A Microanalytic Simulation Procedure for Market Demand Estimation

by

Chung L. Huang, Robert Raunikar, and Holly L. Tyan*

Selected paper presented at the AAEA annual meeting at Cornell University, August 5-8, 1984.

*Authors are Associate Professor, Professor, and Research Coordinator of Agricultural Economics, respectively, University of Georgia, Georgia Experiment Station, Experiment, Georgia 30212.

A Microanalytic Simulation Procedure for Market Demand Estimation

Conventional cross-sectional analyses of household demand for food focused on estimating behavioral relationships between food expenditure (consumption) of a consumer unit and its socioeconomic and demographic characteristics (Prais and Houthakker; Salathe; Haidacher, et al.). Results of these analyses provide investigators with knowledge of the underlying behavioral patterns and a means for assessing the impact of changes in socioeconomic variables on consumer's demand for food. Unfortunately, they are too vague and general to be readily useable to any segment of the food industry. For example, the decision-makers in the food industry are concerned with specific activities; e.g., the development of a potential market, evaluation of the relative position of one market versus another, and the selection of potential locations for production, processing and distribution of their products.

Previous research by Raunikar, et al. contributed to bridging the gap between the empirical demand relationship and the information sought by the decision-makers in the food industry. A procedure was developed to obtain estimates of aggregate market demand among spatially delineated geographic areas based on demand parameters estimated from the Atlanta Consumer Panel. Their studies were widely recognized for their contributions to fill the void of information; however, several limitations for the application of the procedure exist. Blakley suggests the assumption that the results for a particular market segment, i.e. Atlanta, can be applied directly to a wide geographic area, i.e. other markets in the United States, may not be valid.

Projections based on the average population characteristics of a market area also tends to overlook the distributional effects of the population characteristics within the market area.

The objective of the study is to develop a computerized microanalytic simulation model, based on empirically verifiable behavioral relationships, for estimating market requirements among spatially delineated market areas. Results of the model are used as a basis for demand projections under specified distributions of population characteristics. More importantly, the emphasis is focused on testing the reliability and validity of the model.

An Overview

Amstutz (pp. 112-113) defines a microanalytic simulation model as a model that "provides an integrated statement of that which is known and assumed about actions, reactions, and responses within the environment being simulated." He also distinguishes several types of simulation models. The model to be developed in this study is simplistic, econometric and static. It is simple in the sense that the present model does not represent a complete system nor does it simulate the decision-making behaviors of the consumers in the market place. The model is econometric in the context that it uses an econometric model as the underlying building block which describes the functional relationship between a set of observable variables and from which predictions of the probable outcome of alternative representations under specified conditions are examined. The allusion to 'predictions of probable outcomes...' indicates that the model is stochastic, and is used to generate outputs in response to alternative inputs. The model is static because it describes and represents the occurence of an event through a single change of state rather than over time.

The first step toward satisfying the objective involves quantifying a postulated behavioral relationship which provides an overall description and generalization of the salient and important attributes within the simulated environment.

The second step encompasses the process of data generation wherein input requirements are differentiated between parameter values and simulated exogenous variable values. The parameter values are obtained by implementing the first step. Other inputs to be used in the simulation process are generated via a random number generation function.

The final step involves discussions on the reliability and validity of the model results. While reliability may be assessed using normal statistical techniques, there are no objective measures for validity. Amstutz (p. 379) cautions that "rigorus analysis and precise mathematical representation convey an aura of validity.... Assumptions embodied in authoritative equations and conclusions summarized in computer-prepared graphs often go unchallenged on the premise that 'the scientific method' is predestined to reveal profound truths."

