

Aggregation, Flexible Forms, and Estimation of Food Consumption Parameters*

CYRUS RAMEZANI

University of Wisconsin-Madison

DONALD ROSE

Economic Research Service, USDA

and

SUZANNE MURPHY

University of California-Berkeley

February 19, 1995

*An earlier versions of this paper was presented at the American Agricultural Economic Association's meetings (1991). We are indebted to Jim Chalfant for suggesting the demand model used in this paper and to Tom Cox for contributing valuable comments. Any remaining errors are our own. Financial support from the U.S. Department of Agriculture (CA #58-3198-7-169) is gratefully acknowledged. Comments should be directed to the senior author at 1300 Linden Dr., University of Wisconsin, Madison WI 53706.

Aggregation, Flexible Forms, and Estimation of Food Consumption Parameters

ABSTRACT

Grouping schemes, commodity aggregation and the choice of functional specification significantly influence the results of empirical demand studies. This paper aims to assess the importance of these factors for estimating aggregate food consumption parameters. We use recent cross-section data and a new model (Lewbel 1989) that offers functional flexibility and nests alternative specifications to estimate a food demand system for the United States. Foods are aggregated based on a new grouping scheme that is adopted from the widely used “Dietary Guidelines for Americans”. Monte Carlo methods are used to construct confidence intervals for the estimated elasticities. Nutrient intake elasticities with respect to food prices and expenditure are calculated. The influence of socioeconomic variables on consumption and nutrient intake is analyzed. Price, income, and demographic effects are found to be highly significant. Our findings are invariant to the choice of functional form.

Key Words: The Almost Ideal-Translog (AITL) demand system, Food Demand, Nutrient Elasticity, NFCS

Aggregation, Flexible Forms, and Estimation of Food Consumption Parameters

Improving diets and promoting good nutrition is an integral element of successful health care. The development of food and nutritional policies that are consistent with this aim requires detailed knowledge of the interaction between household socioeconomic characteristics, food prices, and food and nutrient choice. Such information is also of importance to food producers, health professionals, private and public educational organizations, and a host of social agencies (Institute of Medicine 1991, National Research Council 1989). Using recent U.S. data we study how socioeconomic characteristics, prices and expenditures influence food consumption and nutrient intake.

In the literature, there are two conceptual methods of linking nutrient intake to socioeconomic characteristics of individuals (Behrman & Deolalikar 1988, Senauer, Asp & Kinsey 1991, Morgan 1986). The ‘direct’ method correlates the intake of selected nutrients, or a composite measure of intake, to socioeconomic and demographic variables (Ramezani 1994, Adrian & Daniel 1976). Food choice and food prices are generally absent from this type of analysis. The ‘indirect’ approach uses a demand model to identify the determinants of food choice and, in turn, nutrient intake. Food choice is central to this approach providing implications concerning nutrient intake (Senauer et al. 1991, Rose 1992). Both methods have specific advantages and shortcomings. The latter seems more appropriate, particularly when price data are available. Concentrating on food choice, however, leads to difficulties that are standard to demand analysis. The most important concerns the selection of a demand model that is consistent with consumer theory. The choice of food groups is equally important, as aggregation mitigates the problems associated with estimation of a large demand system (Deaton & Muellbauer 1980*b*).

This study relies on the demand system approach. We utilize a recently developed model (Lewbel 1989) that is consistent with microeconomic theory and, additionally, offers a flexible functional form that nests competing alternatives. We consider food aggregation based on nutritional considerations and propose a grouping scheme constructed from the

“Basic Four” – dairy products, protein foods, breads and cereals, fruits and vegetables – which has been fundamental to nutrition education since the 1950s (USDA(1979). The parameters of the estimated food demand model are used to calculate price and expenditure elasticities for eighteen nutrients and food components. The influence of socioeconomic variables is also investigated.

Data and Food Grouping Procedure

The data for this analysis come from the household component of the 1987-88 Nationwide Food Consumption Survey (NFCS), which was conducted by the U.S. Department of Agriculture (USDA). The survey consisted of a probability sample taken from the contiguous United States and included 4,273 households. The food manager of each household was asked to report all foods consumed by the household – “eaten at home” – in the previous week, as well as the price of each food item. Additional data regarding household socioeconomic characteristics were also collected (Hama & Riddick 1988).

