Selecting Functional Form in Production Function Analysis

Ronald C. Griffin, John M. Montgomery, and M. Edward Rister

Functional form selection is a sometimes neglected aspect of applied research in production analysis. To provide an improved and uniform basis for form selection, a number of traditional and popular functional forms are catalogued with respect to intrinsic properties. Guidelines for the conduct of form selection are also discussed.

Key words: flexibility, functional form.

The art and science of applied economics critically depends on model building, and the practicing economist makes many decisions in the course of constructing an appropriate model. Within this process, the practitioner must often adopt a functional form as the pattern for one or more continuous (possibly piecewise) physical or economic relationships. Common examples of such relationships are production, profit, and cost functions as well as systems of input and/or output supply and demand. The continuity of these relationships is rarely, if ever, proven but seems sufficiently compelling to be embraced without question in nearly all cases. The researcher, however, is never in a position to know the true functional form, and, as noted by Hildreth, "the principal disadvantage of continuous models lies in the biases which may accrue if an inappropriate [functional] form is used" (p. 64).

The model builder's task is complicated by the growing number of available functional forms. A compilation of alternative functional forms and a comparison of these alternatives on the basis of selected criteria are presented in this paper. Available selection criteria pertain to mathematical, statistical, and economic properties and are useful for formalizing the selection of functional form(s) during model building processes. The use of functional forms in production function applications is emphasized in order to limit the discussion. Because the concept of flexibility is important to form selection, we begin with a brief summary of this topic.

Flexibility and Maintained Hypotheses

Recent advances in developing new functional forms have been dominated by efforts to conceive "flexible" forms, and different technical definitions of flexibility have arisen as a result of these pursuits. Because flexibility is a multidimensioned concept, a given technical definition of flexibility may not be adequate in all situations. Local flexibility (sometimes Diewert flexibility or, simply, flexibility) implies that an approximating functional form conveys zero error (perfect approximation) for an arbitrary function and its first two derivatives at a particular point (Fuss, McFadden, and Mundlak). The locally flexible form places no restrictions on the value of the function or its first or second derivatives at this point. Therefore, no restrictions are imposed on properties that can be expressed in terms of derivatives of second-order or less.

Second-order Taylor series expansions have dominated the field of locally flexible forms but are not unique in the ability to offer local flexibility (Barnett). Ignoring the complications of statistical estimation by momentarily presuming that we wish functionally to approximate a known relationship, locally flex-

The authors are an associate professor, a research assistant, and an associate professor, respectively, in the Department of Agricultural Economics, Texas A&M University.

This research as conducted by the Texas Agricultural Experiment Station (Projects 6447 and 6507) with support from the Texas Rice Research Foundation (Econo-Rice Project).

Appreciation is expressed to anonymous reviewers of this *Journal* and the *AJAE* for helpful comments during various stages of this article's development.

ible forms can impose large errors in the approximating function and its derivatives away from the point of perfect approximation (Despotakis). Global flexibility (sometimes Sobolev flexibility) is preferred to local flexibility in that second-order restrictions are everywhere absent (Gallant 1981, 1982).

Recent attention to the influence of estimation on model development has severely reduced the attractiveness of locally flexible forms. Because an estimated Taylor series expansion is a fit to data from the true form, rather than an expansion of the true form, there may be no actual point where the true function and its first two gradients are perfectly approximated. Inquiry in this area has produced some discomforting results. The examples provided by White demonstrate that ordinary least squares estimators of Taylor series expansions are not reliable indicators of the parameter vector for the true expansion of a known function. As a consequence of these and other findings, predictive properties of locally flexible forms have been found to be satisfactory (for large samples), but inferences involving single parameter estimates and functional combinations of these estimates are not reliable. Because estimation is driven by a global (at least throughout the data domain) criterion such as least sum of squared error, it has been a bit unnatural to tout local flexibility as an advantage in economic analyses. The implications here are severe for typical applications in production function analysis.

To gauge global flexibility in a manner which also acknowledges derivative values, a measure of error is needed which incorporates errors in derivatives as well as errors in the approximating function. The Sobolev norm (a distance measure) satisfies this requirement and has been employed to assess global flexibility (Gallant 1981). In applied work, however, the Sobolev norm is intractable for obtaining parameter estimates, so estimation of globally flexible forms still uses traditional distance measures like least squares. Evidence provided by Elbadawi, Gallant, and Souza suggests that this produces satisfactory results. This finding supports the judgment expressed by Rice that form exceeds norm in importance for approximation (pp. 1, 20). It is still true, though, that the effects of selecting a particular form are jointly determined by the inseparable influence of form restrictiveness and estimation technique.

The concept of maintained hypotheses is fundamental to any concept of flexibility. As pointed out by Fuss, McFadden, and Mundlak, maintained hypotheses "are not themselves tested as part of the analysis, but are assumed true" (p. 222). The choice of functional form immediately renders some hypotheses untestable (maintained) while others remain testable. "Sometimes the question of whether a specific hypothesis should be tested or maintained is critical" (Ladd, p. 9). As an example, King's 1979 address to the AAEA membership is most certainly rooted in his disappointment with the maintained hypotheses of linear models.

Attention to the issue of maintained hypotheses aids the researcher in developing a more careful and richer analysis. Popular concerns pertaining to potential maintained hypotheses within economic production models include homogeneity, homotheticity, elasticity of substitution, and concavity. Less restrictive functional forms would always be desirable, were it not for the greater information needed to adequately specify such relationships. Greater flexibility can usually be achieved by adding arbitrary and nonredundant terms to any given function, but to do so reduces degrees of freedom and may increase collinearity and the expense of parameter estimation. Because reductions in maintained hypotheses come at a cost, added flexibility is not always desirable, and there are likely to be cost-effective opportunities to achieve particular dimensions of flexibility.

