COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 582

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

CAKE SLICING AND REVEALED GOVERNMENT PREFERENCE

John Beggs and Samuel Strong

March 17, 1981

CAKE SLICING AND REVEALED GOVERNMENT PREFERENCE

by

John Beggs* and Samuel Strong**

1. Introduction

In earlier work McFadden [3,4] examined models for revealing an organization's decision making criteria by examining the decision maker's choice when confronted with discrete competing projects. The model envisaged decision makers undertaking an implicit benefit-cost comparison of alternatives, so that, after stochastic elements have been accounted for, there is some probability of selecting each one of the possible alternatives. The formal development of this approach led to the now-familiar multinomial logit choice model. McFadden illustrated the model with empirical evidence on the highway routing decisions of the California Division of Highways. In the same spirit Barton [2] has undertaken an investigation of the FCC's decisions in comparative broadcast licensing cases to assess whether FCC decisions are consistently related to the stated policy objectives of that Commission.

Though the contribution of these works both from a modelling and empirical point of view has been important, it is also true that economists do not usually have data available at the degree of disaggregation required by the above approach. Only rarely is it possible to gain

---

*John Beggs, Cowles Foundation for Research in Economics, Yale University.

**Samuel Strong, Department of Economics, University of Queensland.

sufficient access to the decision making process that data can be obtained on a case by case basis for specific competing projects. Most often the data available to economists on government decisions are in the form of *ex poste* expenditure share allocations. To witness, data on highway funding is more often available in the form of dollar allocations to states, counties and municipalities, rather than in the form of detailed evidence on individual competing road projects. Consequently there is a need for models which can explicitly utilize such share data.

The model employed here conceives of the problem as a "cake slicing" problem, that of dividing up a given bundle of funds among competing categories of use. The categories might be municipal districts, or age cohorts of the population, or educational attainment, etc. Funds are then shared among the categories according to the characteristics of each category. For example, Federal unemployment assistance might be given to states on the basis of the unemployment in the state, the per capita income of the state, the amount of unemployment which is among the young, etc. The formal requirements on the model are that the shares add up to unity and that the shares be valued between zero and one. This paper offers an appealing and computationally amenable cake slicing model, framed in the context of the economist's familiar utility concepts.

2. The Model

Some notation must first be introduced. Let s_t be a k-dimensional vector $(s_{1t}, s_{2t}, \ldots, s_{kt})$ containing the proportions of funds allocated to each of k categories in period t. Let X_{it} $(i = 1, 2, \ldots, k)$ be an m-dimensional vector of exogenous variables describing the characteristics of the i^{th} category in period t. The decision maker is assumed to divide the available funds in such a way as to maximize a utility function defined over the expenditure shares. Denote this function by $U(s_t; X_{it}; i = 1, 2, \ldots, k; \theta)$. The decision maker's problem is then to:

(1) Maximize: $U(\cdot)$

Subject to:

(2) $s_t'e = 1$ and

(3) $1 \geq s_{it} \geq 0$

where $e = (1, 1, \ldots, 1)'_{1\times k}$, and θ is an m-dimensional vector of unknown parameters describing the importance assigned to each characteristic.

An acceptable and widely used form of $U(\cdot)$ is the Cobb-Douglas utility function

$$U = \prod_{i=1}^{k} s_{it}^{\alpha_{it}} .$$

The weights α_{it} assigned to each category in time t are assumed to be determined by the characteristics of each category, contained in an m-dimensional vector in X_{it} . Problems (1)-(3) can be solved for the utility maximizing allocation, yielding the familiar solution,

(4) $$s_{it} = \frac{\alpha_{it}}{\sum_{j=1}^{k} \alpha_{jt}} .$$

This result satisfies the constraint that $1 \geq s_t > 0$ under the restriction that α_{it} be positive valued.

The category weights, α_{it} , contain information on both the exogenous time varying characteristics X_{it} , and the fixed parameter, θ , as well as a stochastic element. The stochastic element is included to account for either the presence of some random behavior by the decision maker, or the fact that the analyst does not have available all the information used by the decision maker.