Aggregation of the large number of reported foods is essential in order to perform demand analysis. The aim of our proposed food grouping is to assess the nutritional impact of changes in food prices and expenditures. For this purpose the food groups must yield enough information on the nutrient content of foods to accurately represent the nutrient consumption (availability) of households. At the same time, the food grouping scheme needs to be concise enough to accommodate multi-equation demand analyses with nonlinearities.

Nutritionists have created broad food grouping schemes to be used as educational tools and for general policy formulation. Economists have commonly aggregated foods based on economic criteria and for the purpose of price and income analysis. Similarity has been the basis for aggregation in both nutrition and economics. Nutritionists base similarity on nutrient content, while economists use criteria such as marketing characteristics or the degree of complementarity among the food items (see Moschini et al. (1994) and

references therein for a discussion of economic methods such as separability procedures and cluster analysis for aggregating commodities). The aggregation scheme we propose is based on both nutritional and economic criteria. It was motivated by a recent study by Murphy et al. (1993), who compared a number of food grouping schemes, including marketing groups, to determine how accurately each reflected the nutritional content of individual diets. The aggregation scheme we use was among the categories studied and performed quite well relative to all other schemes.

The starting point for our proposed aggregation is the USDA's Daily Food Guide (USDA(1979), which is a widely known document, used by USDA and others for planning diets, as a nutrition education tool, and as the basis for menu planning in the current Dietary Guidelines for Americans (USDA 1990). Using these guidelines and the results of the study by Murphy et al. (1993) we aggregate all foods into six distinct groups. The six-group scheme is a modification of the traditional "Basic Four," which includes the original groups (dairy products, protein foods, breads and cereals, fruits and vegetables), as well as two added groups that are reflective of recent nutritional concerns. Fats and oils are separated into a group due to current nutritional advice to reduce the quantity of fat in the diet (USDA(1990). A sixth group, miscellaneous foods, is comprised of all other foods including, sugar and alcoholic beverages.

There are several advantages to this particular food-grouping scheme. First, use of the six aggregated groups reduces the total number of parameters in the model, thus making demand system estimation more manageable. This choice also avoids the estimation difficulties associated with corner solutions, which can result with overly disaggregated groups. Second, this grouping allows for the discussion of results, including demographic effects and the price and expenditure responsiveness of consumers, within the context of a commonly used nutrition education tool. This focus may be more useful to policy makers and nutrition program administrator.

The NFCS respondents reported quantities of foods as purchased in various units, such as quarts, which were all converted to pounds by the USDA. We summed the quantities for

each food group and then deflated by household size to provide the per capita quantity of food consumed in each food group for each household. Household size was defined in terms of 21-meal-equivalent persons which is the total adjusted meals eaten from household food supplies in the seven days previous to the survey, divided by 21. This corrects for differences in household size composition as well as the number of foods consumed at home versus foods outside the home (Smallwood & Blaylock 1984).

Each food group price is a weighted average of prices on specific items faced by the household. The variation in *food group* prices is due to differences in consumed items in each group and variation in prices of each item across households. The latter is due to quality differentials, seasonal effects, and regional market conditions. Using the 1977-78 NFCS data on specific commodities, Cox & Wohlgenant (1986) have shown that failing to control for quality differentials leads to small differences in parameter estimates provided that seasonality induces variation in prices. This is the case with the NFCS as data were collected during all seasons. The model we estimate also includes regional dummy variables which partly control for quality differentials and local market conditions. The remaining variation in food group prices is due to inter-group substitution which reflects relative prices and household preferences. Overall, the variation in food group prices (Table 1) is large, allowing us to obtain reasonable estimates of the aggregate price effects.

Six household demographic variables are used in the analysis. Four dichotomous variables indicate whether both a male and female head were present, whether household heads were of black or white race, and whether households were located in central cities, or non-metropolitan areas (suburban residents are the reference group). Two continuous variables measure the average years of education of the heads of the household and the household's income relative to federal poverty thresholds, which are based on family size. These variables are defined so as to facilitate comparisons with previous studies, which discuss the rationale for their selection (Senauer et al. 1991, Morgan 1986). The final sample used for estimation consists of 3845 household with non-missing observations for all variables. Table 1 provides definitions and statistics for all variables considered.