Model Specification and Procedure

To implement the first step of the proposed simulation process, a model of consumer demand for a specific food product is postulated. The model follows the conventional demand specification based on cross-sectional analysis (Haidacher, et al.; Prais and Houthakker; Raunikar, et al.; Salathe). Household demand for two types of broiler meat, whole and parts, are used to illustrate the estimation procedure. Specifically,

(1) $Y_{ij} = X_i\beta_j + e_{ij}$ $\qquad i = 1,2,\ldots n$, $\; j = 1,2$

where Y_{ij} is quantity of j^{th} broiler product consumed by the i^{th} household. X_i represents a set of socioeconomic characteristics associated with the i^{th} household, such as income, race, urbanization, age-sex composition of the household members; and e_{ij} is a normally distributed random disturbance.

To obtain the parameter estimates for equation (1), the 1977-78 USDA Nationwide Food Consumption Survey (NFCS) data are utilized. A sample of 10,964 households for which complete data were reported was selected from approximately 15,000 NFCS households.

The Tobit regression procedure is used to estimate the parameter values of equation (1). The procedure, pioneered by James Tobin, addressed the problem of censored normal distribution of the dependent variable in the regression context. Amemiya shows that the expected value of Y under the condition of censored normal distribution is no longer equal to $X\beta$. The unconditional expected value, E(Y), according to Amemiya is

(2) $E(Y) = X\beta\ F(z) + \sigma\ f(z)$

Where $z = X\beta/\sigma$, σ is the standard error of estimates from regression, and f(z) and F(z) are the unit normal density function and cumulative normal distribution, respectively. An important property of equation (2) is that the Tobit analysis simultaneously accounts for the level of consumption and the probability of such a consumption occurring.

The next step uses equation (2) to predict average household consumption of a particular broiler meat. Parameters estimated from equation (1) are used as the inputs for this phase of the procedure. The underlying assumption is that equation (1) conveys the empirical generalization of households consumption behaviorial relationships with regard to the impacts of various socio-economic characteristics of the sample households. Hence, the resulting regression model states that household consumption of a particular broiler meat is identical for households with the same characteristics.

Assuming that the population characteristics within a defined market area are known, the procedure is then to simulate a household unit with certain characteristics according to the distributions of the population characteristics within that market area. For example, the age-sex composition of a simulated household is generated in two steps. The procedure is carried out by first determining the size of the household according to the probability distribution of household size within the market area, and then determining the compositions of the household.

The simulation process generates a large number of households of differing types which are representative of a sample population. Specifically, the simulated results become a weighted average of the consumption levels of the household types in the market area, the weighting being proportional to the numbers of such households in the studied market area.

The final aspect of the study focuses on testing the reliability and validity of the simulated results by procedures which are to a large extent equivalent to conducting sensitivity testing. Given that the estimation procedure is based on the total U.S. sample, one would then be interested in examining how the model performs when used to predict broiler meat consumption on subdivided geographic market areas.

The performance of the model is evaluated by simulating household consumption of whole broilers and broiler parts for regional markets, i.e., Northeast, North Central, South and West. For each region, sample sizes of 100, 500, and 1,000 households are generated at a single run for each broiler product. The sample statistics which provide a basis for evaluating the statistical accuracy of the simulation output are computed at each run. The reliability of the model's performance is assessed by examining whether successive replications of a given system will produce results within acceptable limits. Therefore, a sequential estimation procedure as described in Fishman is employed.

Fishman (p. 69) defines the stopping rule for determining the number of replications run in a sequential estimation as

(3) $$k^* = \min\,[k: S_k^2\,(Y) \leq kd^2/t_{k-1,\alpha}^2]$$

and

$$S_k^2(Y) = \left(\sum_{i=1}^{k} (Y_i - \overline{Y}_k)^2\right)/k-1, \quad \overline{Y}_k = \left(\sum_{i=1}^{k} Y_i\right)/k$$

where k is the number of replications, d is a user-specified quantity for a tolerance level, and $t_{k-1,\alpha}$ corresponds to the $1-\alpha/2$ quantile of the Student t distribution with k-1 degree of freedom. Essentially, the procedure is to collect one observation Y_i at each run with the objective of obtaining a result with a given accuracy. Specifically, if k* is determined by equation (3), then it is suggested that

$$\text{Prob}\,(\overline{Y}_{k*} - d \leq \mu \leq \overline{Y}_{k*} + d) \geq 1 - \alpha$$

where μ is the true population mean. The sequential estimation not only provides evidence of reliability but also provides a means for obtaining the estimate at a prescribed level of accuracy.

