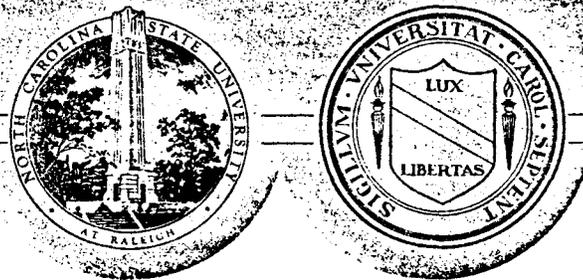


THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM



AN ELASTICITY CAN BE ESTIMATED CONSISTENTLY WITHOUT A
PRIORI KNOWLEDGE OF FUNCTIONAL FORM

by

Ibrahim El Badawi, A. Ronald Gallant, and Geraldo Souza

Institute of Statistics Mimeograph Series No. 1396

March 1982

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

An Elasticity Can Be Estimated Consistently Without A
Priori Knowledge of Functional Form^{***}

by

Ibrahim El Badawi,^{*} A. Ronald Gallant,^{*} and Geraldo Souza^{*,**}

^{*} North Carolina State University, Raleigh, North Carolina 27650 USA

^{**} Embrapa, Brasilia, Brazil

^{***} This research was supported by National Science Foundation Grant SES 8014239
and by North Carolina Agricultural Experiment Station Project NC03641.

Abstract

We consider an open question in applied price theory: Without a priori knowledge of a firm's cost function or a consumer's indirect utility function, is it possible to estimate price and substitution elasticities consistently by observing a demand system? As the work of White (1980), Guilkey, Lovell, and Sickles (1981), and others has shown, ordinary flexible functional forms such as the Translog cannot achieve this objective. We find that if one is prepared to assume that elasticities of substitution cannot oscillate wildly over the region of interest then consistent estimation is possible using the Fourier flexible form provided the number of fitted parameters increases as the number of observations increases. This result obtains with any of the commonly used statistical methods, as examples: multivariate least squares, maximum likelihood, and three-stage least squares. It obtains if the number of fitted parameters is chosen adaptively by observing the data or chosen deterministically according to some fixed rule. We approach the problem along the classical lines of estimability considerations as used in the study of less than full rank linear statistical models and thereby discover that the problem has a fascinating structure which we explore in detail.

1. Introduction and Main Result

Let q denote an N -vector of commodities when a consumer's demand functions are observed or an N -vector of factor inputs when a firm's derived demand functions are observed; let p denote the vector of corresponding prices; and let $c = p'q$ be either the consumer's cost necessary to achieve utility level u or the firm's cost to achieve output level u .

Suppose that the problem is that of consumer demand and let $g(x)$ with $x = p/c$ be the consumer's indirect utility function. Then cost shares are determined by Roy's identity

$$s_i(x) = \left[\sum_{i=1}^N x_i (\partial/\partial x_i) g(x) \right]^{-1} x_i (\partial/\partial x_i) g(x)$$

for $i = 1, 2, \dots, N$. This system is fitted to observed cost shares s_{it} corresponding to observed cost normalized prices $x_{it} = p_{it}/c_t$ with an eye toward estimating the elasticities of substitution

$$\sigma_{ij}(x) = \frac{(\sum_k x_k g_k) g_{ij}}{g_i g_j} - \frac{\sum_k x_k g_{jk}}{g_j} - \frac{\sum_k x_k g_{ik}}{g_i} + \frac{\sum_m \sum_k x_k g_{km} x_m}{\sum_n x_n g_n}$$

where $g_i = (\partial/\partial x_i) g(x)$ and $g_{ij} = (\partial^2/\partial x_i \partial x_j) g(x)$. Other quantities of interest -- compensated price elasticities, uncompensated price elasticities, and income elasticities -- can be obtained from formulas similar to the above in that x , $(\partial/\partial x_i) g(x)$, and $(\partial^2/\partial x_i \partial x_j) g(x)$ are all that must be known. Closer inspection of the formula for $\sigma_{ij}(x)$ reveals that it would be enough to determine the first derivatives of $g(x)$ within a scalar multiple. That is if $g^o(x)$ can be determined such that for some function $\alpha(x) > 0$

$$(\partial/\partial x_i) g^o(x) = \alpha(x) (\partial/\partial x_i) g(x) \quad i = 1, 2, \dots, N$$

then $(\partial/\partial x_i) g^o(x)$ and $(\partial^2/\partial x_i \partial x_j) g^o(x)$ can be used in the formula for $\sigma_{ij}(x)$ in

place of the derivatives of $g(x)$ without changing the result. The same is true of price and income elasticities. Evidently, the object of the empirical exercise is the determination of some function $g^o(x)$ whose derivatives are proportional to those of $g(x)$.

Suppose that the problem is that of factor demand and let $g(x)$ with $x = (p', u)'$ be the firm's cost function. Input cost shares are determined by Shephard's lemma

$$s_i(x) = \left[\sum_{i=1}^N x_i (\partial/\partial x_i) g(x) \right]^{-1} x_i (\partial/\partial x_i) g(x)$$

for $i = 1, 2, \dots, N$. The elasticities of substitution are computed as

$$\sigma_{ij}(x) = \frac{(\sum_{k=1}^N x_k s_k) s_{ij}}{s_i s_j}$$

where the notation is as before. Again, as before, the quantities of interest can be computed for a given x from the first and second derivatives of $g(x)$. Again, knowledge of first derivatives to within a scalar multiple will suffice as if $g^o(x)$ is linear homogeneous in its first N arguments and if

$(\partial/\partial x_i) g^o(x) = \alpha(x) (\partial/\partial x_i) g(x)$ then it follows that

$(\partial^2/\partial x_i \partial x_j) g^o(x) = \alpha(x) (\partial^2/\partial x_i \partial x_j) g(x)$ for $i, j = 1, 2, \dots, N$. To show this, use

the fact that a linear homogeneous function satisfies

$$g(x) = \sum_{i=1}^N x_i (\partial/\partial x_i) g(x)$$

which implies that $\alpha(x) = g^o(x)/g(x)$.