The Demand Model

Functional flexibility, appropriate aggregation properties, and ease of estimation are among the most desirable features of an empirical demand model (Deaton & Muellbauer 1980*b*). The Almost Ideal-Translog (AITL) model (Lewbel 1989) has a flexible functional form and is compatible with exact aggregation over consumers. Furthermore, its functional specification nests the Almost Ideal (AI) (Deaton & Muellbauer 1980*a*) and the Translog (TL) (Christensen, Jorgenson & Lau 1975) models, which have been widely used in applied literature.

This paper reports the results from the *cross-sectional* estimation of the AITL system. Our overall strategy is to obtain estimates of prices, income, and substitution elasticities for the six groups of commodities and to study the influence of functional form on the estimated elasticities. We also compare the precision of the elasticities estimated from each specification using Monte Carlo methods. Denote the vector of n food prices by p and total expenditures by E . The AITL expenditures share for i -th group, S_i , is:

$$S_i = \Delta^{-1} * \left\{ \alpha_i + \sum_{j=1}^n \gamma_{ij} \ln \left(\frac{p_i}{E} \right) + \beta_i [a(p) \ln(E) - \ln P] \right\} \quad (1)$$

where

$$a(p) = \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \ln(p_i) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \ln(E),$$

$$\Delta = \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \ln \left(\frac{p_i}{E} \right),$$

$$\ln P = \alpha_0 + \sum_{i=1}^n \alpha_i \ln p_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} \ln p_i \ln p_j.$$

and α_0, α_i and γ_{ij} s are parameters. The price (ε_{ij}) and expenditure (η_i) elasticity formulas associated with the AITL are:

$$\varepsilon_{ij} = -\kappa_{ij} + (S_i * \Delta)^{-1} * \left[\gamma_{ij} - \beta_i * \left(\alpha_i + \sum_{j=1}^n \gamma_{ij} \ln p_j \right) + \sum_{j=1}^n \gamma_{ij} * (\beta_i \ln(E) - S_i) \right]$$

$$\eta_i = 1 + (S_i * \Delta)^{-1} * \left[\sum_i \sum_j \gamma_{ij} S_i - \sum_j \gamma_{ij} + \beta_i [a(p) + 0.5 \sum_i \sum_j \gamma_{ij} \ln(E)] \right]$$

where $\kappa_{ij} = 1$ for $i = j$ and zero otherwise. Parametric restrictions can be used to test consumer theory and to obtain each nested model:

	TL	AI	AITL
	$\beta_i = 0$	$\sum_{i=1}^n \beta_i = 0$	$\sum_{i=1}^n \beta_i = 0$
Adding-up	$\sum_{i=1}^n \alpha_i = 1$	$\sum_{i=1}^n \alpha_i = 1$	$\sum_{i=1}^n \alpha_i = 1$
Symmetry	$\gamma_{ij} = \gamma_{ji}$	$\gamma_{ij} = \gamma_{ji}$	$\gamma_{ij} = \gamma_{ji}$
Homogeneity	$\sum_{i=1}^n \gamma_{ij} = 0$	$\sum_{i=1}^n \gamma_{ij} = 0$	$\sum_{i=1}^n \sum_{j=1}^n \gamma_{ij} = 0$

Homogeneity and symmetry restrictions are imposed throughout the estimation. This considerably reduces the number of parameters that must be estimated, simplifies the elasticity formulas, and ensures model convergence. Depending on the data and the number of demand equations, tests of demand restrictions may not be possible with highly nonlinear models such as AITL since model convergence may not be achieved.

