UNKNOWN REGRESSION FUNCTIONS AND EFFICIENT FUNCTIONAL FORMS: AN INTERPRETATION

Esfandiar Maasoumi

ABSTRACT

For a given preference structure, utility functions can be represented by many functional forms which are compatible with the preferences. Likewise desirable technologies can be represented by different functional forms. Any choice of functional form determines a "distribution" of utilities (outputs, etc.) with, among others, important welfare and efficiency implications. In this paper we identify criteria according to which some popular functional forms (aggregates) have distributions which "most closely" mimic the distributions of commodities and inputs among the members of a population. Such distributive criteria may be regarded as additional restrictions on index numbers or in aggregation. Some econometric implications are also discussed.

Advances in Econometrics, vol. 5, pages 301–309

ISBN: 0-89232-686-7

I. INTRODUCTION

General discussion of exchange, whether in a pure exchange setting or a competitive equilibrium with prices, identifies "optimal" allocation(s) of goods and services. The solution may be a unique competitive equilibrium or an element of the core of the economy. The optimality or efficiency of such solutions is in the sense of Pareto, with an important distinction arising between Pareto optima and "individually rational" Pareto optima.[1] In any event, the "final" distribution of goods and services from the consumption (or the production) set, is an allocation which may be subjected to social assessment through the medium of social welfare functionals. When, for instance, inequality is measured by evaluative statistics the welfare comparison is implicit. Indeed one may argue that such comparisons should be made explicit so as to bring out the role played by the particular welfare function employed in the analysis. Some welfare functions are more "equality preferring" than others; see, e.g., Atkinson (1970). Thus the distribution of utility matters. To focus on the role played by functional forms I will first consider the case of utility functions since they are the commonly favored arguments of welfare functions. In discussing production a similar role is played by the "full production frontier" or the aggregate input. Representing preferences by utility functions (indices) immediately determines an allocation or distribution of utilities. Although social welfare is not a pie, such as income, to be distributed among individuals, it is clear that any index formally determines a distribution according to the functional form selected for that index.

Whether the allocation of goods is considered optimal or not, in this paper I adopt the position that the choice of utility function, while compatible with preferences and having other desirable properties, should suggest a distribution of welfare that does not distort the actual distribution of goods and services. Otherwise welfare analyses which hope to be sensitive to distributional aspects will be distorted to the extent that they do not accurately incorporate the *actual* distribution of goods and services. Formally, this is an aspect of the index number problem which has been largely neglected in the relevant literature.

Given an allocation of a set of commodities, we have one distribution associated with each commodity that gives the share of each individual in that commodity. The only complete ordering of individuals is possible through preferences and the associated utility function (index). We pose the question: which index implies an ordering (distribution) of individuals which "most closely" represents their orderings (distributions) by each and all the commodities? This requirement restricts the choice of utility functions and may be a desirable axiom for social welfare analyses. Other popular axioms discussed in the literature include additivity, differentiability, and concavity; see, e.g., Debreu (1964, 1972) and Koopmans (1972). These axioms restrict the class of admissible preference relations and thus imply important behavioral restrictions.

As is common to all fields of science, any notion of "closeness" is associated with a criterion of distance or divergence. In the case of the question above we are therefore concerned with measures of divergence between distributions. In mathematical statistics many such measures are proposed and utilized in a variety of seemingly disparate problems; see, e.g., Fisher (1925), Shannon (1948), Rao (1949), Jeffreys (1961), and Burbea and Rao (1982a, 1982b). Such measures have begun to be employed in the fields of income inequality (Cowell, 1980, 1982), multidimensional inequality (Maasoumi, 1979, 1986; Maasoumi and Nickelsburg, 1983), and international trade (Theil, 1979).

