1975

AGRICULTURAL ECONOMICS

department of agricultural economics
virginia polytechnic institute and state university
blacksburg, virginia 24061

Staff Paper SP-75-6   August 1975

THE ESTIMATION OF PRICE ELASTICITIES FROM PANEL DATA

by

William T. Boehm

and

Joseph Havlicek, Jr.

Department of Agricultural Economics
Virginia Polytechnic Institute and State University

Paper presented at the American Agricultural Economics Association Annual Meetings August 11, 1975 Columbus, Ohio

THE ESTIMATION OF PRICE ELASTICITIES
FROM PANEL DATA

Information which indicates the probable consumer response to changes in retail price is often sought for policy purposes as well as for firm decision making. Such information is usually communicated in the form of estimated price elasticities which indicate the expected relative change in quantity demanded associated with a relative change in price. While point estimates are usually reported by researchers, those using such information for decision making often, at least implicitly, assume some distribution for the estimate. Often, neither the method employed in obtaining the estimate nor the source of the price/quantity data is questioned. Such subtleties are usually left to the discretion of the analyst.

The purpose of this paper is to call attention to the important role that the statistical model and the data source play in the determination of a numerical estimate of price response. Price "elasticities" are estimated from both time-series and cross-sectional viewpoints using the same data base and essentially the same method of estimation, ordinary least squares regression (OLS). In each case the data are organized in a slightly different manner. The results obtained are reasonable from a statistical point of view, and yet each set suggests vastly different policy implications.

The focus of the paper in the final section is on the estimation of response parameters from continuing micro-unit data systems. It is argued that if the rate of adjustment from the short run to the fully adjusted long run response can be determined, the cross-sectional estimate can be used for answering policy questions which require an estimate of both long and short-run impacts. Methods of estimation proposed in earlier papers by Kuh, Balestra and Nerlove, and Wallace and Hussain to estimate models of cross-sectional data over time area discussed.

Data and Research Methods

Individual purchase records from the approximately 7500 households of the Market Research Corporation of America (MRCA) National Consumer Panel (NCP) provide the data base for this research.1/ In total more than 1.6 million individual purchase records for various dairy products were available. These records included information about the specific product type, product price, quantity purchased, size and type of container, and other attributes of the purchase. Information regarding the demographic characteristics and the socio-economic status of each NCP household was also available.

Three statistical models were developed to obtain estimates of the consumer response to retail price changes. The models included a cross-sectional model, a time-series model and a combined model which incorporated an estimate of the income effect from the cross-sectional model.

The price response estimate from the cross-sectional model was expected to approximate the long-run response to price changes. In a cross-sectional sample disequilibrium effects are expected to be synchronized to common market forces so that the estimated coefficient will typically show a higher response than the partially adjusted response from a time-series model (Kuh, p. 208). The cross-sectional model is as follows:

Q = f (P, DV, HDV, ED, OCC, R, HES, CS, HC, INC)

where:

Q = Quantity purchased in half-gallon equivalents by a consuming panel household during the 90 week time period under study,

P = the average price paid by a consuming panel household for a half-gallon equivalent,

DV = the percent volume purchased through special deal2/,

HDV = the percent volume purchased through home delivery,

ED = educational level of the household head, by categories of years of schooling completed,

OCC = occupational status of the household head, by employment categories,

R = race, either Caucasian or other,

HES = housewife employment status, either employed or unemployed,

CS = size of city of residence by size category,

HC = household age/sex composition, by numbers in age/sex categories,

INC = annual household income in current dollars.

Educational level of the household, occupation of the household head, race, employment status of the housewife and city size were all entered into the regression as sets of zero-one variables. The age/sex composition of the household was specified by including as variables the actual number of members in each of nine age/sex classifications. A second order polynomial term was specified for the income variable in order to permit the identification of maximum levels of household purchases as incomes increased.

The second model was a time-series model in which regional per capita consumption figures were specified as the dependent variable. Observations were generated by aggregating all purchases by NCP households within each MRCA geographic region for each two week time period. This total purchase figure was then divided by the panel population figures to obtain a per capita consumption rate. It was expected that the results of this model would yield price response estimates of a short-run nature and would be comparable to results obtained using aggregate market data. The general form of this model is

$$Q = f\,(P_0;\ P_1, \ldots,\ P_n;\ P_{nd};\ DV;\ HDV;\ S;\ RGN;)$$

where:

Q = the per capita quantity in half-gallon equivalents purchased by MRCA panel households per each two-week period,

P_0 = the average price paid by panel households for half-gallon equivalent units,

$P_1, \ldots P_n$ = the average price paid by panel households for n substitute and/or complement dairy products,[3/]

P_{nd} = the index of prices paid for all foods,

DV = the percent volume purchased through special deals,

HDV = the percent volume purchased through home delivery,

S = the season of the year during which the purchased were made, and

RGN = the MRCA region of the county: Northeast, South, North Central, Mountain and Southwest or Pacific.