To allow the consumer's indirect utility function to depend on demographic or other characteristics as well as cost normalized prices p/c , let x have dimension $k \geq N$, it being understood that the first N components of x contain the vector p/c . Let $\nabla g(x)$ denote the N -vector obtained by differentiating with respect to the first N components of $g(x)$,

$$\nabla g(x) = [(\partial/\partial x_1)g(x), \dots, (\partial/\partial x_N)g(x)]' .$$

Similarly for the firm's cost function in which case the first N components of x contain the vector p and the rest of x contains output and other covariates. Let \mathcal{X} denote that subset of \mathbb{R}^k from which the observed x_t are to be drawn. We shall assume that \mathcal{X} is bounded, open and convex and that no vector in the closure $\bar{\mathcal{X}}$ corresponds to a zero price for some commodity.

In the same vein, it is possible that only a subvector of the cost share vector s_t is observed, or in the case of factor demand that costs c_t as well as cost shares are observed. In either event, denote by y_t that which is observed; y_t is an M -vector where M may be larger, smaller, or equal to N .

Many plausible data generating mechanisms can be envisaged. Consider factor demand. The simplest assumption is that shares follow the model

$$s_t = \text{diag}(p_t) \nabla g(x_t) / g(x_t) + e_t \quad t = 1, 2, \dots$$

with x_t independent of the errors e_t . Another possibility is that a firm produces on the basis of a forecast \hat{x}_t but that which obtains is $x_t = \hat{x}_t + e_t$ whence shares follow the model

$$s_t = \text{diag}(p_t - e_{(1)t}) \nabla g(x_t - e_t) / g(x_t - e_t) \quad t = 1, 2, \dots$$

with e_t being unobservable. In general, assume that the observed data (y_t, x_t) follows some reduced form

$$y_t = Y(e_t, x_t, g^*) \quad t = 1, 2, \dots$$

and impose:

Assumption 1. The errors are independently and identically distributed having common distribution $P(e)$ with support \mathcal{E} ; that is $P(\mathcal{E}) = 1$. Let $Y(e, x, g^*)$ be continuous on $\mathcal{E} \times \mathcal{X}$ for any fixed g^* that generates continuous

cost share functions upon application of either Roy's identity or Shephard's lemma.

It is necessary to impose restrictions on the limiting behavior of the sequence x_1, x_2, \dots . We use the notion of Cesaro sum generators which had its beginnings in Jennrich (1969), and Malinvaud (1970), and reached its present form in Gallant and Holly (1980).

Definition (Gallant and Holly, 1980) A sequence $\{v_t\}$ of points from a Borel set \mathcal{V} is said to be a Cesaro sum generator with respect to a probability measure ν defined on the Borel subsets of \mathcal{V} and a dominating function $b(v)$ with $\int b(v)d\nu(v) < +\infty$

$$\lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n f(v_t) = \int f(v) d\nu(v)$$

for every real valued continuous function f with $|f(v)| \leq b(v)$.

Assumption 2. (Gallant and Holly, 1980) Almost every realization of $\{v_t\}$ with $v_t = (e_t, x_t)$ is a Cesaro sum generator with respect to the product measure

$$\nu(A) = \int_{\mathcal{X}} \int_{\mathcal{E}} I_A(e, x) dP(e) d\mu(x)$$

and dominating function $b(e, x)$. The sequence $\{x_t\}$ is a Cesaro sum generator with respect to μ and $b(x) = \int_{\mathcal{E}} b(e, x) dP(e)$. For each $x \in \mathcal{X}$ there is a neighborhood N_x such that $\int_{\mathcal{E}} \sup_{N_x} b(e, x) dP(e) < \infty$. \mathcal{X} is a bounded, open, convex set, if A is an open subset of \mathcal{X} then $\mu(A) > 0$, and if x is in $\bar{\mathcal{X}}$ then $x_i > 0$ for $i=1, 2, \dots, N$ where the overbar denotes the closure of \mathcal{X} .

The practical consequence of this restriction is that the limit of an average can be computed as an integral. For example, suppose that cost shares follow

$$s_t = \text{diag}(p_t) \nabla g^*(x_t) / g^*(x_t) + e_t$$

and that all shares are observed

$$y_t = s_t \cdot$$

Let $\hat{\Sigma}$ be an estimator of scale that converges almost surely to a positive semi-definite matrix Σ that has rank $N-1$; the vector $\mathbf{1} = (1, 1, \dots, 1)'$ is always an eigenvector with root zero because shares add to unity. Let $\hat{\Sigma}^+ = \sum_{i=1}^{N-1} \mathbf{1}/\lambda_i \sigma_i \sigma_i'$ where the λ_i are the non-zero eigenvalues of $\hat{\Sigma}$ and σ_i the corresponding orthonormal eigenvectors. Suppose that \hat{g} to estimate g^* is chosen from some class \mathcal{G} of functions that are linear homogeneous in the first N arguments by minimizing

$$s_n(g) = (1/n) \sum_{t=1}^n \left[y_t - \frac{\text{diag}(p_t) \nabla g(x_t)}{g(x_t)} \right]' \hat{\Sigma}^+ \left[y_t - \frac{\text{diag}(p_t) \nabla g(x_t)}{g(x_t)} \right].$$

For example, \mathcal{G} might be all functions of the Translog form

$$\ln g(x) = \alpha_0 + \sum_{i=1}^k \alpha_i \ln(x_i) + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \beta_{ij} \ln(x_i) \ln(x_j)$$

with $\sum_{i=1}^N \alpha_i = 1$, $\beta_{ij} = \beta_{ji}$, and $\sum_{i=1}^N \beta_{ij} = 0$. Then what will happen (Burguete, Gallant, and Souza, 1982) is that \hat{g} will converge almost surely to that function g^0 that minimizes the almost sure limit ^{1/}

$$\begin{aligned} \bar{s}(g, g^*) &= \lim_{n \rightarrow \infty} s_n(g) \\ &= \int_{\mathcal{E}} e' \Sigma^+ e \, dP(e) + \int_{\mathcal{X}} \left[\frac{\nabla g^*(x)}{g^*(x)} - \frac{\nabla g(x)}{g(x)} \right]' \text{diag}(p) \Sigma^+ \text{diag}(p) \left[\frac{\nabla g^*(x)}{g^*(x)} - \frac{\nabla g(x)}{g(x)} \right] d\mu(x) \end{aligned}$$

If both $g(x)$ and $g^*(x)$ are bounded away from zero, linear homogeneous in the first N components of x , both continuous, \mathcal{X} open, and μ puts positive mass on every open subset of \mathcal{X} then

$$\bar{s}(g^0, g^*) \leq \bar{s}(g^*, g^*) \quad \text{implies} \quad \frac{\nabla g^0(x)}{g^0(x)} = \frac{\nabla g^*(x)}{g^*(x)},$$

at every point in \mathcal{X} .

