On Asymptotic Normality when the Number of Regressors Increases and the Minimum Eigenvalue of X′X/n Decreases*

by

A. Ronald Gallant
Department of Statistics
North Carolina State University
Raleigh, NC 27696-8203 USA

July 1989

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

On Asymptotic Normality when the Number of Regressors Increases and the Minimum Eigenvalue of $X'X/n$ Decreases*

by

A. Ronald Gallant

ABSTRACT

Conditions under which a regression estimate based on p regressors is asymptotically normally distributed when the minimum eigenvalue of $X'X/n$ decreases with p are obtained. The results are relevant to the regressions on truncated series expansions that arise in neural networks, demand analysis, and asset pricing applications.

1. INTRODUCTION

The p leading terms of a series expansion $\{\varphi_j\}_{j=1}^{\infty}$ are often used in regression analysis to either represent or approximate the conditional expectation with respect to x of a dependent variable y. Either explicitly or implicitly, p usually grows with the sample size n in these applications. The growth may follow some deterministic rule $\{p_n\}$ or it may be adaptive with p increased when a t-test rejects or some model selection rule such as Schwarz's (1978) criterion, Mallow's (1973) C_p, or cross validation suggests an increase.

The most familiar examples of expansions viewed as representations are experimental designs, which are Hadamard expansions in the levels of factors. In this case, which we refer to as the first paradigm, the data are presumed to have been generated according to the regression

$$y_t = \sum_{j=1}^{p} \theta_j^o \varphi_j(x_t) + e_t \qquad t = 1, 2, \ldots, n.$$

Perhaps the most familiar examples of expansions viewed as approximations occur in response surface analysis when the observed response is regressed on a polynomial in the control variables. In this case, which we refer to as the second paradigm, the data are presumed to have been generated according to the regression

$$y_t = g^o(x_t) + e_t \qquad t = 1, 2, \ldots, n.$$

and $g_p(x|\theta) = \sum_{j=1}^{p} \theta_j \varphi_j(x_t)$ is regarded as an approximation to g^o. While our analysis covers both paradigms, the second provides the primary motivation.

In fact, it is three specific applications that motivate this work, although our results obviously have wider applicability. These three applications are: (i) the statistical interpretation of neural networks (White, 1989) as used in robotics, navigation aids, speech interpretation, and other artificial intelligence applications [feedforward networks can be viewed as series expansion regressions (Gallant and White, 1988)], (ii) flexible functional forms as used in consumer and factor demand analysis (Barnett, Geweke and Yue, 1989), and (iii) Hermite expansions as used in asset pricing and sample selection applications (Gallant and Tauchen, 1989; Gallant and Nychka, 1987). For a variety of reasons -- mimicking biological structure, avoiding unrealistic boundary conditions, or appending a leading special case to the expansion either to improve the approximation or test hypotheses -- these three applications share a common feature: the eigenvalues of the $p \times p$ matrix

$$X'X/n = (1/n)\sum_{t=1}^{n} [\varphi_1(x_t), \ldots, \varphi_p(x_t)][\varphi_1(x_t), \ldots, \varphi_p(x_t)]'$$

(or its first order counterpart if the analysis is nonlinear) decline as n and p increase together.

In these applications, the independent variables $\{x_t\}$ are almost invariably obtained by sampling from some common distribution $\mu(x)$ so that results are more usefully stated in terms of the $p \times p$ matrix

$$G_p = \int [\varphi_1(x), \ldots, \varphi_p(x)][\varphi_1(x), \ldots, \varphi_p(x)]' \, d\mu(x).$$

If one tried to state results in terms of conditions on the eigenvalues of $X'X/n$ (or other characteristics such as the diagonal elements of the hat matrix) results would be conditional upon the particular sequence $\{x_t\}$ that obtained of which, presumably, only the first n terms are known. Moreover, even if one were presumed to have the entire sequence $\{x_t\}_{t=1}^{\infty}$ available for inspection, one would still be involved in the conceptual circularity of having to propose a rule p_n, check the eigenvalues of $X'X/n$ as n tends to infinity, and if that rule didn't work, to try again. Without our results, which we believe to be new, one could not state a rule $\{p_n\}$ *a priori* which would satisfy the conditions on $X'X/n$ for every sequence $\{x_t\}$ encountered in applications. Providing deterministic rules that achieve asymptotic normality under either paradigm and that hold for (almost) every sequence $\{x_t\}$ is the main contribution of the paper. An extension to adaptive rules using results due to Eastwood (1987) is possible; see Eastwood and Gallant (1987) or Andrews (1988) for examples.

To help fix ideas, we illustrate the rates of decline one might encounter in applications with an example taken from Gallant (1984) regarding the log cost function of a firm when the firm's output is the only free variable. The log cost function of a firm gives the logarithm of the cost to a firm over a year, say, to produce x units of log output at log prices p of the factors of production; x is a scalar and p is a vector. Since we shall hold p fixed, that argument is suppressed and we write a log cost function as $g(x)$. Units of measurement are irrelevant, and the capital stock of the firm is fixed (thus bounding feasible output from above and below), so we can assume that $0 \le x \le 2\pi$ without loss of generality. Data is generated according to

$$y_t = g^o(x_t) + e_t \qquad\qquad t = 1, 2, \ldots, n.$$

If one approximated $g^o(x)$ by the Fourier series

$$g_p(x|\theta) = u_0 + 2\sum_{j=1}^{K} u_j \cos(jx_t) - v_j \cos(jx_t) \qquad\qquad p = 2K + 1$$

and μ were uniform over $[0, 2\pi]$ then the eigenvalues of G_p would be bounded from above and below for all p (Tolstov, 1962). But this would imply that the log cost function of a firm is periodic so that conditions at the lowest feasible output are the same as at the highest feasible output which is silly. (At the minimum, one would expect the first derivative of g^o to be negative at 0 and positive at 2π.)

