Unbiased Determination of Production Technologies*

by

A. Ronald Gallant

Department of Statistics
North Carolina State University
Raleigh, NC 27650-5457

*Much of this work was done while visiting the Centro de Estadistica, Colegio de Postgraduados, Chapingo, Mexico and Empresa Brasileria de Pesquisa Agropecuária, Brasília, Brasil. I wish to thank Ernst R. Berndt, Dale W. Jorgenson, William E. Schworm, V. Kary Smith, and John A. Weymark for helpful discussions. Special thanks are due Janice J. Gaddy. This research was supported by North Carolina Agricultural Experiment Station Project NC03641 and by National Research Foundation Grant SES80-14239.

Abstract

To determine whether an industry exhibits constant returns to scale, whether the production function is homothetic, or whether inputs are separable, a common approach is to specify a cost function, estimate its parameters using data such as prices and quantities of inputs, and then test the parametric restrictions corresponding to constant returns, a homothetic technology, or separability. Statistically, such inferences are valid if the true cost function is a member of the parametric class considered, otherwise the inference is biased. That is, the true rejection probability is not necessarily adequately approximated by the nominal size of the statistical test. The use of fixed parameter flexible functional forms such as the Translog, the generalized Leontief, or the Box-Cox will not alleviate this problem.

The Fourier flexible form differs fundamentally from other flexible forms in that it has a variable number of parameters and a known bound, depending on the number of parameters, on the error, as measured by the Sobolov norm, of approximation to an arbitrary cost function. Thus it is possible to construct statistical tests for constant returns, a homothetic technology, or separability which are asymptotically unbiased by letting the number of parameters of the Fourier flexible form depend on sample size. That is, the true rejection probability converges to the nominal size of the test as sample size tends to infinity. The rate of convergence depends on the smoothness of the true cost function; the more times is differentiable the true cost function, the faster the convergence.

The method is illustrated using the data on aggregate US manufacturing of Berndt and Wood (1975, 1979) and Berndt and Khaled (1979).

1. Introduction

Recently there has been an interest in testing assumptions that are routinely imposed in the estimation of consumer or factor demand systems. An example of the former is Christensen, Jorgenson and Lau (1975) and of the latter Berndt and Khaled (1979). Flexible functional forms are used in this work so as to impose minimal <u>a priori</u> assumptions. The idea is to be certain that it is the economic proposition that is being tested and not some obscure consequence of specification error. Restated in statistical terminology, the idea is to use a flexible functional form so that the actual rejection probability of the test agrees closely with the nominal rejection probability of the test under any true state of nature that satisfies the economic proposition. That is, the objective is the construction of an unbiased test of the economic proposition. Unfortunately, tests constructed from flexible functional forms are less successful in controlling bias than one might hope. As an example, computations reported in Section 7 of Gallant (1981) show that the Translog test of an additive indirect utility function can be seriously biased in favor of rejection.

In this paper we study the question: Can unbiased tests actually be constructed? Addressing this question in the context of fitting factor demand systems, we find that it is possible to construct tests based on a logarithmic version of the Fourier flexible form that are asymptotically unbiased. The development of the ideas is as follows.

Setting previous conceptions aside and proceeding from first principles, we are led to consider which errors in the approximation of the true cost function $g(x)$ by a flexible form $g_K(x|\theta)$ are important and which can be neglected. Having determined the relevant approximation errors, we search for a useful measure of distance -- one which is large under relevant approximation errors and which neglects others. We find that the Sobolov norm is the required measure of distance. Our task now is to find a function $g_K(x|\theta)$ which can approximate $g(x)$ as closely as desired in Sobolov norm. We find that a logarithmic version of the Fourier flexible form has this property. The Fourier flexible form of order K can approximate g to within

$$||g - g_K(\theta)||_{\ell,p,\mu} = o(K^{-m+\ell+\varepsilon})$$

for any $\varepsilon > 0$ where m is the number of times is differentiable g, ℓ is the largest order partial derivative regarded as important in the approximation, and $p = 1$ for an L_1 type norm, $p = 2$ for an L_2 norm, $p = \infty$ for a sup norm, and so on.

To derive tests for a homothetic production technology, or constant returns to scale, or the separability of a set of factor inputs, our method is to impose the parametric constraints which cause the Fourier flexible form to exhibit homotheticity, or constant returns, or separability and then test these restrictions statistically, the same as is done with any other flexible form. For the method to be valid according to the preceeding considerations, the imposition of the parametric restriction should not destroy the ability of the Fourier form to approximate any $g(x)$ that satisfies homotheticity, or constant returns, or separability to within

$$||g-g_K(\theta)||_{\ell,p,\mu} = o(K^{-m+\ell+\varepsilon})$$

We find that the Fourier flexible form satisfies this requirement.

Included in this discussion is a study of the limiting behavior of the tests for a homothetic technology and for constant returns when the sample size n_K and the number of parameters p_K increase with the order K of the fitted Fourier form $g_K(x|\theta)$. This is equivalent to letting the number of parameters p_K depend on the sample size n_K . We show that, as K increases, the actual rejection probability tends to the nominal rejection probability if the null hypothesis is true,

$$\lim_{K\to\infty} P(T_K > c_K | H) = \alpha.$$

The algebraic form of the error bound we obtain indicates that if one's objective is bias reduction then it is best to fit as large an order of K as the number of available observations will permit. We assume that the variance-covariance matrix Σ of the errors is known in the proof. As yet we have not overcome the technical difficulties of an extension to estimated Σ or to the nonlinear separability test. Nevertheless, we accept the theoretical results obtained to date as indicative of the validity of the general principle of large K values and move on to some applications.

In the first application, we use the data of Berndt and Wood (1975) and Berndt and Khaled (1979) to test for a homothetic production technology and constant returns in United States manufacturing. We reject both homotheticity and constant returns given homotheticity as do Berndt and Khaled (1979) who use a Box-Cox flexible form. The lack of fit due to imposing constant returns after homotheticity accounts for an overwhelming proportion of the total lack of fit. Comparatively speaking, homotheticity is a mild restriction.

In the second application, with the same data, we test whether capital (K), labor (L), and energy (E) inputs are separable from materials (M) in United States manufacturing. The relevance of this hypothesis in the interpretation of σ_{KE} estimates based on KLE data obtained by Griffen and Gregory (1976) and others is discussed in Berndt and Wood (1979). Using a Translog flexible form, Berndt and Wood (1975) reject the hypothesis of KLE-M separability at a level of $\alpha = .01$. We accept KLE-M separability at a level of $\alpha = .10$.

In the third application, with the same data, we estimate σ_{KE}, the elasticity of substitution of capital for energy, using the Fourier flexible form. We find for the year 1959 that

$$\hat{\sigma}_{KE} = .666134 \qquad SE(\hat{\sigma}_{KE}) = 8.55145$$

and, with homotheticity imposed, that

$$\hat{\sigma}_{KE} = -4.20666 \qquad SE(\hat{\sigma}_{KE}) = 3.49984 \ .$$

Taking account of the standard errors, we conclude that these data do not permit a decision that capital and energy are complements.

2. Arbitrarily Accurate Approximations to a Cost Function

The producers 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 $p=(p_1,p_2,\ldots,p_N)'$. For mathematical convenience we shall assume throughout that the cost function has continuous partial derivatives at least up to the third order.

The theory of the firm implies restrictions on the functional form of the cost function. A list of these restrictions and some plausible hypotheses is as follows [Diewert (1974) and Blackorby, Primont, and Russell (1978)].

R_0. Positive linear homogeneity, $c(p,u)$ is a positive valued function defined on the positive orthant and $c(p,u)$ is homogeneous of degree one in p, $\lambda c(p,u) = c(\lambda p,u)$.

R_1. Constant returns to scale, $c(p,u) = u\, c(p)$.

R_2. Homothetic production technology, $c(p,u) = b(u)c(p)$.

R_3. Homothetically weakly separable, $c(p,u) = \bar{c}[c_1(p_{(1)}),p_{(2)},u]$ where $\bar{c}$ and c_1 satisfy R_0.

R_4. Monotonicity, $(\partial/\partial p_i)\, c(p,u) > 0$ where $i = 1,2,\ldots, N$, $p'(\partial/\partial p)c(p,u) = c(p,u)$, and $(\partial/\partial u)c(p,u) > 0$.

R_5. Concavity, $(\partial^2/\partial p\partial p')c(p,u)$ is a negative semi-definite matrix of rank $N-1$ with p being the eigenvector of root zero.

The set of cost functions that satisfy R_0 is denoted by $\mathfrak{M}_0$, $\mathfrak{M}_{02}$ are those that satisfy R_0 and R_2 and so on. If c in $\mathfrak{M}_0$ is the true model then the derived demand function will satisfy (Shephard's lemma)

$$q = (\partial/\partial p)\, c(p,u) \quad .$$

One can attempt to determine c statistically by fitting equations of this sort to observed inputs q_t, prices p_t, and outputs u_t, $t=1,2,\ldots, n$. In such a study, the quantities of interest are usually the elasticities of

$$\sigma_{ij} = \frac{c(p,u)\ (\partial^2/\partial p_i \partial p_j)\ c(p,u)}{[(\partial/\partial p_i)\ c(p,u)][(\partial/\partial p_j)\ c(p,u)]}$$

and the price elasticities (Allen, 1938)

$$\eta_{ij} = \frac{\partial \ell n q_i}{\partial \ell n p_j} = \sigma_{ij}(p,u) \frac{p_j (\partial/\partial p_j)\ c(p,u)}{c(p,u)} \quad .$$

We find that the notational burden is considerably reduced if the problem is restated in logarithmic quantities. Accordingly, let

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

$$v = \ell n\ u + \ell n\ a_{N+1} ,$$

and

$$g(\ell, v) = \ell n\ c\left(\frac{e^{\ell_1}}{a_1}, \frac{e^{\ell_2}}{a_2}, \ldots, \frac{e^{\ell_N}}{a_N}, \frac{e^{v}}{a_{N+1}}\right)$$

where $\ell = (\ell_1, \ell_2, \ldots, \ell_N)'$. The $\ell n a_i$ are location parameters to be determined later.