The Choice Set

Based on a review of traditional and popular literature, twenty functional forms are identified for consideration. The names of each of these functions and their algebraic form are listed in the first two columns of table 1. Noteworthy reference material for these functions and some of their properties include Heady and Dillon (quadratic, square root, Mitscherlich, Spillman, resistance, modified resistance); Lau (square root[1]); Halter, Carter, and Hocking (transcendental); Uzawa (CES); Silberberg (CES); Diewert (generalized Leontief);

[1] Lau calls this the "generalized version of the 'Generalized Linear Function' " (p. 410). Fuss, McFadden, and Mundlak call it the generalized Leontief function.

Table 1. Properties of Selected Functional Forms

Function	Functional Form ($i, j, k = 1, \ldots, n$)	(A) Does $x_i = 0$ for all i Imply $y = 0$?	(B) Does $x_i = 0$ for any one i Imply $y = 0$?	(C) $\partial y/\partial x_i$	(D) Asymptotic Convergence
Leontief	$y = \min[\beta_1 x_1, \beta_2 x_2, \ldots, \beta_n x_n] \quad \beta \gg 0$	yes	yes	0 or UBC	no
Linear	$y = \alpha + \sum_i \beta_i x_i$	no	no	UBC	no
Quadratic[a]	$y = \alpha + \sum_i \beta_i x_i + \sum_i \sum_j \delta_{ij} x_i x_j$	no	no	U	no
Cubic[a,b]	$y = \alpha + \sum_i \beta_i x_i + \sum_i \sum_j \delta_{ij} x_i x_j + \sum_i \sum_j \sum_k \gamma_{ijk} x_i x_j x_k$	no	no	U	no
Generalized Leontief[a]	$y = \sum_i \sum_j \delta_{ij} x_i^{1/2} x_j^{1/2}$	yes	no	UBN	no
Square root[a]	$y = \alpha + \sum_i \beta_i x_i^{1/2} + \sum_i \sum_j \delta_{ij} x_i^{1/2} x_j^{1/2}$	no	no	U	no
Logarithmic	$y = \alpha + \sum_i \beta_i \ln x_i$	undefined	undefined	UBN	no
Mitscherlich	$y = \alpha \prod_i (1 - \exp(\beta_i x_i))$	yes	yes	UBN	if $\beta \ll 0$
Spillman	$y = \alpha \prod_i (1 - \beta_i^{x_i})$	yes	yes	UBN	if $0 \ll \beta \ll 1$
Cobb-Douglas	$y = \alpha \prod_i x_i^{\beta_i}$	yes	yes	UBN	no
Generalized Cobb-Douglas	$\ln y = \alpha + \sum_i \sum_j \delta_{ij} \ln((x_i + x_j)/2)$	undefined	no	U	no
Transcendental	$y = \alpha \prod_i x_i^{\beta_i}(\exp(\delta_i x_i))$	yes	yes	U	no
Resistance	$y^{-1} = \alpha + \sum_i \beta_i(\delta_i + x_i)^{-1}$	no	no	UBN	yes
Modified resistance[a]	$y^{-1} = \alpha + \sum_i \beta_i x_i^{-1} + \sum_i \sum_{j \neq i} \delta_{ij} x_i^{-1} x_j^{-1}$	undefined	undefined	UBN	yes
CES	$y = \left[\alpha + \sum_i \beta_i x_i^{-\rho}\right]^{-\nu/\rho}$	undefined	undefined	UBN	no
Translog[a]	$\ln y = \alpha + \sum_i \beta_i \ln x_i + \sum_i \sum_j \delta_{ij}(\ln x_i)(\ln x_j)$	undefined	undefined	U	no
Generalized quadratic	$y = \left[\sum_i \sum_j \beta_{ij} x_i^{\delta\gamma} x_j^{\delta(1-\gamma)}\right]^{\nu/\delta}$	yes	no	UBN	no
Generalized power	$y = \alpha \prod_i x_i^{\beta_i(x)} \exp(g(x))$	yes	yes	U	no
Generalized Box-Cox[a]	$y(\theta) = \alpha + \sum_i \beta_i x_i(\lambda) + \sum_i \sum_j \delta_{ij} x_i(\lambda) x_j(\lambda)$, where $y(\theta) = (y^{2\theta} - 1)/2\theta$ and $x_i(\lambda) = (x_i^{\lambda} - 1)/\lambda$	no	no	U	no
Augmented Fourier[a]	$y = \sum_i \beta_i x_i + \sum_i \sum_j \delta_{ij} x_i x_j + \sum_{\lvert h \rvert^* \leq H} \gamma_h \exp(i \sum_i h_i x_i)$, where all $x_i \in [0, 2\pi]$, $\gamma_h = \gamma_h^r + i\gamma_h^c$, $i^2 = -1$	no	no	U	no

Note: U is unrestricted sign (+, 0, or −); UBC, unrestricted in sign but constant; UBN, unrestricted but nonswitching in sign (e.g., if $\partial y/\partial x_k > 0$ for some $x_k > 0$, then $\partial y/\partial x_k > 0$ for all $x_k > 0$); and NG, not in general.

[a] Assuming $\delta_{ij} = \delta_{ji}$ for all i, j.

[b] Assuming $\gamma_{ijk} = \gamma_{ikj} = \gamma_{kij} = \gamma_{kji} = \gamma_{jki} = \gamma_{jik}$ for all i, j, k.

[c] Some of the stated conditions are sufficient but not necessary for local concavity (see text).