The main result of this paper, given in the next section, is that some of the most popular functional forms in economics (CES, Cobb-Douglas, Leontief, etc.) do indeed define welfare distributions which are the "closest" to the commodity distributions. We identify the criteria according to which this is so. It will be seen that such criteria are rather arbitrary and incorporate value judgments on such things as the relative importance of commodities, their substitutability, and the degree of aversion to (valuation of) distributional distortion associated with any choice of utility function or any aggregate of miscellaneous inputs (Bliss, 1975). In the next section we define a multivariate generalization of a generalized information (entropy) measure which serves as our divergence criterion. We then find the class of functional forms that includes many of the popular functions employed in economic analysis, and which minimizes the distributional divergence mentioned earlier. The last two sections offer some remarks on econometric implications and the utility of these measures in evaluating approximate regression functions.

II. OPTIMAL FUNCTIONAL FORMS

Given an allocation of m goods among N individuals, X_{if}; $i = 1, \ldots, N$, and $f = 1, \ldots, m$, denote the ith individual's index of utility by S_i, $i = 1, \ldots, N$. The vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{im})'$ represents the allocation of m commodities to the ith individual. $X^f = (X_{1f}, X_{2f}, \ldots, X_{Nf})'$ is the allocation of commodity f. Without loss of generality and in order to speak of our *convex* criteria as measures of divergence between distributions, we shall work with normalized variables $x_{if} \geq 0$ such that $x_{if} = X_{if}/\sum_i X_{if}$, $\sum_i x_{if} = 1$, and redefine S_i such that $\sum_i S_i = 1$.

A popular, *convex* measure of divergence between any two distributions, $\{Z_i\}$ and $\{y_i\}$, $Z_i, y_i \geq 0$, is the generalized entropy (β-order entropy) given by

$$\Delta_\beta = \frac{1}{\beta(\beta+1)} \sum_i y_i \left[\left(\frac{y_i}{Z_i} \right)^\beta - 1 \right]. \tag{1}$$

Some special forms of (1) are

$$\Delta_0 = \sum_i y_i \log \left(\frac{y_i}{Z_i} \right) \tag{2}$$

and

$$\Delta_{-1} = \sum_i Z_i \log \frac{Z_i}{y_i}. \tag{3}$$

Δ_0 and Δ_{-1} are well-known measures of "expected information" in going from distribution $\{Z_i\}$ to $\{y_i\}$. They are related to Theil's two indices of inequality, which have a central place in the analysis of single-dimensioned (e.g., income) inequality (Theil, 1967). Cowell (1980) provides a useful analysis of the properties of (1) for analyzing distributional change.

As in Maasoumi (1979, 1986), we consider a multivariate generalization of (1) to measure divergence between S_i and all the other m distributions $x^f, f = 1, \ldots, m$.[2]

$$\Delta_{\beta,m} = \frac{1}{m} \sum_f \sum_i \frac{1}{\beta(\beta+1)} S_i \left[\left(\frac{S_i}{x_{if}} \right)^\beta - 1 \right]. \tag{4}$$

Minimizing $\Delta_{\beta,m}$ with respect to S_i, we find the distribution of S_i, $i = 1, \ldots, N$, which is the "closest" to those of x^f in the sense of (4):

$$S_i = \frac{(\sum_f x_{if}^{-\beta})^{-1/\beta}}{\sum_i (\sum_f x_{if}^{-\beta})^{-1/\beta}}. \tag{5}$$

If $\sum_i S_i = 1$ is not imposed, (5) will be

$$S_i \propto \left[\frac{\beta+1}{m} \sum_f x_{if}^{-\beta} \right]^{-1/\beta} \tag{5'}$$

with $\propto$ denoting proportionality.

The solution (5′) may be recognized as a constant elasticity of substitution (CES) function which contains the linear case ($\beta = -1$) and, upon the application of L'Hospital's rule, the Cobb-Douglas function ($\beta = 0$) which is a geometric mean here, and the Leontief case of fixed coefficients ($\beta \to +\infty$).

