \frac{U\, U_1^2}{L^{3/2}\, L_0^2}$$

$$= \frac{\max_{1 \le t \le n} |S_t - T_t|}{n^{1/4+\gamma/2}} \; O\{\frac{[C(p_n)]^{1/2}}{n^{1/4-\gamma/2}}\}.$$

By Lemma 3.3 the first term is o(1) a.s. and the final term is bounded by Assumption P7. ∎

REFERENCES

Andrews, Donald W. K. (1988) "Asymptotic Normality of Series Estimators for Various Nonparametric and Semiparametric Models," Cowles Foundation Discussion Paper No. 874, Yale University.

Cox, Dennis D. (1984) "Gaussian Approximation of Smoothing Splines," *Dept of Statistics Technical Report* 743, University of Wisconsin, Madison.

Cox, Dennis D. (1988) "Approximation of Least Squares Regression on Nested Subspaces," *The Annals of Statistics* 16, 713-732.

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

Eastwood, Brian J. (1987) *Confidence Interval Construcion in Semi-Nonparametric Regression Estimation*. North Carolina State University, Ph.D. dissertation.

Einmahl, U. (1987) "Strong Approximations for Partial Sums of iid B-valued r.v.'s in the Domain of Attraction of a Gaussian Law," *Probabiliy Theory and Related Fields* 77, 65-85.

Eubank, R. (1988) *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

Gallant, A. Ronald (1987) *Nonlinear Statistical Models*. New York: John Wiley and Sons.

Gallant, A. Ronald (1989) "On Asymptotic Normality when the Number of Regressors Increases and Estimates the Minimum Eigenvalue of X'X/n Decreases," *North Carolina Institute of Statistics Mimeograph Series* No. 1955.

Luenberger, David G. (1984) *Linear and Nonlinear Programming.* Reading, Massachusetts: Addison-Wesley.

Portnoy, Stephen (1985) "Asymptotic Behavior of M-Estimators of p Regression Parameters when p^2/n is Large. II. Normal Approximation," *The Annals of Statistics* 13, 1403-1417.

Severini, T.A. and W.H. Wong (1987) "Convergence Rates of Maximum Likelihood and Related Estimates in General Parameter Spaces," unpublished manuscript, Department of Statistics, University of Chicago.

Szego, Gabor (1975) *Orthogonal Polynomials.* 4th edition American Mathematical Society Colloquium Publications, Providence, R.I.