All commonly used econometric estimators exhibit the properties of this example. To each estimator corresponds some sample objective function $s_n(g)$. The estimator itself may be defined as

\hat{g} in \mathcal{G} that minimizes $s_n(g)$.

The sample objective function $s_n(g)$ has an almost sure limit $\bar{s}(g, g^*)$ and \hat{g} is consistent for

g° in \mathcal{G} that minimizes $\bar{s}(g, g^*)$.

If $\mathcal{G} = \{g(x|\theta) : \theta \in \Theta\}$ where corresponding cost shares $s_i(x|\theta)$ are continuous in (x, θ) and Θ is a compact metric space then the almost sure convergence of $s_n(g)$ to $\bar{s}(g, g^*)$ is uniform over \mathcal{G} in typical cases. Examples of estimators with these properties are single equation and multivariate (nonlinear) least squares, single equation and multivariate maximum likelihood, maximum likelihood for (nonlinear) simultaneous systems, iteratively rescaled M-estimates, scale invariant M-estimates, two-and three-stage (nonlinear) least squares (Burguete, Gallant, and Souza, 1982). Whether or not

$$\bar{s}(g^\circ, g^*) \leq \bar{s}(g^*, g^*) \text{ implies } \nabla g^\circ(x) = \alpha(x) \nabla g^*(x)$$

for some $\alpha(x) > 0$ depends on the interaction between the sample objective function $s_n(g)$ and the reduced form $Y(e, x, g^*)$. But it is obviously the minimal identification condition to require of a statistical estimation procedure.

Identification Condition. Assumptions 1 and 2 suffice for the sample objective function $s_n(g)$ to have almost sure limit $\bar{s}(g, g^*)$ for any g that generates continuous cost share functions upon application of either Shephard's lemma or Roy's identity. Moreover, this convergence is uniform in g over any family

$$\mathcal{G} = \{g(x|\theta) : \theta \in \Theta^*\}$$

that is indexed by a compact metric space Θ^* and for which corresponding cost shares $s_i(x|\theta)$ and derivatives $(\partial/\partial x)s_t(x|\theta)$ are continuous in (x, θ) . Then the condition is that

$$\bar{s}(g^0, g^*) \leq \bar{s}(g^*, g^*) \text{ implies } \nabla g^0(x) = \alpha(x) \nabla g^*(x)$$

for all $x \in X$ where $\alpha(x) > 0$ on X .

To have a compact notation for high order partial derivatives, let

$$D^\lambda g(x) = \frac{\partial^{|\lambda|}}{\partial x_1^{\lambda_1} \partial x_2^{\lambda_2} \dots \partial x_k^{\lambda_k}} g(x)$$

where

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$$

has non-negative integers as components and

$$|\lambda| = \sum_{i=1}^k |\lambda_i|$$

is the order of the partial derivatives; when λ is the zero vector take $D^0 g(x) = g(x)$.

We shall assume that the elasticities of substitution $\sigma_{ij}(x)$ do not oscillate wildly over X . This seems innocuous as if one really believed that elasticities were subject to extreme variation then there would be little point to trying to estimate them. The easiest way to impose this restriction is to require that the first order partial derivatives of $g(x)$ are bounded away from zero and that all derivatives through the third order are bounded from above in absolute value.

For either consumer or factor demand, prices and quantities can be rescaled arbitrarily as long as cost shares remain invariant. This rescaling amounts to no more than a change in the units in which a commodity is

measured. Also, the units in which output and other covariates are measured is irrelevant. Thus, recalling that \mathcal{X} is a bounded, open, convex set, we can assume without loss of generality that the closure of \mathcal{X} is contained in the open cube $X_{i=1}^k(0, 2\pi)$.

Assumption 3. The closure $\bar{\mathcal{X}}$ of \mathcal{X} is a subset of the open cube $X_{i=1}^k(0, 2\pi)$. For some $m \geq 3$, $D^\lambda g^*(x)$ is continuous on \mathcal{X} for all λ with $|\lambda| \leq m+1$, $\sup_x |D^\lambda g^*(x)| < \infty$ for λ with $|\lambda| = m+1$, $\sup_x |D^\lambda g^*(x)| \leq b < \infty$ for λ with $|\lambda| \leq m$, and $0 < a \leq \inf_{x \in \mathcal{X}} |(\partial/\partial x_i)g^*(x)|$ for $i=1, 2, \dots, N$.

Our main result is that if the Fourier flexible form is used with an estimation procedure that satisfies the Identification Condition then consistent estimation of price, income, and substitution elasticities is possible. The Fourier flexible form was introduced in Gallant (1981) and may be written as

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

with $C = -\sum_{\alpha=1}^A u_{0\alpha} k_\alpha k'_\alpha$. The sequence $\{k_\alpha\}_{\alpha=1}^\infty$ is a sequence of multi-indexes, multi-indexes being vectors k_α of dimension k whose elements are integers. The rule of formation of the sequence $\{k_\alpha\}$ need not concern us here, it is set forth in Gallant (1981) and can be automated (Monahan, 1981). K is the degree of the trigonometric polynomial meaning that

$$|jk'_\alpha| = j \sum_{i=1}^k |k_{i\alpha}| \leq K, \quad \alpha = 1, 2, \dots, A, \quad j = 1, 2, \dots, J.$$

Thus A and J can be viewed as depending on K and could be written as A_K and J_K .

