# On the Asymptotic Normality of Fourier Flexible Form Estimates*

by

A. Ronald Gallant *and* Geraldo Souza

Department of Statistics
North Carolina State University
Raleigh, NC 27696-8203 USA

Departamento Tecnico Cientifico
EMBRAPA, C.P. 04,0315
70770 Brazilia, DF Brazil

August 1989

*Research supported in Brazil by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) and in the USA by National Science Foundation Grant SES-8808015, North Carolina Agricultural Experiment Station Projects NCO-5593, NCO-3879, and the PAMS Foundation.

# On the Asymptotic Normality of Fourier Flexible Form Estimates*

by

A. Ronald Gallant *and* Geraldo Souza

Department of Statistics
North Carolina State University
Raleigh, NC 27696-8203 USA

Departamento Tecnico Cientifico
EMBRAPA, C.P. 04,0315
70770 Brazilia, DF Brazil

August 1989

## ABSTRACT

Rates of increase in the number of parameters of a Fourier factor demand system that imply asymptotically normal elasticity estimates are characterized. The main technical problem in achieving this characterization is caused by the fact that the eigenvalues of the sample sum of squares and cross products matrix of the generalized least squares estimator are not bounded from below. This problem is addressed by establishing a uniform strong law with rate for the eigenvalues of this matrix so as to relate them to the eigenvalues of the expected sum of squares and cross products matrix. Because the minimum eigenvalue of the latter matrix considered as a function of the number of parameters decreases faster than any polynomial, the rate at which parameters may increase is slower than any fractional power of the sample size.

*Research supported in Brazil by Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) and in the USA by National Science Foundation Grant SES-8808015, North Carolina Agricultural Experiment Station Projects NCO-5593, NCO-3879, and the PAMS Foundation.

# 1 INTRODUCTION

The focus of this paper is the determination of a class of rules for increasing the number of parameters of the Fourier factor demand system that imply asymptotically normal elasticity estimates. The consistency of Fourier flexible form elasticity estimates was established in Elbadawi, Gallant and Souza (1983) and asymptotic distributions of test statistics under normal errors were derived in Gallant (1982).

The Fourier factor demand system is a linear, multivariate model of the form

$$y_t = \Phi'(x_t)\theta + u_t \qquad t = 1, 2, \ldots, n$$

where $\Phi'(x)$ is an M by p matrix whose leading columns correspond to a Translog factor demand system (Christensen, Jorgenson and Lau, 1975) and remaining columns are derived from a multivariate Fourier series expansion. The true data generating model is presumed to be

$$y_t = f^o(x_t) + e_t \qquad \mathcal{E}(e_t) = 0, \quad \mathcal{E}(e_t e_t') = \Omega$$

where $f^o$ is derived from a log cost function $g^o$ using Shephard's lemma (Deaton and Muellbauer, 1980). The (N+1)-dimensional vector x contains log factor prices and log output as elements; $g^o$ is defined over a closed, bounded rectangle $\mathcal{X} \subset \mathbb{R}^{N+1}$. Our methods of proof will accommodate drift so that one can write $f_n^o$ and $g_n^o$ if desired.

Each equation of the system is a linear series expansion of the sort studied by Andrews (1989) which is the most recent and comprehensive paper on the subject. Andrews' paper contains an extensive bibliography and some of the sharpest results available.

In order to address our problem, we need to extend Andrews' analysis in two directions. We must make the multivariate extension. And, we must explicitly account for the fact that the sequence of minimum eigenvalues corresponding to the sequence of p by p matrices

$$G_p = \mathcal{E}\, \Phi(x)\Omega^{-1}\Phi'(x) \qquad\qquad p = 1, 2, \ldots$$

is rapidly decreasing; that is, $p^m\lambda_{min}(G_p)$ tends to zero for every positive integer m as p increases. Andrews bounds this sequence from below so his results do not apply to our problem. The first extension is reasonably straightforward. The second is more delicate. To accomplish it, we prove a uniform strong law with rate for the eigenvalues of the matrix

$$G_{np} = \frac{1}{n} \sum_{t=1}^{n} \Phi(x_t)\Omega^{-1}\Phi'(x_t)$$

using results from the empirical process literature. For the technically inclined, this uniform strong law is one of the more interesting aspects of the paper. We study deterministic rules but these are easily extended to adaptive rules using results due to Eastwood and Gallant (1987); see Andrews (1989) for examples.

We derive sufficient conditions such that elasticity estimates are asymptotically normal and examine the class $\mathcal{P}$ of rules $\{p_n\}_{n=1}^{\infty}$ and cost functions $\mathfrak{b}$ that satisfy them. Not surprisingly, given that $\lambda_{min}(G_p)$ is rapidly decreasing, rules in $\mathcal{P}$ grow slowly, slower than any fractional power of n. The functions in $\mathfrak{b}$ are infinitely many times differentiable. This is not as restrictive an assumption as might first appear as the collection of infinitely many times differentiable functions defined on $\mathcal{X}$ is a dense subset of any Sobolev space $\mathcal{W}_{m,q,\mathcal{X}}$ (Adams, 1975); $\mathcal{W}_{m,q,\mathcal{X}}$ is defined in the next section. Thus, it is a very rich collection with which to describe production technologies. Alternatively, the drift mechanism can be exploited to expand the class $\mathcal{D}$; see Section 5. This amounts to an assumption that as data is acquired, the true model becomes an increasingly rich departure from the Translog which is the leading special case. Some would regard this assumption as a realistic description of the attitudes of practitioners and others would find it unacceptable.

The reason that the minimum eigenvalues of $G_p$ decline is that the Fourier flexible form incorporates two modifications to the classical multivariate Fourier expansion. To get rid of unacceptable boundary conditions, the domain $\mathcal{X}$ is effectively a subset of the natural domain $Q = \prod_{i=1}^{N+1}[0, 2\pi]$ of a multivariate Fourier series expansion of the form $\mathcal{X} = \prod_{i=1}^{N+1}[\epsilon, 2\pi-\epsilon]$. To improve performance in finite samples and to provide a means to test interesting hypotheses, a Translog model is added as the leading term of the expansion. The rates of decrease due to various combinations of these two modifications to the classical expansion are displayed in Table 1. We work out the implications of the last entry in Table 1 because it is the expansion used in practice. This analysis also covers the third entry. The other two admit fractional powers of n as rules $\{p_n\}$.

The major limitations of the paper are twofold. We assume homogeneous errors. While it is clear that we could accommodate heterogeneity by adapting Eicker's analysis (1967, p. 77) to our situation, as does Andrews to his, we do not do so because what is at present a tidy, clean analysis would become a cluttered mess, distracting from the main focus of the paper which is to gain a qualitative feel for the rates of expansion that declining eigenvalues permit.

We consider the generalized least squares estimator rather than the seemingly unrelated estimator because the latter is not essential to the objective of the paper and the technical problems in treating a random scale estimate $\hat{\Omega}$ that depends the error process appear formidable. Apparently, a specialized collection of uniform strong laws with rates would need to be established. Were an independent estimate available, our analysis would cover seemingly unrelated estimates. An estimate computed from a holdout sample would satisfy this condition.