Because the expenditure shares sum to one, only $n - 1$ shares equations are used for estimation. In this case, an iterated, seemingly unrelated regression method provides maximum likelihood estimates of the parameters and the results are invariant to the choice of the deleted equation (Barten 1969). Throughout the analysis, we delete the miscellaneous food group. Socioeconomic variables that account for differences in preferences across households are incorporated into each demand equation through the parameter α_i :

$$\alpha_i = \alpha_{0i} + \sum_{k=1}^K \delta_{ik} * D_k$$

where D_k is the k -th demographic variable and δ_{ik} measures its influence on the expenditure share of group i . Incorporating the demographic variables in this manner, rather than additively, ensures that symmetry is not violated. Note that the demographic variables also enter the elasticity formulas through the parameters α_i .

The price and income elasticity estimates in conjunction with the actual share in the total diet of specific nutrients are used to calculate nutrient price and nutrient expenditure elasticities. These estimates are based on the assumption that nutrients are obtained from foods via a fixed coefficient linear technology similar to that used in Heien & Wessels (1988). In particular, it is assumed that the total intake of nutrient i , N_i , is ‘produced’

via a production technology of the form $N_i = \sum_j n_{ji} q_j(p, E)$, where n_{ji} is the amount of nutrient i per unit of food j and q_j is the total consumption of food j . Differentiating this expression with respect to prices and expenditure we obtain nutrient price, $\phi_{in} = \sum_i s_{in} * \varepsilon_{ij}$, and expenditure, $\psi_n = \sum_i s_{in} * \eta_i$, elasticities where s_{in} is the share of nutrient n from food group i , and the values for s_{in} are taken from Murphy et al. (1993).

An important attribute of flexible demand models is that they place few restrictions on demand elasticities. Comparison of these models, however, has often been based on statistical tests of their fit rather than the statistical significance of the estimated elasticities. In fact researchers have often drawn the untested inference that statistically significant parameter estimates automatically lead to significant elasticity estimates. Determining the precision of elasticity estimates in the AITL type models is complicated because the elasticity formulas are nonlinear functions of the estimated parameters. Anderson & Thursby (1986) have shown that even for the TL specification, the finite sample distribution of elasticity estimates can not be derived analytically.

Three methods have been suggested to deal with this problem (Krinsky & Robb 1991). First, approximation techniques may be used to linearize these formulas. This approach will not provide a good approximation given the extent of nonlinearity in the AITL elasticity formulas. Second, Green, Hahn & Rocke (1987) suggested the bootstrap method for constructing confidence intervals. Although this suggestion offers a reasonable solution to the problem, in practice the bootstrap may not be computationally feasible; the AITL system is highly nonlinear, and depending on the number of equations and the specific data, convergence may not be achieved for each bootstrapped sample and computing costs may be significant. A third possibility is to generate a simulated distribution for the elasticities by drawing many samples from the asymptotic multivariate normal distribution of the parameters of the model. We implement this approach by sampling (N=5000) from the joint distribution of the estimated parameters for each model. For each draw, the elasticities are calculated at the mean predicted expenditure shares.

Results

The model was estimated using the iterative, nonlinear, seemingly unrelated regressions procedure implemented by the PROC SYNLIN module of *SAS*, which uses the Gauss-Newton method of optimization (SAS Institute Inc. 1990). Starting values for the parameters were obtained by first estimating the special cases of the AITL. The final model converged using a conservative convergence criteria (0.00001). To achieve convergence the parameter α_0 was fixed at zero. The likelihood function and the elasticity estimates were invariant to fixing this parameter at other values (± 1 , ± 10 , etc.). The estimated parameters, though highly significant, have no direct economic interpretation and therefore are not reported.

The likelihood ratio test is used to consider the statistical evidence in support of the nested alternatives. Let Θ and Θ^* be the vector of parameters for the AITL and each competing model, respectively. Then $LR = -2[\log L(\Theta^*) - \log L(\Theta)]$ has an asymptotic $\chi^2(k)$ distribution where $L(\cdot)$ is the likelihood function and k is the number of required restrictions to obtain the alternative model from AITL (Judge, Griffiths, Hill & Lee 1980). Examining this statistic we reject the TL relative to the AI and the AITL, and both the AI and the TL in favor of the AITL ($\alpha = 0.05$). Lewbel (1989), Bollino & Violi (1990), and Rose (1992) reported similar results favoring the more parametrized AITL.

