Journal of Agricultural and Applied Economics, 35,2(August 2003):415–421

Impacts of Sample Size and Quality-Adjusted Imputed Prices on Own-Price Elasticities Estimated Using Cross-Sectional Data

Matthew C. Stockton

Cross-sectional data sets containing expenditure and quantity information are typically used to calculate quality-adjusted imputed prices. Do sample size and quality adjustment of price statistically alter estimates for own-price elasticities? This paper employs a data set pertaining to three food categories—pork, cheese, and food away from home—with four sample sizes for each food category. Twelve sample sizes were used for both adjusted and unadjusted prices to derive elasticities. No statistical differences were found between own-price elasticities among sample sizes. However, elasticities that were based on adjusted price imputations were significantly different from those that were based on unadjusted prices.

Key Words: cross-sectional data, imputed prices, quality-adjusted prices

JEL Classifications: B41, D12, C21, C24

The ability to conduct sound empirical work in economics relies on the abundance and reliability of the data, but data collection is a time-consuming and costly process. Efficient use of information requires employing the smallest sample and doing the least amount of manipulation possible without compromising outcome. This desire for efficiency is reflected by researchers' endeavors to extract as much knowledge from as small a set of data as possible while maintaining the integrity of the results, which modern business culture refers to as "data mining."

One output of cross-sectional data with quantity and expenditure information is the calculation of imputed price. However, this calculation does not come without a cost. Typically, cross-sectional surveys have a wide variety of respondents who consume many different commodities as well as different qualities of the same type of good. This variation in qualities is referred to as "quality difference." Quality difference can be illustrated by an example of two consumers who purchase selected cuts of the aggregate commodity beef. Consumer 1 purchases 5 lbs. of ground beef at $2.00/lb. and 1 lb. of T-bone steak at $6.00/lb., while consumer 2 purchases 2 lbs. of ground beef at $2.00/lb. and 10 lbs. of T-bone steak at $6.00/lb. Both consumers purchase beef, but the beef each consumer buys is of different quality.

The author is a graduate student, Department of Agricultural Economics, Texas A&M University, College Station, TX.

The author wishes to thank those who have been mentors and teachers, without whom this work would never have been done.

For consumer 1, the imputed price of beef is \$2.67/lb., whereas for consumer 2, the imputed price is \$5.33/lb.

Imputed prices are in turn used to derive elasticities of demand. If the differences in quality are not adjusted for, this omission may lead to estimations of own-price and/or cross-price elasticities that misrepresent the true nature of the demand relationship. To compensate for quality differences, it is generally accepted that quality adjustments on imputed prices are necessary (Cox and Wohlgenant). Adjusted prices then are used in estimating own-price and possibly cross-price elasticities of the aggregate good. To use limited time and resources most efficiently, it is essential to understand whether to use adjusted or unadjusted prices in the estimation of these elasticities.

A second issue of concern when using cross-sectional data is sample size. What effect does sample size have on the estimation of elasticities? Assuming that research resources are limited, spending the budget on an oversized sample would be unwise; resources would be better used for more precise and complete data collection. This paper will investigate two issues: (1) is the extra work associated with quality-adjusted imputed prices necessary in calculating own-price elasticities from cross-sectional data, and (2) what effect does sample size have on the estimation of the own-price elasticities?

Literature Review

We first conducted a study of existing research related to cross-sectional survey information. As mentioned in the introductory section, cross-sectional survey information with quantity and expenditure data can be used to estimate imputed prices (Allen and Bowley; George and King; Prais and Houthakker). However, Griliches found quality variation among brands of a single commodity. Furthermore, Nelson states: "The simple sum of physical quantities is found to be theoretically arbitrary and a potentially misleading measure of demand when goods are heterogeneous." These works have led to the widespread practice of applying price adjustment processes to most cross-sectional studies.

The primary reasons for price adjustment are demographic differences in consumers and quality differences in products. Demographic factors add information to demand analysis (Cox and Wohlgenant).

Deaton further justifies the need for a quality adjustment of price through an examination of geographically clustered household data. Polinsky shows that the failure to adequately specify cross-sectional price effects could result in biased and misleading demand elasticities. Hence, quality differences in goods purchased by demographically varied consumers suggest the need to use adjusted prices in elasticity calculations.