Table 1. Rates at which eigenvalues decrease*

| Expansion | Minimum eigenvalues of $G_p$ |
|---|---|
| Fourier series on Q | Bounded away from zero |
| Translog + Fourier series on Q | Decrease at a polynomial rate |
| Fourier series on $\mathcal{X}$ | Rapidly decreasing |
| Translog + Fourier series on $\mathcal{X}$ | Rapidly decreasing |

*See Section 5 for a verification.

## 2. ESTIMATION ENVIRONMENT

The producer's cost function $c(p,u)$ gives the minimum cost of producing output $u$ during a given period of time using inputs $q = (q_1, q_2, \ldots, q_N)'$ at prices $p = (p_1, p_2, \ldots, p_N)'$. It is more convenient to work in terms of log cost as a function of log prices and log output. Accordingly, let

$$\ell_i = \ell n\ p_i + \ell n\ a_i, \qquad i = 1, 2, \ldots, N$$

$$v = \lambda_{N+1}(\ell n\ u + \ell n\ a_{N+1})$$

and put

$$g(x) = g(\ell,v)$$

$$= \ell n\ c[(1/a_1)\exp(\ell_1), \ldots, (1/a_N)\exp(\ell_N), (1/a_{N+1})\exp(v/\lambda_{N+1})]$$

where $\ell = (\ell_1, \ell_2, \ldots, \ell_N)'$ and $x = (\ell', v)'$.

The $a_1, \ldots, a_{N+1}, \lambda_{N+1}$ are positive scaling factors. Their choice has no substantive impact since a choice of $a_i$ other than unity for $i = 1, 2, \ldots, N$ only amounts to changing the units of measurement in which factor prices are stated. If one works with quantities as data, the ith quantity in original units would be divided by $a_i$ to get the quantity in the new units. If one works with cost shares as data, there is no need to make any compensating adjustments as the price and quantity adjustments would cancel, viz. $s_i = \exp(\ell_i)(q_i/a_i) = (a_i p_i)(q_i/a_i)$. The choice of $\lambda_{N+1}$ and $a_{N+1}$ amounts to a choice of a scale of measurement for output and is also irrelevant to the substance of the discussion. We shall restrict attention to some closed rectangle $\mathcal{X}$ in $\mathbb{R}^{N+1}$ that contains the observed data and shall not attempt to approximate g off $\mathcal{X}$.

The scaling factors are to be chosen so that $\prod_{i=1}^{N+1}[\epsilon,\ 2\pi-\epsilon] \subset \lambda\mathcal{X} \subset \prod_{i=1}^{N+1}[0,\ 2\pi]$ for some small $\epsilon$ and some positive $\lambda$ where $\lambda\mathcal{X}$ indicates multiplication of every element of $\mathcal{X}$ by $\lambda$; see Gallant (1982) for additional details. We shall assume that the observed sequence $\{x_t\}_{t=1}^{n}$ is a random sample from a distribution function $\mu(x)$ defined over $\mathcal{X}$ with a continuous, bounded density function.

The cost function itself is assumed to possess the properties of linear homogeneity, monotonicity, and concavity. Letting 1 denote a vector whose entries are all ones, letting $\nabla g = (\partial/\partial\ell)g(\ell,v)$, and letting $\nabla^2 g = (\partial^2/\partial\ell\partial\ell')g(\ell,v)$, the equivalent conditions on the log cost function $g(\ell,v)$ are [Gallant (1982)]

$R_0$. Linear homogeneity: $g(\ell + \tau 1,\ v) = \tau + g(\ell,v)$,

$R_1$. Monotonicity: $\nabla g > 0,\ 1'\nabla g = 0,\ (\partial/\partial v)g(\ell,v) > 0$,

$R_2$. Concavity: $\nabla^2 g + \nabla g \nabla' g - \mathrm{diag}(\nabla g)$ is a negative semi-definite matrix of rank N-1 with 1 being the eigenvector with root zero.

Letting $s' = (p_1q_1,\ p_2q_2,\ \ldots,\ p_Nq_N)/(\sum_{t=1}^{N} p_iq_i)$ be the N-vector of input cost shares, the firm's factor demand system is given by Shephard's lemma as

$$s = (\partial/\partial\ell)g(\ell,v).$$

Elasticities of substitution $\sigma_{ij}$ at a point $x^0 = (\ell^0,v^0)$ are elements of the matrix

$$\Sigma = [\mathrm{diag}(\nabla g)]^{-1}[\nabla^2 g + \nabla g\nabla' g - \mathrm{diag}(\nabla g)][\mathrm{diag}(\nabla g)]^{-1},$$

evaluated at $x^0$ and price elasticities $\eta_{ij}$ are elements of the matrix

$$\eta = \Sigma \, \mathrm{diag}(\nabla g),$$

evaluated at $x^0$ where the rows of $\eta$ correspond to inputs and the columns to prices [see Gallant (1982)]. We shall let $\sigma(g)$ represent a generic entry from these matrices.

This is a deterministic version of factor demand theory and can be regarded as implying observed factor cost shares follow some distribution with location parameter $(\partial/\partial\ell)g(\ell,v)$. When just factor cost shares are observed, the most common distributional assumption in applied work is that

$$\tilde{y}_t = \tilde{f}(x_t) + \tilde{e}_t, \qquad \mathcal{E}(\tilde{e}_t) = 0, \qquad \mathcal{C}(\tilde{e}_t,\tilde{e}_t') = \tilde{\Omega}$$

where $\tilde{y}_t$ is the vector of observed factor cost shares $s_t$ with the last element discarded, and $\tilde{f}(x_t)$ is $(\partial/\partial\ell)g(\ell_t,v_t)$ with the last element discarded. Since shares sum to one, any distributional fact regarding the last share can be gotten from the identity $s_{Nt} = 1 - \sum_{j=1}^{N-1}\tilde{y}_{jt}$. Were the last share not discarded, the variance of the errors would be singular. Since the statistical methods that we discuss below have the property that estimates are invariant to which share is designated last, deletion is the simplest way to handle the singularity.

When, in addition, total factor cost is observed, the model is

$$y_t = f(x_t) + e_t, \qquad \mathcal{E}(e_t) = 0, \qquad \mathcal{C}(e_t,e_t') = \Omega$$

where: $y_t$ is an N-vector with observed total factor cost as its first element and $\tilde{y}_t$ filling in the remaining elements; $f(x_t)$ is an N-vector with $g(x_t)$ as its first element and $\tilde{f}(x_t)$ at the tail; and, similarly, $e_t$ has an additional error at the head and $\tilde{e}_t$ at the tail.