Consistent estimation obtains by letting the fitted number of parameters depend on the sample size n . Equivalently, let the degree K of $g_K(x|\theta)$ depend

on n . Now this dependence may be according to some fixed rule such as $K_n = \sqrt{n}$ or may be adaptive such as letting K_n depend on the outcome of a statistical test. For example, increase K when a lack of fit test rejects the current model. We refer to the former as a deterministic rule and the latter as an adaptive rule. In either event, consistency obtains.

Theorem 1. Let Assumptions 1 through 3 hold and let the estimation procedure $s_n(g)$ satisfy the Identification Condition. Let K_n be either a deterministic or an adaptive rule for choosing the number of parameters of the Fourier flexible form $g_{K_n}(x|\theta)$. Let $\hat{\theta}_n$ minimize $s_n[g_{K_n}(\theta)]$ subject to $\sup_{x \in \mathcal{X}} |D^\lambda g_{K_n}(x|\theta)| \leq b$ for all $|\lambda| \leq m$ and $a \leq \inf_{x \in \mathcal{X}} |(\partial/\partial x_i) g_{K_n}(x|\theta)|$ for $i = 1, 2, \dots, N$. Let $\hat{\sigma}_n(x)$ be an elasticity of substitution computed from $g_{K_n}(x|\hat{\theta}_n)$ and $\sigma^*(x)$ be computed from $g^*(x)$. If $\lim_{n \rightarrow \infty} K_n = \infty$ almost surely then

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} |\hat{\sigma}_n(x) - \sigma^*(x)| = 0$$

almost surely. A similar result holds for shares, for price elasticities, and for income elasticities.

When the Fourier form is used to approximate a cost function, a log transformation of the data is convenient (Gallant, 1982); let

$$l_i(x) = \ln x_i + \ln a_i \quad i = 1, 2, \dots, k$$

and set

$$l(x) = [l_1(x), l_2(x), \dots, l_k(x)]'$$

In this case

$$\begin{aligned} \ln \xi_K(x|\theta) = & u_0 + b'l(x) + \frac{1}{2} l'(x) C l(x) \\ & + \sum_{\alpha=1}^A \{u_{0\alpha} + 2\sum_{j=1}^J [u_{j\alpha} \cos(j\lambda_k' l(x)) - v_{j\alpha} \sin(j\lambda_k' l(x))]\} \end{aligned}$$

with $C = -\sum_{\alpha=1}^A u_{0\alpha} \lambda^2 k_{\alpha} k'_{\alpha}$; λ is scaling factor chosen so that $\lambda l(x)$ is in the cube $X_{i=1}^k(0, 2\pi)$. Theorem 1 remains valid if this alternative form is used.

The reader who has no interest in detail may stop at this point as the rest of the paper is primarily a proof of Theorem 1.

2. Structure of the Problem

The Fourier flexible form is linear in its parameters and can be written as

$$g_K(x|\theta) = \sum_{j=1}^{P_K} \theta_j \varphi_j(x).$$

The parameters θ_j represent u_0 , the b_i , the $u_{j\alpha}$, and the $v_{j\alpha}$ written down in some order, exactly what order need not concern us. The $\varphi_j(x)$ represent the constant function, the linear terms x_i , the quadratic terms $x_i x_j$, and the sine and cosine terms $\cos(jk'_{\alpha} x)$ and $\sin(jk'_{\alpha} x)$ written down in an order to correspond with the θ_j .^{2/} The process of constructing these $\varphi_j(x)$ and θ_j can be continued indefinitely as K increases so we can imagine a parameter vector θ of infinite extent

$$\theta = (\theta_1, \theta_2, \dots)$$

and view the Fourier flexible form as depending on the leading p_K terms of this infinite dimensional vector

$$g_K(x|\theta) = \sum_{j=1}^{P_K} \theta_j \varphi_j(x).$$

The parameter θ is to be thought of as being infinite dimensional hereafter.

Assumptions 1 through 3 permit application of the Corollary of Theorem 1 of Gallant (1981) to conclude that there is a parameter vector θ^* such that

$$\lim_{K \rightarrow \infty} \max_{|\lambda| \leq m} \sup_{x \in \mathcal{X}} |D^\lambda g^*(x) - D^\lambda g_K(x|\theta^*)| = 0 .$$

This result implies that for our purposes we can accept θ^* as representing $g^*(x)$. Knowing θ^* we can compute the pointwise limit

$$g_\infty(x|\theta^*) = \lim_{K \rightarrow \infty} g_K(x|\theta^*)$$

and recover $g^*(x)$ and its derivatives as $D^\lambda g(x) = D^\lambda g_\infty(x|\theta^*)$ for all $|\lambda| \leq m$. Alternatively, we can differentiate and then take the limit since

$$\lim_{K \rightarrow \infty} D^\lambda g_K(x|\theta^*) = D^\lambda g^*(x) .$$

Moreover, truncated expansions $D^\lambda g_K(x|\theta^*)$ can be used to obtain approximations to $D^\lambda g^*(x)$ that are uniformly accurate over \mathcal{X} . The parameter θ^* is not unique and we accept any infinite dimensional vector $\theta^\#$ with

$$\lim_{K \rightarrow \infty} \max_{|\lambda| \leq m} \sup_{x \in \mathcal{X}} |D^\lambda g^*(x) - D^\lambda g_K(x|\theta^\#)| = 0$$

as equivalent to θ^* . In summary, any information that we require of $g^*(x)$ can be recovered from θ^* , and in several different ways to suit our convenience. ^{3/}

Intuition is guided by the classical analysis of the less than full rank linear model

$$y = X\beta + e$$

where β is a p -vector, X has n rows and p columns, and $\text{rank}(X) < p$. Let x_1, x_2, \dots, x_r be a basis for the row space of X ; that is, any row x'_t of X can be written as $x'_t = \sum_{i=1}^r a_i x_i$. In the classical analysis, attention is focused on the estimation space

$$\left\{ \sum_{i=1}^r a_i x_i : (a_1, a_2, \dots, a_r) \in R^r \right\} .$$

The classical problem is this: One is permitted to observe at selected points x'_t in the estimation space the corresponding values $x'_t \beta$ plus error

$$y_t = x_t' \beta + e_t, \quad t = 1, 2, \dots, n.$$

One wishes to find some function

$$\hat{\beta} = \beta(y_1, \dots, y_n)$$

such that $x' \hat{\beta}$ is unbiased for any x in the estimation space.