Further generalizations are straightforward. For instance, we may consider different weights for different goods distributions. Then the solution is the more general CES,

$$S_i \propto \left[\frac{(\beta+1)}{m} \sum_f \alpha_f x_{if}^{-\beta} \right]^{-1/\beta}. \tag{6}$$

The generalized version of $\Delta_{\beta,m}$ corresponding to (6) is the weighted sum of the m distances defined in (4) with weights $(\alpha_1, \ldots, \alpha_m)$: we consider divergence from the distribution in X^1 to be more deleterious than that from the distribution of good k, $\mathbf{X}^k$, if $\alpha_1 > \alpha_k$, with $\sum_f \alpha_f = m$.[3]

The constant elasticity of substitution in (5′) is given by $\sigma = 1/(1 + \beta)$ which reduces to unity for the simple Cobb–Douglas function ($\beta = 0$) and to zero for the Leontief production or utility functions.

For production functions the interpretation is different, but the solutions (S_i) are the same. In their case the interpretation is as follows: Given a distribution of inputs x_{if}, S_i defines the optimal *technology* for a typical firm $i = 1, \ldots, N$. Coefficients such as α_f assign relative values to various inputs which in turn determine their productivity (relative values) in the outputs S_i. S_i are regarded as the full production frontiers. The information criterion measures "inefficiency," and the solutions here represent the most efficient technology. Alternatively, the solution S_i may be regarded as an aggregation of the miscellaneous inputs; see Bliss (1975, Ch. 7). Here our criterion amounts to the requirement that the aggregate's distribution among the firms should be "close" to the distribution of the miscellaneous inputs. As is clear from Bliss (1975, Ch. 7), traditional aggregation by functions which have similar properties to S_i of this paper require severe conditions on preference relations and/or production possibility sets. This is so, however, since the traditional methods sacrifice or ignore an important source of information: the composition (distribution) of the miscellaneous items to be aggregated. By focusing on the distributional aspects of allocations *and* by utilizing the available compositional "information," our criterion finds aggregates (indices) such as S_i without imposing the usually severe restrictions on preference relations, production possibility sets, or the indices themselves.

It is worthwhile to note that what we have done is "anonymous" with respect to time. It is possible to consider the same commodity, or sets of commodities, at different points in time. Also different values of β in (4) for different (groups of) individuals will produce group-specific utilities, technologies, aggregates, etc.

III. EVALUATING APPROXIMATIONS TO UNKNOWN REGRESSION FUNCTIONS

The minimum value of the criterion function $\Delta_{\beta,m}$ may be calculated and is estimable. Denoting the minimum of $\Delta_{\beta,m}$ by Δ_β^*, we find that

$$\Delta_\beta^* = (L^{-\beta} - 1)/\beta(\beta + 1), \tag{7}$$

$$\Delta_0^* = -\log S_0, S_0 = \sum_{t=1}^{N} \prod_k x_{tk}^{\alpha_k}, \tag{8}$$

$$\Delta_{-1}^* = \sum_t^N \sum_k^m \alpha_k x_{tk} \log \frac{x_{tk}}{\sum_k \alpha_k x_{tk}} \tag{9}$$

$$= \sum_{t=1} \sum_k \alpha_k x_{tk} \log x_{tk} - \sum_t S_t^* \log S_t^*$$

$$= \sum_k \alpha_k H(x_k) - H(S_t). \tag{10}$$

Here $H(Z)$ denotes the entropy of the distribution of Z, $L = \sum_{t=1}^{N} S_t$, with S_t defined in (6), and S_t^* is (6) evaluated at $\beta = -1$.

A number of approximations to the CES form have been studied by White (1980), Byron and Bera (1983), and Gallant (1981). Under certain restrictions, estimates of α_k and β may be found. Only Gallant's Fourier flexible form provides "essentially" consistent estimates of the elasticity parameters, while bounds on the inconsistency of the estimates have been given by White (1980) and further analyzed by Byron and Bera (1983) for the Cobb-Douglas and the translog approximations of CES.