The index of all prices paid for food was the food item component of the consumer price index (CPI). Region and season of the year were both specified as sets of zero-one variables.

The final model was a "mixed" regression model. Information about the income response parameter obtained from the cross-sectional model was incorporated in re-estimating the time-series parameters. The income response parameter was introduced into the time-series regression model as if known with certainty. The estimates obtained are thus conditional upon the validity of that estimate. Many authors have suggested methods which relax the "known with certainty" assumption (Chetty, Durbin, Theil). From a theoretical viewpoint, at least, it would appear that such methods are a noteworthy improvement.

The present example is somewhat unique in that the same data base is used to obtain a cross-sectional estimate of the income response parameter and then use that estimate to adjust the dependent variable in a time-series model. In that sense, the method may legitimately be thought of as a method of pooling information from a time-series of cross-sections. The idea is as follows. In the cross-section sample, price variability only results from price dispersion in the market. Consumer reaction to changes in price cannot be measured in the cross-section sample. Price changes can only occur over time. On the other hand, it is operationally quite difficult to obtain a measure of income uncorrelated with prices in aggregate time-series data. Further, an estimate of "regional" income from one point in time generally does not permit changes in the distribution of income. There is also some question as to the relationship between such a measure and the theoretical construct of a budget constraint. The cross-sectional sample of households, conceptually at least, should provide the "better" income response estimate. Once this estimate of the income response is obtained it is substituted into the structural demand equation and estimates of the price responses are then obtained from the time-series model.

When the cross-sectional estimate of the income response is substituted into the time-series model it is important to make the models compatable. Cross-sectional models are estimated using household consumption or expenditure figures at one point in time as the dependent variable. Often it is necessary to include only purchasing households when estimating parameters in this type of a model. On the other hand, time-series models often use aggregate per capita quantities as the dependent variable. In the present paper the adjustment process to achieve compatibility is as follows.

Let:

Q = the per capita quantity consumed in half-gallon equivalents during two-week time periods,

N = the total number of persons in panel households in each MRCA region,

I = the average annual income of all panel households, in current dollars, adjusted for monthly changes in U. S. disposable income,

P = the percent of all panel households consuming the product during the two-week time period,

A = the average number of persons per panel household in the region,

$\hat{b}_1$ = the cross-sectional estimate of the linear component of income response, and

$\hat{b}_2$ = the cross-sectional estimate of the quadratic component of income response.

Thus, the average household consumption by consuming households during the entire 45 two-week time period, Q^*, may be represented as

$$Q^* = (Q/P) \cdot (A) \cdot (45).$$

The adjusted household consumption, Q^{**}, is

$$Q^{**} = (Q^* - \hat{b}_1 I - \hat{b}_2 I^2)$$

Finally, the adjusted two-week per capita consumption of the product being considered is

$$\tilde{Q} = Q^{**}/(A) \cdot (45).$$

The vector Q is then used as the dependent variable in the time-series regression. The model in its general form is

$$\tilde{Q} = f\ (P_0;\ P_1, \ldots P_{nj} P_{nd};\ DV;\ HDV;\ S;\ RGN)$$

where the variables are the ones previously defined.

Results

Parameter estimates from the three statistical models were obtained for five fluid milk products. The products included regular whole milk, one percent milk, two percent milk, buttermilk and a composite product, total fluid milk. Since 11 models were linear in the variates, elasticity estimates were calculated at the mean values of price and quantity.

Estimates of the direct price elasticities obtained from the three alternative model specifications are presented in Table 1. It should be noted that the variation accounted for by each of the equations for all three models was statistically significant. F tests were all significant at the .001 level or better and R^2 ranged from 85-99 percent for the time-series and combined models. There was, however, some evidence of serial correlation in the time-series model. By introducing the effects of the previously excluded income effect into the combined model, the severity of the serial correlation problem was reduced.