We shall need a compact notation for high order partial derivatives:

$$D^{\lambda}g(x) = (\partial^{\lambda_1}/\partial x_1^{\lambda_1})\cdots(\partial^{\lambda_{N+1}}/\partial x_{N+1}^{\lambda_{N+1}})g(x)$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{N+1})$ has nonnegative integers as components. The order of the partial derivative is $|\lambda| = \sum_{j=1}^{N+1}|\lambda_j|$ and when $|\lambda| = 0$ take $D^0 g(x) = g(x)$. In terms of this notation, the Sobolev norm is:

**Sobolev norm.** For q with $1 \le q < \infty$ the Sobolev norm of g(x) with respect to a distribution $\mu$ defined over $\mathcal{X}$ is

$$\|g\|_{m,q,\mu} = \left[\sum_{|\lambda|\le m} \int_{\mathcal{X}} |D^{\lambda}g(x)|^q d\mu(x)\right]^{1/q}.$$

For $q = \infty$ the norm is

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

To avoid clutter, let $\|g\|_{m,\infty,\mu}$ represent $\|g\|_{m,\infty,\mathcal{X}}$ in a statement such as "$\|g\|_{m,q,\mu} \le B$ for all $1 \le q \le \infty$". □

The advantage of the Sobolev norm is that if the norm $\|e\|_{m,q,\mu}$ of the error $e = g - \hat{g}$ is small for m = 2 then the error when approximating g-elasticities by $\hat{g}$-elasticities will be small. Stated differently, elasticities are continuous in the Sobolev topology. This being the case, it is natural to restrict attention to cost functions contained in some Sobolev space

$$\mathcal{W}_{m,q,\mu} = \{g\colon \|g\|_{m+1,q,\mu} < \infty\}$$

that has $m \ge 2$. The Fourier flexible form can be used to construct dense subsets of $\mathcal{W}_{m,q,\mu}$. It is defined in terms of elementary multi-indexes:

**Elementary multi-indexes.** A multi-index is a vector whose elements are integers. Let

$$\mathcal{K}_{N+1} = \{k: |k| \le K\}$$

denote the set of multi-indexes of dimension N+1 whose length $|k| = \sum_{i=1}^{N+1} |k_i|$ is bounded by K. First, delete from $\mathcal{K}_{N+1}$ the zero vector and any k whose first non-zero element is negative. Second, delete any k whose elements have a common integer divisor. Third, arrange the k which remain into a sequence

$$\mathcal{K}'_{N+1} = \{k_\alpha: \alpha = 1, 2, \ldots, A\}$$

such that $k_1, k_2, \ldots, k_{N+1}$ are the elementary vectors and $|k_\alpha|$ is non-decreasing in $\alpha$. Define J to be the smallest positive integer with

$$\mathcal{K}_{N+1} \subset \{jk_\alpha: \alpha = 1, 2, \ldots, A; j = 0, \pm 1, \pm 2, \ldots, \pm J\}. \qquad \square$$

This construction can be automated using FORTRAN code in Monahan (1981) or using PROC FOURIER in SAS (1986). In terms of this notation, the Fourier form is written as:

**Fourier flexible form.** Define

$$g_K(x|\theta) = u_0 + b'x + (1/2)x'Cx + \sum_{\alpha=1}^{A} \left\{ u_{0\alpha} + 2 \sum_{j=1}^{J} [u_{j\alpha}\cos(j\lambda k'_\alpha x) - v_{j\alpha}\sin(j\lambda k'_\alpha x)] \right\}$$

where $C = -\sum_{\alpha=1}^{A} u_{0\alpha}\lambda^2 k_\alpha k'_\alpha$ and

$$\theta = (u_0, \theta'_{(0)}, \theta'_{(1)}, \ldots, \theta'_{(A)})',$$

$$\theta_{(0)} = b' = (c', b_{N+1})',$$

$$\theta_{(\alpha)} = (u_{0\alpha}, u_{1\alpha}, v_{1\alpha}, \ldots, u_{J\alpha}, v_{J\alpha})' \qquad \square$$

The length of the parameter vector $\theta$ is $p = 1 + (N + 1) + A(1 + 2J) \approx K^N$ where $a(K) \approx b(K)$ means that there exist two positive constants such that $c_1 b(K) \le a(K) \le c_2 b(K)$ for all K. The following result implies that

$$\mathcal{M}' = \{g_K(\cdot|\theta) : \theta \in \mathbb{R}^p,\ K = 0, 1, \dots \}$$

is a dense subset of $\mathcal{W}_{m,q,\mu}$.

**THEOREM 0.** (Gallant, 1982) If g(x) is continuously differentiable up to the order m on an open set containing the closure of $\mathcal{X}$ then there is a sequence $g_K(\cdot|\bar{\theta}_K)$ from $\mathcal{M}'$ such that

$$\| g - g_K(\cdot|\bar{\theta}_K) \|_{\ell,q,\mu} = o(K^{-m+\ell+\epsilon}) \quad \text{as} \quad K \to \infty$$

for every q with $1 \le q \le \infty$, every $\ell$ with $0 \le \ell < m$, and every $\epsilon > 0$. If g is also linear homogeneous and $\mathcal{M}'_0$ denotes the linear homogeneous functions in $\mathcal{M}'$ the result is true with $\mathcal{M}'_0$ replacing $\mathcal{M}'$. Similarly, if $\mathcal{M}'_{02}$ denotes the functions in $\mathcal{M}'$ that are both linear homogeneous and concave and g satisfies both restrictions and so on for other combinations of the three restrictions listed above. □

It is easy to impose linear homogeneity on the Fourier flexible form: Restrict the coefficients $\theta_{(0)} = b' = (c', b_{N+1})'$ so that the leading N elements sum to unity, that is, $1'c = 1$, and put $\theta_{(\alpha)} = 0$ if the leading N elements of $k_\alpha = (r'_\alpha, k_{N+1})'$ do not sum to zero, that is, if $1'r_\alpha \ne 0$. The latter restriction is equivalent to restricting the choice of elementary multi-indexes to the set

$$\mathcal{K}'_{0,N+1} = \{k_\alpha \in \mathcal{K}'_{N+1} : k_\alpha = (r'_\alpha, k_{N+1}),\ 1'r_\alpha = 0\}$$

when constructing $g_K(x|\theta)$. Methods for imposing the other restrictions are discussed in Gallant and Golub (1984).

We shall assume throughout that the linear homogeneity restriction $R_0$ has been imposed, as is usually done in practice. This is an important assumption as it amounts to reducing the dimension of x from N + 1 to N; see Lemma 1 of Gallant (1982). Were this assumption dropped, N would have to be replaced by N + 1 in every result regarding a rate that we report.