[d] Same as for quadratic form plus $\gamma_{ijk} = 0$ for all i, j, k.

[e] If any two of the following conditions are satisfied: (1) all $\beta_i = 0$, (2) all $\delta_{ij} = 0$, (3) all $\gamma_{ijk} = 0$.

[f] Yes, but x_i may equal zero for only one i.

[g] $\frac{1}{2}(n^2 + 3n + 3)$ assuming θ and λ are free.

[h] See appendix.

[i] Linear, quadratic, generalized Leontief, square root, logarithmic, Cobb-Douglas, modified resistance, CES, translog.

Table 1. Extended

(E) Linearly Homogenous	(F) Homothetic	(G) Constant Elasticity of Substitution	(H) Concave[c]	(I) No. of Distinct Parameters	(J) Linearly Separable	(K) Linearly Separable if any $x_i = 0$	(L) Subsumes What Other Functions
yes	yes	$\sigma = 0$	yes	n	no	no	
if $\alpha = 0$	yes	$\sigma = \infty$	affine	$n + 1$	yes	yes	
if $\alpha = 0$, all $\delta_{ij} = 0$	if $\beta = 0$ or all $\delta_{ij} = 0$	NG	NG	$\frac{1}{2}(n + 1)(n + 2)$	yes	yes	Linear
[d]	[e]	NG	NG	$(n + 3)(n + 2) \cdot (n + 1)/6$	yes	yes	Linear, quadratic
yes	yes	NG	if all $\delta_{ij} \geq 0$	$\frac{1}{2}n(n + 1)$	yes	yes	
if $\alpha = 0$, $\beta = 0$	if $\beta = 0$	NG	if $\beta \gg 0$, all $\delta_{ij} \geq 0$	$\frac{1}{2}(n + 1)(n + 2)$	yes	yes	Generalized Leontief
no	yes	NG	if $\beta \gg 0$	$n + 1$	yes	undefined	
no	no	NG	if $\alpha > 0, \beta \ll 0$, $\sum \exp(\beta_i x_i) < 1$	$n + 1$	no	no	Spillman
no	no	NG	if $\alpha > 0$, $0 \ll \beta \ll 1$, $\sum \beta_i^{x_i} < 1$	$n + 1$	no	no	Mitscherlich
if $\sum \beta_i = 1$	yes	$\sigma = 1$	if $\alpha > 0$, $0 \ll \beta \ll 1$, $\sum \beta_i < 1$	$n + 1$	yes	no	
if $\sum_i \sum_j \delta_{ij} = 1$	yes	no	NG	$\frac{1}{2}(n^2 + n + 2)$	yes	[f]	Cobb-Douglas
if $\sum \beta_i = 1$, $\delta = 0$	if $\delta = 0$	if $\delta = 0$	if $\alpha > 0$, $0 \ll \beta \ll 1$, $\sum \beta_i < 1$	$2n + 1$	yes	no	Cobb-Douglas
if $\alpha = 0, \delta = 0$	if $\delta = 0$	NG	if $\beta \gg 0, \delta \gg 0$	$2n + 1$	yes	yes	
if $\alpha = 0$, all $\delta_{ij} = 0$	if all $\delta_{ij} = 0$	NG	NG	$\frac{1}{2}(n^2 + n + 2)$	yes	undefined	
if $\alpha = 0$, $v = 1$	yes	$\sigma = 1/(1 + \rho)$	if $\alpha > 0, \beta \gg 0$, $0 < v < 1$	$n + 3$	NG	undefined	Linear, Cobb-Douglas
if $\sum \beta_i = 1$, all $\sum_i \delta_{ij} = 0$	if all $\sum_i \delta_{ij} = 0$	NG	if $\beta \gg 0$, all $\delta_{ij} \geq 0$	$\frac{1}{2}(n + 1)(n + 2)$	yes	undefined	Cobb-Douglas
if $v = 1$	yes	NG	if all $\beta_{ij} \geq 0$, $0 \leq \gamma \leq 1$, $\delta \leq 1$, $v \leq 1$	$n^2 + 3$	NG	NG	Generalized Leontief, CES
NG	NG	NG	NG	indeterminate	NG	NG	Transcendental
if $\sum \beta_i = 2a\lambda + 1$, all $\lambda\beta_i = 2 \sum_j \delta_{ij}$	if all $\delta_{ij} = 0$ or all $\delta_{ij} = 0$, $i \neq j$ and all $\lambda\beta_i = 2\delta_{ii}$	NG	NG	[g]	NG	NG	[i]
no	no	NG	NG	[h]	yes	yes	Linear, quadratic

Christensen, Jorgenson, and Lau (translog — transcendental logarithmic); Fuss, McFadden, and Mundlak (translog, generalized Cobb-Douglas, square root); Denny (generalized quadratic); de Janvry (generalized power); Applebaum (generalized Box-Cox); Berndt and Khaled (generalized Box-Cox); and Gallant (augmented Fourier).

The augmented Fourier form is globally flexible, asymptotically, in that no second-order restrictions are imposed anywhere in the domain. It is also nonparametric; the number of parameters to be estimated varies with sample size. Because the augmented Fourier form is not commonly used in production function analyses, a few essential details are presented in an appendix.

The Criteria

As stated above, determination of the true functional form of a given relationship is impossible, so the problem is to choose the best form for a given task. This problem leads to consideration of choice criteria, that is, how one functional form may be judged better or more appropriate than another. These criteria may be grouped in four categories according to whether they relate to maintained hypotheses, estimation, data, or application.