One could improve the approximation and have the means to test against the leading special case by adding a Translog cost function (Christensen, Jorgenson, and Lau, 1975) to the expansion

$$g_p(x|\theta) = u_0 + bx + cx^2 + 2\sum_{j=1}^{K} u_j\cos(x_t) - v_j\cos(x_t) \qquad p = 2K + 3.$$

By minimizing $\theta' G_p\theta/\theta'\theta$ over all θ in $\mathbb{R}^p$ one gets the smallest eigenvalue of G_p. For given b and c, the solution to this minimization problem is gotten by putting $\theta = (-\bar{u}_0, b, c, -\bar{u}_1, \bar{v}_1, \ldots, -\bar{u}_K, \bar{v}_K)$ where $\bar{u}_j$ and $\bar{v}_j$ denote the Fourier coefficients of $bx + cx^2$ (Tolstov, 1962). The minimum itself is $\sum_{j=K+1}^{\infty}(\bar{u}_j^{\,2} + \bar{v}_j^{\,2})/\sum_{j=1}^{K}(\bar{u}_j^{\,2} + \bar{v}_j^{\,2})$ which by direct computation can be shown to decline at the rate K^{-1} when $c \neq 0$ and $b \neq 2c\pi$ ($bx + cx^2$ is symmetric about π when $b = 2c\pi$). But this would imply that the log cost function of an arbitrary firm has periodic second and higher derivatives which is implausible.

This difficulty can be overcome by assuming that the support of μ is $[\epsilon, 2\pi-\epsilon]$ for some small $\epsilon > 0$. The minimum eigenvalue $\underline{\lambda}_K$ is gotten by: extending $bx + cx^2$ from $[\epsilon, 2\pi-\epsilon]$ to $[0, 2\pi]$ such that the extension is periodic and infinitely many times differentiable, computing the Fourier coefficients $\bar{u}_j$ and $\bar{v}_j$ of the extended function, and putting $\underline{\lambda}_K = \sum_{j=K+1}^{\infty}(\bar{u}_j^{\,2} + \bar{v}_j^{\,2})/\sum_{j=1}^{K}(\bar{u}_j^{\,2} + \bar{v}_j^{\,2})$. In this case $\underline{\lambda}_K$ is rapidly decreasing; that is, $(K^m)(\underline{\lambda}_K)$ tends to zero with K for every m.

In summary, as p increases one could have the minimum eigenvalues of G_p bounded from below, declining at a polynomial rate, or rapidly decreasing.

In the literature related to our problem within the first paradigm, Huber (1973) gives the basic result used to obtain asymptotic normality. Portnoy (1985) extends Huber's and Yohai and Maronna's (1979) results to M-estimators and gives a rich discussion of the history of research relating p to n so as to achieve asymptotic normality when the eigenvalues of $X'X/n$ are bounded from above and below. His methods of proof are related to ours in that metric entropy and an exponential inequality are used to get a uniform strong law with a rate. We use these same ideas but at a more macro level by citing results in the empirical process literature as found, for instance, in Pollard (1984). Closely related work within the second paradigm is found in Severini and Wong (1987) and Andrews (1988). Ours, Severini and Wong's, and Andrews' proof strategies are similar. Severini and Wong consider maximum likelihood and similarly structured sieve estimators. Their regularity conditions would basically require the eigenvalues of $X'X/n$ to be bounded above and below when their results are specialized to regression. Andrews (1988) provides an excellent history of work related to this problem and an extensive list of nonparametric estimation strategies encompassed within the second paradigm. His regularity conditions are not as closely related to applications as ours, operating at a level closer to the results. Verification of Andrews' regularity conditions is reasonably straightforward when the eigenvalues of $X'X/n$ are bounded above and below. He does not explicitly relate the eigenvalues of $X'X/n$ to those of G_p and so verification of regularity conditions is conditional on $\{x_t\}$ and subject to the problems discussed above.

He extends his results on deterministic rules to adaptive rules using results of Eastwood (1987). This same extension covers our results as well. Overall, one might view our work as moving Andrews' results closer to applications.

2. ASYMPTOTIC NORMALITY UNDER THE FIRST PARADIGM

We consider a regression model with p parameters

$$y_t = \sum_{j=1}^{p} \theta_j \varphi_j(x_t) + u_t \qquad t = 1, 2, \ldots, n$$

Putting $\varphi(x) = [\varphi_1(x), \varphi_2(x), \ldots, \varphi_p(x)]'$, the least squares estimator is

$$\hat{\theta} = G_{np}^{-1} \left[\frac{1}{n} \sum_{t=1}^{n} y_t \varphi(x_t) \right]$$

where

$$G_{np} = X'X/n = \frac{1}{n} \sum_{t=1}^{n} \varphi(x_t)\varphi'(x_t) \ .$$

The objective of this section is to find rules relating p to n such that

$$\frac{\rho'[\hat{\theta} - \mathcal{E}(\hat{\theta}|\{x_t\})]}{\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}}$$

is asymptotically normally distributed when the data is generated according to $y_t = g_n^o(x_t) + e_t$ where $\{e_t\}$ is an iid sequence with common distribution P(e), $\{x_t\}$ is an iid sequence with common distribution $\mu(x)$, $\{x_t\}$ is independent of $\{e_t\}$, ρ is an arbitrary vector in $\mathbb{R}^p$, and $\mathrm{Var}(\rho'\hat{\theta}|\{x_t\}) = (1/n)\rho' G_{np}^{-1} \rho$.

The dependence of g^o on n is to include the first paradigm within the analysis. Under the first paradigm, $\rho'\mathcal{E}(\hat{\theta}|\{x_t\})$ is presumed to be an unbiased estimate of the parametric function of interest and the results of this section are all that is required for asymptotic normality. Under the second paradigm, there is a bias term to deal with. An analysis of the bias term is deferred to the next section.

Put

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

We assume that B(p) is finite for each p. For instance, if $y_t = g^o(x_t) + e_t$ and $\mathcal{X} = [0,2\pi]$ then one might fit

$$y_t = a + bx_t + cx_t^2 + 2\sum_{j=1}^{K} [u_j \cos(jx_t) - v_j \sin(jx_t)] + u_t$$

by least squares using

$$\rho'\hat{\theta} = \hat{a} + \hat{b}x_o + \hat{c}x_o^2 + 2\sum_{j=1}^{K} [\hat{u}_j \cos(jx_o) - \hat{v}_j \sin(jx_o)]$$

to estimate $g^o(x^o)$ in which case $p = 2K + 3$ and $B(p) = 1 + 2\pi + 4\pi^2 + 2K$. If instead $y_t = (d/dx)g^o(x_t) + e_t$ one might fit

$$y_t = b + 2cx_t - 2\sum_{j=1}^{K} j[u_j \sin(jx_t) + v_j \cos(jx_t)] + u_t$$

by least squares, using

$$\rho'\hat{\theta} = 2\hat{c} + 2\sum_{j=1}^{K} j^2[-\hat{u}_j \cos(jx_o) + \hat{v}_j \sin(jx_o)]$$

to estimate $(d^2/dx^2)g^o(x^o)$ whence $p = 2K + 2$ and $B(p) = 1 + 4\pi + K(K+1)$.