Now, one can think of a point x in the estimation space either as a point in R^p or as a point a in R^r given by $a = (a_1, a_2, \dots, a_r)$ with $x = \sum_{i=1}^r a_i x_i$; it makes no difference. When passing to the infinite dimensional case it is essential to take the latter view and focus on the coordinates (a_1, \dots, a_r) of a point in the estimation space. Also, one can think of β as being a vector in R^p or as being a linear functional defined over the estimation space by

$$\langle a, \beta \rangle = \sum_{i=1}^r a_i x_i' \beta.$$

Again, when passing to the infinite dimensional case, it is the latter view that is the more useful. If we adopt these two points of view: a point in the estimation space is represented by its coordinates a , not by its components x , and β is to be thought of as a linear functional, not as a point in R^p , then the classical problem is restated as: One is permitted to observe a linear functional β evaluated at selected points a_t in the estimation space plus error

$$y_t = \langle a_t, \beta \rangle + e_t, \quad t = 1, 2, \dots, n.$$

One wishes to find some mapping $\hat{\beta}$ of the observations into the space of linear functionals on the estimation space such that $\langle a, \hat{\beta} \rangle$ is unbiased for $\langle a, \beta \rangle$ at all points a in the estimation space.

These are the thoughts that guide intuition and lead us to pose the problem of consistent estimation of a price, income, or substitution elasticity as follows.

Define the infinite dimensional vectors

$$D^\lambda \varphi_x = (D^\lambda \varphi_1(x), D^\lambda \varphi_2(x), \dots)$$

recalling that when λ is the zero vector

$$D^0 \varphi_x = \varphi_x = (\varphi_1(x), \varphi_2(x), \dots)$$

Thinking of the $D^\lambda \varphi_x$ as basis vectors, represent a point d in the estimation space D by

$$d = \sum_{|\lambda| \leq m} \sum_{i=1}^I a_{x_i, \lambda} \circ D^\lambda \varphi_{x_i}$$

A point d in D may be thought of as a function $d(x, \lambda)$ defined on $\mathcal{X} \times I^k$ that is zero everywhere except at a finite number of points (x_i, λ) where it takes on the value $a_{x_i, \lambda}$. A linear combination $d = \alpha d_1 + \beta d_2$ would be that function $d(x, \lambda)$ having a finite number of jumps as determined by $d(\lambda, x) = \alpha d_1(\lambda, x) + \beta d_2(\lambda, x)$

Clearly D is a linear space and, by defining

$$\|d\| = \sum_{|\lambda| \leq m} \sum_{i=1}^I |a_{x_i, \lambda}|$$

D becomes a normed linear space. Now $D^\lambda \varphi(x)$ can be thought of as an infinite dimensional vector or as a basis vector. Strictly speaking, in the latter case we should write $1 \circ D^\lambda \varphi(x)$ but we shall not and we take $D^\lambda \varphi(x)$ to mean either as determined by context.

Let D^* denote the bounded linear functionals on D . A linear functional d^* is determined by the values it takes on at the basis vectors since

$$\langle d, d^* \rangle = \sum_{|\lambda| \leq m} \sum_{i=1}^I a_{x_i, \lambda} \langle D^\lambda \varphi_{x_i}, d^* \rangle$$

The norm of a linear functional d^* is, in general,

$$\|d^*\| = \sup_{\|d\|=1} |\langle d, d^* \rangle|$$

but by defining

$$\|d^*\|_{m, \infty, \mu} = \max_{|\lambda| \leq m} \sup_{x \in X} |\langle D^\lambda \varphi_x, d^* \rangle|$$

we have

$$\text{Lemma 1. } \|d^*\| = \|d^*\|_{m, \infty, \mu} .$$

Proof.

$$\begin{aligned} \|d^*\| &= \sup_{\|d\|=1} |\langle d, d^* \rangle| \\ &\leq \sup_{\|d\|=1} \sum_{|\lambda| \leq m} \sum_{i=1}^I |a_{x_i, \lambda}| |\langle D^\lambda \varphi_{x_i}, d^* \rangle| \\ &\leq \left(\sup_{\|d\|=1} \sum_{|\lambda| \leq m-1} \sum_{i=1}^I |a_{x_i, \lambda}| \right) \|d^*\|_{m, \infty, \mu} \\ &= \|d^*\|_{m, \infty, \mu} . \end{aligned}$$

Now $D^\lambda \varphi_x$ is a particular instance of d in D with $\|d\|=1$ whence

$$\|d^*\|_{m, \infty, \mu} \leq \|d^*\| . \quad \square$$

We are only concerned with those linear functionals that correspond to an admissible indirect utility function or cost function. Accordingly, let Θ denote those infinite dimensional vectors θ such that

$$\lim_{K \rightarrow \infty} \max_{|\lambda| \leq m} \sup_{x \in X} \left| \sum_{j=1}^{P_K} \theta_j D^\lambda \varphi_j(x) \right| < \infty$$

and define

$$\langle D^\lambda \varphi_x, \theta \rangle = \lim_{K \rightarrow \infty} \sum_{j=1}^{P_K} \theta_j D^\lambda \varphi_j(x) .$$

Note that

$$g_\theta(x|A) = \langle \varphi_x, \theta \rangle ,$$

that

$$D^\lambda g_\infty(x|\theta) = \langle D^\lambda \varphi_x, \theta \rangle$$

and that

$$\|\theta\|_{m,\infty,\mu} = \max_{|\lambda| \leq m} \sup_{x \in \mathcal{X}} |D^\lambda g_\infty(x|\theta)|.$$

We see that when the norm $\|d^*\|$ of a linear functional in D^* is applied to θ in Θ its value is the Sobolev norm (Gallant, 1981) $\|g_\infty(\theta)\|_{m,\infty,\mu}$ of the cost or indirect utility function that corresponds to θ .