We propose to fit the model

$$y_t = f_K(x_t|\theta) + u_t \qquad t = 1, 2, \ldots, n$$

to data that was actually generated according to

$$y_t = f(x_t) + e_t, \qquad \mathcal{E}(e_t) = 0, \qquad \mathcal{C}(e_t, e_t') = \Omega$$

using generalized least squares. Above, $f_K(x|\theta)$ is constructed from $g_K(x|\theta)$ exactly as f(x) is constructed from g(x). Generalized least squares estimates estimates are gotten by minimizing

$$s(\theta,\Omega) = \frac{1}{n}\sum_{t=1}^{n}[y_t - f_K(x_t|\theta)]'\Omega^{-1}[y_t - f_K(x_t|\theta)].$$

As $f_K(x|\theta)$ is a linear function of $\theta$ it can be represented as $\Phi'(x)\theta$ where $\Phi'$ is an N by p matrix. Thus $s(\theta,\Omega)$ is a quadratic form in $\theta$

$$s(\theta,\Omega) = \frac{1}{n}\sum_{t=1}^{n}[y_t - \Phi'(x_t)\theta]'\Omega^{-1}[y_t - \Phi'(x_t)\theta].$$

with minimum

$$\hat{\theta} = \left[\frac{1}{n}\sum_{t=1}^{n} \Phi(x_t)\Omega^{-1}\Phi'(x_t)\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n} \Phi(x_t)\Omega^{-1}y_t\right]$$

From this estimate, g can be estimated by $\hat{g}_K = g_K(\cdot|\hat{\theta})$ and an elasticity $\sigma = \sigma(g)$ by $\hat{\sigma}_K = \sigma[g_K(\cdot|\hat{\theta})]$. If the cost function is not estimated together with factor demands, then $\tilde{f}$ and $\tilde{\Omega}$ replace f and $\Omega$ above and $\Phi'(x)$ has N - 1 rows. In order to have a generic notation that represents either case, let $\Phi'$ have M rows.

We shall henceforth adopt the convention that as K is increased, columns are appended to the end of $\Phi'$ rather than inserted. This implies a rearrangement of the elements of $\theta$ from that given above but we have no need to devise a notation to keep track of it because any finite dimensional calculation that concerns us is invariant to the ordering of the columns of $\Phi'$. Also, because of the way that the matrix C is parameterized, some columns of $\Phi'(x)$ will be exact linear combinations of predecessors for every x in $\mathcal{X}$. We assume that these have been deleted in forming $\Phi'$ and $\theta$.

**Infinite dimensional representation.** With these conventions, one can choose a point $\theta_\infty = (\theta_1, \theta_2, \ldots)$ in $\mathbb{R}^\infty$ and use its leading elements $\theta_K = (\theta_1, \theta_2, \ldots, \theta_p)$ to compute $g_K(\cdot|\theta_K)$. If there is a g in $\mathcal{W}_{m,\infty,\mathcal{X}}$ such that

$$\lim_{K\to\infty} \|g_K(\cdot|\theta_K) - g\|_{m,\infty,\mathcal{X}} = 0$$

write $g_\infty(\cdot|\theta_\infty)$ to represent g. Every g in $\mathcal{W}_{m+1,\infty,\mathcal{X}}$ has such a representation (Edmunds and Moscatelli, 1977). We shall usually suppress the subscripts $\infty$ and K and let the context determine whether $\theta_\infty$ or its truncation $\theta_K$ is intended. $\square$

Our methods of proof can accommodate drift. As drift may be relevant to hypothesis testing or other applications, we will assume that the true cost function is indexed by the sample size n and denote it by $g_n^0$. The corresponding data generating mechanism is denoted as $y_{nt} = f_n^o(x_t) + e_t$ with $e_t$ as above and $f_n^o$ constructed from $g_n^o$ as above.

An elasticity $\sigma(g)$ is called an evaluation functional since it is necessary to be able to evaluate g, $\nabla g$, and $\nabla^2 g$ at the point $x^o$ in order to compute it. Thus, our interest is focused on estimates of $D^\lambda g_n^o(x^o)$ for $|\lambda| \le 2$. The estimate $D^\lambda g_K(x^o|\hat{\theta})$ is a linear function in $\hat{\theta}$ of the form $\rho'\hat{\theta}$. As an elasticity estimate is a rational function of such estimates its asymptotic distribution can be gotten by the delta method if the asymptotic normality of $\rho'\hat{\theta} - D^\lambda g_n^o(x^o)$ can be established for each $\lambda$ with $|\lambda| \le 2$. In the next section we determine rules $p_n$ relating p to n, equivalently rules $K_n$ relating K to n, such that $\rho'\hat{\theta}$ is asymptotically normal when centered about its conditional expectation $\mathcal{E}(\rho'\hat{\theta}|\{x_t\})$. In the section after that, we determine the subset of rules $K_n$ that drive the bias term $\mathcal{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_n^o(x^o)$ to zero rapidly enough to be negligible relative to the error term $\rho'\hat{\theta} - \mathcal{E}(\rho'\hat{\theta}|\{x_t\})$. The sum of the error and bias terms is the estimate centered about the object of interest $\rho'\hat{\theta} - D^\lambda g_n^o(x^o)$ which will in consequence be asymptotically normal.

# 3. ASYMPTOTIC NORMALITY OF THE RELATIVE ERROR

Of the assumptions of the previous section, the subset required for the results of this section are as follows.

**Assumptions.** The observed data is generated according to

$$y_{nt} = f_n^o(x_t) + e_t \qquad t = 1, 2, \ldots, n; \quad n = 1, 2, \ldots$$

where $\{f_n^o\}$ is a sequence of functions that map $\mathcal{X}$, a subset of $\mathbb{R}^{N+1}$, into $\mathbb{R}^M$. Throughout we shall write $y_t$ instead of $y_{nt}$; N and M are finite and fixed throughout. The error process $\{e_t\}_{t=1}^{\infty}$ is an iid sequence of random variables that have common distribution P(e) with mean zero and variance-covariance matrix

$$\Omega = \int ee'dp(e).$$

We assume that $\Omega$ is nonsingular and factor its inverse as

$$\Omega^{-1} = P'P.$$

The process $\{x_t\}_{t=1}^{\infty}$ is an iid sequence of random variables with common distribution $\mu(x)$ defined on $\mathcal{X}$; $\{x_t\}$ and $\{e_t\}$ are independent processes. We consider the random variable

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

where

$$G_{np} = \frac{1}{n}\sum_{t=1}^{n}\Phi(x_t)\Omega^{-1}\Phi'(x_t) \ .$$

and $\Phi'(x)$ maps each x in $\mathcal{X}$ to an M by p matrix. Note that $\Phi$ depends only on p and that $\hat{\theta}$ depends on both n and p. The objective of this section is to find rules $p_n$ relating p to n such that

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

is asymptotically normally distributed where $\text{Var}(\rho'\hat{\theta}|\{x_t\}) = (1/n)\rho' G_{np}^{-1}\rho$. In this section, but not elsewhere, $\rho$ is an arbitrary, nonzero vector in $\mathbb{R}^p$, p can take on any positive integer value, and the choice of $\rho$ for one value of (p,n) need not have any relation to the choice for another. □

The ambiguity that arises when $G_{np}^{-1}$ does not exist is resolved by putting the linearly dependent rows and columns of $G_{np}$ to zero to get $G_0$, putting the corresponding elements of $\rho$ to zero, and letting $G_{np}^{-1}$ be a g-inverse of $G_0$. Thus defined, $G_{np}^{-1}$ is unique and can be factored. When the conditions of Theorem 5 are in force, $G_{np}$ is nonsingular with probability one for all n large enough so one could adopt any resolution of the ambiguity. However, with this construction the algebra below is correct whether $G_{np}$ is nonsingular or not.