Using

$$(\partial/\partial p)\ c(p,u) = c(p,u)\ P^{-1}\ \nabla g$$

$$(\partial^2/\partial p \partial p')\ c(p,u) = c(p,u)\ P^{-1}[\nabla^2 g + \nabla g \nabla' g - \mathrm{diag}(\nabla g)]P^{-1}$$

where

$$\nabla g = (\partial/\partial \ell)\ g(\ell, v)$$

$$\nabla^2 g = (\partial^2/\partial \ell \partial \ell')g(\ell, v)$$

$$P = \mathrm{diag}(p) ,$$

a list of conditions on $g(\ell,v)$ equivalent to those on $c(p,u)$ is:

R_0. Positive linear homogeneity, $g(\ell,v)$ is a real valued function defined on R^{N+1} and $g(\ell+\tau 1,v) = \tau + g(\ell,v)$ where $1 = (1,1,\ldots,1)'$.

R_1. Constant returns to scale, $g(\ell,v) = \ln u + g(\ell)$

R_2. Homothetic production technology, $g(\ell,v) = h(v) + g(\ell)$.

R_3. Homothetically weakly separable, $g(\ell,v) = \bar{g}[g_1(\ell_{(1)}), \ell_{(2)}, v]$ where $\bar{g}$ and g_1 satisfy R_0.

R_4. Monotonicity, $(\partial/\partial\ell_i)g(\ell,v) > 0$ where $i = 1,2,\ldots, N$, $1'(\partial/\partial\ell)g(\ell,v) = 1$, and $(\partial/\partial v)g(\ell,v) > 0$.

R_5. Concavity, $\nabla^2 g + \nabla g \nabla' g - \text{diag}(\nabla g)$ is a negative semi-definite matrix of rank $N-1$ with 1 being the eigenvector of root zero.

Letting $s = (p_1q_1, p_2q_2, \ldots, p_Nq_N)'/(\Sigma_{i=1}^N p_iq_i)$ be the N-vector of input cost shares, Shephard's lemma becomes

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

the elasticities of substitution σ_{ij} become the elements of the matrix

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

and the price elasticities η_{ij} become the elements of the matrix

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

where the rows of η correspond to inputs and the columns to prices.

The result of the logarithmic transformation is to shift the focus to a determination of g in $\mathfrak{m}_0$ by fitting

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

and possibly

$$\ln c = g(\ell,v)$$

to observed shares s_t, ℓog prices ℓ_t, ℓog outputs v_t, and possibly ℓog costs $\ln c_t$, $t=1,2,\ldots,n$.

There are several approaches to the determination of $c(p,u)$ or equivalently $g(\ell,v)$. One approach is to determine the derived demand curve statistically and then solve back to the cost function. A famous study of this genre is Arrow, Chenery, Minhas and Solow (1965) where the function

$$\ell n\, s_i = a + b\, \ell n\, p_i$$

was identified statistically $(N=2)$. It implies the CES production function. A CES production function has a cost function of the CES form. Another approach has been to argue that after a suitable monotonic transformation of the prices, $z = (\theta(p_1),\ldots,\theta(p_N))'$ a quadratic expansion in the transformed prices $c(p,u) = f(b'z + \frac{1}{2}z'Bz, u)$ ought to give a sufficiently accurate approximation in applications. Some cost functions generated in this fashion are the Translog $\theta(p_i) = \ell n(p_i)$, the generalized Leontief $\theta(p_i) = \sqrt{p_i}$, and the Box-Cox $\theta(p_i) = (p_i^{\lambda} - 1)/\lambda$. The result of these and other approaches has been to generate a finite collection of (logarithmic) cost functions

$$\mathfrak{m}_0^{\#} = \{g_1, g_2, \ldots, g_L\}$$

each of which has performed well in some sense in some application(s). However, if one assumes that, say, g_1 in $\mathfrak{m}^{\#}$ is the true cost function then all other cost functions in $\mathfrak{m}_0^{\#}$ which do not generate g_1 as a special case will bias statistical inferences in some undesirable direction (Guilkey, Lovell and Sickeles, 1981). Contin attacks on the problem along these lines can only lead to an ever larger class of models each leading to biased inferences if some plausible alternative is in fact true. The basic problem leading to this state of affairs is that the problem: Find $g \in \mathfrak{m}_0$ by observing $s = (\partial/\partial\ell)\, g(\ell,v)$ is statistically intractable because $\mathfrak{m}_0$ is too large a class of models. We need some sensible method to reduce $\mathfrak{m}_0$ to manageable size.

One method of reducing $\mathfrak{m}_0$ to a manageable class of models is as follows.

First, determine which approximation errors are important and which can be neglected. That is, g^* is to be approximated by g and one chooses a norm $\|e\|$ which is sensitive to important approximation errors $e = g^* - g$. Second, find a functional form $g_K(\ell, v|\theta)$ with a variable number p_K of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_{p_K})'$ such that g_K is in $\mathfrak{m}_0$ for all $\theta \in \Theta$ and

$$\ell im_{K\to\infty} \min_{\Theta} \|g - g_K\| = 0$$

for every g in $\mathfrak{m}_0$. That is one finds a conveniently indexed dense subset of $\mathfrak{m}_0$,

$$\mathfrak{m}_0' = \{g_K(\ell, v|\theta) : \theta \in \Theta \subset R^{p_K}; K=1,2,\ldots\} \subset \mathfrak{m}_0 .$$

Mathematically, the smaller class $\mathfrak{m}_0'$ can be regarded as equivalent to $\mathfrak{m}_0$ as its closure contains $\mathfrak{m}_0$. Statistically, the problem shifts from trying to find g in $\mathfrak{m}_0$ to trying to determine an adequate value of K. Finding an adequate K is a much more tractable statistical problem. To follow this scheme, we are led to consider approximation errors.

As noted earlier, the quantities of interest in an empirical study are the derived demand functions

$$s = \nabla g ,$$

The elasticities of substitution

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

and possibly the cost function itself. One can see from these formulas that it is necessary to approximate $g(\ell, v)$ accurately, the first partial derivatives $(\partial/\partial\ell_i)\, g(\ell, v)$ accurately, and the second partial derivatives $(\partial^2/\partial\ell_i \partial\ell_j)\, g(\ell, v)$ accurately. Other errors of approximation are irrelevant.

We anticipate that an approximation over a rectangle in the positive orthant will suffice. Accordingly, let the boundaries of this rectangle be

$$0 < p_i^L < p_i^H < \infty \qquad i = 1,2,\ldots, N,$$

$$0 < u^L < u^H < \infty \quad .$$

These choices are at the arbitrary discretion of the investigator. However, all data values must lie within these limits as well as any values (p,u) for which predictions are planned. Setting limits slightly larger than anticipated predictions and the extreme values observed in the data would be a reasonable choice. Choose the location parameters $a_1, a_2, \ldots, a_{N+1}$ of the transformation[1/]

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

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

so that

$$\ell_i^L = \ell n\ p_i^L + \ell n\ a_i > 0 \qquad i = 1,2,\ldots, N\ ,$$

$$v^L = \ell n\ u^L + \ell n\ a_{N+1} > 0\ .$$

Then letting

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

the region of approximation is

$$\mathfrak{X} = \{x = (\ell', v)' : \ell_i^L < \ell_i < \ell_i^H,\ v^L < v < v^H\}$$

which is an open rectangle in the positive orthant of R^{N+1} . The closure $\bar{\mathfrak{X}}$ is also a proper subset of the positive orthant.

Next we shall define a measure of distance, the Sobolov norm, which is sensitive to relevant approximation errors over $\mathfrak{X}$. First we need some additional notation.

1/ The definition of v is modified in the last paragraph of this section.

A multi-index is an N + 1 - vector with integer components. The length of a multi-index k is defined as

$$|k|^* = \Sigma_{i=1}^{N+1} |k_i| \ .$$

Letting λ be a multi-index with non-negative components, partial differentiation of a function g(x) is denoted as

$$D^\lambda g(x) = \frac{\partial^{|\lambda|^*}}{\partial x_1^{\lambda_1} \partial x_2^{\lambda_2} \cdots \partial x_{N+1}^{\lambda_{N+1}}} g(x) \ .$$

We assume that $D^\lambda g(x)$ is a continuous function of x whenever this notation is used. Let $\mu(x)$ be a continuous distribution function with bounded density function giving the relative frequency with which values of the independent variable x_t occur as sample size n tends to infinity [Malinvaud (1970), Gallant and Holly (1980)]. We use this notation to define the Sobolov norm.

<u>Sobolov norm</u>. If $1 \le p < \infty$ let $W^{m,p}(\mathcal{X})$ denote the collection of all complex valued functions g(x) with $\int_{\mathcal{X}} |D^\lambda g| \, dx < \infty$ for all λ up to $|\lambda|^* \le m$. Given $g \in W^{m,p}(\mathcal{X})$ define its norm as

$$\|g\|_{m,p,\mu} = (\Sigma_{|\lambda|^* \le m} \int_{\mathcal{X}} |D^\lambda g|^p d\mu)^{1/p} \ .$$

If $p = \infty$ let $W^{m,\infty}(\mathcal{X})$ denote the collection of all complex valued functions g with $\sup_{x \in \mathcal{X}} |D^\lambda g(x)| < \infty$ for all λ up to $|\lambda|^* \le m$. Given $g \in W^{m,\infty}(\mathcal{X})$ define its norm as

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

We see from the definition that the Sobolov norm is the relevant notion of distance. Next we show how to construct dense subsets of $\mathfrak{M}_0$ with respect to this norm. The construction is based on a modification

of a result due to Edmunds and Moscatelli (1977). The discussion will be brief, see Gallant (1981) for full details.

The construction of $\mathfrak{m}_0'$ is based on a sequence of elementary multi-indexes

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

obtained as follows.

<u>Elementary Multi-Indexes</u>. Let

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

the set of multi-indexes of dimension N+1 and length $|k|^* = \Sigma_{i=1}^{N+1} |k_i| \le K$. 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 integral 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, ..., k_{N+1} are the elementary vectors and $|k_\alpha|^*$ is non-decreasing in α. Define J to be the smallest positive integer with

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

This construction is tedious if attempted by hand for $N+1>3$. FORTRAN code for constructing $\{k_\alpha\}$ is available in Monahan (1981).