Methodology

One of the difficulties encountered in using survey data containing both household expenditures for food items and quantities purchased is that some respondents report zero expenditures and purchases for the respective products. Using a censored model estimation procedure such as the Heckman two-step method solves this zero observation dilemma (Nelson). In the probit component of the Heckman procedure, a variable is estimated accounting for the bias created by the nonparticipation of consumers, which is known as the inverse Mills ratio or IMR, which is unique for every observation. Probit analysis requires that the dependent variable be either 0 or 1. The value for nonparticipation observations is 0, and for all positive observations, the value is 1. In the second step of the Heckman procedure, the demand equation is estimated using the IMR as an additional regressor. The Heckman procedure is not the only way to address the nonparticipation response bias. Alternatives to this procedure include a maximum likelihood approach, which simultaneously estimates both the probit and demand equations.

In our analysis, the probit model is given as follows:

Probit model

$$PI_i = \alpha_0 + \alpha_1 D_{1k} + \alpha_2 D_{2l} + \alpha_3 D_{3m} + \beta_1 ED_i + \beta_2 HS_i + \beta_3 I_i + v_i$$

where $PI_i = 1$ if the imputed price exists because of nonzero expenditure and quantity information and is otherwise 0.

- D_{1k} Dummy variable for region (k = north, midwest, and south)
- D_{2l} Dummy variable for race (l = black and white)
- D_{3m} Dummy variable for area (m = city and suburban)
- ED Dummy variable for education
- HS Household size
- I Income
- IMR (Inverse) Mills ratio

The Mills ratio is used as a regressor in an auxiliary regression with price as the dependent variable and the appropriate demographic variables as the independent variables. The adjusted price for the ith observation is equal to the intercept term of the auxiliary regression plus the associated residual. Expenditure divided by quantity equals unadjusted price.

The calculation of adjusted prices is based on a least-squares estimation of the following auxiliary regression:

$$P_i = \alpha_0 + \alpha_1 D_{1k} + \alpha_2 D_{2l} + \alpha_3 D_{3m} + \beta_1 ED_i + \beta_2 HS_i + \beta_3 I_i + \gamma_1 IMR + \mu_i$$

$$\text{Adjusted price}_i = \text{constant} + \text{error term}$$

Unadjusted and adjusted prices are used to estimate demand equations, employing the price, expenditure, and demographic dummy variables with the Mills ratio. These models were assumed to be linear equations. A natural logarithmic transformation cannot be used, because quality-adjusted prices could be negative. This process is repeated for three different products (pork, cheese, and food away from home) using four different sample sizes. Demand models

FAFHQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + PRICE + BETAA + e_1

FAFHQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + ADJPRICE
+ BETAA + e_2

PORKQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + PRICE + BETAB + e_3

PORKQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + ADJPRICE
+ BETAB + e_4

CHSEQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + PRICE + BETAC + e_5

CHSEQR
= NORTHE + MIDWEST + SOUTH
+ BLACK + WHITE + SIZER
+ INCOMER + ADJPRICE
+ BETAC + e_6

Variable interpretation

- FAFHQR Quantity of food away from home
- PORKQR Quantity of pork
- CHSEQR Quantity of cheese
- NORTHE Dummy variable for the northeast region
- MIDWEST Dummy variable for the midwest region
- SOUTH Dummy variable for the southeast region
- BLACK Dummy variable for race group black region
- WHITE Dummy variable for race group white region

SIZER	Household size
INCOMER	Total expenditure
PRICE	Unadjusted price
ADJPRCE	Adjusted price
BETAA	Mills ratios for food away from home model
BETAB	Mills ratios for pork model
BETAC	Mills ratios for cheese model

The dummy variables used as the base are WEST for region and OTHER for race. The race variable OTHER is a combination of Asian and all other races.

Ten sample sets of four sizes were compared: (1) a full sample, (2) three samples of one third of the total observations, (3) three samples of one fifth of the total observations, and (4) three samples of one tenth of the total observations. Random samples were drawn without replacement; thus, all observations were used for the full sample and the one-third–sized samples. There were 5 one-fifth–sized samples and 10 one-tenth–sized samples; hence, the three samples of these sizes were randomly drawn. A Mills ratio was calculated for each sample size and each commodity group. For each commodity group and each sample, imputed unadjusted prices and imputed quality-adjusted prices were used to estimate the respective linear demand functions. The resulting coefficients were then used to estimate own-price elasticities.

Subsequently, the own-price elasticities were used as the dependent variable for a series of regressions employing four sample sizes, two price types, and three commodity (quality) levels for food groups. This analysis permitted a comparison of the respective treatments relating to sample size and quality adjustment.