Fix a realization of $\{x_t\}$. Then $\text{RelErr}(\rho'\hat{\theta}|\{x_t\})$ is a linear function of the errors, *viz.*

$$\text{RelErr}(\rho'\hat{\theta}|\{x_t\}) = \sum_{t=1}^{n}\left[\frac{(1/n)\rho' G_{np}^{-1}\Phi(x_t)P'}{\sqrt{[(1/n)\rho' G_{np}^{-1}\rho]}}\right] Pe_t \ ,$$

We have the following result:

**THEOREM 1.** If $\{x_t\}$ is a fixed sequence (a nonrandom sequence) then

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

implies that $\mathrm{RelErr}(\rho'\hat{\theta}|\{x_t\})$ is asymptotically normally distributed.

**Proof.** We can write the relative error more compactly as

$$\mathrm{RelErr}(\rho'\hat{\theta}|\{x_t\}) = \sum_{t=1}^{n}(s_{nt}'u_t)/s_n$$

where $s_{nt}' = (1/n)\rho' G_{np}^{-1}\Phi(x_t)P'$, $u_t = Pe_t$, and $s_n^2 = \sum_{t=1}^{n} s_{nt}' s_{nt}$. The result will follow if we verify Lindeberg's condition (Billingsley, 1979, p.310)

$$\lim_{n\to\infty} \sum_{t=1}^{n} (s_n^2)^{-1} \int I(|s_{nt}'u| \ge \epsilon s_n)\ (s_{nt}'u)^2\ dP(e) = 0$$

for every $\epsilon > 0$ where $u = Pe$ and $I(A)$ denotes the indicator function of a set A. By the Cauchy-Schwartz inequality

$$\sum_{t=1}^{n} (s_n^2)^{-1} \int I(|s_{nt}'u| \ge \epsilon s_n)\ (s_{nt}'u)^2\ dP(e)$$

$$\le \sum_{t=1}^{n} (s_n^2)^{-1} \int I(\|s_{nt}\|\|u\| \ge \epsilon s_n)\ (\|s_{nt}\|\|u\|)^2\ dP(e)$$

$$\le \sum_{t=1}^{n} (\|s_{nt}\|/s_n)^2 \int I\left[\sqrt{e'\Omega^{-1}e} \ge \epsilon \left(\sup_{1\le t\le n} \|s_{nt}\|/s_n\right)^{-1}\right] e'\Omega^{-1}e\ dP(e)$$

$$= \int I\left[\sqrt{e'\Omega^{-1}e} \ge \epsilon \left(\sup_{1\le t\le n} \|s_{nt}\|/s_n\right)^{-1}\right] e'\Omega^{-1}e\ dP(e)$$

which tends to zero with n because $\lim_{n\to\infty}\sup_{1\le t\le n} \|s_{nt}\|/s_n = 0$ and $e'\Omega^{-1}e$ is integrable. $\square$

Let $Z(x) = \Phi(x)P'$ with elements denoted as $z_{j\alpha}(x)$ where $j = 1, \ldots, p$ is the row index and $\alpha = 1, \ldots, M$ is the column index. Put

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

The following result relates the condition for asymptotic normality to the ratio of $B(p)$ to the smallest eigenvalue of $G_{np}$. It is stated and proved in terms of the following notation: $\lambda_{min}(G)$ and $\lambda_{max}(G)$ denote the smallest and largest eigenvalues of a matrix $G$, $tr(G)$ the trace of $G$, $G_{(\alpha)}$ the $\alpha$th column of $G$, and $G^{-1/2}$ the Cholesky factor of $G^{-1}$; that is, $G^{-1} = (G^{-1/2})'G^{-1/2}$.

**THEOREM 2.**

$$\sup_{1 \le t \le n} \frac{\| (1/n)\rho' G_{np}^{-1} \Phi(x_t) P' \|}{\surd[(1/n)\rho' G_{np}^{-1} \rho]} \le \left[ \frac{B(p)}{n \, \lambda_{min}(G_{np})} \right]^{1/2}$$

**Proof.** The square of the left hand side is

$$\frac{\rho' G_{np}^{-1} \Phi(x_t) \Omega^{-1} \Phi'(x_t) G_{np}^{-1} \rho}{n \rho' G_{np}^{-1} \rho}$$

$$\le \frac{\rho' G_{np}^{-1} \rho \; \lambda_{max}[G_{np}^{-1/2} \Phi(x_t) \Omega^{-1} \Phi'(x_t)(G_{np}^{-1/2})']}{n \rho' G_{np}^{-1} \rho}$$

$$\le (1/n) tr[G_{np}^{-1/2} \Phi(x_t) \Omega^{-1} \Phi'(x_t)(G_{np}^{-1/2})']$$

$$= (1/n) tr[P \Phi'(x_t)(G_{np}^{-1}) \Phi(x_t) P']$$

$$= (1/n) \sum_{\alpha=1}^{M} \{[Z(x_t)]_{(\alpha)}\}'(G_{np}^{-1})[Z'(x_t)]_{(\alpha)}$$

$$\leq (1/n) \sum_{\alpha=1}^{M} \{[Z(x_t)]_{(\alpha)}\}'[Z(x_t)]_{(\alpha)} \lambda_{max}(G_{np}^{-1})$$

$$\leq \frac{B(p)}{n\ \lambda_{min}(G_{np})} \qquad \square$$

The implication of this result is that if $p_n$ is a rule relating p to n with $\ell im_{n\to\infty} B(p_n)/[n\ \lambda_{min}(G_{np_n}) = 0$ then $\mathrm{RelErr}(\rho'\hat{\theta}|\{x_t\})$ is asymptotically normal 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 $\mathrm{RelErr}(\rho'\hat{\theta}|\{x_t\})$ is asymptotically normal as well.

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

$$G_p = \int \Phi(x)\Omega^{-1}\Phi'(x)\ d\mu(x)$$

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

$$\mathcal{F}_p = \{\theta'\Phi(x)\Omega^{-1}\Phi'(x)\theta/B(p) : \theta'\theta = 1,\ \theta \in \mathbb{R}^p\}$$

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, $G_{np} = \mathcal{E}_n \Phi\Omega^{-1}\Phi'$ and $G_p = \mathcal{E}\, \Phi\Omega^{-1}\Phi'$. Note, there is a $\theta$ with $\theta'\theta = 1$ and $\theta' G_{np}\theta = \lambda_{min}(G_{np})$ so there is an f in $\mathcal{F}_p$ with $\lambda_{min}(G_{np})/B(p) = \mathcal{E}_n f$.

**LEMMA 1.** The number of $\epsilon$-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 patterned 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 $\theta$ is a point on the surface, then an $\epsilon/2$ neighborhood must intersect one of these balls; hence $\theta$ is within $\epsilon$ 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 $M \le V/v$.

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

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

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

[by the mean value theorem]

$$\le 2p(1 + \epsilon/2)^{p-1}/(\epsilon/2)^{p-1}$$

□

**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 $\mathscr{F}_p$ is bounded and a computation of the metric entropy of $\mathscr{F}_p$. The first two paragraphs take care of these details. The third paragraph has the main argument.