Fix a realization of $\{x_t\}$. Then $\rho'[\hat{\theta} - \mathcal{E}(\hat{\theta}|\{x_t\})]/\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}$ is a linear function of the errors, *viz.*

$$\sum_{t=1}^{n} \frac{(1/n)\rho' G_{np}^{-1}\varphi(x_t)}{\sqrt{[(1/n)\rho' G_{np}^{-1}\rho]}} \ e_t \ ,$$

which is asymptotically normally distributed, conditional on $\{x_t\}$, if and only if (Huber, 1973)

$$\lim_{n\to\infty} \sup_{1\le t\le n} \frac{(1/n)\rho' G_{np}^{-1}\varphi(x_t)}{\sqrt{[(1/n)\rho' G_{np}^{-1}\rho]}} = 0 .$$

Letting $\underline{\lambda}(G)$ denote the smallest eigenvalue of a matrix G, and $G^{-1/2}$ the Cholesky factor of G^{-1} we have, using the Cauchy-Schwartz inequality,

$$\frac{|(1/n)\rho' G_{np}^{-1}\varphi(x_t)|}{\sqrt{[(1/n)\rho' G_{np}^{-1}\rho]}} \le \frac{\|\rho' G_{np}^{-1/2}\| \, \|G_{np}^{-1/2}\varphi(x_t)\|}{\sqrt{n} \, \|\rho' G_{np}^{-1/2}\|}$$

$$\le \left[\frac{B(p)}{n\,\underline{\lambda}(G_{np})}\right]^{1/2}$$

Thus, any rule p_n relating p to n such that $\lim_{n\to\infty} B(p_n)/[n\,\underline{\lambda}(G_{n,p_n})] = 0$ will achieve asymptotic normality, conditional on $\{x_t\}$. If the rule p_n does not depend on knowledge of $\{x_t\}$, other than knowledge that $\{x_t\}$ does not correspond to some null set of the underlying probability space, then the unconditional distribution of $\rho'[\hat{\theta} - \mathcal{E}(\hat{\theta}|\{x_t\})]/\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}$ is asymptotically normal as well.

Our strategy for finding p_n depends on relating the eigenvalues of G_{np} to the eigenvalues of

$$G_p = \int \varphi(x)\varphi'(x)\, d\mu(x)$$

by establishing a strong law of large numbers that holds with rate ϵ_n uniformly over the family

$$\mathcal{F}_p = \{[\theta'\varphi(x)]^2/B(p) \colon \theta'\theta = 1,\ \theta \in \mathbb{R}^p\}$$

when $p = p_n$. First, we need some additional notation and two lemmas.

Let $\mathcal{E}$ denote expectation with respect to $dP \times d\mu$ or $d\mu$, as appropriate, and let $\mathcal{E}_n$ denote expectation with respect to the empirical distribution of $\{(e_t, x_t)\}_{t=1}^n$ or $\{x_t\}_{t=1}^n$, as appropriate. That is, for $f(e,x)$

$$\mathcal{E}_n f = \frac{1}{n} \sum_{t=1}^{n} f(e_t, x_t) \qquad \mathcal{E}\, f = \iint f(x,e)\, dP(e) d\mu(x),$$

and for $f(x)$

$$\mathcal{E}_n f = \frac{1}{n} \sum_{t=1}^{n} f(x_t) \qquad \mathcal{E}\, f = \int f(x)\, d\mu(x) \ .$$

With this notation, $G_{np} = \mathcal{E}_n \varphi\varphi'$ and $G_p = \mathcal{E}\, \varphi\varphi'$. Note, there is a θ with $\theta'\theta = 1$ and $\theta' G_{np} \theta = \underline{\lambda}(G_{np})$ so there is an f in $\mathcal{F}_p$ with $\underline{\lambda}(G_{np})/B(p) = \mathcal{E}_n f$. Letting $\underline{\lambda}(G)$ and $\overline{\lambda}(G)$ denote the smallest and largest eigenvalues of a matrix respectively, put $\underline{\Pi}(p) = B(p)/\underline{\lambda}(G_p)$ and $\overline{\Pi}(p) = B(p)/\overline{\lambda}(G_p)$.

LEMMA 1. The number of ϵ-balls required to cover the surface of a sphere in $\mathbb{R}^p$ is bounded by $2p(2/\epsilon + 1)^{p-1}$.

Proof. The proof is patterened after Kolmogorov and Tihomirov (1959). Let M be the maximum number of non-intersecting balls of radius $\epsilon/2$ and center on the surface of the unit sphere. If θ is a point on the surface, then an $\epsilon/2$ neighborhood must intersect one of these balls; hence θ is within ϵ of its center. If V denotes the volume of a shell in $\mathbb{R}^p$ with outer radius $1 + \epsilon/2$ and inner radius $1 - \epsilon/2$ and v denotes the volume of an $\epsilon/2$-ball, then $M \leq V/v$.

$$V/v = \frac{2\pi^{p/2}[(1 + \epsilon/2)^p - (1 - \epsilon/2)^p]/[p\Gamma(p/2)]}{2\pi^{p/2}(\epsilon/2)^p/[p\Gamma(p/2)]}$$

$$= [(1 + \epsilon/2)^p - (1 - \epsilon/2)^p]/(\epsilon/2)^p$$

$$= [1 + p(1 + \bar{\epsilon}/2)^{p-1}\epsilon/2 - 1 + p(1 - \bar{\bar{\epsilon}}/2)^{p-1}\epsilon/2]/(\epsilon/2)^p$$

[by the mean value theorem]

$$\leq 2p(1 + \epsilon/2)^{p-1}/(\epsilon/2)^{p-1} \quad \blacksquare$$

LEMMA 2. Let $p_n \to \infty$ and $\epsilon_n \to 0$ as $n \to \infty$. If $n\epsilon_n^2 > 1/8$ then

$$P\left(\sup_{f \in \mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E} f| > 8\epsilon_n \right) < 16\ p_n\ (4/\epsilon_n + 1)^{p_n - 1} \exp(- \tfrac{1}{2}\, n\epsilon_n^2).$$

Proof. The proof uses results from Pollard (1984, Chapter II) which require a demonstration that $\mathcal{F}_p$ is bounded and a computation of the metric entropy of $\mathcal{F}_p$. The first two paragraphs take care of these details. The third paragraph has the main argument.