Data

The 1987–1988 Nationwide Food Consumption Survey (NFCS) was used in this analysis. Although dated, the NFCS allows the imputation of prices because of the reporting of expenditure and quantity information. The data set contains 4,068 observations of expenditure, quantity, and demographic information for 12 product groups. Of the 12 groups, three commodity groups were selected for testing: (1) food away from home, (2) pork, and (3) cheese. The general selection criteria for the commodity groups were based on the degree of heterogeneity or quality differences within the aggregate commodity group. The theoretical foundation of quality differences requires a degree of heterogeneity of the products that comprise the commodity aggregate. The three commodity groups were chosen to represent levels of quality diversity. Cheese was presumed to be the most homogeneous or the least heterogeneous of the 12 choices. Cheese is either a natural product (e.g., American, Italian, Swiss) or a processed product. Pork was selected to represent semiheterogeneous foods. Pork ranges in type from less expensive cuts such as ribs and shoulders to more expensive cuts such as chops or tenderloins. Some pork products are further processed, such as bacon, ham, and sausage. Food away from home was selected as the most heterogeneous food group. Consumers who choose to eat away from home may select the type of restaurant, the entrée, and the particular meal, which leads to a great number of quality difference possibilities.

Of the 4,068 available observations, some were eliminated from the analysis because demographic information had not been reported. Other observations also were eliminated because either reported expenditures or quantities were beyond five standard deviations from their respective means. On the basis of these variable screens, 3,901, 3,903, and 3,909 observations were used in the analysis of pork, cheese, and food away from home.

Empirical Results

Ten observations pertaining to the sample size for each of the three different commodity groups with two price treatments yielded a total of 60 own-price elasticity estimates (Table 1). Each sample size required the estimation of a unique Mills ratio, mean price, and mean quantity. To calculate the own-price elasticity

Table 1. Unadjusted and Quality-Adjusted Own-price Elasticities for Cheese, Pork, and Food Away From Home

	Unadjusted Own-price Elasticities Homogeneous, Semi-Heterogeneous, and Heterogeneous			Adjusted Own-price Elasticities		
Sample Size	Cheese	Pork	Food Away From Home	Cheese	Pork	Food Away From Home
Full	−0.1157	−0.2914	−0.1823	−0.1790	−0.4572	−0.2834
One Third	−0.1675	−0.2741	−0.1785	−0.2065	−0.4164	−0.2736
One Third	−0.0603	−0.3302	−0.1832	−0.0940	−0.5158	−0.2863
One Third	−0.1163	−0.2752	−0.2450	−0.1785	−0.4382	−0.3913
One Fifth	−0.1311	−0.2635	−0.1487	−0.1749	−0.2886	−0.2076
One Fifth	−0.0121	−0.3478	−0.1594	−0.0162	−0.5444	−0.2514
One Fifth	−0.1353	−0.2305	−0.2714	−0.2053	−0.3643	−0.4299
One Tenth	−0.0869	−0.2755	−0.2274	−0.1342	−0.4318	−0.3543
One Tenth	−0.0911	−0.3981	−0.1739	−0.1375	−0.6259	−0.2825
One Tenth	−0.1847	−0.2101	−0.2325	−0.2820	−0.3439	−0.3623

linear demand functions, the coefficient associated with own-price elasticity is multiplied by the mean right-hand side of the mean price of the mean quantity. Each of the 60 own-price elasticities is negative, in accordance with theory. However, they are not uniform, either across food commodities or within a particular food commodity.

Attempts then were made to discern whether own-price elasticities differ between unadjusted imputed prices and quality-adjusted imputed prices, taking sample size into account, and to discern whether own-price elasticities differ by sample size, taking price adjustment into account. Statistically, 20 own-price elasticities associated with a particular commodity group are regressed on a dummy variable pertaining to the price adjustment treatment (unadjusted price as the base category) and on dummy variables pertaining to the sample size treatment (full sample as the base category). The empirical results associated with these auxiliary regressions are shown in Table 2. From Table 2, own-price elasticities that are based on adjusted imputed prices are larger in magnitude (in absolute value) than are those that are based on unadjusted imputed prices. For cheese, the own-price elasticities based on adjusted prices are larger by .0507; for pork, they are larger by .1530; and for food away from home, they are greater by .1120. These differences are statistically significant for pork and food away from home at the .01 level and are statistically significant for cheese at the .10 level. Bottom-line, quality-adjusted imputed price affects the estimation of own-price elasticities from cross-sectional data. On the basis of these results, if prices are not adjusted for quality, then corresponding own-price elasticities will be underestimated. On the other hand, from Table 2, sample size treatments do not statistically affect the estimation of own-price elasticities for cheese, pork, and food away from home, regardless of whether from the full sample (a full sample is nearly 4,000 observations), one third of the full sample, one fifth of the full sample, or one tenth of the full sample.