If f is in $\mathscr{F}_p$ then $\sup_{x\in\mathscr{X}} |f(x)| \le 1$ since

$$\begin{aligned}\theta'\Phi\Omega^{-1}\Phi'\theta/B(p) &= \theta'ZZ'\theta/B(p)\\ &= \textstyle\sum_{\alpha=1}^{M}(\theta'Z_{(\alpha)})^2/B(p)\\ &\le \textstyle\sum_{\alpha=1}^{M}\|\theta\|^2\|Z_{(\alpha)}\|^2/B(p)\\ &= \|\theta\|^2\textstyle\sum_{\alpha=1}^{M}\|Z_{(\alpha)}\|^2/B(p)\\ &= \|\theta\|^2\textstyle\sum_{\alpha=1}^{M}\sum_{j=1}^{p}[z_{j\alpha}(x)]^2/B(p)\\ &\le 1.\end{aligned}$$

A consequence of this bound is $\mathscr{E}_n f^2 \le 1$ and $\mathrm{Var}(\mathscr{E}_n f) \le 1/n$ for each f in $\mathscr{F}_p$.

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 $g_j = \tilde{\theta}_j'\Phi\Omega^{-1}\Phi'\tilde{\theta}_j/B(p)$. Now $f = \theta'\Phi\Omega^{-1}\Phi'\theta/B(p)$ must have $\theta$ in some ball so

$$\begin{aligned}\min_j |g_j - f| &= \min_j |\theta'\Phi\Omega^{-1}\Phi'\theta/B(p) - \tilde{\theta}_j'\Phi\Omega^{-1}\Phi'\tilde{\theta}_j|/B(p)\\ &= \min_j [(\theta - \tilde{\theta}_j)'Z][Z'(\theta + \tilde{\theta}_j)]/B(p)\\ &\le \min_j \|(\theta - \tilde{\theta}_j)'Z\|\|Z'(\theta + \tilde{\theta}_j)\|/B(p)\end{aligned}$$

$$= \min_j \sqrt{\left\{\sum_{\alpha=1}^M [(\theta - \tilde\theta_j)'Z_{(\alpha)}]^2 \sum_{\alpha=1}^M [(\theta + \tilde\theta_j)'Z_{(\alpha)}]^2\right\}}/B(p)$$

$$\le \min_j \sqrt{\left\{\left[\sum_{\alpha=1}^M \|\theta - \tilde\theta_j\|^2 \|Z_{(\alpha)}\|^2\right]\left[\sum_{\alpha=1}^M \|\theta + \tilde\theta_j\|^2 \|Z_{(\alpha)}\|^2\right]\right\}}/B(p)$$

$$= \min_j \|\theta - \tilde\theta_j\| \|\theta + \tilde\theta_j\| \left[\sum_{\alpha=1}^M \|Z_{(\alpha)}\|^2\right]/B(p)$$

$$\le \min_j \|\theta - \tilde\theta_j\| [\|\theta\| + \|\tilde\theta_j\|] \sum_{\alpha=1}^M \sum_{j=1}^p [z_{j\alpha}(x)]^2/B(p)$$

$$\le (\epsilon/2)(2)B(p)/B(p) \;\; \le \epsilon.$$

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

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

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

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

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

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:

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

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

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

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

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

We can now state and prove the main result of this section.

**THEOREM 4.** If $p_n$ satisfies

$$B(p_n)/\lambda_{min}(G_{p_n}) \le n^{\beta} \qquad 0 \le \beta < 1/2$$

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

then

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

**Proof.** Suppose $\sup_{f\in\mathcal{F}_{p_n}} |\mathcal{E}_n f - \mathcal{E} f| \le n^{-\beta}/2$. There is an $f = \theta'\Phi\Omega^{-1}\Phi'\theta/B(p_n)$ in $\mathcal{F}_{p_n}$ such that $\lambda_{min}(G_{n,p_n})/B(p_n) = \mathcal{E}_n f$ whence

$$\begin{aligned}\lambda_{min}(G_{n,p_n})/B(p_n) &= \mathcal{E}_n \theta'\Phi\Omega^{-1}\Phi'\theta/B(p_n)\\ &\ge \mathcal{E}\theta'\Phi\Omega^{-1}\Phi'\theta/B(p_n) - n^{-\beta}/2\\ &\ge \lambda_{min}(G_{p_n})/B(p_n) - n^{-\beta}/2\\ &\ge n^{-\beta}/2.\end{aligned}$$

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

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

Apply Theorem 3. $\square$

Asymptotic normality follows immediately.

**THEOREM 5.** If $p_n$ satisfies

$$B(p_n)/\lambda_{min}(G_{p_n}) \le n^{\beta} \qquad 0 \le \beta < 1/2$$

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

then

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

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

**Proof.** By Theorem 4 $P[\, B(p_n)/[n\, \lambda_{min}(G_{n,p_n})] > n^{\beta-1}/2 \;\; i.\, o.] = 0$ whence $\ell im_{n\to\infty}\, B(p_n)/[n\, \lambda_{min}(G_{n,p_n})] = 0$ except for realizations of $\{x_t\}$ that correspond to an event in the underlying probability space that occurs with probability zero. □

# 4. A BOUND ON THE RELATIVE BIAS.

In addition to the assumptions listed in the previous section we require the following assumptions for the results of this section.

**Assumptions (continued).** The log cost functions $g_n^o(x)$ are (m+1)-times continuously differentiable on an open set containing $\mathcal{X}$ which is a closed, nonempty rectangle in $\mathbb{R}^{N+1}$. Linear homogeneity is imposed on $g_n^o$ and $g_K(\cdot|\theta)$. The distribution $\mu(x)$ has a continuous density function defined over $\mathcal{X}$. The objective of this section is to find rules $K_n$ relating K to n such that the relative bias

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

tends to zero with n where $|\lambda| = \ell \le m$. Unlike the previous section where p could be any positive integer and $\rho$ was arbitrary, in this section each p will correspond to some K so that p takes discrete jumps as K increases and $\rho$ is defined by the relation: $D^\lambda g_K(x^o|\theta) = \rho'\theta$ for all $\theta$ in $\mathbb{R}^p$. □

The bound on relative bias that we derive is stated in terms of the error in a Fourier flexible form approximation to a log cost function:

**Truncation error.** For the sequence of cost functions $\{g_n^o\}$ above define

$$T_K = \sup_{1\le n\le\infty} \inf_{\theta\in\mathbb{R}^p} \|g_n^o - g_K(\cdot|\theta)\|_{\ell,\infty,\mathcal{X}}. \qquad □$$

When $g_n^o \ne g^o$ for all n, Theorem 0 is not enough to deliver a polynomial rate of decay for $T_K$. One can impose polynomial decay using the following construction: In the representation $g_\infty(\cdot|\theta)$, note that each element $\theta_i$ of $\theta$

except the first few corresponds to a sine or cosine term with argument $(\lambda j k_\alpha' x)$. Put $\alpha_i = \lambda|j||k_\alpha|$ for these and $\alpha_i = \lambda$ for the first few. Choose $\bar{\theta}_i \geq 0$ to satisfy $\sum_{i=1}^\infty (\alpha_i)^{m+1}\bar{\theta}_i < \infty$. In forming the sequence $\{g_n^0\}$, restrict attention to $\theta_n^0$ in $\mathbb{R}^\infty$ with $|\theta_{in}^0| \leq \bar{\theta}_i$ and put $g_n^0 = g_\infty(\cdot|\theta_n^0)$. With this construction, $T_K = o(K^{-m+\ell+\epsilon})$ for every $\epsilon > 0$. Similarly, to force the sequence $\{T_K\}$ to decrease more rapidly than any polynomial (called a rapidly decreasing sequence) choose $\bar{\theta}_i \geq 0$ to satisfy $\sum_{i=1}^\infty (\alpha_i)^m \bar{\theta}_i < \infty$ for every integer $m > 0$ and form the sequence $\{g_n^0\}$ as above.