The parallel with linear models theory is nearly exact. The values $\langle D^\lambda \varphi_{x_t}, \theta^* \rangle$ or some function of the values that a linear functional θ^* takes on at selected points $D^\lambda \varphi_{x_t}$ in the estimation space D can be observed subject to error. An estimator $\hat{\theta}_n$ mapping the observed data into the space of bounded linear functionals D^* is obtained by minimizing a sample objective function $s_n[g_K(\theta)]$. But instead of unbiasedness, the property we seek is uniform strong consistency

$$\lim_{n \rightarrow \infty} \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |\langle D^\lambda \varphi_x, \hat{\theta}_n \rangle - \langle D^\lambda \varphi_x, \theta^* \rangle| = 0$$

almost surely.

As mentioned earlier, typical estimation methods have a sample objective function $s_n(g)$ that converges uniformly over an indexed family

$$\mathcal{G} = \{g_\infty(x|\theta) : \theta \in \Theta^*\}, \text{ viz}$$

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta^*} |s_n[g_\infty(\theta)] - \bar{s}[g_\infty(\theta), g^*]| = 0 \text{ almost surely,}$$

provided that Θ^* is compact in a metrizable topology \mathcal{J} and that the vector of shares $s(x|\theta)$ corresponding to $g_\infty(x|\theta)$ are continuous in the product topology

generated by the Euclidean norm on \mathcal{X} and this topology \mathcal{J} on Θ^* . Maximum likelihood methods may require that $(\partial/\partial x')$ $s(x|\theta)$ be continuous as well as the Jacobian of the shares can appear in the likelihood.

Lemma 2 which follows gives the construction of the set Θ^* and extends the definition of $g_\omega(x|\theta)$ to this set so that the requisite compactness of the parameter space and continuity of the shares obtains. Basically what Lemma 2 says is that the Fourier form with derivatives to the third order bounded above by b and first derivatives bounded below by a can be suitably indexed provided some annoying details regarding limiting operations are accounted for.

One should note that as we progress derivatives are lost. In Assumption 3 we assumed that an admissible indirect utility function or cost function $g(x)$ possessed derivatives to order $m+1$. To obtain the representation $D^\lambda g(x) = D^\lambda g_\omega(x|\theta)$ one order was lost and equality holds only for λ with $|\lambda| \leq m$. This fact requires that we work with the norm $\|d^*\|_{m,\infty,\mu}$ on D^* . In reading Lemma 2, note that to obtain a compact parameter space Θ^* we lose one more order of differentiation and the norm that we shall work with thereafter is

$$\|d^*\|_{m-1,\infty,\mu} = \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |\langle D^\lambda \phi_x, d^* \rangle|.$$

Another consequence of Lemma 2 is that if d^* is in Θ^* then $\|d^*\|_{m-1,\infty,\mu}$ can also be written as

$$\|d^*\|_{m-1,\infty,\mu} = \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |D^\lambda g_\omega(x|d^*)|$$

Lemma 2. Let

$$\Theta_0 = \{A \in \Theta : \|A\|_{m,\infty,\mu} \leq b, \inf_{x \in \mathcal{X}} |\langle D^{\epsilon_i} \phi_x, A \rangle| \geq a, \quad i=1,2,\dots,N\}$$

where a and b are given by Assumption 3 and ϵ_i denotes the i^{th} elementary

vector. Let $\bar{\Theta}_0$ denote those $d^* \in D^*$ such that d^* is the pointwise limit of some sequence $\{\theta_n\}$ from Θ_0 in the sense that

$$\lim_{n \rightarrow \infty} \langle D^\lambda \varphi_x, \theta_n \rangle = \langle D^\lambda \varphi_x, d^* \rangle$$

for all $x \in \mathcal{X}$ and all $|\lambda| \leq m$; stated differently, $\bar{\Theta}_0$ is the closure of Θ_0 in the weak* topology on D^* . For d^* in $\bar{\Theta}_0$ define

$$g_\infty(x|d^*) = \langle \varphi_x, d^* \rangle.$$

Suppose that $g_\infty(x|d^*)$ is not continuously differentiable to order $|\lambda| \leq m-1$ or does not satisfy $D^\lambda g_\infty(x|d^*) = \langle D^\lambda \varphi_x, d^* \rangle$ for all x in \mathcal{X} and all λ with $|\lambda| \leq m-1$. Then d^* can be replaced by d_0^* in D^* with $g_\infty(x|d_0^*)$ continuously differentiable to order $|\lambda| \leq m-1$ and with $D^\lambda g_\infty(x|d_0^*) = \langle D^\lambda \varphi_x, d_0^* \rangle$ a.e. μ for all λ with $|\lambda| \leq m-1$. Let Θ^* be the set $\bar{\Theta}_0$ with these replacements made as necessary. Then Θ^* is compact in the relative topology on Θ^* generated by the norm $\|d^*\|_{m-1, \infty, \mu}$. The shares

$$s_i(x|d^*) = \left[\sum_{i=1}^N x_i (\partial/\partial x_i) g_\infty(x|d^*) \right]^{-1} x_i (\partial/\partial x_i) g_\infty(x|d^*)$$

and their first partial derivatives in x are continuous in the product topology generated by the Euclidean norm on \mathcal{X} and the norm $\|a\|_{m-1, \infty, \mu}$ on Θ^* .

Proof. By Alaoglu's theorem (Royden, 1963, Ch. 10) the ball $\{d^* \in D^* : \|d^*\|_{m, \infty, \mu} \leq 1\}$ is compact in the weak* topology on D^* . As $\bar{\Theta}_0$ is a weak* closed subset of this ball, $\bar{\Theta}_0$ is compact. In the event that d^* is in $\bar{\Theta}_0$ but not in Θ_0 , $g_\infty(x|d^*)$ may not possess derivatives in the conventional sense. What we shall do is prove that $g_\infty(x|d^*)$ does possess weak derivatives (Adams, 1975, Ch. 1) and moreover if $D^\lambda g_\infty(x|d^*)$ denotes the weak derivative of $g_\infty(x|d^*)$ then $D^\lambda g_\infty(x|d^*) = \langle D^\lambda \varphi_x, d^* \rangle$. Let $t(x)$ be a test function, that is a function that is infinitely many times continuously differentiable and has compact support in the Euclidean topology

on \mathcal{X} . Let θ_n in Θ_0 converge pointwise to d^* . Since $t(x) D^\lambda g_\infty(x|\theta_n)$ is dominated by $b \cdot t(x)$ for all λ with $|\lambda| \leq m$ we have by the Dominated Convergence Theorem and integration by parts that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{\mathcal{X}} t(x) D^\lambda g_\infty(x|\theta_n) dx \\ &= \lim_{n \rightarrow \infty} (-1)^{|\lambda|} \int_{\mathcal{X}} [D^\lambda t(x)] g_\infty(x|\theta_n) dx \\ &= (-1)^{|\lambda|} \int_{\mathcal{X}} [D^\lambda t(x)] g_\infty(x|d^*) dx . \end{aligned}$$