A suitable form for α_{it} is

$$\alpha_{it} = \exp(X_{it}\theta + \varepsilon_{it}) \tag{5}$$

where ε_{it} is a stochastic disturbance term. This function has the desirable property that it is positive valued and is monotone in the category characteristic variable, X_{it} . The solution of the maximizing problem shown in (4) can now be written as

$$s_{it} = \frac{\exp(X_{it}\theta + \varepsilon_{it})}{\sum_{j=1}^{k} \exp(X_{jt}\theta + \varepsilon_{jt})} , \quad i = 1, 2, \ldots, k . \tag{6}$$

It can be seen that the monotonicity of the α_{it} function with respect to each category characteristic ensures that the category shares are also monotone with respect to each category characteristic, enabling straightforward interpretation of the θ parameter. Furthermore, the elasticities with respect to the various category characteristics can be easily derived from (6), enabling the relative significance of each characteristic to be examined. Two possible definitions of elasticity are offered in (7a) and (7b). The first, ξ_1 , corresponds to the standard definition of elasticity. The concept measured by ξ_2 gives the change, in percentage points, of funds going to the i^{th} category for a one percent change in the exogenous ℓ^{th} characteristic of the i^{th} category. From the point of view of interpreting the model ξ_2 may, in many instances, be more meaningful.

(7a) $$\xi_1 = \frac{\partial s_i}{\partial X_{i\ell}} \cdot \frac{X_i}{s_i} = \theta_\ell(1 - s_i)X_{i\ell}$$

(7b) $$\xi_2 = \frac{\partial s_i}{\partial X_{i\ell}} \cdot X_{i\ell} = \theta_\ell s_i(1 - s_i)X_{i\ell}$$

It is commonly the case in cake slicing situations that the category characteristics are themselves subject to adding up constraints. For example, consider a case where the categories are defined by State or Provincal boundaries and an exogenous characteristic is the geographical area of each state or province. In this situation, caution must be taken to ensure that a redefinition of any state boundary is accommodated by the redefinition of other state boundaries. That is, a direction derivative will be required in view of the fact that one is moving on the face of a simplex.

3. Estimation

Assuming the stochastic terms, ε_{it}, are distributed independently, both serially and across categories, the model (6) is then that of a k-dimensional random variable generating category shares on a k-1 dimensional unit simplex. The points on the simplex can be represented by the k-1 shares ratios, s_1/s_2, s_1/s_3, ..., s_1/s_k. The usefulness of this representation is seen in equation (8).

$$(8) \qquad \frac{s_{1t}}{s_{jt}} = \frac{\exp(X_{1t}\theta + \varepsilon_{1t})}{\exp(X_{jt}\theta + \varepsilon_{jt})} , \quad j = 2, 3, \ldots, k .$$

This is rewritten as,

$$(8a) \qquad \ln\left(\frac{s_{1t}}{s_{jt}}\right) = (X_{1t} - X_{jt})\theta + \eta_{jt} , \quad j = 2, 3, \ldots, k$$

where $\eta_{it} = \varepsilon_{1t} - \varepsilon_{it}$ is a zero mean random variable with variance covariance structure

$$(9) \qquad E(\eta_t \eta_t') = \sigma_\varepsilon^2 [I_{k-1} + ee'] = \Omega$$

where e is a (k-1)-dimensional vector of units. Since the variance structure is known the parameter estimates for θ can be obtained with the application of generalized least squares. The necessary inversion of the variance-covariance matrix, Ω , is given by $\frac{1}{\sigma_\varepsilon^2}[I_{k-1} - \frac{1}{k}ee']$. The m-dimensional parameter vector θ is identified when m is less than $k-2$, where k is the number of categories. When the cake slicing is observed on T occasions, the identification of θ requires m to be less than $T(k-1) - 1$.

When time series data are available it is also possible to estimate category specific effects, either by introducing a dummy variable or by reformulating the variance-covariance matrix to allow for a category specific error component. In this latter case, the model (8a) appears as

$$(10) \qquad \ln \frac{s_{it}}{s_{it}} = (X_{it} - X_{it})\theta + (\varepsilon_{it} - \varepsilon_{it}) + (\mu_1 - \mu_i) ,$$

$$i = 2, 3, \ldots, k ,$$

where the μ are assumed to be independent identically distributed $(0,\ \sigma_\mu^2)$. The variance-covariance matrix of the disturbance term in (10) is given by

$$\text{(11)} \qquad \Omega = I_T \otimes \sigma_\varepsilon^2 \cdot [I'_{k-1} + e_{k-1}e_{k-1}] + e_T e'_T \otimes \sigma_\mu^2 \cdot [I_{k-1} + e_{k-1}e'_{k-1}] \ .$$