Incidentally, a bound on $B(p)$ can be deduced from these $\alpha_i$ since $\sup_{x\in\mathcal{X}}|(\partial/\partial x_w)\cos(\lambda j k_\alpha' x)| \leq \alpha_i$ for $w = 1, \ldots, N$ and $\sum_{i=1}^p (\alpha_i)^2 \approx \lambda^2 \sum_{i=0}^K (i)^{N+1} \approx K^{N+2}$. For additional details, see Gallant and Monahan (1985).

The next theorem bounds the relative bias. Notice that this bound does not depend on $\{x_t\}$ so the relative bias is bounded both conditionally and unconditionally.

**THEOREM 6.** $\text{RelBias}(\rho'\hat{\theta}|\{x_t\}) \leq \sqrt{n}T_K[2\sqrt{\lambda_{max}}\Omega^{-1} + B(p)/\rho'\rho]$.

**Proof.** $\theta_n^* = \mathcal{E}(\hat{\theta}|\{x_t\})$ minimizes $\mathcal{E}_n[g_K(x|\theta) - g_n^0]'\Omega^{-1}[g_K(x|\theta) - g_n^0]$ where, recall, $\mathcal{E}_n$ denotes expectation with respect to the empirical distribution $\mu_n$ of $\{x_t\}_{t=1}^n$. Let $g_\infty(\cdot|\theta_n^0)$ represent $g_n^0$. First we derive two inequalities:

$$
\begin{aligned}
&|\mathcal{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_\infty(x^0|\theta_n^0)| \\
&\quad \leq |\mathcal{E}(\rho'\hat{\theta}|\{x_t\}) - D^\lambda g_K(x^0|\theta_n^0)| + |D^\lambda g_K(x^0|\theta_n^0) - D^\lambda g_\infty(x^0|\theta_n^0)| \\
&\quad = |\rho'(\theta_n^* - \theta_n^0)| + |D^\lambda g_K(x^0|\theta_n^0) - D^\lambda g_\infty(x^0|\theta_n^0)|
\end{aligned}
$$

$$\leq |\rho'(G_{np}^{-1/2})'(G_{np}^{1/2})'(\theta_n^* - \theta_n^o)| + T_K$$

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

$$= [\surd\rho' G_{np}^{-1}\rho]\surd\mathcal{E}_n[g_K(x|\theta_n^*) - g_K(x|\theta_n^o)]'\Omega^{-1}[g_K(x|\theta_n^*) - g_K(x|\theta_n^o)] + T_K.$$

$$\surd\mathcal{E}_n[g_K(x|\theta_n^*) - g_K(x|\theta_n^o)]'\Omega^{-1}[g_K(x|\theta_n^*) - g_K(x|\theta_n^o)]$$

$$\leq \surd\mathcal{E}_n[g_K(x|\theta_n^*) - g_n^o)]'\Omega^{-1}[g_K(x|\theta_n^*) - g_n^o)]$$

$$+ \surd\mathcal{E}_n[g_K(x|\theta_n^o) - g_n^o)]'\Omega^{-1}[g_K(x|\theta_n^o) - g_n^o)]$$

$$\leq 2\surd\mathcal{E}_n[g_K(x|\theta_n^o) - g_n^o)]'\Omega^{-1}[g_K(x|\theta_n^o) - g_n^o)]$$

[because $\theta_n^*$ is the minimizing value]

$$\leq 2[\surd\lambda_{max}\Omega^{-1}]\surd\mathcal{E}_n[g_K(x|\theta_n^o) - g_n^o)]'[g_K(x|\theta_n^o) - g_n^o)]$$

$$= 2[\surd\lambda_{max}\Omega^{-1}]\|g_K(\cdot|\theta_n^o) - g_n^o\|_{0,2,\mu_n}$$

$$\leq 2[\surd\lambda_{max}\Omega^{-1}]\|g_K(\cdot|\theta_n^o) - g_n^o\|_{m,\infty,\mathcal{X}}$$

$$\leq 2[\surd\lambda_{max}\Omega^{-1}]T_K.$$

Substituting the second inequality into the first we have:

$$\text{RelBias}(\rho'\hat{\theta}|\{x_t\})$$

$$\leq [(\surd\rho' G_{np}^{-1}\rho)(2)(\surd\lambda_{max}\Omega^{-1})T_K + T_K]/\surd[(1/n)\rho'(G_{np}^{-1})\rho]$$

$$= \surd n T_K[2\surd\lambda_{max}\Omega^{-1} + 1/\rho'(G_{np}^{-1})\rho].$$

The argument is completed using the inequality $\lambda_{max}(G_{np})/B(p) \le 1$ from the proof of Lemma 2 to get

$$1/\rho'(G_{np}^{-1})\rho \le \lambda_{max}(G_{np})/\rho'\rho = [B(p)/\rho'\rho][\lambda_{max}(G_{np})/B(p)] \le B(p)/\rho'\rho. \qquad \square$$

# 5. ASYMPTOTIC NORMALITY OF ELASTICITY ESTIMATES

As seen from the results of Section 3, the behavior of the sequence of minimum eigenvalues $\{\lambda_{min}(G_p)\}_{p=1}^{\infty}$ of the matrices $G_p = \int \Phi(x)\Omega^{-1}\Phi'(x)\, d\mu(x)$ determines the class $\mathcal{P}$ of sequences $\{p_n\}_{n=1}^{\infty}$ that imply asymptotic normality of the relative error in an estimate of a derivative. Given a sequence $\{p_n\}$ from $\mathcal{P}$, the results of Section 4 determine the rate at which the sequence of truncation errors $\{T_K\}_{K=1}^{\infty}$ must decline in order that the relative bias in an estimate of a derivative tends to zero with n. The rate of decrease in the sequence $\{T_K\}$ determines the class $\mathcal{G}$ of log cost functions that admit of asymptotically normal elasticity estimates. In this section, we work through this construction to determine $\mathcal{P}$ and $\mathcal{G}$. Throughout, every p corresponds to some K; recall that $p \approx K^N$.