The Fourier flexible form can be used to generate dense subsets of a Sobolov space. It is defined as follows

<u>Fourier flexible form</u> (Gallant, 1981). Consider as an approximation to a real valued function $g \in W^{m,p}(\mathcal{X})$

$$g_K(x|\theta) = a_0 + b'x + \tfrac{1}{2}x'Cx + \Sigma_{\alpha=1}^{A}\Sigma_{j=-J}^{J} a_{j\alpha}\, e^{ij\lambda k_\alpha' x}$$

where

$$a_{j\alpha} = \bar{a}_{-j\alpha}\,, \quad C = -\Sigma_{\alpha=1}^{A}\, a_{0\alpha}\, \lambda^2\, k_\alpha k_\alpha'$$

and a_0 , $a_{0\alpha}$, and b are real valued. The overbar denotes complex conjugation and $i = \sqrt{-1}$. The scaling factor λ is computed as

$$\lambda = (2\Pi - \epsilon)/\max\{x_i^H : i = 1,2,\ldots,N+1\}$$

for some ϵ with $0 < \epsilon < 2\Pi$. A reasonable choice is $(2\Pi - \epsilon) = 6$. To obtain the derivatives of $g_K(x|\theta)$, partition k_α and b as

$$k_\alpha = (r_\alpha', k_{N+1})', \quad b = (c', b_{N+1})'$$

where r_α and c are N-vectors. Then

$$(\partial/\partial\ell) g_K(x|\theta) = c + \lambda\Sigma_{\alpha=1}^{A}(-a_{0\alpha}\lambda k_\alpha' x + i\Sigma_{j=-J}^{J}\, j a_{j\alpha} e^{ij\lambda k_\alpha' x}) r_\alpha\,,$$

$$(\partial^2/\partial\ell\partial\ell') g_K(x|\theta) = -\lambda^2\Sigma_{\alpha=1}^{A}(a_{0\alpha} + \Sigma_{j=-J}^{J} j^2 a_{j\alpha} e^{ij\lambda k_\alpha' x}) r_\alpha r_\alpha'\,.$$

The purpose of the scaling factor λ is to be sure that $\lambda\mathfrak{X} = \{\lambda x : x \in \mathfrak{X}\}$ is a rectangle with no edge longer than 2Π , that is $\lambda[\max\{x_i : x \in \mathfrak{X}\} - \min\{x_i : x \in \mathfrak{X}\}] < 2\Pi \quad i = 1,2,\ldots,N+1$. This is the essential feature of the construction. The location of $\mathfrak{X}$ in the positive orthant by choice of a_i is only for convenience in applications. It does not affect the theory as any shift in location can be absorbed into the coefficients a_0 and b of the quadratic part and the coefficients $a_{j\alpha}$ of the exponential part. What is essential is that $\lambda\mathfrak{X}$ has no edge longer than 2Π .

Although complex valued exponential representations are more convenient in discussion, in applications sine/cosine representations are easier to work with. Writing

$$a_{0\alpha} = u_{0\alpha} \qquad \alpha = 1,2,\ldots,A$$

$$a_{j\alpha} = u_{j\alpha} + iv_{j\alpha} \qquad \alpha = 1,2,\ldots,A\ , \quad j = 1,2,\ldots,\ J$$

$$a_{-j\alpha} = u_{j\alpha} - iv_{j\alpha} \qquad \alpha = 1,2,\ldots,A\ , \quad j = 1,2,\ldots,\ J$$

and using $e^{ij\lambda k_\alpha' x} = \cos(j\lambda k_\alpha' x) + i\sin(j\lambda k_\alpha' x)$ one has that

$$g_K(x|\theta) = u_0 + b'x + \tfrac{1}{2}x'Cx$$

$$+ \Sigma_{\alpha=1}^{A}\{u_{0\alpha} + 2\Sigma_{j=1}^{J}[u_{j\alpha}\cos(j\lambda k_\alpha' x) - v_{j\alpha}\sin(j\lambda k_\alpha' x)]\}$$

$$(\partial/\partial \ell)\ g_K(x|\theta) =$$

$$c - \lambda\Sigma_{\alpha=1}^{A}\{u_{0\alpha}\lambda k_\alpha' x + 2\Sigma_{j=1}^{J} j[u_{j\alpha}\sin(j\lambda k_\alpha' x) + v_{j\alpha}\cos(j\lambda k_\alpha' x)]\}r_\alpha$$

$$(\partial^2/\partial \ell \partial \ell')g_K(x|\theta) =$$

$$-\lambda^2\Sigma_{\alpha=1}^{A}\{u_{0\alpha} + 2\Sigma_{j=1}^{J} j^2[u_{j\alpha}\cos(j\lambda k_\alpha' x) - v_{j\alpha}\sin(j\lambda k_\alpha' x)]\}r_\alpha r_\alpha' \ .$$

Letting

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

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

the parameters of $g_K(x|\theta)$ are

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

which is a vector of length $1 + N + 1 + A(1 + 2J)$.

A verification that the Fourier flexible form can be used to construct a dense subset of $\mathfrak{m}_0$ depends on the following result.

Theorem 1. [Edmunds and Moscatelli (1977), Corollary 1]. Let the real valued function $g(x)$ be continuously differentiable up to order m on an open set containing the closure of $\mathcal{X}$. Then it is possible to choose a triangular array of coefficients $\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_K, \ldots$ such that for every p with $1 \le p \le \infty$, every ℓ with $0 \le \ell < m$, and every $\epsilon > 0$

$$\|g - g_K(\bar{\theta}_K)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$$

as $K \to \infty$.

We are now in a position to construct $\mathfrak{M}_0'$ but first we need the following lemma.

Lemma 1. A function which satisfies R_0 can be written as

$$g(\ell,v) = \Sigma_{i=1}^N w_i \ell_i + g_0(\ell_1 - \ell_N, \ldots, \ell_{N-1} - \ell_N, v)$$

where $\Sigma_{i=1}^N w_i = 1$.

Proof. The transformation

$$d = (d_1, d_2, \ldots, d_{N+1})$$
$$= (\Sigma_{i=1}^N w_i \ell_i, \ell_1 - \ell_N, \ldots, \ell_{N-1} - \ell_N, v)$$

is invertable so that there is a function $h(d)$ such that

$$g(\ell,v) = h(\Sigma_{i=1}^N w_i \ell_i, \ell_1 - \ell_N, \ldots, \ell_{N-1} - \ell_N, v).$$

If $g(\ell,v)$ satisfies R_0 then

$$h(\Sigma_{i=1}^N w_i \ell_i, \ell_1 - \ell_N, \ldots, \ell_{N-1} - \ell_N, v)$$
$$= g(\ell, v)$$
$$= \tau + g(\ell - \tau 1, v)$$
$$= \tau + h(\Sigma_{i=1}^N w_i \ell_i - \tau, \ell_1 - \ell_N, \ldots, \ell_{N-1} - \ell_N, v).$$

In particular set $\tau = \Sigma_{i=1}^N w_i \ell_i$ and the result follows. □

The construction of $\mathfrak{M}_0'$ is given by the following theorem. It states that a cost function that satisfies R_0 can be approximated by the Fourier flexible form with all terms with $\Sigma_{i=1}^{N} k_{i\alpha} \neq 0$ deleted and the constraint $\Sigma_{i=1}^{N} b_i = 1$ imposed.

Theorem 2. Let the real valued function $g(x)$ satisfy R_0 and be continuously differentiable up to order m on an open set containing the closure of $\mathcal{X}$. Let

$$g_K(\ell, v|\theta) = a_0 + b'x + \tfrac{1}{2}x'Cx + \Sigma_{\alpha=1}^{A}\Sigma_{j=-J}^{J} a_{j\alpha} e^{ij\lambda k_\alpha' x}$$

where $x = (\ell', v)'$ subject to the following parametric restrictions

$$R_0' \quad \begin{cases} 1'c = 1 & \text{where } b = (c', b_{N+1})' \\ a_{j\alpha} = 0 & \text{if } 1'r_\alpha \neq 0 \text{ where } k_\alpha = (r_\alpha', k_{N+1,\alpha})' . \end{cases}$$

Then $g_K(\ell, v|\theta)$ satisfies R_0 and it is possible to choose a triangular array of coefficients $\bar\theta_1, \bar\theta_2, \ldots, \bar\theta_K, \ldots$ such that for every p with $1 \leq p \leq \infty$, every ℓ with $0 \leq \ell < m$, and every $\epsilon > 0$

$$\|g - g_K(\bar\theta_K)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$$

as $K \to \infty$. Recall that

$$a_{j\alpha} = \bar a_{-j\alpha}, \quad C = -\Sigma_{\alpha=1}^{A} a_{0\alpha}\lambda^2 k_\alpha k_\alpha'$$

and that a_0, $a_{0\alpha}$, and b are real valued.

Proof. By Lemma 1

$$g(x) = \Sigma_{i=1}^{N} w_i x_i + g_0(x_i - x_N, \ldots, x_{N-1} - x_N, x_{N+1})$$

where $\Sigma_{i=1}^{N} w_i = 1$. Let

$$d = d(x) = (x_1 - x_N, \ldots, x_{N-1} - x_N, x_{N+1})' .$$

Now $\Sigma_{|k|^* \leq K} a_k e^{i\lambda k'd}$ and $\Sigma_{\alpha=1}^{A}\Sigma_{j=-J}^{J} a_{j\alpha} e^{ij\lambda k_\alpha' d}$ are equivalent representations of the exponential part of the Fourier form, setting coefficients to zero as necessary. Similarly $C = -\Sigma_{|k|^* \leq K} a_{0k} k k' = -\Sigma_{\alpha=1}^{A} a_{\alpha 0} k_\alpha k_\alpha'$ for suitably chosen coefficients. Now apply Theorem 1 to $g_0(d)$ to obtain the result that

$$ {}_\circ g_K(d|\theta) = {}_\circ b'd + \tfrac{1}{2}d'{}_\circ C\, d + \Sigma_{|{}_\circ k|^* \le {}_\circ K}\, a_{{}_\circ k}\, e^{i\lambda_\circ k'd} $$

satisfies $\|g - g_K\|_{\ell,p,\nu} = o(K^{-m+\ell+\epsilon})$ where $\nu(d)$ is the distribution of $d = d(x)$, that is, $\int_D g(d)\, d\nu(d) = \int_{\mathcal{X}} g[d(x)]\, d\mu(x)$ for integrable g. In this expression ${}_\circ k$ is a multi-index of dimension N and $\lambda D = \{\lambda d(x) : x \in \mathcal{X}\}$ is a rectangle in R^N with edges shorter than 2Π. Then using $\int_D g(d)\, d\nu(d) = \int_{\mathcal{X}} g[d(x)]\, d\mu(x)$ for integrable g we have that

$$ h_K(x|\theta) = \Sigma_{i=1}^{N} w_i x_i + {}_\circ g_K[d(x)|\theta] $$

satisfies $\|g - h_K\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$. What has to be shown is that it is possible to choose the coefficients of the Fourier flexible form subject to $1'c = 1$ and $1'r = 0$ so that $g_K(x|\theta) = h_K(x|\theta)$.