If f is in $\mathcal{F}_p$ then $\sup_{x \in \mathcal{X}} |f(x)| \leq 1$ because the Cauchy-Schwartz inequality implies $(\theta'\varphi)^2/B(p) \leq \|\theta\|^2\|\varphi\|^2/B(p) \leq [\sum_{j=1}^p \varphi_j^2(x)]/B(p) \leq 1$. A consequence of this bound is that $\mathcal{E}_n f^2 \leq 1$ and $\mathrm{Var}(\mathcal{E}_n f) \leq 1/n$ for each f in $\mathcal{F}_p$.

From Lemma 1, the number of $\epsilon/2$-balls required to cover the surface of the unit sphere in $\mathbb{R}^p$ is bounded by $N_1(\epsilon,p) = 2p(4/\epsilon + 1)^{p-1}$. Let $\tilde{\theta}_j$ denote the centers of these balls and put $g_j = (\tilde{\theta}_j'\varphi)^2/B(p)$. Since $f = (\theta'\varphi)^2/B(p)$ must have θ in some ball we have

$$\begin{aligned}
\min_j |g_j - f| &= \min_j |(\theta'\varphi)^2 - (\tilde{\theta}_j'\varphi)^2|/B(p) \\
&= \min_j |\varphi'(\theta - \tilde{\theta}_j)||\theta'\varphi + \tilde{\theta}_j'\varphi|/B(p) \\
&\leq \min_j \|\varphi\|\ \|\theta - \tilde{\theta}_j\|\ (\|\theta\|\|\varphi\| + \|\tilde{\theta}_j\|\|\varphi\|)/B(p) \\
&\leq 2[\|\varphi\|^2/B(p)](\epsilon/2) \\
&\leq \epsilon.
\end{aligned}$$

Let $\mathscr{E}_n^o = (1/n)\sum_{t=1}^n \sigma_t f(x_t)$ where σ_t takes on the values ±1 with equal probability, independently of $\{x_t\}_{t=1}^n$. From Pollard (1984, p. 31)

$$P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n f - \mathscr{E} f| > 8\epsilon_n\right) \le 4\, P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n^o f| > 2\epsilon_n\right)$$

$$P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n^o f| > 2\epsilon_n \,\middle|\, \{x_t\}\right) \le 2\, N_1(\epsilon_n, p_n)\, \exp\left[-\tfrac{1}{2}\, n\epsilon_n^2/(\max_j \mathscr{E}_n g_j^2)\right]$$

provided $\mathrm{Var}(\mathscr{E}_n f)/(4\epsilon_n)^2 \le 1/2$. By the bound above we have $\max_j \mathscr{E}_n g_j^2 \le 1$ and $\mathrm{Var}(\mathscr{E}_n f) \le 1/n$ whence, substituting for N_1, the second inequality becomes

$$P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n^o f| > 2\epsilon_n \,\middle|\, \{x_t\}\right) \le 4p_n(4/\epsilon_n + 1)^{p_n - 1} \exp\left[-\tfrac{1}{2}\, n\epsilon_n^2\right].$$

provided $\mathrm{Var}(\mathscr{E}_n f)/(4\epsilon_n)^2 \le (16n\epsilon_n^2)^{-1} \le 1/2$. Since the right hand side does not depend on the conditioning random variables $\{x_t\}$ we have

$$P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n^o f| > 2\epsilon\right) \le 4p_n(4/\epsilon_n + 1)^{p_n - 1} \exp\left[-\tfrac{1}{2}\, n\epsilon_n^2\right].$$

provided $n\epsilon_n^2 > 1/8$. Substitution into the first inequality yields the result.∎

Lemma 2 can be used to establish a uniform strong law with rate:

LEMMA 3. Let $p_n \le n^\alpha$ for some α with $0 \le \alpha < 1$. If $0 \le \beta \le (1 - \alpha)/2$ then

$$P\left(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n f - \mathscr{E} f| > n^{-\beta}/2 \ \text{ infinitely often}\right) = 0.$$

Proof. If $\sum_{n=1}^\infty P(\sup_{f\in\mathscr{F}_{p_n}} |\mathscr{E}_n f - \mathscr{E} f| > 8\epsilon_n) < \infty$ for $\epsilon_n = n^{-\beta}/16$ then the result will follow by the Borel-Cantelli lemma. With this choice of ϵ_n, $n\epsilon_n^2 = n^{1-2\beta}/256 \ge n^\alpha/256$ which exceeds 1/8 for n large enough. By Lemma 2, we will have $\sum_{n=1}^{\infty} P(\sup_{f\in\mathcal{F}p_n} |\mathcal{E}_n f - \mathcal{E}f| > 8\epsilon_n) < \infty$ if

$$\left[p_n(4/\epsilon_n + 1)^{p_n - 1} \exp(-\tfrac{1}{2} n\epsilon_n^2)\right] n^{1+c} \le B .$$

for some B, c > 0. Taking the logarithm of the left hand side we have for large n that

$$\begin{aligned}
&\ell og\, p_n + (p_n - 1)\ell og(4/\epsilon_n + 1) - n\epsilon_n^2/2 + (1 + c)\ell og\, n \\
&\quad \le \ell og\, n^{\alpha} + (n^{\alpha} - 1)\ell og(64n^{\beta} + 1) - n^{1-2\beta}/256 + (1 + c)\ell og\, n \\
&\quad \le (1 + \alpha + c)\ell og\, n + n^{\alpha}\ell og(65n^{\beta}) - n^{1-2\beta}/256 \\
&\quad = \ell og(65)n^{\alpha} + (1 + \alpha + c + \beta n^{\alpha})\ell og\, n - n^{1-2\beta}/256 \\
&\quad < 2\beta n^{\alpha}\ell og\, n - n^{1-2\beta}/256 .
\end{aligned}$$

The right hand side is negative for n large enough because $0 \le \alpha < 1 - 2\beta$. ∎

We can now state and prove the main result of this section; recall that $\underline{\Pi}(p) = B(p)/\underline{\lambda}(G_p)$ which will be a polynomial in p or rapidly decreasing in typical applications.