The estimated elasticities for the AITL are reported in Tables 2. The reported standard errors for the elasticities are obtained from the Monte Carlo samples. The sign of price, expenditure, and substitution elasticities are consistent with the theory and their magnitude is within the expected range. The eigenvalues of the matrix of substitution elasticities indicate that this matrix is negative semidefinite suggesting that the estimated AITL model satisfies the curvature restriction implied by demand theory (Chalfant, Gray & White 1991). The mean and the range of the estimated elasticities from the alternative models are very similar to the AITL; the differences – mostly in the third decimal point – are of no economic significance. Thus, based on this data set we can conclude that the

estimated elasticities are essentially invariant to which specification is used.

Considering the precision of the elasticities, the AI performed better, since it had a much larger number of statistically significant elasticity estimates. This may lead one to select that the AI, particularly given the added programming effort required to estimate the more general AITL. As a referee has pointed out, however, selecting models based on the significance of elasticity estimates is ad hoc and inferior to the statistical procedures such as the likelihood ratio test, which, for this data set, indicates that the AITL is better than its nested alternatives.

The demographic effects are also reported in Table 2. Overall, the results are invariant to the choice of specification; education, household income relative to the poverty threshold, and race influence the expenditure share of most food groups significantly, while influences of urbanization are generally insignificant. It appears that an increase in the education of the household heads reduces the share of food expenditures on meat and other protein- rich foods, and on fats and oils, and increases expenditure shares for all other groups. Assuming that an increase in nutrition knowledge (which was not directly measured in this study) would be at least as effective as an increase in general education, these analyses suggest that nutrition education programs should be considered as a method for improving diets. Increases in a household's relative income also should improve diets by increasing consumption of fruits and vegetables.

The mean of nutrient price and expenditure elasticities for eighteen nutrients and food components are reported in Table 3. These elasticities are based on the AITL price and expenditure elasticities and the assumption that nutrients are obtained from food via a linear technology (the estimates from the AI and TL are almost identical). All elasticities with respect to food prices have negative sign, indicating that food price increases lead to lower intake of most nutrients. Overall, a change in the price of the protein foods has the largest impact on the consumption of most nutrients. Changes in the prices of fruits and vegetables has the greatest impact on intake of fiber, vitamin A and vitamin C. Calcium intake is most sensitive to changes in dairy prices. The nutrient expenditure elasticities

are generally around one, indicating that virtually all nutrients and food components are normal. Waterfield (1985) discusses the importance of nutrition elasticity estimates for the design of price and income policies that aim to improve the nutritional intake of the poor.

Summary and Conclusions

Using a flexible demand model, the influence of factors affecting food consumption and nutrient intake in the United States were analyzed. It is well known that the overall performance of any empirical demand model is influenced by the choice of food groups, the extent of commodity aggregation and the functional specification of the model. Our analysis attempted to control for these factors by combining recent advances in demand analysis and nutritional sciences. We used flexible and nested demand models to assess the importance of functional specification, Monte Carlo methods to obtain standard error for the estimated elasticities, a food grouping scheme that is relevant to current policy concerns, and provided nutrient elasticity estimates for the United States based on a consistent demand model.

The analysis of NFCS data suggests a strong link between socioeconomic variables, household food choices and nutrient intake. These results are based on a new food grouping scheme that builds upon widely used nutrition guidelines. Overall, the model estimated suggests that income policies may be more effective in influencing consumption patterns than price policies. Demographic variables, particularly education and household head status, significantly influence the consumption of foods with high nutritional value such as fruits and vegetables, as well as those that are less desirable such as fats and oils. Income policies in conjunction with nutrition education efforts appear to be the most promising tools for changing consumption behavior of those nutritionally at risk.