Considering the linear term, we have

$$ \Sigma_{i=1}^{N} w_i x_i + {}_\circ b'd(x) $$
$$ = \Sigma_{i=1}^{N-1} (w_i + {}_\circ b_i)x_i + (w_N - \Sigma_{i=1}^{N-1}{}_\circ b_i)x_N + {}_\circ b_N x_{N+1} $$
$$ = \Sigma_{i=1}^{N} c_i x_i + b_{N+1}\, x_{N+1} $$

making the obvious associations. Then $\Sigma_{i=1}^{N} c_i = \Sigma_{i=1}^{N} w_i + \Sigma_{i=1}^{N-1}{}_\circ b_i - \Sigma_{i=1}^{N-1}{}_\circ b_i = 1$ as required. Now consider the term $\lambda_\circ k'd(x)$. If $\lambda\mathcal{X}$ is a rectangle with edges shorter than 2Π then so is λD. Now

$$ {}_\circ k'd(x) = \Sigma_{i=1}^{N-1}{}_\circ k_i x_i - (\Sigma_{i=1}^{N-1}{}_\circ k_i)x_N + {}_\circ k_N x_{N+1} $$
$$ = \Sigma_{i=1}^{N} r_i x_i + k_{N+1}\, x_{N+1} $$
$$ = k'x $$

where $k = (r, k_{N+1}) \in R^{N+1}$ and $1'r = 0$. Now $\Sigma_{i=1}^{N-1}{}_\circ k_i$ is an integer so k is a multi-index. For some K we will have $|k|^* \le K$ provided $|{}_\circ k|^* \le {}_\circ K$ as required. Considering the quadratic term we have

$$d'(x) \circ Cd(x) = -\Sigma_{|\circ k|^* \leq \circ K} \circ a_k [\circ k' d(x)]^2$$

$$= -\Sigma_{|k|^* \leq K} a_k (k'x)^2$$

for suitably chosen a_k as required. ▯

For notational convenience, we have made some overly restrictive assumptions. These are that v is to be combined with $\ell_1, \ell_2, \ldots, \ell_N$ in computing the common scaling factor λ and that output is univariate. Actually, the only variables which require a common scaling factor are those affected by R_0, namely $\ell_1, \ell_2, \ldots, \ell_N$. Output can be measured according to any scale of measurement without affecting the analysis so that v can be defined as

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

for any positive λ_{N+1}; $\ell n\ a_{N+1}$ is to be chosen as described earlier. We suggest that λ_{N+1} be chosen so that

$$v^H \doteq \max\ \{\ell_1^H, \ \ell_2^H, \ldots, \ell_N^H\}.$$

The conversion from univariate v to a multivariate measure of log output is straightforward increase in dimension from $N+1$ to, say, $N+L$. The multivariate vector v can include variables other than log outputs such as log time. Of course, these non-output variables would not be subject to the monotonicity restriction, R_4.

3.1. Returns to Scale

As noted previously, the Fourier flexible form

$$g_K(\ell,v|\theta) = a_o + b'x + \tfrac{1}{2}x'Cx + \Sigma_{\alpha=1}^{A}\Sigma_{j=-J}^{J} a_{j\alpha}e^{ij\lambda k_\alpha' x}$$

will satisfy R_o , positive linear homogeneity, if

$$R_o' \begin{cases} 1'c = 1 \quad \text{where } b = (c', b_{N+1})' \\ a_{j\alpha} = 0 \text{ if } 1'r_\alpha \neq 0 \quad \text{where } k_\alpha = (r_\alpha', k_{N+1,\alpha})' . \end{cases}$$

If, in addition to R_o' ,

$$R_1' \begin{cases} b_{N+1} = \left(\dfrac{1}{\lambda_{N+1}}\right) \quad \text{where } b = (c', b_{N+1})' \\ a_{j\alpha} = 0 \quad \text{if } k_{N+1,\alpha} \neq 0 \quad \text{where } k_\alpha = (r_\alpha', k_{N+1,\alpha})' \end{cases}$$

is imposed then the Fourier flexible form will satisfy R_1 , constant returns to scale. Let the (logarithmic) cost function $g(x)$ satisfy the hypotheses of Theorem 2, linear homogeneity

$$R_0 . \quad g(\ell + \tau 1, v) = \tau + g(\ell, v) ,$$

and correspond to a constant returns to scale production technology

$$R_1 . \quad g(\ell, v) = \ell n\, u + g(\ell) .$$

Then the Fourier flexible form subject to R_0' and R_1' can approximate $g(x)$ to within

$$\|g - g_K(\hat{a}_K)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$$

for every $\epsilon > 0$ as $K \to \infty$ in the notation of Theorem 2. The argument is as follows.

A cost function which exhibits constant returns to scale is written as

$$g(\ell,v) = \ell n\ u + g(\ell)\ .$$

By Lemma 1,

$$g(\ell,v) = \ell n\ u + \Sigma_{i=1}^{N} w_i \ell_i + g_o(\ell_1 - \ell_N, \ell_2 - \ell_N, \ldots, \ell_{N-1} - \ell_N)$$

where $\Sigma_{i=1}^{N} w_i \ell_i = 1$. Then, exactly as in the proof of Theorem 2, Theorem 1 is applied to $g_o(\ell_1 - \ell_N, \ell_2 - \ell_N, \ldots, \ell_{N-1} - \ell_N)$ to obtain the result that

$$\|g - g_K(\bar\theta_K)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})\ .$$

Tests of constant returns to scale that are asymptotically free of specification bias can be constructed using this result. That is, one can construct a test statistic T_K based on, say, share data and find a critical point c_K such that if the true function $g(x)$ satisfies R_o and R_1 then

$$\ell im_{K\to\infty} P(T_K > c_K) = \alpha\ .$$

A significant test statistic can thus be attributed to violation of constant returns to scale rather than specification bias. The specification bias in tests of this sort using a fixed parameter functional form such as a Translog can be substantial [Gallant (1981), Section 7]. We shall give a construction in the case where Σ is known. With a little extra work this analysis covers the case where Σ can be estimated consistently from, say, replicated observations. We anticipate that these results will extend to Σ estimated from regression residuals but, since the number of parameters is increasing, the technical details will not be trivial.

Assume that the observed input cost shares follow

$$s_t = f(x_t) + u_t \qquad t = 1,2,\ldots, n$$

where $f(x) = (\partial/\partial\ell)g(\ell,v)$ and the u_t are independently and normally distributed with mean zero and variance-covariance matrix Σ having rank $(\Sigma) = N-1$. The approximation to $f(x)$ is $(\partial/\partial\ell)g_K(\ell,v|\theta)$ subject to R_o which is a linear function of θ and can be written as

$$(\partial/\partial\ell)g_K(\ell,v|\theta) = z'\theta \ .$$

The restrictions R_1 are zero restrictions on elements of θ and can be represented as

$$(\partial/\partial\ell)g_K(\ell|\rho) = z'_{(1)}\rho \ .$$

Let

$$a(K) \approx b(K)$$

mean that there exist two positive constants such that

$$c_1 b(K) \le a(K) \le c_2 b(K)$$

for large enough K. Then from Edmunds and Moscatelli (1977) we have that

$$\text{length }(\theta) = p_\theta \approx K^N \ ,$$

$$\text{length }(\rho) = p_\rho \approx K^{N-1} \ .$$

A vector notation is convenient. Let

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad f = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix},$$

each vectors of length nN, and let

$$Z = \begin{pmatrix} z_1' \\ z_2' \\ \vdots \\ z_n' \end{pmatrix} \quad , \quad Z_{(1)} = \begin{pmatrix} z_{1(1)}' \\ z_{2(1)}' \\ \vdots \\ z_{n(1)}' \end{pmatrix} \quad ,$$

matrices of order $nN \times p_\theta$ and $nN \times p_\rho$ respectively. In this notation

$$s = f + u$$

where $u \sim N(0, I \otimes \Sigma)$. Let Σ^+ denote the Moore-Penrose g-inverse of Σ [Rao (1973), Sec. 16.5], and let

$$P = (I \otimes \Sigma^+) Z[Z'(I \otimes \Sigma^+)Z]^{-1} Z'(I \otimes \Sigma^+) \quad ,$$

$$P_{(1)} = (I \otimes \Sigma^+) Z_{(1)}[Z_{(1)}'(I \otimes \Sigma^+)Z_{(1)}]^{-1} Z_{(1)}'(I \otimes \Sigma^+) \quad .$$

Consider the statistic

$$T_K = s'(P - P_{(1)}) s \quad .$$

We assume that there are more observations than parameters $n_K N > p_\theta \approx K^N$ and that the number of observations grow at some polynomial rate with K

$$n_K \approx K^{N+\delta} \quad \text{with } \delta \geq 0 \ .$$

Then T_K follows the non-central chi-square distribution with

$$v_K = \text{rank}\ (P - P_{(1)}) \approx K^N$$

degrees of freedom and non-centrality parameter

$$\tau_K = f'(P - P_{(1)}) f \quad .$$

Let the two-tailed α-level critical point of the standard normal distribution be denoted by $z_{\frac{\alpha}{2}}$. We shall show that if $c_K = \sqrt{2v_K}\, z_{\frac{\alpha}{2}}$ and that if the true cost function $g(x)$ is differentiable to an order greater than $1+(s + N/2)/2$ then

$$\lim_{K\to\infty} P(T_K > c_K) = \alpha \ .$$

Now T_K may be written as [Rao (1973), Sec. 3b.2]

$$T_K = (z_o - \sqrt{\tau_K})^2 + \Sigma_{i=1}^{v_K - 1} z_i^2$$

where the z_i are independent standard normal variates. Suppose that we can show that $(z_o - \sqrt{\tau_K})^2/\sqrt{v_K}$ converges in probability to zero. Then by the central limit theorem $T_K/\sqrt{2v_K}$ converges in distribution to the unit normal and the result will follow. As z_o is bounded in probability, it suffices to show that $\lim_{K\to\infty} \tau_K/\sqrt{v_K} = 0$.