Finally, the experiments are replicated by using different seedings in the simulation process. These results form the additional basis for evaluating the stability and consistency of the model's performance.

Results

For brevity, the regression results of equation (1) on household consumption of whole broilers and broiler parts based on the 1977-78 NFCS data are not presented. In general, most household characteristics included in the model exhibit significant impacts on broiler meats consumption. The magnitudes of the coefficients associated with the age-sex composition variables indicate that greater impacts are exerted by adults than children and by male than female members of the household unit. Results suggest a distinctive income effect on consumption of whole broilers and broiler parts with a positive and a negative income effect, respectively. Similar finding were reported by Salathe using the 1972-73 BLS Consumer Expenditure Diary Survey and by Haidacher, et al. using the USDA 1977-78 NFCS data.

The simulation results are summarized and presented in Tables 1 and 2, and Figure 1. Table 1 provides an example of a single run results on simulating the household characteristics for the U.S. sample. The simulated results are fairly close to the actual sample. As the number of households being simulated increases, further improvement in the accuracy of the results in approximating the actual sample would be expected.

Figure 1 depicts the results of predicting the average U.S. household consumption of broiler meats based on simulating 100 households with 100 replications. The simulation process was quite successful in reproducing the results with successive replications, i.e., approximating the observed

Table 1. Sample Statistics and Simulated Results of Household Socio-economic Characteristics, U.S.

Characteristics	Sample[a]			Simulated[b]		
	Mean	Min	Max	Mean	Min	Max
Northeast region	0.25	0	1	0.25	0	1
North Central region	0.24	0	1	0.20	0	1
West region	0.17	0	1	0.16	0	1
Black	0.12	0	1	0.07	0	1
Rural	0.28	0	1	0.24	0	1
Log (income)	9.27	5.70	12	9.32	6.96	10.78
Household size	2.94	1	15	2.87	1	14
Male adult $\geq$ 35 years	0.54	0	4	0.50	0	2
Female adult $\geq$ 35 years	0.64	0	4	0.60	0	3
19 $\leq$ male adult $\leq$ 34 years	0.34	0	4	0.33	0	3
19 $\leq$ female adult $\leq$ 34 years	0.39	0	4	0.47	0	3
13 $\leq$ male $\leq$ 18 years	0.19	0	4	0.15	0	2
13 $\leq$ female $\leq$ 18 years	0.18	0	5	0.23	0	5
6 $\leq$ child $\leq$ 12 years	0.38	0	6	0.32	0	5
Child $\leq$ 5 years	0.28	0	5	0.27	0	3

[a]Compiled from the 1977-78 USDA Nationwide Food Consumption Survey.

[b]Based on a single run with 100 simulated households.