References

- Adrian, J. & Daniel, R. (1976), 'Impact of socioeconomic factors on consumption of selected food nutrients in the United States', *American Journal of Agricultural Economics* **58**(1), 31–38.
- Anderson, R. & Thursby, J. (1986), 'Confidence intervals for elasticity estimators in Translog models', *The Review of Economics and Statistics* **68**, 647–656.
- Barten, A. P. (1969), 'Maximum likelihood estimation of complete system of demand functions', *European Economic Review* **1**, 7–73.
- Behrman, J. & Deolalikar, A. (1988), Health and nutrition, in Chenery & Srinivasan, eds, 'Handbook of Development Economics', North Holland.
- Bollino, C. & Violi, R. (1990), 'Gaitl: A generalised version of the Almost Ideal and Translog demand system', *Economics Letters* **33**(2), 127–129.
- Chalfant, J. A., Gray, R. S. & White, K. J. (1991), 'Evaluating prior beliefs in a demand system: The case of meats demand in Canada', *American Journal of Agricultural Economics* **73**(2), 476–490.
- Christensen, L. R., Jorgenson, D. W. & Lau, L. J. (1975), 'Transcendental logarithmic utility functions', *American Economic Review* **65**(3), 367–383.
- Cox, T. L. & Wohlgenant, M. K. (1986), 'Prices and quality effects in cross-sectional demand analysis', *American Journal of Agricultural Economics* **68**, 908–910.
- Deaton, A. & Muellbauer, J. (1980a), 'An Almost Ideal Demand System', *American Economic Review* **70**(3), 312–326.
- Deaton, A. & Muellbauer, J. (1980b), *Economics and Consumer Behavior*, Cambridge University Press, New York.
- Green, R., Hahn, W. & Roche, D. (1987), 'Standard errors for elasticities: A comparison of bootstrap and asymptotic standard errors', *Journal of Business and Economic Statistics* **5**, 145–149.
- Hama, M. & Riddick, H. (1988), 'Nationwide food consumption survey, 1987', *Family Economic Review* **2**, 24–27.
- Heien, D. & Wessels, C. (1988), 'The nutritional impact of the dairy price support program', *Journal of Consumer Affairs* **22**(2), 201–219.
- Institute of Medicine (1991), *Improving America's Diet and Health: From recommendation to action*, National Academy Press, Washington, D.C.
- Judge, G. G., Griffiths, W. E., Hill, R. C. & Lee, T. C. (1980), *The Theory and Practice of Econometrics*, John Wiley and Sons, Inc.

- Krinsky, I. & Robb, A. (1991), 'Three methods for calculating the statistical properties of elasticities: A comparison', *Empirical Economics* **16**, 199–209.
- Lewbel, A. (1989), 'Nesting the AIDS and Translog demand systems', *International Economic Review* **30**, 349–356.
- Morgan, K. J. (1986), 'Socioeconomic factors affecting dietary status: An appraisal', *American Journal of Agricultural Economics* **68**(5), 1240–1246.
- Moschini, G., Moro, D. & Green, R. (1994), 'Maintaining and testing separability in demand systems', *American Journal of Agricultural Economics* **76** (1), 61–73.
- Murphy, S. P., Ramezani, C. A., Rose, D. & Hudes, M. (1993), *The effect of economic, demographic, and food choice variables on household nutritional adequacy in the United States*, USDA-HNIS Technical Report.
- National Research Council (1989), *Diet and Health: Implications for Reducing Chronic Disease Risk*, National Academy Press, Washington, D.C.
- Ramezani, C. A. (1994), 'Determinants of nutrient demand: A nonparametric analysis', Working Paper, *Dept. of Consumer Science, Univ. of Wisconsin-Madison*.
- Rose, D. (1992), Planning for nutrition in rural Mexico: A case study in household food consumption behavior, Ph.D. Dissertation, University of California, Berkeley.
- SAS Institute Inc. (1990), *SAS User's Guide: Economic Time Series, Version 6*, SAS Institute Inc., Cary, NC.
- Senauer, B., Asp, K. & Kinsey, J. (1991), *Food Trends and the Changing Consumer*, Eagan Press, St. Paul, Minnesota.
- Smallwood, D. M. & Blaylock, J. R. (1984), 'Scaling household nutrient data', *Agricultural Economic Research*.
- United States Department of Agriculture (1979), *Daily Food Guide*, Home and Garden Bulletin No. 288, Washington, D.C.
- United States Department of Agriculture (1990), *Nutrition and your health: Dietary Guidelines for Americans*, Government Printing Office, Washington, D.C.
- Waterfield, C. (1985), 'Disaggregating food consumption parameters', *Food Policy* **2**, 337–350.