Now let $\bar{\rho}$ be as given by Theorem 1 and let δ be the largest eigen value of Σ^+ , then

$$\begin{aligned}
\tau_K &= f'(P - P_{(1)})f \\
&= (f - Z_{(1)}\bar{\rho})'(P - P_{(1)})(f - Z_{(1)}\bar{\rho}) \\
&\leq (f - Z_{(1)}\bar{\rho})'(I \otimes \Sigma^+)(f - Z_{(1)}\bar{\rho}) \\
&= \Sigma_{t=1}^{n_K}[(\partial/\partial\ell)g(x_t) - (\partial/\partial\ell)g_K(\ell_t|\bar{\rho})]'\Sigma^+[(\partial/\partial\ell)g(x_t) - (\partial/\partial\ell)g(\ell_t|\bar{\rho})] \\
&\leq \delta \ \Sigma_{t=1}^{n_K} \Sigma_{i=1}^{N}[(\partial/\partial\ell_i)g(x_t) - (\partial/\partial\ell_i)g_K(\ell_t|\bar{\rho})]^2 \\
&\leq \delta \ n_K N \, \| g - g_K(\bar{\rho})\|^2_{1,\infty,\mu} \\
&= n_K o(K^{-m+1+\epsilon})^2
\end{aligned}$$

and

$$\tau_K/\sqrt{v_K} \approx \tau_K\, K^{-N/2} = n_K o(K^{-m+1-N/2+\epsilon}) \approx K^{N+\delta}\; o(K^{-2m+2-N/2+\epsilon})$$

whence

$$\tau_K/\sqrt{v_K} = o(K^{-2m+2+\delta+N/2+\epsilon}).$$

The homothetic technology restriction

R_2. $g(\ell,v) = h(v) + g(\ell)$

is a slight generalization of the constant returns to scale restriction; one writes $h(v)$ for $\ell n\, u$. The restrictions for constant returns to scale were derived by applying Lemma 1 and Theorem 2 to $g(\ell)$ and, since $h(v) = \ell n\, u$ was assumed, there was no need to approximate $h(v)$. To obtain the generalization to a homothetic production technology, Theorem 2 is applied to $h(v)$. The result may be stated as follows.

The Fourier flexible form

$$g(\ell,v|\theta) = a_0 + b'x + \tfrac{1}{2}x'Cx + \Sigma_{\alpha=1}^{A}\, \Sigma_{j=-J}^{J}\, a_{j\alpha}\, e^{ij\lambda k_\alpha' x}$$

will satisfy R_0 , positive linear homogeneity, and R_2 , homothetic production technology, if in addition to R_0' , the restriction

R_2'. $a_{j\alpha} = 0$ if $k_{N+1,\alpha} \neq 0$ and $r_\alpha \neq 0$ where $k_\alpha = (r_\alpha',\; k_{N+1,\alpha})'$

is imposed. Let the (logarithmic) cost function $g(x)$ satisfy the hypotheses of Theorem 2, positive linear homogeneity R_0 , and correspond to a homothetic production technology R_2 . Then the Fourier flexible form subject to R_0' and R_2' can approximate $g(x)$ to within

$$\|g - g_K(\theta)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$$

for every $\epsilon > 0$ as $K \to \infty$.

To comment further on this result, the error bound $o(K^{-m+\ell+\epsilon})$ was identified in the previous paragraphs as being the critical feature of the Fourier flexible form that permits asymptotically unbiased inference. If inference is based on share data then the non-centrality parameter of the test statistic based on the Fourier flexible form of order K decreases to zero at the rate

$$\tau_K/\sqrt{v_K} = o(K^{-2m+2+\delta+N/2+\epsilon})$$

for every $\epsilon > 0$ when observations are obtained at the rate $n = K^{N+\delta}$, recall that N is the number of inputs. For asymptotically unbiased inference to be possible at all, the cost function g(x) must be differentiable to an order greater than $1 + (\delta+N/2)/2$. Thus, keeping δ small permits unbiased inference over a wider class of g(x) . Consequently, contrary to the usual wisdom, one should attempt to keep the number of parameters large relative to the number of observations if one's intention is to defend against biased inference. Also, note that if g(x) is differentiable to a very high order of m then $\tau_K/\sqrt{v_K}$ will be small for low orders of K .

3.2. A Test of Separability

Let the vector s of cost shares be partitioned into two groups

$$s = (s'_{(1)}, s'_{(2)})'$$

where $s_{(1)}$ is an N_1-vector and $s_{(2)}$ is an N_2-vector with $N = N_1 + N_2$. Let the vector ℓ of log input prices be partitioned similarly

$$\ell = (\ell'_{(1)}, \ell'_{(2)})' .$$

As we show next, one practical aspect of separability is that cost shares in the first sector, reexpressed as cost shares relative to expenditures in the first sector, can be modeled independently of $\ell_{(2)}$. This fact forms the basis for a test of separability.

A homothetically weakly separable cost function has the form

$$g(\ell,v) = \bar{g}[g_1(\ell_{(1)}) , \ell_{(2)} , v]$$

where $\bar{g}(g_1, \ell_{(2)}, v)$ and $g_1(\ell_{(1)})$ satisfy R_0. Thus

$$(\partial/\partial\ell_{(1)})g(\ell,v) = (\partial/\partial g_1)\bar{g}(g_1, \ell_{(2)}, v)(\partial/\partial\ell_{(1)})g_1(\ell_{(1)})$$

and, using Shepard's lemma,

$$s_{(1)}/1's_{(1)} = (\partial/\partial\ell_{(1)})g_1(\ell_{(1)})$$

which is independent of $\ell_{(2)}$. But in general

$$s_{(1)}/1's_{(1)} = (\partial/\partial\ell_{(1)})g(\ell,v)/1'(\partial/\partial\ell_{(1)})g(\ell,v) \quad .$$

The idea is to substitute $g_K(\ell,v|\theta)$ for $g(\ell,v)$, in the general equation, fit

$$s_{(1)}/1's_{(1)} = (\partial/\partial\ell_{(1)})g_K(\ell,v|\theta)/1'(\partial/\partial\ell_{(1)})g_K(\ell,v|\theta) ,$$

and test the hypothesis that all terms in $g_K(\ell,v|\theta)$ involving $\ell_{(2)}$ and v have zero coefficients. In terms of further parametric restrictions on the form given in Theorem 2, the restrictions to be tested in the subsystem are[2/]

$$R_3' \begin{cases} (c_{(2)}', b_{N+1}) = 0 \text{ where } b = (c_{(1)}', c_{(2)}', b_{N+1})' \\ a_{j\alpha} = 0 \text{ if } (r_{(2)\alpha}', k_{N+1,\alpha}) \neq 0 \text{ where } k_\alpha = (r_{(1)\alpha}', r_{(2)\alpha}', k_{N+1,\alpha})' . \end{cases}$$

The parameters $c_{(2)}$, b_{N+1} and those $a_{j\alpha}$ for which $r_{(1)\alpha} = 0$ are not identified in the subsystem regression

$$s_{(1)}/1's_{(1)} = (\partial/\partial\ell_{(1)})g_K(\ell,v|\theta)/1'(\partial/\partial\ell_{(1)})\, g_K(\ell,v|\theta) .$$

These parameters are set to zero in the estimation [Gallant (1981), Section 5]. Thus, the only testable restrictions are

$$R_3'' .\quad a_{j\alpha} = 0 \text{ if } (r_{(2)\alpha}', k_{N+1,\alpha}) \neq 0 \text{ and } r_{(1)\alpha} \neq 0 \text{ where } k_\alpha = (r_{(1)\alpha}', r_{(2)\alpha}', k_{N+1,\alpha})'$$

If the sub-cost function $g_1(\ell_{(1)})$ satisfies the hypotheses of Theorem 2 then the Fourier flexible form subject to R_0' and R_3' can approximate $g_1(\ell_{(1)})$ to within

$$\|g_1(\ell_{(1)}) - g_K(\ell|\bar{\theta}_K)\|_{\ell,p,\mu} = o(K^{-m+\ell+\epsilon})$$

for every $\epsilon > 0$ as $K \to \infty$ in the notation of Theorem 2. This result is a direct consequence of Theorem 2 applied in a lower dimension.

Another approach to testing separability is to begin with the condition that

$$(\partial/\partial\ell_{(2)})\frac{(\partial/\partial\ell_i)g_K(\ell,v|\theta)}{(\partial/\partial\ell_j)g_K(\ell,v|\theta)} = 0 \quad ,$$

2/ R_1' and R_3' taken together imply $1'c_{(1)} = 1$.

deduce the corresponding parametric restrictions on θ , fit the full system, and then test these restrictions [Blackorby, Primont, and Russell(1978), Sec. 8.2.2]. This approach leads to messy algebra with the Fourier flexible form and so we abandoned it in favor of the simpler approach of fitting only to $s_{(1)}/1's_{(1)}$. There may be an efficiency penalty but simplicity and tidiness of the theoretical defense of the test are the dominant considerations in our opinion. If we had an interest in estimating the full system subject to a separability constraint, which we do not, it would probably be more sensible to attack the problem directly with an approximating cost function of the form $g_K[g_K(\ell_{(1)}|\rho), \ell_{(2)}, v|\theta]$.

3.3. Curvature Restrictions

The imposition of monotonicity

R_4'. $(\partial/\partial x_i) g_K(x|\theta) > 0$ all $x \in \mathcal{X}$, $i = 1,2,\ldots,N+1$

or concavity

R_5'. $\nabla^2 g_K(\theta) + \nabla g_K(\theta)\, \nabla' g_K(\theta) - \mathrm{diag}(\nabla g_K(\theta))$ negative semi-definite for all $x \in \mathcal{X}$

or both will not affect the ability of the Fourier flexible form to approximate a function which satisfies R_4 or R_5 or both respectively on $\bar{\mathcal{X}}$. The argument depends on continuity and consists of a verification that eventually the constraint is not binding.

Consider a function $g(x)$ which satisfies R_4 on $\bar{\mathcal{X}}$ and the hypotheses of Theorem 2 for $m \geq 2$. Then, setting $\ell = 1$, the triangular array $\{\bar{\theta}_K\}$ of Theorem 2 satisfies

$$\sup_{\mathcal{X}} |(\partial/\partial x_i) g(x) - (\partial/\partial x_i) g_K(x|\bar{\theta}_K)| = o(K^{-1+\epsilon}) \quad .$$

As $(\partial/\partial x) g(x)$ is continuous on the compact set $\bar{\mathcal{X}}$, $(\partial/\partial x_i) g(x) > \delta > 0$ for some δ on $\mathcal{X}$. Thus for all K large enough $(\partial/\partial x_i) g_K(x|\bar{\theta}_K) > \delta/2 > 0$. The argument for the restriction R_5' is the same. Let $m \geq 3$ and note that the determinants of the principal minors of $\nabla^2 g + \nabla' g \nabla g - \mathrm{diag}(\nabla g)$ are continuous functions of ∇g and $\nabla^2 g$.