In every case except for one percent milk and buttermilk the cross-sectional estimate is more elastic than the time-series estimate. This result tends to add support to the contention that estimates from a cross-sectional study should be thought of as long-run responses. However, these approximate long-run responses should be interpreted with care. For total fluid milk, for example, they indicate simply that those <u>households</u> consuming milk and paying 10 percent more than the mean price consumed on the average 16 percent less milk. This is a quite different implication than if these estimates were interpreted as short-run aggregate market responses. The relative magnitude of the cross-sectional estimates is also revealing.

Table 1. A Comparison of Price Elasticity Estimates from Household Panel Data for Five Fluid Milk Products, United States, 1972-73.

Product	Statistical Model		
	Cross-Section	Time Series	Combined
1. Regular Whole Milk	-1.7008*	-0.3795*	-0.2058
2. One Percent Milk	-.8334*	-1.1706*	-3.4220*
3. Two Percent Milk	-1.3279	-0.5484*	-0.7030*
4. Buttermilk	-1.5191	-1.7756*	-3.5146*
5. Total Fluid Milk[1/]	-1.6282*	-0.1767*	-0.3173*

*Indicates that the estimated coefficient was statistically significant at the 10 percent level or better. All estimates were calculated at the mean of the data.

[1/] Includes the consumption of regular whole milk, chocolate milk, one percent, skim milk, two percent and buttermilk.

The lowest price per unit product, one percent milk, has the lowest long-run elasticity. This result supports the contention that purchase responses to price changes are perhaps more correctly thought of in a discrete sense rather than as the continuous process of adjustment suggested by the theory. Once households have made the decision to purchase the lowest price per unit alternative, consumption rates for that product are more stable than for the higher priced per unit product.

The short-run estimates obtained from the time-series model are quite consistent with a priori expectations and previous studies. The price response is highest for buttermilk and lowest for the composite product, total fluid milk. Given these results one would expect a one percent increase in the weighted average price of total fluid milk to yield by a 0.17 percent reduction in per capita consumption during a two-week time period.

Results obtained from the combined model also appear "reasonable" in a relative sense. The magnitude of the estimates, if presented alone, would find a rather wide acceptance. In those cases where the income response was positive and neglected as a variable it was expected that the time-series price response would be underestimated. This appears to have been the case for one percent, two percent and total fluid milk. A priori we expected that the unadjusted time-series response for buttermilk was an overestimate since the income response estimate was negative. This hypothesis was, however, not supported by the data. The fact that the low income Southern region has a rather high per capita consumption rate for the product may be part of the explanation. The magnitude of the response for one percent milk (-3.42) was also quite surprising.

When the magnitudes of these various estimates are compared, the policy implications of the results become quite clear. While consumers appear to be somewhat passive to price increases for fluid milk products in the short-run they are quite responsive to such increases when given time to find substitutes or otherwise adjust their consumption patterns. For the dairy industry it appears that the longer-run consequences of the current industry practice of placing disproportionate price increases on the fluid products to cover the increased costs of production and processing may not be as "painless" as the short-run elasticity estimates imply.

The estimated cross elasticities for the time-series and combined models are presented in Table 2. While the suggested length of the paper does not permit a detailed discussion of these results it is important to note that there is a good deal of substitutability among the various fluid milk products. In all cases except one, where the estimated cross price coefficients were statistically significant at the .10 level or better, the cross-price elasticities displayed the expected relationship. We would of course raise serious questions about the magnitude of the cross-price responses for one percent milk.

From a methodological viewpoint it would be possible to make use of the estimated effect of the substitute prices in reformulating and then re-estimating the cross-sectional model. This is an extension of the present work and we hope to complete it shortly.

The important conceptual issue emphasized by the presentation of these results is that both the magnitude of the estimates and their interpretation are quite different depending on which model is used. It should serve as a warning that when the data of a given set are combined in different ways

Table 2. Calculated Direct and Cross Price Elasticities for Five Fluid Milk Products from Two Alternative Model Specifications, United States, 1972-73

Quantity of:	Model	Retail Price of:								
		Regular Whole Milk	One Percent Milk	Two Percent	Buttermilk	Total Fluid	American Cheese	Evaporated Milk	Ice Cream	Butter
Regular Whole Milk	Time Series	-.379*	.295*	.205*						
	Combined	-.206	.326*	-.048						
One Percent Milk	Time Series	-1.166	-1.171*	3.061*						
	Combined	-12.076*	-3.422*	+16.691*						
Two Percent Milk	Time Series	.857*	.052*	-.548*						
	Combined	.981*	.292	-.703*						
Buttermilk	Time Series	1.355*			-1.775*					
	Combined	2.854*			-3.5146*					
Total Fluid Milk	Time Series					-.1767*	.528*	-.197	-.305	.163*
	Combined					-.3173*	.589*	-.208	-.222	.178