Put $\lambda_K = \lambda_{min}(G_p)$. We claim that the sequence $\{\lambda_K\}_{K=1}^{\infty}$ is rapidly decreasing; recall that rapidly decreasing means $\lim_{K\to\infty} K^m \lambda_K = 0$ for every m. This claim is verified as follows: Without loss of generality, assume that putting the scale factor in the definition of the Fourier form to unity makes the closed rectangle $\mathcal{X}$ a proper subset of $Q = \prod_{i=1}^{N+1}[0, 2\pi]$ without boundary points in common with Q. Given values for the elements of $\theta$ in $g_K(\cdot|\theta)$ that correspond to the Translog part of the Fourier form, $u_0 + b'x + (1/2)x'Cx$, there is a periodic function h defined on Q that possesses partial derivatives of every order and agrees with the Translog part on $\mathcal{X}$ (Edmunds and Moscatelli, 1977). A Fourier series expansion $h_K$ of h will have $\|h - h_K\|_{1,2,\mu} \le \|h - h_K\|_{1,\infty,Q} = o(K^{-m+1+\epsilon})$ for every m and every $\epsilon > 0$ (Edmunds and Moscatelli, 1977). Put the negatives of these coefficients in the corresponding entries of $\theta$ in $g_K(\cdot|\theta)$ and one has

$$\lambda_K = \lambda_{min}(G_p)$$
$$\le \theta' G_p \theta / \theta'\theta$$
$$\le \lambda_{max}(\Omega^{-1})[\|h - h_K\|_{1,2,\mu}]^2/\theta'\theta$$
$$= o(K^{-m+1+\epsilon})/\theta'\theta.$$

Parseval's equality implies that $\theta'\theta$ has a finite limit as $K \to \infty$.

A useful characterization of a rapidly decreasing sequence is obtained as follows. If $K^m\lambda_K \to 0$ then $\ell og\lambda_K + m\ell ogK \to -\infty$ for any m. This implies that for each m, $\ell og\lambda_K/\ell ogK < -m$ for all K large enough. Thus $\ell og\lambda_K/\ell ogK = -a(K)$ for some positive, increasing function a(K). Equivalently $\lambda_K = K^{-a(K)}$. Conversely, $\lambda_K = K^{-a(K)}$ for some positive, increasing function a(K) implies $\lambda_K$ is rapidly decreasing. As an example, $\lambda_K = e^{-bK}$ corresponds to $a(K) = bK/\ell ogK$.

We can now determine the class $\mathcal{P}$; recall that $B(p) \le K^{N+2}$. The first condition of Theorem 5 requires that $B(p)/\lambda_K \le n^{-\beta}$ for some $\beta$, $0 \le \beta < 1/2$, which implies $[N + 2 + a(K)]\ell og(K) \le \beta\ell og(n)$. Thus, $\{K_n\}$ must satisfy

$$a(K_n)\ell og(K_n) \le \beta\ell og(n) \qquad \text{for some } \beta,\ 0 \le \beta < 1/2.$$

The implication of this relation is that $K_n$ must grow slower than any fractional power of n; more precisely, $n^\alpha/K_n \to 0$ as $n \to \infty$ for any $\alpha > 0$. Since $p \approx K^N$, the same is true of $p_n$. As $p_n$ increases slower than any fractional power of n, the second condition of Theorem 5 is always satisfied and we have that

$$\mathcal{P} = \left\{(p_n):\ p_n \approx K_n^N,\ a(K_n)\ell og(K_n) \le \beta\ell og(n) \text{ for some } \beta,\ 0 \le \beta < 1/2\right\}$$

is the set of rules that satisfy the conditions of Theorem 5.

To determine $\mathcal{G}$, we first note that the sequence $\{p_n^*\}$ with $(K_n^*)^{a(K_n^*)} = \sqrt{n}$ bounds every sequence in $\mathcal{P}$. Since $(K_n^*)^{a(K_n^*)} T_{K_n^*} = \sqrt{n} T_{K_n^*} \leq \sqrt{n} T_{K_n}$ for every $\{p_n\}$ in $\mathcal{P}$, the bound $\sqrt{n} T_K [2\sqrt{\lambda_{max}} \Omega^{-1} + B(p)/\rho'\rho]$ of Theorem 6 cannot decrease if $K^{a(K)} T_K = T_K/\lambda_K$ does not tend to zero as $K \to \infty$. Thus the set $\mathcal{G}$ of sequences of log cost functions $\{g_n^o\}$ for which we can make use of Theorem 6 (having already invoked Theorem 5) satisfies

$$\mathcal{G} \subset \left\{ \{g_n^o\} : T_K/\lambda_K \to 0 \text{ as } K \to \infty \right\}.$$

If one is not willing to exploit drift and holds $g_n^o$ fixed at some $g^o$ for every n then $g^o$ must be infinitely many times differentiable if $T_K$ is to decrease fast enough to damp $1/\lambda_K$.

If one is willing to work within a paradigm that assumes that the true cost function $g_n^o$ moves slowly away from the Translog as data is acquired, then one can always choose a sequence $\{g_n^o\}$ that will drive $T_K$ to zero as rapidly as required by the choice of $\{p_n\}$ from $\mathcal{P}$. The extreme form of this view is that the fitted model is correct (Huber, 1973) in which case $T_K \equiv 0$ regardless of the tail behavior of $\theta_n^o$ in the representations $g_n^o = g_\infty(\cdot|\theta_n^o)$.

# REFERENCES


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

Andrews, Donald W. K. (1989), "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Models," Cowles Foundation Discussion Paper No. 874R, Yale University.

Billingsley, Patrick (1979), *Probability and Measure*, Wiley, New York.

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

Deaton, Angus, and John Muellbauer (1980) *Economics and Consumer Behavior*, Cambridge University Press, Cambridge.

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, University of Chicago.

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

Eicker, Friedhelm (1967), "Limit Theorems for Regressions with Unequal and Dependent Errors," *Proceedings of the Fifth Berkeley Symposium on Probability and Mathematical Statistics* 1, 59-82.

Elbadawi, Ibrahim, A. Ronald Gallant, and Geraldo Souza (1983), "An Elasticity Can be Estimated Consistently Without A Priori Knowledge of Functional Form," *Econometrica* 51, 1731-1752.

Gallant, A. Ronald (1982), "Unbiased Determination of Production Technologies," *Journal of Econometrics* 20, 285-323.

Gallant, A. Ronald, and Gene H. Golub (1984), "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics* 26, 295-321.

Gallant, A. Ronald, and John F. Monahan (1985), "Explicitly Infinite Dimensional Bayesian Analysis of Production Technologies," *Journal of Econometrics* 30, 171-201.

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), "$\epsilon$-entropy and $\epsilon$-capacity of Function Spaces," *Uspehi Mat. Nauk* 14, 3-86; English translation (1961) *Mathematical Society Translations* 17, 277-364.

Monahan, John F. (1981) "Enumeration of Elementary Multi-indices for Multivariate Fourier Series," Institute of Statistics Mimeograph Series No. 1338, North Carolina State University, Raleigh, NC.

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

SAS (1986) "SUGI Supplemental Library User's Guide," SAS Institute Inc. Cary, NC.