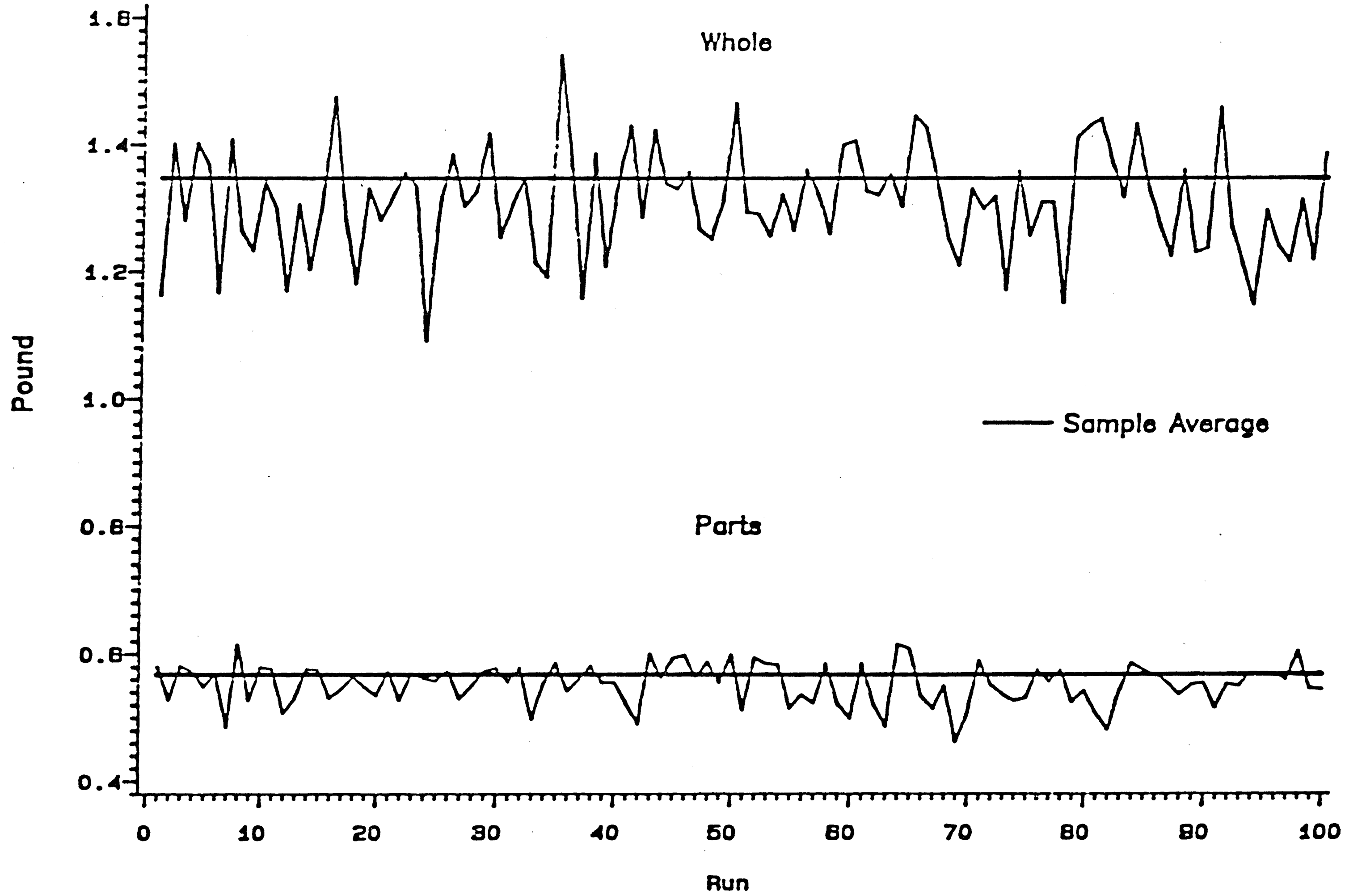

Figure 1. Simulation of U.S. Average Broiler Meats Consumption per Household per week, NFCS 1977-78.

sample average within a narrow band of variations. The replications seem quite stable and consistent with no indications of any apparent trend of departure from the sample averages.

The results of predicting the regional broiler meats consumption are summarized in Table 2. Regional consumption was simulated by applying the regional socioeconomic and demographic characteristics distributions as inputs in the model. The results were obtained by the sequential simulation procedure. The simulation was specified to estimate the average broiler meats consumption to within ±0.25 pounds with a 95 percent probability level. Applying the stopping rule of equation (3), the results indicate that in most cases only two replications are required. In addition, the resulting estimates are very similar regardless of the use of different random seedings in the simulation process.