*Indicates that the estimated coefficient was statistically significant at the 10 percent level or better.

different results are obtained. When a point estimate is used for prediction or for assessing the impact of changes in a given variable, the magnitude of the estimate is, in fact, the issue. Applying numerical estimates of estimated relations to models quite different from the one responsible for their generation can be misleading.

Cross-Sectional Models and Policy Analysis

The pooling principle employed in this paper is rather simplistic in nature and, of course, is not new in econometric estimation. What is important, however, is that once again we remind ourselves of the insightful comment by Fred Waugh that "there is probably no such thing as the elasticity of demand" for any product. The nature of the data and the method of estimation in large part determine the point estimate we obtain and then use for policy analysis and business decision making.

Kuh argues that for policy purposes the numerical value of the difference between the cross-sectional and time-series estimates should be ascertained. He states, "If the time-series estimate is some function of the typical cross-sectional estimate, one estimate can be translated into the other irrespective of the causal factors that determined the discrepancy" (p. 210). The parameter which relates the time-series and the cross-section estimate may be thought of as approximately equal to the rate of adjustment from the short-run or immediate response to the new long-run equilibrium position. Given an estimate for this parameter the cross-sectional response may be used to provide estimates of either long-run or short-run consequences to price changes.

This suggested procedure has practical importance. Often policy analysts want information about probable changes in the aggregate market response. Such an estimate is ordinarily obtained from time-series data -

quite often with limited sample sizes. Generating the aggregate response from cross-section data and then, with an estimate of the adjustment rate, deriving the short-run response appears preferable to the procedure suggested by Nerlove for obtaining long-run responses from time-series models for at least two reasons. First, the base response (the equilibrium response) can be generated using a data base which permits structural change. That is, changes in the parameters over time can be observed. Secondly, in most studies, cross-sectional data permit a closer correspondence of the empirical estimation to the theory being used to build the model.

Methods of estimation which appear to permit structural estimation from continuing micro-unit data systems have been developed (Kuh, Balestra and Nerlove, Wallace and Hussain). Seldom however have such methods been employed. The lack of available price/quantity data, available computer software or even the physical limits imposed by computer hardware systems are often difficult hurdles to overcome.

Basically, such methods entail the estimation of parameters from a cross-section of independent micro-units (firms or households) making adjustments over time to changing economic variables. The regression model is expressed as follows:

$$Y_{jt} = a + \sum_{i=1}^{k} \beta_{ijt} X_{ijt} + \varepsilon_{jt}$$

where:

Y_{jt} = the observation of the j<u>th</u> cross-sectional unit in the t<u>th</u> time period,

X_{ijt} = the observation of the i<u>th</u> independent variable associated with the j<u>th</u> cross-sectional unit in the t<u>th</u> time period, and

ε_{jt} = the error term.

It is generally recognized that, under the usual assumptions at least, a straight forward application of OLS to such a model yields parameter estimators which are unbiased and consistent. Such estimates are inefficient however, since the error term, ϵ_{jt}, incorporates an individual micro-unit effect and a time effect in addition to the random error component. If it can be assumed that each of the separate components of ϵ_{jt} are additive, random, have zero means, are independent of each other and have constant variances (which may be different from each other) and that unbiased and efficient estimates of the individual micro-unit effect and the time effect can be obtained, a variance-covariance matrix can be developed and the model estimated via generalized least squares (GLS). Regression coefficients obtained in this way have the same properties as the Atkin estimator. They are consistent, asympotically efficient and asympotically normal. Once these estimates are obtained it should be possible to apply the cross-section estimates in a straight forward fashion to aggregate prediction problems for policy purposes.

Two potential problems with this approach become immediately apparent. First, if any one of the error components is correlated with any of the explanatory variables, so that it is not random or does not have a zero mean, the estimated coefficients from the model will be biased and inefficient. Secondly, and from a methodological viewpoint, one must find a way to find unbiased and efficient estimates of the error components so that an estimate of the variance-covariance matrix can be developed. Wallace and Hussain show that when there are no lagged values of the dependent variables present in the matrix of explanatory variables, the calculated residuals from a first round OLS estimation may be used to compute asympotically unbiased and consistent estimates of the error variance components. To our knowledge this model has never been applied to data.