THEOREM 1. If p_n satisfies

$$\underline{\Pi}(p_n) \le n^{\beta} \qquad 0 \le \beta < 1/2$$

$$p_n \le n^{\alpha} \qquad 0 \le \alpha < 1 - 2\beta$$

then

$$P[\, B(p_n)/\underline{\lambda}(G_{n,p_n}) > 2n^{\beta} \text{ infinitely often }] = 0.$$

Proof. Suppose $\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E}f| \le n^{-\beta}/2$. There is an $f = (\theta'\varphi)^2/B(p_n)$ in $\mathcal{F}_{p_n}$ such that $\underline{\lambda}(G_{n,p_n})/B(p_n) = \mathcal{E}_n f$ whence

$$\begin{aligned}\underline{\lambda}(G_{n,p_n})/B(p_n) &= \mathcal{E}_n(\theta'\varphi)^2/B(p_n)\\ &\ge \mathcal{E}(\theta'\varphi)^2/B(p_n) - n^{-\beta}/2\\ &\ge \underline{\lambda}(G_{p_n})/B(p_n) - n^{-\beta}/2\\ &= 1/\underline{\Pi}(p_n) - n^{-\beta}/2\\ &\ge n^{-\beta}/2.\end{aligned}$$

Thus, $\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E}f| \le n^{-\beta}/2$ implies $\underline{\lambda}(G_{n,p_n})/B(p_n) \ge n^{-\beta}/2$. The contrapostitive is $B(p_n)/\underline{\lambda}(G_{n,p_n}) > 2n^{\beta}$ implies $\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E}f| > n^{-\beta}/2$. Thus

$$P[B(p_n)/\underline{\lambda}(G_{n,p_n}) > 2n^{\beta} \text{ i.o. }] \le P\left(\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E}f| > n^{-\beta}/2 \text{ i.o.}\right).$$

Apply Lemma 3. ∎

Asymptotic normality follows immediately.

THEOREM 2. If p_n satisfies

$$\underline{\Pi}(p_n) \le n^{\beta} \qquad 0 \le \beta < 1/2$$

$$p_n \le n^{\alpha} \qquad 0 \le \alpha < 1 - 2\beta$$

then

$$\frac{\rho'[\hat{\theta} - \mathcal{E}(\hat{\theta}|\{x_t\})]}{\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}} \xrightarrow{\mathcal{L}} N(0, 1)$$

both conditionally on $\{x_t\}$ and unconditionally.

Proof. By Theorem 1 $P[\ B(p_n)/[n\ \underline{\lambda}(G_{n,p_n})] > n^{\beta-1}/2\ \text{ i. o.}] = 0$ whence $\ell im_{n\to\infty}\ B(p_n)/[n\ \underline{\lambda}(G_{n,p_n})] = 0$ except for realizations of $\{x_t\}$ that correspond to an event in the underlying probability space that occurs with probability zero. ∎

Under assumptions (i) $0 < b \le \underline{\lambda}(G_{n,p_n}) < \overline{\lambda}(G_{n,p_n}) \le B < \infty$, and (ii) $0 < b' \le B(p)/p \le B' < \infty$, which are often imposed in studies that relate p to n, asymptotic normality will hold with p_n growing as fast as $p_n/n \to 0$. When a strong law is invoked, the rate typically deteriorates to $p_n^2/n \to 0$ (Portnoy, 1985). Theorem 2 would require $p_n^3/n \to 0$ under these assumptions. The reason for this slower rate is the use of the bound $\max_j\ \mathcal{E}_n g_j^2 \le 1$ in the proof of Lemma 2. Our method of proof will provide better rates when better bounds on $\mathcal{E}_n g_j^2$ are available. For example, under assumptions (i) and (ii) above, recalling that $g_j \le 1$, the bound on $\mathcal{E}_n g_j^2$ is $\mathcal{E}_n g_j^2 \le \mathcal{E}_n g_j \le (B/b')/p_n$ and the conclusion of Lemma 2 would read: If $np_n^2\epsilon_n^2 > (B/b')/8$ then

$$P\left(\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E} f| > 8\epsilon_n \right) < 16\ p_n\ (4/\epsilon_n + 1)^{p_n-1} \exp[-\tfrac{1}{2}\ np_n\epsilon_n^2/(B/b')]$$

The last line of the proof of Lemma 3 would have $0 \le \alpha < 1 + \alpha - 2\beta$ instead of $0 \le \alpha < 1 - 2\beta$. Since $0 < b \le \underline{\lambda}(G_{n,p_n})$ implies $0 < b'' \le \underline{\lambda}(G_{p_n})$ (Lemma 3), $B'p_n/b'' \le n^{\beta}$, $p_n \le n^{\alpha}$, and $0 \le \alpha < 1 + \alpha - 2\beta$ would imply asymptotic normality. That is, under assumptions (i) and (ii) our method of proof delivers asymptotic normality for $p_n^2/n \to 0$.

3. ASYMPTOTIC NORMALITY UNDER THE SECOND PARADIGM

In the previous section we were able to find rates for p_n such that

$$\mathrm{RelErr}(\rho'\hat{\theta}\,|\,\{x_t\}) = \frac{\rho'[\hat{\theta} - \mathcal{E}(\hat{\theta}\,|\,\{x_t\})]}{\sqrt{\mathrm{Var}(\rho'\hat{\theta}\,|\,\{x_t\})}}$$

is asymptotically normally distributed when the data is generated according to

$$y_t = g_n^o(x_t) + e_t \qquad t = 1, 2, \ldots, n.$$

without putting conditions on g_n^o or describing what $\rho'\hat{\theta}$ is intended to estimate. Here we must be more specific.

In the applications that motivate this work, usually g_n^o does not depend on n and has domain $\mathcal{X}$ which is a subset of $\mathbb{R}^M$. Usually an evaluation functional such as

$$D^\lambda g^o(x^o) = (\partial\lambda_1/\partial x_1^{\lambda}1) \cdot \ldots \cdot (\partial\lambda_M/\partial x_M^{\lambda}M)\; g^o(x^o)$$

is the object of interest. Above, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_M)$ and the absolute value $|\lambda| = \sum_{i=1}^M |\lambda_i|$ gives the order of the partial derivative; the zero order derivative is the function itself, $D^0 g \equiv g$. The dependence of g^o on n does not make the proofs more difficult and might be useful if one wanted to consider a notion analogous to a Pitman drift, so we shall retain it. These considerations make it natural to regard g_n^o as a sequence of points in a (weighted) Sobolev space and to assume that the sequence $\{\varphi_j\}$ is a dense subset of that space.