The restriction R_4 is usually irrelevent in applications. The coefficients of the Fourier form are estimated, at least in part, from the regression

$$s_t = \nabla g_K(x_t|\theta) + e_t .$$

Since the components of the vector of shares s_t are between zero and one a reasonably close fit to data will require that the predicted shares $\nabla g_K(x_t|\theta)$ be between zero and one as well. In the applications considered to date, this has been the case. Monotonicity follows as a matter of course.

Concavity is not of much interest in hypothesis testing, the subject of this paper, so we have not devoted a great deal of energy to finding necessary and sufficient conditions for the concavity of the Fourier form which are simple enough to be used in applications. Tractable sufficient conditions can be obtained by noting that

$$\nabla^2 g_K + \nabla g_K \nabla' g_K - \text{diag}(\nabla g_K)$$

is negative semi-definite provided that the components of ∇g_K are between zero and one and that $\nabla^2 g_K$ is negative semi-definite. $\nabla^2 g_K$ can be made negative semi-definite by using the restrictions of Section 5 of Gallant (1981) and then changing sign. However, the data used for illustration, described next, strongly reject these restrictions, so we have been unable to follow this approach here.

4.1. An Illustration: Determining Returns to Scale

To illustrate the statistical methods, we reexamine some issues that are discussed in Berndt and Wood (1975, 1979) and Berndt and Khaled (1979) using their annual data on the U.S. manufacturing sector from 1947 to 1971. Total input cost (C), input prices of capital (K), labor (L), energy (E), and materials (M), and the corresponding cost shares are taken from Tables 1 and 2 of Berndt and Wood (1975). The output series (Y) is taken from Table 1 of Berndt and Khaled (1979). These data are transformed as shown in Table 1.

We shall ignore technical change in the analysis which is equivalent to imposing the maintained hypothesis that there has been none. The reasons are as follows. Technological progress is primarily a stochastic phenomenon and is probably best modeled by viewing the parameters of $g_K(x_t|\theta)$ as the realization of a stochastic process indexed by time. Thus observed data would follow $s_t = \nabla g_K(x_t|\theta_t)$ for $t = 1,2,\ldots,n$. An adequate formulation of this approach is beyond the scope of this paper. If technological change is modeled as depending deterministically on time then our approach can be applied directly by including time as a variable treated similarly to output as noted earlier. However, this would lead to a model with more parameters than these data can support and is thus not practical here. A model such as that used by Berndt and Khaled (1979), $g_K(x|\theta) + t(\tau + \Sigma\tau_i \ell_i)$, is feasible but will lead to the sort of biases that our approach seeks to avoid. Moreover, Berndt and Khaled found only marginal significance for $\tau, \tau_1, \ldots, \tau_N$ with these data and with the generalized Box-Cox cost function replacing $g_K(x|\theta)$. Thus, even if we followed their approach, the results would probably be uninteresting.

Table 1. Data and Scaling Factors

Endogeneous Variables	Exogeneous Variables	Scaling Factors
$s_0 = \ell n(C)$	$x_1 = \ell_1 = \ell n(P_K) - \ell n(.74371) + \epsilon$	$\epsilon = 10^{-5}$
s_1 = K cost share	$x_2 = \ell_2 = \ell n(P_L) + \epsilon$	$\lambda_5 = \frac{\ell n(2.76025) + \epsilon}{\ell n(466.82965/182.82936) + \epsilon}$
s_2 = L cost share	$x_3 = \ell_3 = \ell n(P_E) + \epsilon$	$\lambda = \frac{6}{\ell n(2.76025) + \epsilon}$
s_3 = E cost share	$x_4 = \ell_4 = \ell n(P_M) + \epsilon$	
s_4 = M cost share	$x_5 = v = \lambda_5 [\ell n(Y) - \ell n(182.82936) + \epsilon]$	

With technical change ignored, the statistical model is

$$s_{0t} = g_K(x_t|\theta) + e_{0t}$$

$$s_{1t} = (\partial/\partial\ell_1)\, g_K(x_t|\theta) + e_{1t}$$

$$s_{2t} = (\partial/\partial\ell_2)\, g_K(x_t|\theta) + e_{2t}$$

$$s_{3t} = (\partial/\partial\ell_3)\, g_K(x_t|\theta) + e_{3t}$$

where the share equation for s_{4t} is discarded due to the restriction that $\Sigma_{i=1}^{4} s_{it} = 1$ [Theil (1971), Section 7.7]. Recall that

$$g_K(x|\theta) = u_0 + b'x + \tfrac{1}{2}x'C\,x$$

$$+ \Sigma_{\alpha=1}^{A}\{u_{0\alpha} + 2\Sigma_{j=1}^{J}[u_{j\alpha}\cos(j\lambda k_\alpha' x) - v_{j\alpha}\sin(j\lambda k_\alpha' x)]\}$$

$$(\partial/\partial\ell)\, g_K(x|\theta) =$$

$$c - \lambda\Sigma_{\alpha=1}^{A}\{u_{0\alpha}\lambda k_\alpha' x + 2\Sigma_{j=1}^{J} j[u_{j\alpha}\sin(j\lambda k_\alpha' x) + v_{j\alpha}\cos(j\lambda k_\alpha' x)]\}r_\alpha$$

where

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

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

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

and

$$C = -\Sigma_{\alpha=1}^{A} \quad u_{0\alpha}\,\lambda^2 k_\alpha k_\alpha' \quad .$$

The restriction R_0' is imposed as a maintained hypothesis

$$R_0' \quad \begin{cases} \Sigma_{i=1}^{4} b_i = 1 \ , & \text{and} \\ u_{j\alpha} = v_{j\alpha} = 0 \ \text{ if } \ \Sigma_{i=1}^{4} k_{i\alpha} \neq 0 \ . \end{cases}$$

The set of multi-indexes that satisfy $\Sigma_{i=1}^{4} k_{i\alpha} = 0$ and have norm $|k_\alpha|^* \leq 3$ are displayed in Table 2. For this set $A = 19$, and we take $J = 1$, whence θ is a vector of nominal length 63. The effective number of parameters is 53 due to the following restrictions.

The nonhomogeneous restriction $\Sigma_{i=1}^{4} b_i = 1$ reduces the number of effective parameters by one. The remaining restrictions are due to overparameterization of the matrix C. To see this, think of the model as the sum of a Translog

$$u_0 + b'x + \frac{1}{2} x' C x$$

plus the sum of $A = 19$ univariate Fourier expansions along the directions k_α

$$\Sigma_{\alpha=1}^{A} \left[u_{0\alpha} + 2\Sigma_{j=1}^{J} u_{j\alpha} \cos(j \lambda k_\alpha' x) - v_{j\alpha} \sin(j \lambda k_\alpha' x) \right] .$$

The matrix C of the Translog portion is a 5 x 5 symmetric matrix which satisfies five homogeneous restrictions $\Sigma_{j=1}^{4} c_{ij} = 0$ $(i = 1,2,3,4,5)$. Thus C can have at most ten free parameters and in the parameterization $C = -\Sigma_{\alpha=1}^{19} u_{0\alpha} \lambda^2 k_\alpha k_\alpha'$, ten of the $u_{0\alpha}$ are free parameters and nine must be set to zero. In sum, θ is subject to one nonhomogeneous restriction and nine homogeneous restrictions which reduces the number of parameters from 63 to 53.

The model may be written in a vector notation

$$y_t = f(x_t|\theta) + e_t \qquad t = 1,2,\ldots,25$$

with $y_t = (s_{0t}, s_{1t}, s_{2t}, s_{3t})'$ and similarly for f and e_t where we assume that the errors are independently distributed each with mean zero and variance-covariance matrix Σ. As $f(x_t|\theta)$ is linear in the parameters,

$$f(x_t|\theta) = z_t' \theta$$

where z_t' is of order $N \times 1 + N + 1 + A(1+2J)$, this is a multivariate linear

Table 2. The Sequence $\{k_\alpha\}_{\alpha=1}^{19}$

	Main Effects							Interactions											
		Input ratios						Ratios x low level of u						Ratios x high level of u					
	u	K/L	K/E	L/E	K/M	L/M	E/M	K/L	K/E	L/E	K/M	L/M	E/M	K/L	K/E	L/E	K/M	L/M	E/M
K	0	1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	0
L	0	-1	0	1	0	1	0	-1	0	1	0	1	0	-1	0	1	0	1	0
E	0	0	-1	-1	0	0	1	0	-1	-1	0	0	1	0	-1	-1	0	0	1
M	0	0	0	0	-1	-1	-1	0	0	0	-1	-1	-1	0	0	0	-1	-1	-1
u	1	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$\|k_\alpha\|^*$	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3

regression model.[3/] We use the seemingly Unrelated Nonlinear Regression estimation method (Zellner, 1962) but, as a priori identification of those $u_{0\alpha}$ to be set to zero in order to impose the nine homogeneous restrictions would be tedious, we identify them automatically in the course of the computations as follows.

Let

$$r = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \hline \underset{\sim}{0} \end{bmatrix} \qquad R = \left[\begin{array}{c|c} \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 \end{matrix} & \underset{\sim}{0} \\ \hline \underset{\sim}{0} & \underset{\sim}{I}_{58} \end{array}\right]$$

and let A^- denote that g_2-inverse of a positive semi-definite matrix A described in Goodnight (1979). Then

$$\bar{\theta} = r + R\left(\Sigma_{t=1}^{n} R' z_t \Sigma^{-1} z_t' R\right)^- \Sigma_{t=1}^{n} R' z_t \Sigma^{-1}(y_t - z_t' r)$$

minimizes

$$s(\theta,\Sigma) = \frac{1}{n} \Sigma_{t=1}^{n} (y_t - z_t' \theta)' \Sigma^{-1}(y_t - z_t' \theta)$$

subject to the nonhomogeneous constraint $\Sigma_{i=1}^{4} b_i = 1$ and the nine homogeneous constraints. The procedure generalizes; in the general case, r is the $N+1$ st elementary vector of order $1+N+1+A(1+2J)$, the north-west corner of R is the identity of order N bordered below by the N-vector $(0,-1,-1,\ldots,-1)$, and the south-east corner of R is the identity of order $1+A(1+2J)$.

[3/] Machine readable FORTRAN code to compute z_t directly for given λ, A, J and x_t is available from the author at the cost of reproduction and postage.