On the other hand

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{\mathcal{X}} t(x) D^\lambda g_\infty(x|\theta_n) dx \\ &= \int_{\mathcal{X}} t(x) [\lim_{n \rightarrow \infty} D^\lambda g_\infty(x|\theta_n)] dx \\ &= \int_{\mathcal{X}} t(x) [\lim_{n \rightarrow \infty} \langle D^\lambda \phi_x, \theta_n \rangle] dx \\ &= \int_{\mathcal{X}} t(x) \langle D^\lambda \phi_x, d^* \rangle dx . \end{aligned}$$

The equality

$$(-1)^{|\lambda|} \int_{\mathcal{X}} [D^\lambda t(x)] g_\infty(x|d^*) dx = \int_{\mathcal{X}} t(x) \langle D^\lambda \phi_x, d^* \rangle dx$$

which holds for every test function shows that the weak derivative $D^\lambda g_\infty(x|d^*)$ exists and that $D^\lambda g_\infty(x|d^*) = \langle D^\lambda \phi_x, d^* \rangle$. Now let $p > k$ and define the pseudonorm on $\bar{\Theta}_0$ by

$$\|d^*\|_{m,p,\mu} = \left(\sum_{|\lambda| \leq m} \int_{\mathcal{X}} |D^\lambda g_\infty(x|d^*)|^p d\mu(x) \right)^{1/p} .$$

Again by the Dominated Convergence Theorem if d_n^* in $\bar{\Theta}_0$ converges weak* to d^* then $\lim_{n \rightarrow \infty} \|d_n^* - d^*\|_{m,p,\mu} = 0$. Thus, the relative topology on $\bar{\Theta}_0$ generated by the pseudonorm $\|d^*\|_{m,p,\mu}$, call it \mathcal{S} , is weaker than the weak* topology whence $\bar{\Theta}_0$ is compact in this topology as well.

By the Sobolev imbedding theorem (Adams, 1975, Ch. 5), for each d^* in $\bar{\Theta}_0$ the corresponding $g_\infty(x|d^*)$ can be modified (if necessary) on a set of measure zero so that $g_\infty(x|d^*)$ and its derivatives $D^\lambda g_\infty(x|d^*)$ up to order $|\lambda| \leq m-1$ are continuous on \mathcal{X} and

$$\max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |D^\lambda g_\infty(x|d^*)| \leq c \|d^*\|_{m,p,\mu}$$

where the bound c is independent of d^* . Stated differently, there is a continuous mapping from the compact topological space $(\bar{\Theta}_0, \mathfrak{S})$ to the topological space (Θ^*, \mathfrak{J}) where \mathfrak{J} is the topology generated by the norm $\|d^*\|_{m-1,\infty,\mu}$. As this mapping is onto, (Θ^*, \mathfrak{J}) must be compact. Let $\lim_{n \rightarrow \infty} x_n = x$ in the Euclidean norm on \mathcal{X} and let $\lim_{n \rightarrow \infty} \|d_n - d^*\|_{m-1,\infty,\mu} = 0$. As $m-1 \geq 2$ we have that $(\partial/\partial x_i)g_\infty(x|d_n^*)$ converges uniformly to $(\partial/\partial x_i)g_\infty(x|d^*)$. Then $(\partial/\partial x_i)g_\infty(x_n|d_n^*)$ converges to $(\partial/\partial x_i)g_\infty(x|d^*)$. Similarly $(\partial^2/\partial x_i \partial x_j)g_\infty(x_n|d_n^*)$ converges to $(\partial^2/\partial x_i \partial x_j)g_\infty(x|d^*)$. As $x_i(\partial/\partial x_i)g_\infty(x|d^*)$ is bounded away from zero, shares and their first derivatives are continuous. \square

3. Uniform Strong Consistency

We can now prove Theorem 1. Let $\{(e_t, x_t)\}_{t=1}^\infty$ be a realization such that $\lim_{n \rightarrow \infty} K_n = \infty$ and

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta^*} |s_n[g_\infty(\theta)] - \bar{s}[g_\infty(\theta), g^*]| = 0;$$

recall that n denotes the sample size. Almost every realization is such by hypothesis. Recall also that to $g^*(x)$ there corresponds a point $\theta^* \in \Theta^* \cap \Theta$ such that

$$\lim_{n \rightarrow \infty} \max_{|\lambda| \leq m} \sup_{x \in \mathcal{X}} |D^\lambda g^*(x) - D^\lambda g_{K_n}(x|a^*)| = 0$$

and set

$$\theta_n^* = (\theta_1^*, \theta_2^*, \dots, \theta_{p_{K_n}}^*, 0, 0, \dots)$$

where p_{K_n} is the last term of the series $g_{K_n}(x|\theta)$. Similarly let $\hat{\theta}_n$ in $\Theta^* \cap \Theta$ have zero components past p_{K_n} .

$$\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n}, \dots, \hat{\theta}_{p_{K_n}, m}, 0, 0, \dots)$$

and minimize $s_n[g_{K_n}(\theta)]$ over $\Theta^* \cap \Theta$. Note that $g_\infty(x|\hat{\theta}_n) = g_{K_n}(x|\hat{\theta}_n)$ and $g_\infty(x|\theta_n^*) = g_{K_n}(x|\theta_n^*)$.