Estimates of regional average broiler meats consumption based on conventional approach are also presented in Table 2 for purpose of comparison. These results are computed by substituting the average household characteristics of a regional market into the regression equation. As shown in Table 2, this approach consistently under estimates the magnitudes of broiler meats consumption within each regional market. Furthermore, the results tend to predict incorrectly the relative magnitude of whole broilers to broiler parts consumption in the Northeast regional market.

Conclusion

This study proposes a microanalytic simulation process for estimating market demand among spatially delineated market areas. The major advantage of the proposed procedure is that by simulating the household characteristics, population distribution patterns that characterize a specific market

Table 2. Sample Average and Simulated Regional Average Broiler Meats Consumption, Pounds Per Household, Per Week[a]

Region/ broiler type	Sample[b]	Experiment 1[c] Simulation 1	Experiment 1[c] Simulation 2	Experiment 2 Simulation 1	Experiment 2 Simulation 2	Non-simulated
Northeast						
Whole broilers	1.172 (2.209)[d]	1.053	1.063	1.065	1.082	0.920
Broiler parts	0.951 (1.802)	0.960	0.954	0.950	0.958	0.929
North Central						
Whole broilers	1.193 (2.01)	1.175	1.163	1.167	1.172	1.045
Broiler parts	0.396 (1.19)	0.350	0.359	0.359	0.359	0.335
South						
Whole broilers	1.713 (2.401)	1.704	1.732	1.698	1.700	1.573
Broiler parts	0.473 (1.270)	0.453	0.452	0.459	0.460	0.427
West						
Whole broilers	1.086 (1.86)	1.065	1.057	1.083	1.088	0.961
Broiler parts	0.450 (1.14)	0.461	0.458	0.462	0.459	0.441

[a]Simulation results are based on sequential estimations with d=.25 and α=.05.

[b]Compiled from the 1977-78 USDA Nationwide Food Consumption Survey.

[c]Experiments 1 and 2 simulate 1000 and 500 households per single run, respectively. Simulations 1 and 2 differ only with respect to the use of random number seedings.

[d]Numbers in the parentheses are standard deviations.

area can be incorporated in the process for obtaining market demand estimates. Another advantage of the procedure is that more accurate estimates can be obtained through the use of sequential estimation. Predictions obtained directly from the regression equation based on cross-sectional data are usually associated with a large standard error of estimates and hence, a large confidence interval which is often more than desirable.

U.S. household consumption of whole broilers and broiler parts are used to illustrate the estimation procedure employed. Based on the parameters obtained from the U.S. sample, regional consumption of broiler meats are estimated to form a basis for evaluating the viability and performance of the proposed simulation procedure. The study suggests that, in every case, the simulation results outperform the conventional estimates.

References

Amemiya, T. "Regression Analysis when the Dependent Variable is Truncated Normal." Econometrica 41(1973):997-1016.

Amstutz, A. E. Computer Simulation of Competitive Market Response. Massachusetts: The M.I.T. Press, 1970.

Blakley, L. V. "Regional Markets for Agricultural and Food Products: Needed Research." S. J. Agr. Econ. 13(1981):69-78.

Fishman, G. S. Principles of Discrete Event Simulation. New York: John Wiley and Sons, Inc., 1978.

Haidacher, R. C., J. A. Craven, K. S. Huang, D. M. Smallwood, and J. R. Blaylock. Consumer Demand for Red Meats, Poultry, and Fish. USDA, ERS, NED, ERS Staff Rept. No. AGES 820818, September 1982.

Prais, S. J., and H. S. Houthakker. The Analysis of Family Budgets. Cambridge: Cambridge University Press, 1955.

Raunikar, R., J. C. Purcell, and J. C. Elrod. Spatial and Temporal Aspects of the Demand for Food in the United States. IV. Poultry. Georgia Agr. Exp. Sta. Res. Bul. 87, May 1971.

Salathe, L. E. Household Expenditure Patterns in the United States. USDA, ESCS, Tech Bul. No. 1603, April 1979.

Tobin, J. "Estimation of Relationships for Limited Dependent Variables." Econometrica 26(1959):24-36.