A Sobolev space is the set $\mathcal{W}_{m,q,\delta}$ of g with a finite Sobolev norm $\|g\|_{m,q,\delta}$. The Sobolev norm is defined as

$$\|g\|_{m,q,\delta} = \Big[\sum_{|\lambda|\le m} \int |D^{\lambda} g(x)|^{q} \, d\delta(x) \Big]^{1/q} \qquad 1 \le q < \infty$$

$$\|g\|_{m,q,\mathcal{X}} = \max_{|\lambda|\le m} \sup_{x\in\mathcal{X}} |D^{\lambda} g(x)| \qquad q = \infty .$$

To avoid clutter, $\|g\|_{m,\infty,\delta}$ will be understood to represent $\|g\|_{m,\infty,\mathcal{X}}$ in a statement such as "$\|g\|_{m,q,\delta} \le B$ for some $1 \le q \le \infty$"; in these instances $\mathcal{X}$ is the support of δ. $\mathcal{W}_{\infty,q,\delta} = \cap_{m=0}^{\infty} \mathcal{W}_{m,q,\delta}$. For many choices of $\mathcal{X}$ and δ these norms are interleaved in the sense that for each q with $1 \le q < \infty$ the inequality $\|g\|_{m,q,\delta} \le c\|g\|_{m,\infty,\mathcal{X}} \le c\|g\|_{m+M/q+1,q,\delta}$ holds for some c that does not depend on g. Obviously, $\|g\|_{m,q,\delta} \le \|g\|_{m,\infty,\mathcal{X}}$ holds if δ is a P-measure. For details see Adams (1975). Gallant and Nychka (1987) contains some results on interleaving for weighted norms with $\mathcal{X}$ unbounded.

By dense one means that for each g in $\mathcal{W}_{m,q,\delta}$ there exists a sequence of coefficients $\{\theta_j\}$ such that

$$\lim_{p\to\infty} \| g - g_p(\cdot|\theta) \|_{m,q,\delta} = 0$$

where

$$g_p(x|\theta) = \textstyle\sum_{j=1}^{p} \theta_j \varphi_j(x).$$

In this section, we shall take θ to be the infinite dimensional vector

$$\theta = (\theta_1, \theta_2, \ldots .)$$

and we shall let the context determine when its truncation to a p-vector is intended, $\theta' G_p \theta$ for instance. For given θ, if $g_p(\cdot|\theta)$ has limit g in $\bar{\mathcal{W}}_{m,q,\delta}$ in the sense above, write $g_\infty(\cdot|\theta)$. $g_\infty(\cdot|\theta)$ represents g and denseness implies that every g in $\mathcal{W}_{m,q,\delta}$ has such a representation; θ is not necessarily unique.

It would be unnatural to consider a regression on $\{\varphi_j\}$ under the second paradigm if these conditions were not in force for some choice of m, q, and δ.

If $\rho'\hat{\theta}$ is to estimate $D^\lambda g^o(x^o)$ then

$$\begin{aligned}\rho'\hat{\theta} &= D^\lambda g_p(x^o|\hat{\theta}) \\ &= \sum_{j=1}^{p} \hat{\theta}_j D^\lambda \varphi_j(x^o).\end{aligned}$$

In the motivating examples it is the case that

$$D^\lambda \varphi_j \in \text{span}\{\varphi_1, \ldots, \varphi_p\}.$$

As an example, a typical term of a multivariate Fourier expansion is cos(k′x) where k is an M-vector with integer elements and

$$(\partial^2/\partial x_1^2)\cos(k'x^o) = -(k_1)^2\cos(k'x^o);$$

similarly for most polynomial expansions such as the Hermite.

Due to these considerations, we shall consider the case when

$$\rho'\theta = D^\lambda g_p(x^o|\theta)$$

for some x^o in $\mathcal{X}$ and the elements ρ_j of ρ are increasing at some known polynomial rate. The case when ρ_j is decreasing with j is not of much interest because simpler methods of proof would yield stronger results. This situation would arise, for example, if the functional $\int fg^o \, d\mu$ were the object of interest and $f,g^o \in \mathcal{W}_{m,q,\mu}$.

Centering the estimate about the object of interest, rather than the conditional expectation of $\rho'\hat{\theta}$, we have

$$\frac{\rho'\hat{\theta} - D^{\lambda}g^{o}(x^{o})}{\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}} = \mathrm{RelErr}(\rho'\hat{\theta}|\{x_t\}) + \mathrm{RelBias}(\rho'\hat{\theta}|\{x_t\})$$

where

$$\mathrm{RelBias}(\rho'\hat{\theta}|\{x_t\}) = \frac{\mathcal{E}(\rho'\hat{\theta}|\{x_t\}) - D^{\lambda}g_{\infty}(x^{o}|\theta_n^{o})}{\sqrt{\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})}} .$$

Thus, a verification of asymptotic normality requires both the results of the previous section and a determination of the rate at which the relative bias decreases.

The behavior of the relative bias is intimately related to the truncation error

$$T_p = \| g_{\infty}(x|\theta_n^{o}) - g_p(\cdot|\theta_n^{o}) \|_{m,\infty,\mathcal{X}}$$

inherent in the series expansion $\{\varphi_j\}$. To illustrate, if g mapping $\mathbb{R}^M$ into $\mathbb{R}$ is in $\mathcal{W}_{m,q,\delta}$ where μ puts its mass on $\mathcal{X} = \times_{i=1}^{M}[\epsilon,\ 2\pi-\epsilon]$ and $g_p(\cdot|\theta)$ is a multivariate Fourier series expansion of degree K then $p \approx K^M$ and for any q with $1 \le q \le \infty$, and any small $\tau>0$ the order of the truncation error is $T_p = o(K^{-m+\ell+\tau})$ (Edmunds and Moscatelli, 1977). If, instead, $g \in \mathcal{W}_{\infty,q,\delta}$ (which implies g is infinitely many times differentiable) then T_p is rapidly decreasing. In these two examples, there is no dependence of g on n. T_p does not depend on n which is to say that the condition above is uniform in n. Additional structure such as $g_n \in \bar{\mathcal{G}} = \{g: \|g\|_{m+1,\infty,\mathcal{X}} \leq B\}$ would be required to obtain the requisite uniformity for these examples when a dependence on n is permitted.