The Seemingly Unrelated Regressions estimator $\hat{\theta}$ is computed as follows. First compute

$$\bar{\theta} \text{ to minimize } s(\theta, I) \text{ subject to } R_0'$$

using the formula of the previous paragraph with I replacing Σ^{-1}. Next estimate Σ by

$$\hat{\Sigma} = \frac{1}{n} \Sigma_{t=1}^{n} (y_t - z_t' \bar{\theta})(y_t - z_t' \bar{\theta})' .$$

Finally, compute

$$\hat{\theta} \text{ to minimize } s(\theta, \hat{\Sigma}) \text{ subject to } R_0'$$

using the formula of the previous paragraph with $\hat{\Sigma}^{-1}$ replacing Σ^{-1}. For these data we obtain

$$n\, s(\hat{\theta}, \hat{\Sigma}) = 63.6110, \quad 53 \text{ parameters.}$$

The hypothesis of a homothetic production technology

$$R_2' \quad u_{j\alpha} = v_{j\alpha} = 0 \text{ if } k_{N+1,\alpha} \neq 0 \text{ and } r_\alpha \neq 0 \text{ where } k_\alpha = (r_\alpha', k_{N+1,\alpha})$$

is equivalent to the deletion of

$$k_8, k_9, \ldots, k_{19}$$

from the list of multi-indexes given in Table 2. The hypothesis R_2' can be viewed as analogous to the hypothesis of no interaction in an experimental design. To see this write

$$g_K(x|\theta) = u_0 + b'x - \frac{1}{2} \Sigma_{\alpha=1}^{A} u_{0\alpha} \lambda^2 (k_\alpha' x)^2$$

$$+ \Sigma_{\alpha=1}^{A} \left[u_{0\alpha} + 2\Sigma_{j=1}^{J} u_{j\alpha} \cos(j \lambda k_\alpha' x) - v_{j\alpha} \sin(j \lambda k_\alpha' x) \right] .$$

One can view $k_\alpha' x$ as a measure of the covariance of x with the vector k_α and the Fourier flexible form as the sum of additive effects in $(k_\alpha' x)$.

Under homotheticity, it suffices to consider main effects only, those due to $k_1, k_2, \ldots, k_7$. Under nonhomotheticity there is an interaction with output and the interaction effects, those due to $k_8, k_9, \ldots, k_{19}$, must be taken into account. We find an analysis of variance interpretation of our procedures helpful as our approach to bias minimization using Fourier series expansions as large as the data will support comes closer to the ideas of experimental design than it does to parsimonious parametric modeling. The estimate $\tilde{\theta}$ subject to R_0' and R_2' is computed as follows.

To compute

$$\tilde{\theta} \text{ to minimize } s\,(\theta,\hat{\Sigma}) \text{ subject to } R_0', R_2'$$

one uses the previous scheme for the minimization of $s\,(\theta,\Sigma)$ subject to R_0' but using the partial set of multi-indexes $k_1, \ldots, k_7$ instead of full set $k_1, \ldots, k_{19}$; r is a vector of length 27 and R is 27 x 26. As $C = -\frac{1}{2}\,\Sigma_{\alpha=1}^{A}\, u_{0\alpha}\lambda^2 k_\alpha k_\alpha'$ will admit of 11 free parameters and there are only 7 available, there are no homogeneous restrictions. We obtain

$$n\,s(\tilde{\theta},\hat{\Sigma}) = 138.3487, \text{ 26 parameters.}$$

The difference

$$n\,s(\tilde{\theta},\hat{\Sigma}) - n\,s(\hat{\theta},\hat{\Sigma}) = 74.7377, \text{ 27 d.f.}$$

is asymptotically distributed as chi-square random variable provided that the same estimate of Σ is used to compute both $\tilde{\theta}$ and $\hat{\theta}$ (Burguete, Gallant and Souza, 1982). Presumed also is that p_θ is held fixed as sample size increases. As noted earlier, we expect that our results with p_θ depending on n for known Σ will carry over to estimated Σ but we have no proof as yet. A chi-square of 74.7377 with 27 degrees of freedom is significant at a level of $p = .0005$ thus homotheticity is rejected. This agrees with

Berndt and Khaled's (1979) finding (a chi-square of 43.1162 with 3 d.f.). Moreover, the same result obtains if the computations are repeated assuming a first-order diagonal autoregression on the errors.

Constant returns to scale implies homotheticity so it too will be rejected. Letting $\tilde{\tilde{\theta}}$ denote the estimate obtained by minimizing $s(\theta,\hat{\Sigma})$ subject to R_0' and R_1' we obtain

$$n\, s(\tilde{\tilde{\theta}},\hat{\Sigma}) = 1193.6333, \text{ 22 parameters.}$$

This is computed using the same computational scheme by replacing s_0 with $s_0 - \ln u$ and using the six multi-indexes $k_2, k_3, \ldots, k_7$ to define the Fourier flexible form. Testing constant returns to scale against a homothetic technology we have

$$n\, s(\tilde{\tilde{\theta}},\hat{\Sigma}) - n\, s(\tilde{\theta},\hat{\Sigma}) = 1055.2846, \text{ 3 d.f.}$$

Thus, we strongly reject constant returns to scale even assuming homotheticity as the maintained hypothesis. Berndt and Khaled (1979) reject this hypothesis as well but not as strongly (a chi-square of 66.017 with 2 d.f.).

4.2. An Illustration: Testing Separability

Berndt and Wood (1979) list some recent studies which have estimated the cross-price elasticity between capital and energy. These divide into two groups, those estimated from data on capital, labor and energy (KLE) alone, and those estimated from data on capital, labor, energy and materials (KLEM). If the cost function is homothetically weakly separable then cross-price elasticities obtained from KLE data and KLEM data can be related. If not, cross-price elasticities estimated from KLE data would seem to defy interpretation. See Berndt and Wood (1979) for details[4]. These considerations motivate a test that KLE is homothetically weakly separable from M.

Recall that the logic of the inference procedure runs as follows. Under the null hypothesis the regression

$$s_1/(s_1+s_2+s_3) = (\partial/\partial\ell_1)\ g_K(\ell_1,\ell_2,\ell_3|\theta) + e_1$$

$$s_2/(s_1+s_2+s_3) = (\partial/\partial\ell_2)\ g_K(\ell_1,\ell_2,\ell_3|\theta) + e_2$$

will adequately represent the data; Table 1 defines the variables and, as before, we have deleted the equation for $s_3/(s_1+s_2+s_3)$. A plausible alternative model which contains the null model as a special case is

$$s_i/(s_1+s_2+s_3) = (\partial/\partial\ell_i)\ g_K(\ell,v|\theta)/\Sigma_{i=1}^{3}(\partial/\partial\ell_i)\ g_K(\ell,v|\theta) + e_i$$

$(i = 1,2)$. The idea is to test the null model against the plausible alternative. However, the formal logic of a statistical test of a null hypothesis requires only that the alternative model contain the null model as a

---

[4] Berndt and Wood (1979) impose linear homogeneity on the implied sub-production function whereas we only impose homotheticity; see Theorem 3.8 of Blackorby, Primont, and Russell (1978).

special case. Plausibility is not required. Therefore, we can arbitrarily impose homotheticity R_2' on $g_K(\ell,v|\theta)$ without affecting the validity of the test. Power may be lost but not validity. We shall impose R_2' as there is not enough data to support the full alternative model. Moreover, this will permit a direct comparison with Berndt and Wood (1975).

To fit the model, we use the Seemingly Unrelated Nonlinear Regressions method (Gallant, 1975). Write

$$s_1/(s_1+s_2+s_3) = (\partial/\partial\ell_1)\, g_K(\ell,v|\theta)/\Sigma_{i=1}^3\, (\partial/\partial\ell_i)\, g_K(\ell,v|\theta) + e_1$$

$$s_2/(s_1+s_2+s_3) = (\partial/\partial\ell_2)\, g_K(\ell,v|\theta)/\Sigma_{i=1}^3\, (\partial/\partial\ell_i)\, g_K(\ell,v|\theta) + e_2$$

as the multivariate nonlinear model

$$y_t = f(x_t|\theta) + e_t\ .$$

and let

$$s(\theta,\Sigma) = \frac{1}{n}\,\Sigma_{t=1}^n \left[y_t - f(x_t|\theta)\right]'\,\Sigma^{-1}\left[y_t - f(x_t|\theta)\right]\ .$$

Under the alternative, $g_K(\ell,v|\theta)$ is subject to linear homogeneity R_0' and homotheticity R_2'. Additionally, multi-indexes with $r_{(1)\alpha} = 0$ are deleted. Thus, under the alternative $g_K(\ell,v|\theta)$ is computed using the multi-indexes k_2 through k_7 of Table 2. The parameters u_0 and b_4, b_5 of $g_K(x|\theta)$ do not appear in the regression equations. Further, $f_i(x|\theta)$ is the ratio of two linear models and is therefore homogeneous of degree zero in the remaining parameters. Consequently, $f(x|\theta)$ is not identified without a normalization rule. We impose $b_1+b_2+b_3 = 1$. The number of free parameters in the alternative model is thus $2+6(1+2) = 20$ for $J=1$. As before, we compute $\bar{\theta}$ to minimize $s(\theta,I)$, set

$$\hat{\Sigma} = \frac{1}{n} \Sigma_{t=1}^{n} \left[y_t - f(x_t|\bar{\theta})\right] \left[y_t - f(x_t|\bar{\theta})\right]' ,$$

and compute $\hat{\theta}$ to minimize $s(\theta,\hat{\Sigma})$. We obtain

$$n s(\hat{\theta},\hat{\Sigma}) = 40.6307, \text{ 20 parameters} \quad .$$

To fit the null model, fix $\hat{\Sigma}$ and use only the multi-indexes k_2, k_3 and k_4 to compute $g_K(x|\theta)$ the rest is as for the alternative model. We obtain

$$n s(\tilde{\theta},\hat{\Sigma}) = 54.3196, \text{ 11 parameters.}$$

The difference

$$n s(\tilde{\theta},\hat{\Sigma}) - n s(\hat{\theta},\hat{\Sigma}) = 13.6889, \text{ 9 d.f.}$$

is asymptotically distributed as a chi-square random variable (Burguete, Gallant and Souza, 1982). Separability is accepted at a level of $p = .10$. This conclusion differs from that of Berndt and Wood (1975) who reject separability at a level of .01.