With some algebraic manipulation it can be shown that

$$\text{(12)} \qquad \Omega^{-1} = I_T \otimes \frac{1}{\sigma_\varepsilon^2} \cdot A + e_T e'_T \otimes \left[\frac{-\sigma_\mu^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + T\sigma_\mu^2)} \right]$$

where $\qquad A = \left[I_{k-1} + \frac{e_{k-1}e'_{k-1}}{k} \right]$

which enables generalized least squares to be applied without resorting to the numerical inversion of the matrix Ω, which is of order $T(k-1)$. Since the variance components σ_ε^2 and σ_μ^2 are not known, a standard two-step estimation procedure can be employed, first using OLS and estimating the variance components, and then doing a second stage generalized least squares estimation of the parameter θ .

This topic is closely related to a large body of literature on the specification and estimation of systems of demand equations. This literature has typically paid attention to the "adding-up" restriction, that commodity shares add to unity, and has neglected the problem of restricting each share to be on the interval $(0,1)$. Indeed the requirement that the shares lie on the unit simplex creates special problems when stochastic features are added to a model. To see this, consider the model,

(13) $s_t = g(X_t, \theta) + \varepsilon$

where s_t is a vector of category shares, X_t is a vector of stochastic disturbance terms and $g(\cdot)$ is a non-linear vector-valued function in the category characteristics and the parameters. Even if the vector-valued function $g(\cdot)$ lies on the unit simplex, the category shares will only lie on the unit simplex if the stochastic disturbance terms satisfy special conditions, namely that

(14a) $-g(X_t; \theta) \leq \varepsilon \leq 1 - g(X_t; \theta)$

(14b) $t'e = 1$

where e is a k-dimensional vector of units. It is then seen that the distribution of the disturbance terms is not independent of the exogenous variables, and the usual estimation procedures are not applicable. In general, some normalization of the form shown in equation (4) will be required, along with a requirement that the category weights be positive valued (for all values of the exogenous variable and the stochastic disturbance term). This positivity is assured in equation (5) by raising all terms to the exponential. Even powered polynomials, arctans, selected piece-wise functions and integrals of positive valued functions will all satisfy this condition. These functions are usually either not monotone in the exogenous variable (polynomials) or are computationally burdensome to work with.

The parameter estimates obtained from the model are invariant to the chowce of category used for the normalization. The result is different to the well known theorem of Barten [1], namely that maximum likelihood estimates are invariant to the choice of commodity used for normalization. Barten's result applies to a system of equations where the disturbances are subject only to an adding up restriction, such as (14b). The invariance result is demonstrated in the Appendix.

4. Illustrative Example

An interesting public policy question concerns the criteria which are employed for the allocation of federal funds to the states for educational purposes within the United States.

It is hypothesized here that there are three main variables concerned: political representation, state per capita income, and state population changes.* To test this hypothesis, data on the variables were collected for 1976, and analyzed with model (8a). The equation was fitted using generalized least squares.

$$E_1 - E_i = \underset{(15.4)}{0.79}(R_1 - R_i) + \underset{(1.22)}{0.39}(Y_1 - Y_i) + \underset{(2.62)}{0.013}(G_1 - G_i) \tag{14}$$

t-values in parentheses, 46 degrees of freedom, $R^2 = 0.85$

In (14) E_i is the natural log of the share of education expenditure, R_i is the natural log of the number of state members in the House of Representatives, Y_i is the natural log of state per capita income, and G_i is population growth rate, where $i = 2, 3, \ldots, 50$ denotes states. Each variable was appropriately scaled.

Under the assumption of the model structure it seems that the primary determining factor in the allocation of education funds is political

---

*Education expenditure is defined as those expenditures contained under the heading "Federal Grants to State and Local Governments, by Purpose-States and Other Areas, 1976." Source: Statistical Abstract of the United States, 1977, U.S. Department of Commerce, Bureau of the Census, p. 321.

representation. This is hardly surprising since such representation will be almost perfectly colinear with population, and either variable would serve equally well in the equation. The population growth variable was included to capture the fact that growing states have larger capital works requirements and hence have a special need for funds. The empirical results suggest that this is a statistically significant consideration. The income variable was included to determine whether federal funds were of an "augmenting" or "replacing" nature. A negative sign on this coefficient would suggest that poorer states were receiving federal money to replace the deficiency of state funds. A positive sign would suggest that richer states receive more federal money, augmenting the already prosperous education system. The latter hypothesis is supported by the data but only at a low level of statistical significance.