We have considered two paradigms, the first where g_n^o is put equal to $g_p(\cdot|\theta_n^o)$ so as to force the relative bias term to zero leaving the relative error as the only concern, and the second where g_n^o is put equal to $g_\infty(x|\theta^o)$. As alluded to above, one might want to entertain a third paradigm where g_n^o is moved slowly away from some leading special case as n increases. For instance, in demand analysis, one might want to let g_n^o drift slowly away from the Translog model. The effect would be to move T_p to zero with p faster than the natural rate of the series expansion. This is completely analogous to the use of Pitman drift to obtain asymptotic approximations to the power of test statistics. It would also allow one to break free of the confines of Stone (1980) regarding the inherent limits of multivariate nonparametric estimation. Our results are general enough to accommodate this third paradigm.

SETUP. To summarize, we shall study the limiting behavior of the relative bias term

$$\text{RelBias}(\rho'\hat{\theta}|\{x_t\}) = \frac{\mathcal{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_\infty(x^o|\theta_n^o)}{\sqrt{\text{Var}(\rho'\hat{\theta}|\{x_t\})}}$$

when $\rho'\theta = D^\lambda g_p(x^o|\theta)$, the elements ρ_j of ρ increase at some known rate with j, the data are generated according to

$$y_t = g_n^o(x_t) + e_t \qquad t = 1, 2, \ldots, n.$$

as described in the previous section, and the sequence $\{T_p\}$ of truncation errors

$$T_p = \| g_\infty(x|\theta_n^o) - g_p(\cdot|\theta_n^o) \|_{m,\infty,\mathscr{X}}$$

declines at a known rate with p for some $m \ge |\lambda|$.∎

We begin by establishing two results, a lemma and a theorem. The first relates the bias to the truncation error and conditional variance. The second relates the maximum eigenvalue of G_{np} to the maximum eigenvalue of G_p to get an upper bound on $\bar{\lambda}(G_{np})$. It derives its relevance from the inequality

$$n\mathrm{Var}(\rho'\hat{\theta}|\{x_t\}) = \rho'(G_{np}^{-1})\rho \ge \rho'\rho/\bar{\lambda}(G_{np}) = [\rho'\rho/B(p)][B(p)/\bar{\lambda}(G_{np})]$$

which bounds the conditional variance from below.

LEMMA 4. $\mathrm{RelBias}(\rho'\hat{\theta}|\{x_t\}) \le \sqrt{n}T_p\left\{2 + 1/\sqrt{[n\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})]}\right\}$.

Proof. $\theta_n^* = \mathscr{E}(\hat{\theta}|\{x_t\})$ minimizes $\|g_p(\cdot|\theta) - g_n^o\|_{0,2,\mu_n}$ over $\mathbb{R}^p$ where μ_n denotes the empirical distribution of $\{x_t\}_{t=1}^n$.

$$|\mathscr{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_\infty(x^o|\theta_n^o)|$$

$$\le |\mathscr{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_p(x^o|\theta_n^o)| + |D^\lambda g_p(x^o|\theta_n^o) - D^\lambda g_\infty(x^o|\theta_n^o)|$$

$$= |\rho'(\theta_n^* - \theta_n^o)| + |D^\lambda g_p(x^o|\theta_n^o) - D^\lambda g_\infty(x^o|\theta_n^o)|$$

$$\le |\rho' G_{np}^{-1} G_{np}(\theta_n^* - \theta_n^o)| + T_p$$

$$\le \left\{[\rho' G_{np}^{-1}\rho][(\theta_n^* - \theta_n^o)' G_{np}(\theta_n^* - \theta_n^o)]\right\}^{1/2} + T_p$$

$$= \sqrt{[n\mathrm{Var}(\rho'\hat{\theta}|\{x_t\})]}\,\|g_p(\cdot|\theta_n^*) - g_p(\cdot|\theta_n^o)\|_{0,2,\mu_n} + T_p.$$

$$
\begin{aligned}
&\|g_p(\cdot|\theta_n^*) - g_p(\cdot|\theta_n^o)\|_{0,2,\mu_n} \\
&\quad \le \|g_p(\cdot|\theta_n^*) - g_n^o\|_{0,2,\mu_n} + \|g_n^o - g_p(\cdot|\theta_n^o)\|_{0,2,\mu_n} \\
&\quad \le \|g_p(\cdot|\theta_n^o) - g_n^o\|_{0,2,\mu_n} + \|g_n^o - g_p(\cdot|\theta_n^o)\|_{0,2,\mu_n} \\
&\quad \le \|g_p(\cdot|\theta_n^o) - g_n^o\|_{m,\infty,\mathcal{X}} + \|g_n^o - g_p(\cdot|\theta_n^o)\|_{m,\infty,\mathcal{X}} \\
&\quad \le 2T_p. \quad \blacksquare
\end{aligned}
$$

We can now state and prove the main result of this section; recall that $\overline{\Pi}(p) = B(p)/\overline{\lambda}(G_p)$ which in most applications will be a polynomial in p.

THEOREM 3. If p_n satisfies

$$
\begin{aligned}
&\overline{\Pi}(p_n) \ge n^{\gamma} && 0 \le \gamma < 1/2 \\
&p_n \le n^{\alpha} && 0 \le \alpha < 1 - 2\gamma
\end{aligned}
$$

then

$$P[\ B(p_n)/\overline{\lambda}(G_{n,p_n}) < (2/3)n^{\gamma} \text{ infinitely often }] = 0.$$

Proof. Suppose $\sup_{f\in\mathcal{F}_{p_n}}|\mathcal{E}_n f - \mathcal{E} f| \le n^{-\gamma}/2$. There is an $f = (\theta'\varphi)^2/B(p_n)$ in $\mathcal{F}_{p_n}$ such that $\overline{\lambda}(G_{n,p_n})/B(p_n) = \mathcal{E}_n f$ whence

$$
\begin{aligned}
\overline{\lambda}(G_{n,p_n})/B(p_n) &= \mathcal{E}_n(\theta'\varphi)^2/B(p_n) \\
&\le \mathcal{E}(\theta'\varphi)^2/B(p_n) + n^{-\gamma}/2 \\
&\le \overline{\lambda}(G_{p_n})/B(p_n) + n^{-\gamma}/2 \\
&= 1/\overline{\Pi}(p_n) + n^{-\gamma}/2 \\
&\le (3/2)n^{-\gamma}.
\end{aligned}
$$