Concerns regarding maintained hypotheses form one set of objectives whereby appropriateness can be assessed. If the maintained hypotheses implied by a certain function are acceptable, or even useful, then the function may be deemed appropriate. In the absence of a strong theoretical or empirical basis for adopting a given maintained hypothesis, however, a functional form which is unrestrictive with respect to this hypothesis may be considered appropriate.

Second, functional form has implications for statistical processes of parameter estimation. Data availability, data properties, and the availability of computing resources can affect the choice of functional form for statistical estimation. Moreover, some forms do not permit parameter estimation by linear least squares procedures, and alternative procedures typically offer less information concerning estimator properties. Criteria relating to maintained hypotheses and statistical estimation are discussed below, and important findings have been condensed in table 1.

A third category of selection criteria involves data-specific considerations (goodness-of-fit and general conformity to data). Such concerns are necessarily omitted from this discussion because comparisons among function forms require the use of a specific dataset and findings would not be general. In addition, there is a significant body of literature concerning testing of nested and non-nested models (see Judge et al.).

The fourth grouping of selection criteria pertains to application-specific characteristics. For example, if the resulting equation is to be used in simulation or optimization procedures, there may be other desirable properties for functional form. These considerations are largely excluded from the present discussion because the number of potential criteria within this category is quite large and highly customized.

For ease of presentation, the discussion emphasizes selection of a production function rather than a demand, profit, or cost function. The rationale for this choice stems partially from the fact that most recent advances in conceiving flexible forms are published as cost or demand applications. Much of the discussion presented here can be readily applied to cost or demand studies. System-related properties, such as symmetry, are not considered because of the production function emphasis.

Some functional forms allow as special cases the assumption of all properties of other functional forms. The more flexible forms may be appropriate when information regarding the nature of the relationship does not permit certain hypotheses to be maintained. Those functional forms subsumed by others are indicated in the last column of table 1. Employing a nonlinear transformation of coordinates, the generalized Box-Cox form subsumes nine other forms as nested cases. They are: translog ($\theta = \lambda = 0$); CES ($v\theta = \lambda$, all $\lambda\beta_i = 2\delta_{ii}$, $\delta_{ij} = 0$ for all $i \neq j$); modified resistance ($\theta = -\frac{1}{2}$, $\lambda = -1$, all $\delta_{ii} = 0$); Cobb-Douglas ($\theta = \lambda = 0$, $\Sigma\, \beta_i = 1$, all $\delta_{ij} = 0$); logarithmic ($\theta = \frac{1}{2}$, $\lambda = 0$); square root and generalized Leontief ($\theta = \lambda = \frac{1}{2}$); and quadratic and linear ($\theta = \frac{1}{2}$, $\lambda = 1$).

Maintained Hypotheses

Column A of table 1 concerns the assumed effect of inputs specified in the function but not applied (i.e., those inputs with a zero level). Assuming that one or more inputs are hypothesized to be related to output and included as exogenous variables, column A indicates for which functional forms output would be zero if all input levels were zero. Such a maintained hypothesis may not conform to observed phenomena. In agronomic fertility studies, for example, one often encounters experimental data where nutrient applications are observed, but soil nutrient levels are not. Zero yields are imposed by particular functional forms when observations pertain to applied nutrients and no nutrients are applied. The hypothesis represented by column A thereby underscores the relationship between variable specification and functional form.

A similar but more relevant property is described in column B, which indicates for which functions inputs are essential; that is, no output occurs if at least one of the inputs of a multiple-input model is not applied. Again, the importance of this property primarily depends on how the inputs of a specific model are defined.

First derivatives of production functions are important in nearly all applications. Because some are cumbersome to compute, derivatives for all twenty forms are given in table 2. Column C of table 1 indicates maintained hypotheses concerning the sign of marginal product. Functions in table 1 are characterized by marginal products which are unrestricted in sign (*U*), unrestricted but nonswitching in sign (*UBN*), or unrestricted in sign but constant in value (*UBC*). Functional forms with marginal products that are unrestricted but nonswitching in sign allow successive units of applied input to increase, decrease, or not change total output, but any change in total output must be in the same direction as the change produced by the previous unit of input. Thus, the nine *UBN* functional forms (as well as the two *UBC* ones) do not allow model estimation to determine at what input level output begins to decrease (assuming the data supports this testable hypothesis) but rather maintains the hypothesis of everywhere positive (or everywhere negative) marginal productivity.

Column D indicates which functional forms maintain the hypothesis of asymptotic convergence of output towards a maximum as input levels are increased. In such cases, output increases as the level of input increases, regardless of the data employed to estimate the change in output due to a change in the level of input.

Two related properties exhibited by certain forms are homogeneity and homotheticity. Functional forms homogenous of degree 1, said to be linearly homogenous, are indicated in column E. If production is found to be profitable for some input combination in a linearly homogenous production function, profit can be increased without bound by increasing inputs proportionately.

Homothetic forms are indicated in column F of table 1. The interesting characteristic of homothetic production functions is that the marginal rate of technical substitution (*MRTS*) remains constant as all inputs are increased proportionately. When the marginal rate of technical substitution is constant, as for homothetic functions, any change in the value of output will affect the optimum input levels proportionately as long as input prices are unchanged.

Related to the concept of the marginal rate of technical substitution is the concept of the elasticity of substitution. In general, $MRTS_{ij}$ will increase in magnitude as x_i is increased (and x_j is decreased) along any given isoquant. This rate of change can be measured by the elasticity of substitution.[2] Those functional forms assuming constant elasticity of substitution are indicated in column G.