With these estimates, Δ_β^* is estimated to give the minimum divergence value against which any approximation is to be assessed. Using (4) and if $g(x, \theta)$ approximates S_t, we may estimate $\Delta_{\beta,m}$ by replacing S_t with $g(x, \hat{\theta})$, and α_k and β with $\hat{\alpha}_k$ and $\hat{\beta}$ (parameter estimates are denoted by a caret). Denote this estimated value of $\Delta_{\beta,m}$ by $\Delta_{\hat{g}}$. Consequently, a possible criterion for the fit (efficiency) of the approximation $g(\cdot)$ is as follows:

$$\begin{aligned} \text{IC1} &= \Delta_\beta^*/\Delta_{\hat{g}} \\ &\leq 1 \end{aligned} \tag{11}$$

This is similar to the likelihood ratio criterion. The hypothesis underlying (11) concerns the functional forms, however, rather than parameter restrictions per se. In this sense it is akin to Cox's generalization of the likelihood ratio statistic for tests of nonnested hypotheses. Cox's test may be employed in the usual way to choose between S_t and $g(\cdot)$. IC1 is meant as an additional diagnostic criterion of goodness of fit or efficiency. Given that S_t is the maximum entropy function, IC1 indicates the additional degree of *non-sample* information implicitly utilized by predictions based on $g(\cdot)$. Values of IC1 close to unity reflect favorably on $g(\cdot)$.

IV. CONCLUDING REMARKS

It is interesting to note that concave functions of (4) or its generalizations may be regarded as components of welfare functions. This component of welfare is maximized when the divergence criterion is minimized. This is particularly meaningful if the allocation $\{X_{if}\}$ is regarded as optimal, for instance, in the sense of Pareto. In this situation the functional form which corresponds to $\Delta_{\min}$ is referred to as the optimal functional form.

The solutions obtained here are generally differentiable and, for some values of β, are also concave. Thus it would seem that our requirement of distributive representation need not restrict the class of preference relations beyond those discussed, for example, by Debreu (1972).

Under the assumption of utility maximizing agents, systems of demand equations are estimated which obtain the estimates of α_k, β, etc. This will determine the appropriate functional form as well as the maximum value of the associated welfare function.

Data on output of firms are directly available (unlike individual utilities). Estimation of fairly flexible production functions, taking cost minimization conditions into account, can also provide the "optimal" functional form. From these estimated coefficients, the maximum value of the corresponding welfare function (minimum of the divergence measure) may be obtained.

Since the divergence criteria employed here are additive (corresponding to additive welfare functions), we may allow different utility functions (technologies) for different groups of individuals (producers). This is done, for example, in (4) by specifying different values of β for different groups. In econometric estimation, firm (individual) specific coefficients may then be allowed, and if estimable, these coefficients may then be employed to provide estimates of the corresponding welfare index. Measures such as (4) are being considered by the author as indices of technical efficiency in the manner of Caves et al. (1982).

The paper has introduced a notion of optimization implicit in the choice of functional forms in economics. The main objective of the discussion has been one of interpretation and demonstration rather than to provide a necessarily normative view of "optimal" functional forms.

Since a lack of consensus on an appropriate criterion function is inevitable, the correspondence developed in this paper suggests that the functional forms employed in empirical analysis should be as flexible as possible. This is consistent with statistical principles which recommend decisions based on the data information when reliable prior information is lacking. The prior information (restrictions) implicit in the choice of any functional form is here interpreted as induced from the choice of objective (decision) criteria such as (4) and its generalizations.

Finally, we note that in the mathematical literature the S_i solutions discussed in this paper are referred to as generalized weighted means (or β-means). They contain the Box-Cox transformation which has become a popular flexible functional form in econometric literature. The Box-Cox transformation contains the linear and the Cobb-Douglas cases as it is a transformation of the general CES function that is particularly convenient for statistical estimation. See Box and Cox (1964).

ACKNOWLEDGMENTS

Thanks are due my colleagues Robert Becker and Fwu-Ranq Chang for valuable comments and suggestions. Skillful typing by Ms L. Steinwachs, Ms M. Welsh, and Ms A. Gilmore is most appreciated.