Let $\sigma(x|\theta)$ denote an elasticity of substitution computed from $g_\infty(x|\theta)$ for $\theta \in \Theta^*$. Thus $\hat{\sigma}_n(x) = \sigma(x|\hat{\theta}_n)$ and $\sigma^*(x) = \sigma(x|\theta^*)$ with $\hat{\theta}_n$ and θ^* in $\Theta^* \cap \Theta$. Note that by the definition of Θ^*

$$\sup_{\theta \in \Theta^*} \sup_{x \in \mathcal{X}} |s(x|\theta)| < \infty.$$

Thus the sequence

$$\{\sup_{x \in \mathcal{X}} |\hat{\sigma}_n(x) - \sigma^*(x)|\}_{n=1}^\infty$$

lies in a closed and bounded interval and has at least one limit point τ^0 . Let n_i be a sequence such that

$$\lim_{i \rightarrow \infty} \sup_{x \in \mathcal{X}} |\hat{\sigma}_{n_i}(x) - \sigma^*(x)| = \tau^0.$$

Now $\hat{\theta}_{n_i}$ is a sequence in the compact set Θ^* and has at least one limit point θ^0 . Let n_j be a subsequence of n_i with $\lim_{j \rightarrow \infty} \|\hat{\theta}_{n_j} - \theta^0\|_{m-1, \infty, \mu} = 0$.

Then by uniform convergence

$$\begin{aligned} \bar{s}[g_\infty(\theta^0), g^*] &= \lim_{j \rightarrow \infty} s_{n_j}[g_\infty(\hat{\theta}_{n_j})] \\ &= \lim_{j \rightarrow \infty} s_{n_j}[g_{K_{n_j}}(\hat{\theta}_{n_j})] \\ &\leq \lim_{j \rightarrow \infty} s_{n_j}[g_{K_{n_j}}(\theta_{n_j}^*)] \\ &= \lim_{j \rightarrow \infty} s_{n_j}[g_\infty(\theta_{n_j}^*)] \\ &= \bar{s}(g^*, g^*). \end{aligned}$$

where the inequality follows from the fact that $\hat{\theta}_{n_j}$ minimizes $s_{n_j}[g_{K_{n_j}}(\theta)]$ over $\Theta^* \cap \Theta$ while $\theta_{n_j}^*$ does not; for large enough n_j , $\theta_{n_j}^*$ is in Θ^* because θ^* is in Θ^* and $\lim_{n \rightarrow \infty} \|\theta_{n_j}^* - \theta^*\|_{m, \infty, \mu} = 0$. By the Identification Condition $\nabla g_{\infty}(x|\theta^0) = \alpha^0(x) \nabla g^*(x)$ at every point x in \mathcal{X} for some $\alpha^0(x) > 0$. Let $\sigma^0(x)$ be the elasticity of substitution computed from $g_{\infty}(x|\theta^0)$. As the meaning of $\lim_{n \rightarrow \infty} \|\hat{\theta}_{n_j} - \theta^0\|_{m-1, \infty, \mu}$ is that

$$\lim_{n \rightarrow \infty} \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |D^{\lambda} g_{\infty}(x|\hat{\theta}_{n_j}) - D^{\lambda} g_{\infty}(x|\theta^0)| = 0$$

we have that

$$\lim_{j \rightarrow \infty} \sup_{x \in \mathcal{X}} |\hat{\sigma}_{n_j}(x) - \sigma^0(x)| = 0.$$

But $\sigma^0(x) = \sigma^*(x)$ so we have

$$\lim_{j \rightarrow \infty} \sup_{x \in \mathcal{X}} |\hat{\sigma}_{n_j}(x) - \sigma^*(x)| = 0.$$

Since n_j is a subsequence from n_i we must have that $\tau^0 = 0$ which proves the result since every limit point of

$$\left\{ \sup_{x \in \mathcal{X}} |\hat{\sigma}_n(x) - \sigma^*(x)| \right\}_{n=1}^{\infty}$$

must be zero. \square

Footnotes

1. The consistency proof given by Burguete, Gallant, and Souza (1982) requires that the class \mathcal{G} consist of parametric functions of the form $g(x|\theta)$ where θ is restricted to some finite dimensional, compact set Θ . Effectively, the contribution of this paper will be to eliminate the restriction that Θ be finite dimensional.
2. The process of constructing these $\varphi_j(x)$ (and $D^\lambda \varphi_j$) for given x can be automated and FORTRAN code for constructing them can be obtained from the authors at the cost of reproduction and postage.
3. To state this differently, if $\|g\| = \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |g(x)|$ and $\|\theta\| = \lim_{K \rightarrow \infty} \max_{|\lambda| \leq m-1} \sup_{x \in \mathcal{X}} |\sum_{j=1}^K \theta_j D^\lambda \varphi_j(x)|$ then the normed linear space $\{g : \|g\| < \infty\}$ is isometrically isomorphic to the normed linear space $\{\theta : \|\theta\| < \infty\}$.

References

- Adams, Robert A. (1975), Sobolev Spaces. New York: Academic Press.
- Burguete, Jose F., A. Ronald Gallant, and Geraldo Souza (1982), "On the Unification of the Asymptotic Theory of Nonlinear Econometric Models," Econometric Reviews, forthcoming.
- Edmunds, D. E. and V. B. Moscatelli (1977), "Fourier Approximation and Embeddings of Sobolev Spaces," Dissertationes Mathematicae CXLV.
- Gallant, A. Ronald (1977), "Three-Stage Least Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," Journal of Econometrics 5, 71-88.
- Gallant, A. Ronald (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," Journal of Econometrics 15, 211-245.
- Gallant, A. Ronald (1982), "Unbiased Determination of Production Technologies," Journal of Econometrics, forthcoming.
- Gallant, A. Ronald and Alberto Holly (1980), "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," Econometrica 48, 697-720.
- Guilkey, David, C. A. Knox Lovell, and Robin C. Sickles (1981), "A Comparison of Three Flexible Functional Forms," Unpublished manuscript, University of North Carolina, Chapel Hill, NC.
- Jennrich, Robert I. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators," Annals of Mathematical Statistics 40, 633-643.
- Malinvaud, E. (1970), "The Consistency of Nonlinear Regressions," Annals of Mathematical Statistics 41, 965-969.

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.

Royden, H. L. (1963), Real Analysis. New York: The Macmillan Company.

Rudin, Walter (1973), Functional Analysis. New York: McGraw-Hill Book Company.

White, Halbert (1980), "Using Least Squares to Approximate Unknown Regression Functions," International Economic Review 21, 149-170.