The generalized constant elasticity of substitution (CES) production function, which is defined not only to have constant elasticity of substitution but also to be homogenous of degree v when $\alpha = 0$, represents a much-studied class of production functions (Arrow et al.; McFadden; Uzawa). The concept of elasticity of substitution and the CES class of production functions have been applied primarily to the problem of the distribution of income among generally defined or highly aggregated factors of production (Arrow et al.; Behrman; Nerlove). With regard to production function analysis, however, Silberberg states that "knowledge of σ at any point would undoubtedly be a useful technological datum for em-

[2] For any change along an isoquant, $d(x_i/x_j)$ is the change in the use of x_i as compared with that of x_j, and $d(dx_i/dx_j)$ is the corresponding change in the marginal rate of technical substitution. "The ratio of these differentials, expressed in proportional terms to make them independent of units of measurement, is defined as the elasticity of substitution between the factors at the combination of factors considered" (Allen, p. 341). The elasticity of substitution between x_i and x_j is given by

$$\sigma_{ij} = \frac{(f_i/f_j)}{(x_i/x_j)} \frac{d(x_i/x_j)}{d(f_i/f_j)},$$

where $f_i = \partial y/\partial x_i$ and $f_j = \partial y/\partial x_j$.

Table 2. First Derivatives of Selected Functional Forms

Function	$\frac{\partial y}{\partial x_i}$
Leontief	0 or β_i
Linear	β_i
Quadratic	$\beta_i + \sum_j 2\delta_{ij}x_j$
Cubic	$\beta_i + \sum_j 2\delta_{ij}x_j + \sum_{j \neq i} \sum_k 6\gamma_{ijk}x_j x_k + 3\gamma_{iii}x_i^2$
Generalized Leontief	$\sum_j \delta_{ij}x_i^{-1/2}x_j^{1/2}$
Square root	$(1/2\beta_i + \sum_j \delta_{ij}x_j^{1/2})x_i^{-1/2}$
Logarithmic	$\beta_i x_i$
Mitscherlich	$-\alpha\beta_i \exp(\beta_i x_i) \prod_{j \neq i} (1 - \exp(\beta_j x_j))$
Spillman	$-\alpha\beta_i^{x_i}\ln(\beta_i) \prod_{j \neq i} (1 - \beta_j^{x_j})$
Cobb-Douglas	$\alpha\beta_i x_i^{-1} \prod_j x_j^{\beta_j}$
Generalized Cobb-Douglas	$y \sum_j 2\beta_{ij}/(x_i + x_j)$
Transcendental	$\alpha(\delta_i + \beta_i/x_i) \prod_j x_j^{\alpha_j}\exp(\delta_j x_j)$
Resistance	$\beta_i y^2(\delta_i + x_i)^{-2}$
Modified resistance	$y^2(\beta_i x_i^{-2} + \sum_{j \neq i} \delta_{ij}x_i^{-2}x_j^{-1})$
CES	$\beta_i v x_i^{-1-\rho}[\alpha + \sum_i \beta_i x_i^{-\rho}]^{-\frac{v+\rho}{\rho}}$
Translog	$y[\beta_i + 2 \sum_j \beta_{ij}\ln(x_j)]/x_i$
Generalized quadratic	$v x_i^{-1} y^{\frac{v-\delta}{v}} [\gamma \sum_j \beta_{ij}x_i^{\delta\gamma}x_j^{\delta(1-\gamma)} + (1 - \gamma) \sum_j \beta_{ij}x_i^{\delta(1-\gamma)}x_j^{\delta\gamma}]$
Generalized power	———
Generalized Box-Cox	$x_i^{\lambda-1}y^{1-2\theta}[\lambda\beta_i + 2 \sum_j \delta_{ij}(x_i^{\lambda} - 1)]/\lambda$
Augmented Fourier	$\beta_i + \sum_j 2\delta_{ij}x_j + i \sum_{\|h\|^* \leq H} h_i\gamma_h \exp(i \sum_i h_i x_i)$

pirical work" in that the statistic describes the relative ease with which one input may be substituted for another (p. 317).

A final property, concavity, is important in at least two respects. It is in one sense pertinent to description of the production process, reflecting a situation in which output increases at a decreasing rate (or decreases at an increasing rate) as the level of input is increased. The property may be of primary importance, however, in the context of economic optimization; only if the function is concave can input levels that maximize profit be computed from first-order equations.

As indicated in column H of table 1, estimated parameters of the model may need to be examined before the researcher can ascertain whether concavity is present. The importance of concavity for profit maximization implies that such an examination should always be conducted. While the conditions listed for concavity in table 1 are necessary and sufficient for global concavity, that is, for concavity everywhere in the nonnegative orthant, many of these conditions are not necessary for local concavity. In particular, the indicated conditions are merely sufficient for local concavity in the case of the square root, resistance, modified resistance, CES, generalized Leontief, translog, and generalized quadratic forms.

Statistical Estimation

Certain properties embodied in functional forms have important implications for the mathematical procedures employed in the statistical estimation of parameters. Pertinent properties included in table 1 are the number of distinct parameters (column I), and linear separability of the parameters (columns J and K).

Most functions, in their complete forms, require a geometrically increasing number of parameters to be estimated as the number of variables of main effect is increased. This is primarily because of the large number of interactions specified among variables of main effect. For example, if two inputs are hypothesized to affect output, a linear function requires only three parameters to estimate that effect, whereas a complete cubic function requires estimation of ten parameters. If ten inputs are hypothesized to affect output, a linear function requires that eleven parameters be estimated, but a cubic function requires estimation of 286 parameters.[3]

Fuss, McFadden, and Mundlak have expressed concern that "excess" parameters would "exacerbate" multicollinearity present in survey data (p. 224). If multicollinearity is high, the variance of the parameter estimates is increased such that it may be impossible to determine how much variation in the endogenous variable is explained by different exogenous variables. Higher parameter variances also enlarge confidence intervals for applications (such as the computation of an optimal input program) involving subsets of the estimated parameters. If, however, multicollinearity is irrelevant because the primary purpose of the model is for prediction (Maddala, p. 186), then the issue of having many parameters can be quite unimportant. Of possible concern to the researcher may be the expense of estimating a large number of parameters or the loss of degrees of freedom. Either of these problems could actually preclude the use of certain forms, depending on the number of variables of main effect, computing resources available to the researcher, and/or size of the dataset. The total number of parameters to be estimated by a given form may be calculated with the formulas presented in column I.