NOTES

1. The latter inevitably allows a more significant role for initial distribution (of endowments) than the former.

2. Maasoumi (1979) considered this generalization for (2). The measure in (1) can be obtained as a nonunique solution in an "axiomatic" derivation of suitable measures of distributional change. See Cowell (1982) for a most interesting discussion. Axioms of decomposability (by groups of individuals) and "independent homotheticity" are necessary for (1) to be a suitable measure. We note that any increasing function of (1) will be acceptable.

3. It may be the case that we wish to measure divergence of S_i from $x_{if}^{\delta_f}$ so that different goods contribute to S_i in different ways. Then the criterion in (4) is further generalized as follows:

$$\Delta_\beta^* = \frac{1}{m} \sum_f \frac{\alpha_f}{\beta(\beta+1)} \sum_i S_i \left[\left(\frac{S_i}{x_{if}^{\delta_f}} \right)^\beta - 1 \right].$$

The solution for the optimal distribution is a more general CES (Cobb-Douglas) functional:

$$S_i \propto \left[\frac{\beta+1}{m} \sum_f \alpha_f x_{if}^{-\delta_f \beta} \right]^{-1/\beta}.$$

REFERENCES

Atkinson, A. B. (1970). On the Measurement of Inequality, *Journal of Economic Theory* 2, 244-263.

Bliss, C. (1975). *Capital Theory and the Distribution of Income*, Amsterdam: North-Holland.

Box, G. E. P. and D. R. Cox (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B* 26, 211-252.

Burbea, J. and C. R. Rao (1982a). Entropy Differential Metric, Distance and Divergence Measures in Probability Space: A Unified Approach, *Journal of Multivariate Analysis* 12, 575-596.

Burbea, J. and C. R. Rao (1982b). On the Convexity of Some Divergence Measures Based on Entropy Functions, *IEEE Transactions on Information Theory* IT-28, 489-495.

Byron, R. and A. Bera (1983). Using Least Squares to Approximate Unknown Regression Functions: A Comment, *International Economic Review* 24, 255-260.

Caves, D. W., L. R. Christensen, and E. W. Diewert (1982). The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity, *Econometrica* 50, 1393–1414.

Cowell, F. A. (1980). Generalized Entropy and the Measurement of Distributional Change, *European Economic Review* 13, 147–159.

Cowell, F. A. (1982). Measures of Distributional Change. An Axiomatic Approach, Mimeo, The London School of Economics (April).

Debreu, G. (1964). Continuity Properties of Paretian Utility, *International Economic Review* 5, 285–293.

Debreu, G. (1972). Smooth Preferences, *Econometrica* 40, 603–615.

Fisher, R. A. (1925). Theory of Statistical Estimation, *Proceedings of Cambridge Philosophical Society* 22, 700–725.

Gallant, A. R. (1981). On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form, *Journal of Econometrics* 15, 211–246.

Jeffreys, H. (1961). *Theory of Probability*, Third Edition, London: Oxford University Press (especially section 3–10), first edition 1939.

Koopmans, T. (1972). Representation of Preference Orderings with Independent Components of Consumption. In McGuire, G., and R. Radner (eds.), *Decision and Organization*, Amsterdam: North-Holland.

Maasoumi, E. (1979). A Multivariate Index of Inequality Based on Information Theory, Mimeo, Los Angeles, CA: University of Southern California.

Maasoumi, E. (1986). The Measurement and Decomposition of Multi-Dimensional Inequality, *Econometrica* 54.

Maasoumi, E. and G. Nickelsburg (1983). Measuring Joint Inequality in Incomes, Wealth and Schooling and a Decomposition, Mimeo, Bloomington, IN: Indiana University.

Rao, C. R. (1949). On the Distance between Two Populations, *Sankhyā* 9, 246–248.

Shannon, C. E. (1948). A Mathematical Theory of Communication, *Bell System Technical Journal* 27, 379–423, 623–656.

Theil, H. (1967). *Economics and Information Theory*, Amsterdam: North-Holland.

Theil, H. (1979). The Measurement of Inter-Industrial Specialization in International Trade, *Economics Letters* 2, 377–379.

White, H. (1980). Using Least Squares to Approximate Unknown Regression Functions, *International Economic Review* 21, 149–170.