Thus, $\sup_{f\in\mathcal{F}_{p_n}}|\mathcal{E}_n f - \mathcal{E}f| \le n^{-\gamma}/2$ implies $\bar{\lambda}(G_{n,p_n})/B(p_n) \le (3/2)n^{-\gamma}$. The contrapositive is $B(p_n)/\bar{\lambda}(G_{n,p_n}) < (2/3)n^{\gamma}$ implies $\sup_{f\in\mathcal{F}_{p_n}}|\mathcal{E}_n f - \mathcal{E}f| > n^{-\gamma}/2$. Thus

$$P[B(p_n)/\bar{\lambda}(G_{n,p_n}) < (2/3)n^{\gamma} \text{ i.o. }] \le P\left(\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E}f| > n^{-\gamma}/2 \text{ i.o.}\right).$$

Apply Lemma 3. ∎

THEOREM 4. If p_n satisfies

$$\bar{\Pi}(p_n) \ge n^{\gamma} \qquad 0 \le \gamma < 1/2$$

$$p_n \le n^{\alpha} \qquad 0 \le \alpha < 1 - 2\gamma$$

then

$$P\left(\mathrm{RelBias}(\rho'\hat{\theta}|\{x_t\}) > \sqrt{n}T_p\left\{2 + \sqrt{[(3/2)n^{-\gamma}]}\sqrt{[B(p)/\rho'\rho]}\right\} \text{ infinitely often}\right) = 0.$$

Proof. By Lemma 4 and the inequality immediately preceding Lemma 4

$$\mathrm{RelBias}(\rho'\hat{\theta}|\{x_t\}) \le \sqrt{n}T_p\left\{2 + \sqrt{[\bar{\lambda}(G_{np})/B(p)]}\sqrt{[B(p)/\rho'\rho]}\right\}$$

By Theorem 3, for almost every realization of $\{x_t\}$ there is an N such that n>N implies $\sqrt{[\bar{\lambda}(G_{n,p_n})/B(p_n)]} \le \sqrt{[(3/2)n^{-\gamma}]}$. ∎

Note that $\bar{\Pi}(p) \ge 1$. Thus if $[B(p)/\rho'\rho]$ is bounded $\gamma = 0$ is the least stringent choice in an application of Theorem 4.

4. CONCLUSIONS

We conclude with an application. The most interesting case is rapidly decreasing $\underline{\lambda}(G_p)$ because it is representative of the motivating examples cited in Section 1. For specificity, let the smallest eigenvalue of $X'X/n$ decrease exponentially as $\underline{\lambda}(G_p) = e^{-ap}$ where a is positive, which is a rapidly decreasing sequence. Also, let $B(p) = \rho'\rho = p$. With these choices, Theorem 2 admits rules of the form

$$\ell og[\underline{\Pi}(p)] = ap\ell og(p) = \beta\ell og(n)$$

with β in the interval $0 < \beta < 1/2$. Since $B(p)/\rho'\rho$ is bounded and $\overline{\Pi}(p) \geq 1$ for all p, the relevant value of γ in Theorem 4 is $\gamma = 0$. With this choice, the bound on the relative bias is proportional to

$$\sqrt{n}T_p = \sqrt{n}e^{-bp}.$$

Substituting for n using the rule for p above we have

$$\sqrt{n}T_p = pe^{\alpha p/(2\beta)}T_p.$$

If this term is to decrease, the truncation error of the series expansion T_p must decrease exponentially with p.

Within the second paradigm, this severely restricts the class of functions g^o that admit of asymptotically normal estimates. As seen from the examples, in many instances the implication of this restriction is that g^o must be a very smooth function. However, in many applications this restriction may be more palatable than the boundary conditions one would have to accept in order to slow the rate at which the smallest eigenvalue of $X'X/n$ decreases. Moreover, in some applications, notably neural networks, one does not have the option of modifying the expansion or assumptions and must accept the problem as posed.

REFERENCES

Adams, Robert A. (1975), *Sobolev Spaces*, New York: Academic Press.

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

Barnett, William, John Geweke, and Piyu Yue (1989), "Seminonparametric Bayesian Estimation of the Asymptotically Ideal Model: The AIM Consumer Demand System," in William A. Barnett, James Powell, George E. Tauchen, eds. *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Cambridge: Cambridge University Press, forthcoming.

Christensen, Laurits R., Dale W. Jorgenson and Lawrence J. Lau (1975), "Transcendental Logarithmic Utility Functions," *American Economic Review* 65, 367-383.

Eastwood, Brian J. (1987), *Confidence Interval Construction in Seminonparametric Regression Estimation*, North Carolina State University, Ph.D. dissertation.

Eastwood, Brian J., and A. Ronald Gallant (1987), "Adaptive Truncation Rules for Seminonparametric Estimators That Achieve Asymptotic Normality." H. B. Alexander Foundation Paper 87-53, Graduate School of Business, Univeristy of Chicago.

Edmunds, D. E., and V. B. Moscatelli (1977), "Fourier Approximation and Embeddings of Sobolev Spaces, *Dissertationes Mathematicae* CXLV.

Gallant, A. Ronald (1984), "The Fourier Flexible Form," *American Journal of Agricultural Economics* 66, 204-208.

Gallant, A. Ronald, and Douglas W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55, 363-390.

Gallant, A. Ronald, and George Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications, *Econometrica*, forthcoming.

Gallant, A. Ronald, and Halbert White (1988), "There Exists a Neural Network that Does Not Make Avoidable Mistakes," *Proceedings of the Second Annual IEEE Conference on Neural Networks* I, 657-664.

Huber, Peter J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *The Annals of Statistics* 1, 799-821.

Kolmogorov, A. N., and V. M. Tihomirov (1959), "ϵ-entropy and ϵ-capacity of Function Spaces," *Uspehi Mat. Nauk* 14, 3-86; English translation (1961) *Mathematical Society Translations* 17, 277-364.

Mallows, C. L. (1973), Some Comments on C_p," *Technometrics* 15, 661-675.

Pollard, David (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.

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.

Rumelhart, David E., and James L. McClelland (1986), *Parallel Distributed Processing, Volume 1.* Cambridge: The MIT Press.

Schwarz, G., (1978), "Estimating the Dimension of a Model, *Annals of Statistics* 6, 461-464.

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

Stone, C. J. (1980), "Optimum Rate of Convergence for Nonparametric Estimators," *Annals of Statistics* 5, 595-645.

Tolstov, Georgi P. (1962), *Fourier Series*. New York: Dover Publications.

White, Halbert (1989), "Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Network Models," *Journal of the American Statistical Association*, forthcoming.

Yohai, Victor J., and Ricardo A. Maronna (1979), "Asymptotic Behavior of M-Estimators for the Linear Model," *The Annals of Statistics* 7, 258-268.