Ease of analysis and availability of resources may also limit the modeling process to functional forms containing parameters that are all linearly separable (indicated in column J) and can therefore be estimated by common and well-developed linear least squares regression techniques. Whereas direct estimation of nonlinearly separable forms may be possible by such techniques as maximum likelihood estimation, Fuss, McFadden, and Mundlak note that "linear-in-parameter systems have a computational cost advantage, and have, in addition, the advantage of a more fully developed statistical theory" (p. 225). Linear least squares regression provides information regarding the small-sample accuracy and precision of estimates, which may not be available from other techniques. The value of this information must be weighed against any advantages in the use of nonlinearly separable forms. On the other hand, the gain in information for linear least squares estimators may be artificial, since this information results from several important assumptions.

In addition, the nature of the dataset must be considered if a technique such as ordinary least squares is to be employed. Column K of table 1 shows which functional forms are linearly separable in parameters when any input levels take on zero values. For example, even though column J identifies the generalized Cobb-Douglas form as being linearly separable, if the dataset used for estimating the function contains an observation in which none of one input is applied, ordinary least squares cannot be conducted.

It is notable that the generalized Box-Cox form is not linear in parameters, and this fact tends to limit applications. If coefficients θ and λ are chosen a priori, the form becomes linear in parameters. When this is done, the Box-Cox form becomes one of the subsumed functional forms, and its generality is sacrificed.

Formalizing the Selection of Functional Form

This paper describes why the researcher should be concerned with the functional form of a continuous input-output model and provides some guidance for assessing whether some functional forms may be considered more ap-

[3] In some cases, additional maintained hypotheses can be introduced to eliminate certain interactions. For example, any real interaction between the quantities of irrigation water and harvest labor in crop production may be thought, a priori, to be absent.

propriate than others. As a result of the mathematical properties inherent in each functional form, hypotheses are maintained about the production relationship whenever a specific function is selected for estimation.[4] Any potential hypothesis which is not maintained and can be expressed as one or more linear relations among function parameters will be testable.[5] Such relations are indicated in table 1. Inherent mathematical properties of a given functional form can have important implications for statistical estimation and economic application.

A functional form may be appropriate because of the correspondence of maintained hypotheses with generally held theories of the true input-output relationship, possibility and ease of statistical estimation, possibility and ease of application, general conformity to data, or a combination of these criteria. None of the many criteria guarantee that the true relationship will be discovered, nor do any allow a totally objective choice to be made. Subjective judgment is a necessary aspect of choice regarding functional form. Thus, formalization of the selection process requires deliberative choice and frank presentation.

Having selected two or more estimable functional forms with acceptable theoretical and application properties, the researcher may wish to base final selection upon a statistical criterion. Such criteria entail data-specific considerations and may comprise a decision rule as simple as choosing the model with the highest coefficient of multiple determination. Strictly statistical information can be used for the process of model selection since several formal/informal opportunities are available for both nested and nonnested testing. Judge et al. note that tests of nested models measure "how well the models fit the data after some adjustment for parsimony" (p. 862). Tests of nonnested models ask the question: Is the performance of model 1 "consistent with the truth" of model 2 (Pesaran and Deaton, p. 678)? Thus, the testing of each model against the evidence provided by the other will not necessarily allow the investigator to choose one model over the other. While it is clear that such data-related objectives can form an important set of criteria, allowing statistical evidence to dictate model selection is regarded as a questionable and often unnecessary practice.

The following observation by Hildreth is of interest: "It is particularly disconcerting that, in many instances in which several alternative assumptions [as to functional form] have been investigated, alternative fitted equations have resulted which differ little in terms of conventional statistical criteria such as multiple correlation coefficients or F tests of the deviation, but differ much in their economic implication" (p. 64). Given the possible differences in economic implications, it is often advisable to explore the sensitivity of calculated economic optima to the choice of functional form. As an interesting example, in an analysis of a single application of nitrogen and potassium on corn, Bay and Schoney have assessed the costs of "incorrectly" choosing among four possible functional forms. Similarly, in an analysis of multiple nitrogen applications on rice, Griffin et al. have examined the sensitivity of optimal fertilizer programs to functional form.

Conclusions

It is often difficult to ascertain why particular functional forms are chosen for the models presented in published economic research. Possibly, researchers confine their attentions to particular functional forms because they are most comfortable and experienced with these forms or because these forms are in vogue. In fairness, there are probably many examples where the research process is extensive and formal in the selection of functional form, but a description of these procedures is omitted from published results. These omissions may be due to a propensity of authors, reviewers, or editors to consider such material extraneous. Applied economists can enrich their analysis by selecting functional forms from as broad a choice set as possible and by considering a number of selection criteria (other than merely data-related criteria). Formal statement of these procedures in published output is desirable so that readers might better gauge research results (particularly sensitivity to form).