5. Conclusion

The Cobb-Douglas utility function was shown to lead to a most convenient model for the estimation of the cake slicing criteria used by decision maker. The straightforward estimation procedure derived here commends the model for a wide range of economic applications, both with government and corporate decision models. Future research will probably be more econometric in nature and will be directed towards generalizing the function forms in the model while still retaining the logical restriction that the expenditure shares each be positive and less than unity, and that collectively they sum to unity.

REFERENCES

[1] Barten, A. P. (1969). "Maximum Likelihood Estimation of a Complete System of Demand Equations," European Economic Review, Vol. 1, No- 1, P7-73.

[2] Barton, M. F. (1979). "Conditional Logit Analysis of FCC Decision Making," Bell Journal of Economics, Vol. 10, No. 2, P399-411.

[3] McFadden, D. (1975). "The Revealed Preferences of a Government Bureaucracy: Theory," The Bell Journal of Economics, Vol. 6, No. 2.

[4] __________ (1976). "The Revealed Preferences of a Government Bureaucracy: Empirical Evidence," The Bell Journal of Economics, Vol. 7, No. 1, P55-72.

APPENDIX

Consider the model to be estimated, as written in equation (8a). Though (8a) shows normalization with respect to category one, we could alternatively normalize with respect to any of the other commodities. The model to be estimated can be written more generally with the aid of some further notation.

Let

$$y = \begin{bmatrix} \ln s_{1t} \\ \ln s_{2t} \\ \vdots \\ \ln s_{kt} \end{bmatrix}_{k \cdot 1} \qquad X = \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{bmatrix}_{k \cdot m} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}_{k \cdot 1}$$

Consider the writing equation (8a) where the normalization is with respect to the r^{th} commodity. Define the transformation matrix R of order $(k-1)k$. Let the elements of R be,

$$R_{ij}^{(r)} = \begin{cases} -1 & \text{for } j = r \text{ and all } i \\ 1 & \text{for } i = j \text{ except where } i = j = r \\ 1 & i = k-1 \text{ , } j = k \text{ except where } k = r \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1, 2, \ldots, (k-1) \text{ , } j = 1, 2, \ldots, k$$

The generalized least squares estimates of θ can then be written

$$\hat{\theta}_{GLS} = \theta + (X'R^{(r)'}\Omega^{-1}R^{(r)}X)^{-1} X'R^{(r)'}\Omega^{-1}R^{(r)}\varepsilon \text{ .}$$

It can readily be shown that the matrix $R^{(r)'}\Omega^{-1}R^{(r)}$ is of the form $\left[I_k - \frac{\ell_k \ell_k'}{k}\right]$ and hence is invariant to the choice of " r " the category used to normalize the equations. It follows that the $\hat{\theta}_{GLS}$ will be invariant to the normalization.

Before leaving this section some attention should be paid to one alternative specification of the basic model in equations (4), (5), and (6). Our model envisages a situation where the decision maker assigns weights to a category first by assigning a deterministic component (determined by the exogenous variable, X_{it} , and the parameters, θ) to each category and then assigning a stochastic component (ε_{it}) . A simple normalization would be to assume, at the outset, that deterministic components are assigned to each category weight. One category is then selected as the "base case" and its weighting factor is entirely deterministic;* the stochastic elements are added to each of the remaining (k-1) categories so as to adjust for unobserved factors, and this adjustment is "relative" to the base case category. This assumption removes the need for generalized least squares, but creates two further difficulties. First, any random behavior in the model is assigned entirely to the decision maker so that nothing is left for the possibility of a stochastic gap between the data available to the after-the-fact analysts (such as ourselves) and the data that was originally available to the decision maker. Second, it presumes knowledge of which category was in fact the "base case" employed by the decision makers. Arbitarily assigning this role to a category wh ch was not the best case will result in inefficient estimators since there will be terms in the variance-covariance matrix of the disturbances which are not being accounted for.

---

*The disturbance term on the normalizing category is equal to zero.