Concluding Remarks

The results of this analysis support the use of an adjusted price methodology for food groupings that have quality differences. The results for cheese, pork, and food away from home show that if quality adjustments in price are not taken into account, the own-price elasticity estimated from cross-sectional data will be un-

Table 2. Auxiliary Regression of Own-Price Elasticities On Price Adjustment Dummy Variable and Sample Size Dummy Variables

	Coefficient	Std Error	t-Stat	p-Value
Commodity—Homogeneous (Cheese)				
Adjusted Price (Price Adjustment Dummy)[a]	−0.0507	0.0292	−1.7364	0.1030
One Third Sample (Sample Size Dummy)[b]	0.0102	0.0533	0.1908	0.8513
One Fifth Sample (Sample Size Dummy)[b]	0.0349	0.0533	0.6538	0.5232
One Tenth Sample (Sample Size Dummy)[b]	−0.0054	0.0533	−0.1010	0.9209
CONSTANT	−0.122	0.0484	−2.5191	0.0236
F-statistic for Sample Sizes			0.4111	0.7474
Commodity—Semi-Heterogeneous (Pork)				
Adjusted Price (Price Adjustment Dummy)[a]	−0.153	0.0386	−3.9623	0.0013
One Third Sample (Sample Size Dummy)[b]	−0.0007	0.0705	−0.0097	0.9924
One Fifth Sample (Sample Size Dummy)[b]	0.0344	0.0705	0.4885	0.6322
One Tenth Sample (Sample Size Dummy)[b]	−0.0066	0.0705	−0.0937	0.9266
CONSTANT	−0.2978	0.064	−4.6498	0.0003
F-statistic for Sample Sizes			0.2706	0.8456
Commodity—Heterogeneous (Food Away From Home)				
Adjusted Price (Price Adjustment Dummy)[a]	−0.112	0.0269	−4.1617	0.0003
One Third Sample (Sample Size Dummy)[b]	−0.0268	0.0492	−0.5458	0.5932
One Fifth Sample (Sample Size Dummy)[b]	−0.0119	0.0492	−0.2423	0.8118
One Tenth Sample (Sample Size Dummy)[b]	−0.0393	0.0492	−0.7999	0.4362
CONSTANT	−0.1768	0.0446	−3.9603	0.0013
F-statistic for Sample Sizes			0.3497	0.8110

[a] Unadjusted price is the base category.
[b] Full sample size is the base category.

derestimated. Sample size does not statistically alter the own-price elasticity estimates of the three commodity groupings using either quality-adjusted imputed or unadjusted imputed prices.

This analysis was limited to a small number of experiments that were based on three commodity aggregates and four variations in sample size. Additional work with other commodity aggregates and sample sizes is needed to verify the robustness of the results.

The strength or power of any finding can be measured by how widely it may be applied. A confirmation of the sample size findings, adjusted price versus unadjusted price findings, and degree of heterogeneity of commodity group findings in estimation of own-price elasticities would go a long way toward clarifying the procedures necessary to appropriately estimate these important demand parameters using cross-sectional data. This confirmation would also reduce the waste of research resources and ensure a better consistency of results.

References

Allen, R.G.D., and A.L. Bowley. *Family Expenditures: A Study of Its Variation.* London: P.S. King & Sons, 1935.

Cowling, K., and A.J. Rayner. "Price, Quality and Market Share." *Journal of Political Economics* 78,6(November–December 1970):1292–1309.

Cox, T., and M.K. Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68(November 1986):908–19.

Deaton, A. "Quality, Quantity, and Spatial Variation of Price." *American Economic Review* 78,3(June 1988):418–30.

George, P.S., and G.A. King. *Consumer Demand for Food Commodities in the United States with*

Projections for 1980. Berkeley: University of California, Giannini Foundation Monograph No. 26, March 1971.

Griliches, Z. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." *The Price Statistics of the Federal Government.* New York: National Bureau of Economic Research, 1961.

Nelson, J. "Quality Variation and Quantity Aggregation in Consumer Demand for Food." *American Journal of Agricultural Economics* 73(November 1991):1204–12.

Polinsky, A.M. "The Demand for Housing: A Study in Specification and Grouping." *Econometrica* 45(March 1977):447–62.

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