The need "to make research processes overt"

[4] Also, properties which are not inherent to a given form can often be incorporated as maintained hypotheses by the appropriate choice of linear restrictions. Table 1 indicates the appropriate restrictions for the considered criteria. Moreover, different functional forms can be simultaneously developed in order to test different hypotheses (as in Shumway's paper).

[5] However, such tests require the additional (augmenting) hypothesis that the true form is of the class under consideration (Gallant 1984).

(Tweeten, p. 549) pertains as much to the choice of functional form as to other aspects of commonly reported research methodology. That the true form cannot be known and that it is impossible to measure how well any chosen form approximates the true form lends importance to the selection process. Formalization of this process requires the development of an explicit and, hopefully, exhaustive choice set and some acceptable criteria for conducting the selection. The objective of this paper has been to develop such a choice set and offer an initial listing of plausible criteria. Extension of this compilation and discussion to settings other than production function applications remains as a needed endeavor.

[*Received December 1985; final revision received August 1987.*]

References

Allen, R. G. D. *Mathematical Analysis for Economists.* New York: St. Martin's Press, 1938.

Applebaum, E. "On the Choice of Functional Forms." *Int. Econ. Rev.* 20(1979):449–58.

Arrow, K. J., H. B. Chenery, B. S. Minkas, and R. M. Solow. "Capital Labor Substitution and Economic Efficiency." *Rev. Econ. and Statist.* 63(1961):225–50.

Barnett, W. A. "New Indices of Money Supply and the Flexible Laurent Demand System." *J. Bus. and Econ. Statist.* 1(1983):7–23.

Bay, T. F., and R. A. Schoney. "Data Analysis with Computer Graphics: Production Functions." *Amer. J. Agr. Econ.* 64(1982):289–97.

Behrman, J. R. "Selectoral Elasticities of Substitution Between Capital and Labor in a Developing Economy: Time Series Analysis in the Case of Postwar Chile." *Econometrics* 40(1972):311–26.

Berndt, E. R., and M. S. Khaled. "Parametric Productivity Measurement and Choice Among Flexible Functional Forms." *J. Polit. Econ.* 87(1979):1220–45.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau. "Transcendental Logarithmic Production Frontiers." *Rev. Econ. and Statist.* 55(1973):28–45.

de Janvry, A. "The Generalized Power Function." *Amer. J. Agr. Econ.* 54(1972):234–37.

Denny, M. "The Relationship Between Forms of the Production Function." *Can. J. Econ.* 7(1974):21–31.

Despotakis, K. A. "Economic Performance of Flexible Functional Forms." *Eur. Econ. Rev.* 30(1986):1107–43.

Diewert, W. E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." *J. Polit. Econ.* 79(1971):481–507.

Elbadawi, I., A. R. Gallant, and G. Souza. "An Elasticity Can Be Estimated Consistently without A Priori Knowledge of Functional Form." *Econometrica* 51(1983):1731–51.

Fuss, M., D. McFadden, and Y. Mundlak. "A Survey of Functional Forms in the Economic Analysis of Production." *Production Economics: A Dual Approach to Theory and Applications,* ed. M. Fuss and D. McFadden, pp. 219–68. Amsterdam: North-Holland Publishing Co., 1978.

Gallant, A. R. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form." *J. Econometrics* 15(1981):211–45.

———. "The Fourier Flexible Form." *Amer. J. Agr. Econ.* 66(1984):204–15.

———. "Unbiased Determination of Production Technologies." *J. Econometrics* 20(1982):285–323.

Griffin, R. C., M. E. Rister, J. M. Montgomery, and F. T. Turner. "Scheduling Inputs with Production Functions: Optimal Nitrogen Programs for Rice." *S. J. Agr. Econ.* 17(1985):159–68.

Halter, A. N., H. O. Carter, and J. G. Hocking. "A Note on the Transcendental Production Function." *J. Farm Econ.* 39(1957):966–74.

Heady, E. O., and J. L. Dillon, eds. *Agricultural Production Functions.* Ames: Iowa State University Press, 1961.

Hildreth, C. G. "Discrete Models with Qualitative Descriptions." *Methodological Procedures in the Economic Analysis of Fertilizer Use Data,* ed. E. L. Baum, E. O. Heady, and J. Blackmore, pp. 62–75. Ames: Iowa State College Press, 1955.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T. Lee. *The Theory and Practice of Econometrics,* 2nd ed. New York: John Wiley & Sons, 1985.

King, R. A. "Choices and Consequences." *Amer. J. Agr. Econ.* 61(1979):839–48.

Ladd, G. W. "Artistic Tools for Scientific Minds." *Amer. J. Agr. Econ.* 61(1979):1–11.

Lau, L. J. "Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints." *Production Economics: A Dual Approach to Theory and Applications,* ed. M. Fuss and D. McFadden, pp. 409–53. Amsterdam: North-Holland Publishing Co., 1978.

McFadden, D. "Constant Elasticity of Substitution Production Functions." *Rev. Econ. Stud.* 30(1963):73–83.

Maddala, G. S. *Econometrics.* New York: McGraw-Hill Book Co., 1977.

Nerlove, M. "Recent Empirical Studies of the CES and Related Production Functions." *The Theory and Empirical Analysis of Production,* ed. M. Brown, pp. 55–122. New York: Columbia University Press, 1967.

Pesaran, M. H., and A. S. Deaton. "Testing Non-Nested Nonlinear Regression Models." *Econometrica* 46(1978):677–94.

Rees, C. S., S. M. Shah, and C. V. Stanojevic. *Theory and Applications of Fourier Analysis.* New York: M. Dekker, 1981.

Rice, J. R. *The Approximation of Functions,* vol. 1. Reading MA: Addison-Wesley Publishing Co., 1964.