In the years since Nerlove's insightful paper in 1958 there has not been a sustained effort on the part of researchers in our profession to explore methodologies relating estimates obtained from cross-section and time-series data systems. We are in the process of developing the computer software necessary to estimate models using the MRCA data base and the method suggested by Wallace and Hussain. We are also interested in exploring in more detail both the iterative procedure used by Telser and the method proposed by Balestra and Nerlove. It is hoped that this research will increase our understanding of the consumer response to price changes and will eventually result in methods of estimation which may be used to answer a wider range of policy questions.

Concluding Comments

The purpose of this paper has been to re-emphasize the importance of not only the nature of the data but also the model specification in obtaining numerical estimates of price response. A "rich" set of data generated quite different numerical estimates depending on the specification of the model. The results indicate the importance of specifying models which are capable of generating the answers to specific questions. Furthermore, they indicate the practical importance of increasing our understanding of just how cross-sectional and time-series models are related. If the cross-sectional estimate may appropriately be thought of as the long-run response (ie. the fully adjusted equilibrium response) and the rate of adjustment can be determined, such estimates may be useful for answering a wide range of policy questions. A method of estimation originally proposed by Wallace and Hussain appears to provide, with necessary modification of course, a means for obtaining parameter estimates with desirable properties. Conceptually, at least, it appears that a pooled model of cross-sectional data over time permits the estimation of structural parameters consistent with the theory. That is, they are capable of following the reaction of decision making units, rather than geographic regions, over time when prices change.

References

Balestra, P. and M. Nerlove, "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", Econometrica, Vol. 34, No. 3, July 1966, pp. 585-612.

Chetty, V. K., "Pooling of Time Series and Cross-Section Data", Econometrica, Vol. 36, No. 2, April 1968, pp. 279-290.

Durbin, J., "A Note on Regression When There is Extraneous Information About One of the Coefficients", Journal of the American Statistical Association, Vol. 48, 1953, pp. 799-808.

Kuh, E., "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications", Econometrica, Vol. 27, 1959, pp. 197-214.

Nerlove, M., "Distributed Lags and the Estimation of Long-Run Supply and Demand Elasticities: Theoretical Considerations", Journal of Farm Economics, Vol. XL, May 1958, pp. 301-311.

Telser, L. G., "Iterative Estimation of a Set of Linear Regression Equations", Journal of the American Statistical Association, Vol. 59, September 1974, p. 845-862.

Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis", Journal of the American Statistical Association, Vol. 58, 1963, pp. 401-414.

Wallace, T. D. and A. Hussain, "The Use of Error Components Models in Combining Cross-Section With Time Series Data", Econometrica, Vol. 37, No. 1, January 1969, pp. 55-72.

Waugh, F. V., Demand and Price Analysis, Technical Bulletin #1316, ERS, USDA, Washington, D.C., November 1964.

Footnotes

*The authors are Assistant Professor of Agricultural Economics (Boehm) and Professor of Agricultural Economics and Statistics (Havlicek), Virginia Polytechnic Institute and State University. Part of this research is based on Boehm's unpublished 1974 Purdue University Ph.D. thesis, An Econometric Analysis of the Household Demand for Major Dairy Products. The contributions of E. M. Babb in the early stages of this work are acknowledged.

1. The United Dairy Industry Association (UDIA) acquired these data as a client of the Market Research Corporation of America and made them available to us for research purposes. Dr. G. G. Quackenbush, Director of Economic and Marketing Research of UDIA, was instrumental in getting research using these data started and continues to make contributions as the research progresses.

2. Retail purchases made subject to special promotions or "deals" (cents off, coupon sale, etc.) were reported by NCP households. The percent of the total volume purchased subject to such a special promotion was then specified as an independent variable.

3. The number of substitute or complement products varied among the beverage milk equations according to the hypothesis concerning the expected relations among products. Specifically, for each of the four fluid milk types the substitute or complement prices were prices in half-gallon equivalent units for the other three products, when appropriate. For the composite product, total fluid milk, the substitute or complement prices were prices for American cheese and butter in pounds, evaporated milk in 13 ounce containers and ice cream in half gallons. Table 2 contains the exact cross price responses estimated for each equation.