4.3. An Illustration: Estimating Elasticities

Standard errors for estimated price elasticities and estimated elasticities of substitution are fairly easy to obtain using the Fourier flexible form because it is linear in its parameters. Recall that for given ℓog prices ℓ_0,

$$\ell_{i0} = \ell n\, p_{i0} + \ell n\, a_{io} \quad ,$$

and given ℓog output v_0,

$$v_0 = \lambda_{N+1}(\ell n\, u_0 + \ell n\, a_{N+1}) \ ,$$

the first and second order partial derivatives of the Fourier form at

$$x_0 = (\ell_0', v_0)'$$

are

$$(\partial/\partial\ell)\, g_K(x_0|\theta) =$$

$$c - \lambda\Sigma_{\alpha=1}^{A}\{u_{0\alpha}\lambda k_\alpha' x_0 + 2\Sigma_{j=1}^{J} j[u_{j\alpha}\sin(j\lambda k_\alpha' x_0) + v_{j\alpha}\cos(j\lambda k_\alpha' x_0)]\}r_\alpha$$

$$(\partial^2/\partial\ell\partial\ell')g_K(x|\theta) =$$

$$-\lambda^2\Sigma_{\alpha=1}^{A}\{u_{0\alpha} + 2\Sigma_{j=1}^{J} j^2[u_{j\alpha}\cos(j\lambda k_\alpha' x_0) - v_{j\alpha}\sin(j\lambda k_\alpha' x_0)]\}r_\alpha r_\alpha' \quad .$$

Letting

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

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

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

then a first order partial is a linear function of the form

$$\frac{\partial}{\partial x_i} g_K(x_0|\theta) = g_i' \theta$$

and a second order partial is a linear function of the form

$$\frac{\partial^2}{\partial x_i \partial x_j} g_K(x_0|\theta) = h_{ij}'\theta$$

where g_i , h_{ij} , and θ are vectors of length $1+N+1+A(1+2J)$.

Using this notation, an elasticity of substitution and its derivative with respect to θ are for $i \neq j$

$$\sigma_{ij}(\theta) = 1 + (g_i'\theta)^{-1}(g_j'\theta)^{-1}(h_{ij}'\theta)$$

$$\begin{aligned}(\partial/\partial\theta)\sigma_{ij}(\theta) = {} & (g_i'\theta)^{-1}(g_j'\theta)^{-1}h_{ij} \\ & - (g_i'\theta)^{-2}(g_j'\theta)^{-1}(h_{ij}'\theta)g_i \\ & - (g_i'\theta)^{-1}(g_j'\theta)^{-2}(h_{ij}'\theta)g_j\end{aligned}$$

and for $i = j$

$$\sigma_{ii}(\theta) = 1 + (g_i'\theta)^{-2}(h_{ii}'\theta) - (g_i'\theta)^{-1}$$

$$\begin{aligned}(\partial/\partial\theta)\ \sigma_{ii}(\theta) = {} & (g_i'\theta)^{-2}h_{ii} - 2(g_i'\theta)^{-3}(h_{ii}'\theta)g_i \\ & + (g_i'\theta)^{-2}g_i \quad .\end{aligned}$$

A cross price elasticity and its derivative are for $i \neq j$

$$\eta_{ij}(\theta) = (g_i'\theta)^{-1}(h_{ij}'\theta) + (g_j'\theta)$$

$$(\partial/\partial\theta)\eta_{ij}(\theta) = (g_i'\theta)^{-1}h_{ij} - (g_i'\theta)^{-2}(h_{ij}'\theta)g_i + g_j$$

where i indexes factor inputs and j indexes factor prices. An own price elasticity and its derivative are

$$\eta_{ii}(\theta) = (g_i'\theta)^{-1}(h_{ii}'\theta) + (g_i'\theta) - 1$$

$$(\partial/\partial\theta)\eta_{ii}(\theta) = (g_i'\theta)^{-1}h_{ii} - (g_i'\theta)^{-2}(h_{ii}'\theta)g_i + g_i \quad .$$

Let $\hat{\theta}$ denote the Seemingly Unrelated Regression computed as described in Section 4.1. Its estimated variance-covariance matrix is

$$\hat{\Omega} = R(\Sigma_{t=1}^{n} R' z_t \hat{\Sigma}^{-1} z_t' R)^{-} R'$$

in the notation of Section 4.1 . Then an estimate of, say, a cross price elasticity is obtained by evaluating $\eta_{ij}(\theta)$ at $\hat{\theta}$,

$$\hat{\eta}_{ij} = \eta_{ij}(\hat{\theta}) \ ,$$

and its standard error is computed as

$$SE(\hat{\eta}_{ij}) = [(\partial/\partial\theta') \ \eta_{ij}(\hat{\theta}) \ \hat{\Omega} \ (\partial/\partial\theta) \ \eta_{ij}(\hat{\theta})]^{\frac{1}{2}} \ .$$

Similarly for own price elasticities and elasticities of substitution.[5/]

Following this procedure, we obtain estimated elasticities and standard errors as shown in Table 3 for two cases: non-homothetic production technology (multi-indexes k_1, k_2, ..., k_{19}) and a homothetic production technology (multi-indexes k_1, k_2, ..., k_7) . In the latter case, Ω was estimated from residuals from the value of θ that minimizes $s(\theta, I)$ subject to R_0' and R_2' . These results may be compared with Table 5 of Berndt and Khaled (1979).

To comment, with homotheticity imposed, Fourier form estimates are roughly comparable to the Box-Cox form estimates obtained by Berndt and Khaled (1979). With homotheticity relaxed, they are not. As discussed in Berndt and Wood (1979), the question as to whether capital and energy are substitutes or complements is a matter of some interest. We find that,

---

[5/] Machine readable FORTRAN code to compute these estimates and standard errors directly for given x_0, $\hat{\theta}$, and $\hat{\Omega}$ is available from the author at the cost of reproduction and postage.

Table 3. Fourier Flexible Form Estimates of Allen Partial Elasticities of Substitution and Price Elasticities, U. S. Manufacturing, 1959.

	Non-homothetic		Homothetic	
Elasticity	Estimate	Std. Err.	Estimate	Std. Err.
σ_{KK}	- 6.5321	9.9222	-27.4646	6.9830
σ_{KL}	.3288	1.8962	1.2554	.3232
σ_{KE}	.6613	8.5515	- 4.2067	3.4998
σ_{KM}	.4545	1.4758	2.3081	.7301
σ_{LL}	- .2813	.7291	- 1.7437	.1391
σ_{LE}	4.5678	3.5918	- .0787	.4763
σ_{LM}	- .2422	.3611	.6626	.0648
σ_{EE}	-28.5133	36.2705	- 7.5178	10.1295
σ_{EM}	- .0157	3.5003	.9858	.9338
σ_{MM}	.0642	.3583	- .5817	.1135
η_{KK}	- .4013	.6022	- 1.5839	.3892
η_{KL}	.0908	.5238	.3456	.0897
η_{KE}	.0300	.3879	- .1945	.1605
η_{KM}	.2806	.9118	1.4328	.4545
η_{LK}	.0202	.1165	.0724	.0184
η_{LL}	- .0776	.2015	- .4801	.0385
η_{LE}	.2069	.1622	- .0036	.0220
η_{LM}	- .1495	.2230	.4113	.0403
η_{EK}	.0406	.5255	- .2426	.2003
η_{EL}	1.2608	.9911	- .0217	.1311
η_{EE}	- 1.2917	1.6353	- .3477	.4658
η_{EM}	- .0097	2.1606	.6119	.5802
η_{MK}	.0279	.0902	.1331	.0410
η_{ML}	- .0668	.0996	.1824	.0177
η_{ME}	- .0007	.1586	.0456	.0428
η_{MM}	.0396	.2211	- .3611	.0708

for 1959,

$$\hat{\sigma}_{KE} = .666134$$

$$SE(\hat{\sigma}_{KE}) = 8.55145$$

and, subject to homotheticity, that

$$\hat{\sigma}_{KE} = -4.20666 ,$$

$$SE(\hat{\sigma}_{KE}) = 3.49984 .$$

Taking account of the standard errors, we conclude that these data do not answer the question.

References

Allen, R. G. D., 1938, Mathematical analysis for economists (Macmillan, London).

Arrow, K. J., H. B. Chenery, B. S. Minhas, and R. M. Solow, 1961, Capital-labor substitution and economic efficiency, The Review of Economics and Statistics, 43, 225-250.

Berndt, Ernst R. and Mohammed S. Khaled, 1979, Parametric productivity measurement and choice among flexible functional forms, Journal of Political Economy 87, 1220-1245.

Berndt, Ernst R. and David O. Wood, 1975, Technology, prices and the derived demand for energy, Review of Economics and Statistics 57, 259-268.

Berndt, Ernst R. and David O. Wood, 1979, Engineering and econometric interpretations of energy-capital complementarity, The American Economic Review 69, 342-354.

Blackorby, Charles, Daniel Primont, and R. Robert Russell, 1978, Duality, separability, and functional structure: Theory and economic applications (North Holland, New York).

Burguete, Jose F., A. Ronald Gallant, and Geraldo Souza, 1982, On the unification of the asymptotic theory of nonlinear econometric models, Econometric Reviews, forthcoming.

Christensen, Laurits R., Dale W. Jorgenson and Lawrence J. Lau, 1975, Transcendental logarithmic utility functions, The American Economic Review 65, 367-383.

Diewert, W. E., 1974, Applications of duality theory, in: M. D. Intriligator and D. A. Kendrick, eds., Frontiers of quantitative economics, Vol. II (North-Holland, Amsterdam).

Edmunds, D. E. and V. B. Moscatelli, 1977, Fourier approximation and embeddings of Sobolov spaces, Dissertationes Mathematicae, CXLV.

Gallant, A. Ronald, 1975, Seemingly unrelated nonlinear regressions, Journal of Econometrics, 3, 35-50.

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 and Alberto Holly, 1980, Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation, Econometrica, 48, 697-720.

Goodnight, James H., 1979, A tutorial on the sweep operator, The American Statistician, 33, 149-158.

Griffin, James M. and Paul R. Gregory, 1976, An intercountry translog model of energy substitution responses, The American Economic Review 66, 845-857.

Guilkey, David, C. A. Knox Lovell, and Robin C. Sickeles, 1981, A comparison of three flexible functional forms, Unpublished Manuscript (University of North Carolina, Chapel Hill, NC).

Malinvaud, E., 1970, The consistency of nonlinear regressions, The Annals of Mathematical Statistics, 41, 956-969.

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).

Rao, C. Radhakrishna, 1973, Linear statistical inference and its applications, second edition (John Wiley and Sons, New York).

Theil, Henri, 1971, Principles of econometrics (John Wiley and Sons, New York).

Uzawa, H., 1962, Production functions with constant elasticities of substitution, Review of Economic Studies, 29, 291-299.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association, 57, 348-368.