Rudin, W. *Principles of Mathematical Analysis,* 2nd ed. New York: McGraw-Hill Book Co., 1964.

Shumway, C. R. "Supply, Demand, and Technology in a Multiproduct Industry: Texas Field Crops." *Amer. J. Agr. Econ.* 65(1983):748–60.

Silberberg, E. *The Structure of Economics: A Mathematical Analysis.* New York: McGraw-Hill Book Co., 1978.

Tolstov, G. P. *Fourier Series.* Englewood Cliffs NJ: Prentice-Hall, 1962.

Tweeten, L. "Hypothesis Testing in Economic Science." *Amer. J. Agr. Econ.* 65(1983):548–52.

Uzawa, H. "Production Functions with Constant Elasticity of Substitution." *Rev. Econ. Stud.* 29(1962): 291–99.

White, H. "Using Least Squares to Approximate Unknown Regression Functions." *Int. Econ. Rev.* 21(1980):149–70.

Appendix

The Augmented Fourier Form

Because complex-valued trigonometric polynomials of order H,

$$\text{(A.1)} \qquad y(x) = \sum_{|h|^* \leq H} \gamma_h \exp\left(i \sum_i h_i x_i\right),$$

can be used to approximate any known periodic function, and because the approximation is perfect when $H = \infty$, trigonometric polynomials have a number of fruitful applications in physics and the engineering sciences. The following additional information is needed to interpret equation (A.1): h is an n-dimensional vector of integers $(\ldots, -1, 0, 1, \ldots)$ with "length" $|h|^* = \sum_i |h_i|$; H is exogenously chosen by the researcher and serves to limit the number of terms in (A.1). Each coefficient γ_h is complex-valued, $\gamma_h = \gamma_h^r + i \cdot \gamma_h^c$, where $i^2 = -1$ and γ_h^r and γ_h^c are the real and complex components of γ_h, respectively. Thus, each γ_h really represents two separate parameters.

If $y(x)$ is known and the terms γ_h are chosen in a certain manner (see Tolstov, pp. 13–14, 174; or Rees, Shah, and Stanojevic, p. 225), then these terms are Fourier coefficients, and equation (A.1) identifies a Fourier series approximating $y(x)$. For all possible trigonometric polynomials of limited order approximating $y(x)$, the Fourier series provides the best approximation as measured by least sum-of-square differences across the domain (Rudin, pp. 172–73; Tolstov, p. 174).

In production analysis we are not interested in periodic or complex-valued functions. Because production functions are not expected to be periodic, the exogenous factors, the x_i, must be scaled so that they lie in an interval of length less than 2π. This is typically accomplished by scaling or normalizing x_i into $[0, 2\pi]$ (Gallant 1981). In order to guarantee that $y(x)$ in equation (A.1) is real-valued, it is sufficient to impose

$$\text{(A.2)} \qquad \gamma_{h=0}^c = 0, \quad \gamma_h^r = \gamma_{-h}^r, \quad \text{and} \quad \gamma_h^c = -\gamma_h^c.$$

Equation (A.1), conditions (A.2), and the identity

$$\exp(i \sum h_i x_i) \equiv \cos(\sum h_i x_i) + i \cdot \sin(\sum h_i x_i)$$

can be used to obtain

$$\text{(A.3)} \qquad y(x) = \sum_{|h|^* \leq H} [\gamma_h^r \cos(\sum h_i x_i) - \gamma_h^c \sin(\sum h_i x_i)].$$

While (A.1) is a more compact representation of the general trigonometric polynomial, application to production function analysis begins with equation (A.3). Note that equation (A.3) and its derivatives are obviously real-valued, and conditions (A.2) must still be econometrically imposed.

Gallant (1981) is responsible for combining the quadratic form and equation (A.3) to produce the so-called Fourier form or, by our nomenclature, the augmented Fourier form:

$$\text{(A.4)} \qquad y = \sum_i \beta_i x_i + \sum_i \sum_j \partial_{ij} x_i x_j + \sum_{|h|^* \leq H} [\gamma_h^r \cos(\sum h_i x_i) - \gamma_h^c \sin(\sum h_i x_i)].$$

The constant term, α, is omitted to avoid redundancy since it is contained in the Fourier expression ($h = 0$). The number of parameters in the augmented Fourier form depends on n and H, but we have been unable to obtain a completely general formula for the number of parameters in the Fourier portion of (A.4). The following formula applies for $n = 1$, 2, 3, or 4.

$$m = m_1 + m_2 = \frac{n(n+3)}{2} + m_2,$$

where m is the total number of coefficients in the augmented Fourier form, m_1 is the number of terms in the quadratic portion (excluding the constant term), and m_2 is the number of terms (including the constant term) in the Fourier portion; m_2 is given by an nth degree polynomial in H:

$$m_2 = a_0 + a_1 H^1 + a_2 H^2 + \ldots + a_n H^n$$

with coefficients identified in the following table for n up to 4.

Table A.1. Polynomial Coefficients for m_2

		a_0	a_1	a_2	a_3	a_4
n	1	1	2	0	0	0
	2	1	2	2	0	0
	3	1	8/3	2	4/3	0
	4	3/11	536/63	−2/33	164/99	2/3

The number of terms in the Fourier portion can be quite large if either n or H is large. To illustrate the effect of increasing n when H is small: if $H = 2$ and $n = 2$, $m_1 = 5$ and $m_2 = 13$; if $H = 2$ and $n = 4$, $m_1 = 14$ and $m_2 = 41$.

Sufficient (but not necessary) conditions for concavity of the augmented Fourier form can be derived by following the procedure of Gallant (1981). To do so requires substantial additional notation and will not be pursued here.