Prices and Quality Effects in Cross-Sectional Demand Analysis

Thomas L. Cox and Michael K. Wohlgenant

A conceptual clarification of the sources and meaning of cross-sectional price variability is used to motivate a theoretical and econometric framework for the estimation of cross-sectional demand functions. Quality effects are distinguished from supply-related price variability to identify cross-sectional demand for disaggregated food commodities. An empirical application using data from the 1977–78 Nationwide Food Consumption Survey indicates that parameter differences resulting from a failure to adjust cross-sectional prices for quality effects are likely to be small for relatively homogenous, disaggregated food commodities.

Key words: cross-sectional demand functions, quality-adjusted prices, quantity/quality choice.

In demand analysis with cross-sectional, household budget data, it is usually assumed that prices are constant (Allen and Bowley, Prais and Houthakker, George and King). Given this assumption, Engel functions are estimated where expenditure (or quantity) is regressed on income (or total expenditures), family size, and other demographic characteristics. The assumption that cross-sectional price effects are absent or are captured adequately by spatial and temporal dummy variables has not been evaluated empirically. Whether cross-sectional price effects can be treated in this manner has implications for the specification of cross-sectional demand functions as well as the estimation of Engel functions.

There is a considerable literature on estimating price elasticities with cross-sectional data. Most of the applications utilize time-series/cross-sectional data. Annual (or quarterly) household consumption surveys and more narrowly defined panel data are the most commonly used sources. Examples of complete demand systems estimated with a time series of cross sections include Muellbauer, Pollak and Wales, and Ray. Given the scarcity of time-series data, much food policy research in developing countries utilizes price elasticities estimated with cross-sectional data. An insightful summary of this literature is provided by Alderman. Applications of demand models to U.S. cross-sectional food data include Benus, Kmenta, and Shapiro; Capps and Havlicek; Cox, Ziemer, and Chavas; Purcell and Raunikar; Thraen, Hammond, and Buxton.

The existence of cross-sectional price variation raises several important issues. Failure to adequately specify cross-sectional price effects could result in biased and misleading demand elasticities (Polinsky). Thus, traditional Engel analysis may be inappropriate if prices are not constant in the cross section. In addition, prices in cross-sectional data are generally assumed to reflect "quality" effects which should be corrected for prior to estimation (Black, Cramer, George and King, Houthakker, Prais and Houthakker, Theil).

The objectives of this paper are twofold: (*a*) to identify the sources of observed variability of prices in cross-sectional data, and (*b*) to develop a conceptual and empirical framework which captures the observed variability in prices sufficient to estimate price elasticities

Thomas L. Cox and Michael K. Wohlgenant are, respectively, an assistant professor of agricultural economics, University of Wisconsin and an associate professor of agricultural economics, North Carolina State University (formerly at Texas A&M University).

This research was conducted with support from the Texas and Wisconsin Agricultural Experiment Stations. An earlier version of this article appeared as Department of Agricultural Economics Staff Paper No. 241, University of Wisconsin.

Helpful comments from Mike Carter, Kathy Segerson, Jean-Paul Chavas, and anonymous *Journal* reviewers are gratefully acknowledged.

Review was coordinated by Peter Helmberger, associate editor.

of demand. The empirical model is illustrated through an application to cross-sectional data for vegetable consumption from the 1977–78 Nationwide Food Consumption Survey (NFCS).

Conceptual Issues

In the early literature on demand estimation with cross-sectional data, attention focused on the sources and meaning of price variability. Prais and Houthakker (p. 110) argue that the causes of cross-sectional price variation must be identified in order to interpret correctly the effects of prices in the analysis of household budget data. They identify price variation due to region, price discrimination, services purchased with the commodity, seasonal effects, and quality differences caused by heterogenous commodity aggregates. Of these factors, price variation induced by region and season is desirable from the standpoint of estimating commodity demand curves. When the structure of demand is relatively constant, seasonal and regional price variation can be attributed to changed supply conditions and can be used to identify commodity demand curves à la Working. To quote Friedman (p. 33):

> Another kind of contemporaneous data consists of data for different spatial units, such as different states or cities or countries. The problem of constructing a demand curve from such spatial data is essentially similar to that already considered for time-series data. To construct a demand curve, it is essential that conditions of supply vary considerably, conditions of demand vary little. But for any product that has a national market, conditions of supply are about identical except for transportation cost for different states or other subdivisions of the United States. It follows that a demand curve can be constructed readily only for products that have a local market, which would imply that supply conditions are different.

For the vegetable commodities analyzed here, and perhaps for food products in general, supply variability due to regional markets appears to be an important source of demand identification.

The other source of cross-sectional price variation, "quality" differences caused by heterogenous commodity aggregates, may be more problematical to the estimation of demand functions. Quality effects in cross-sectional price variation result mainly from commodity aggregation (Houthakker). Traditional consumer theory assumes homogenous goods with a single price. When separate goods are aggregated into a single commodity, this results in variations in the average price paid for the aggregated commodity, which changes with the quantities of the component goods consumed.

Black, Theil, and Houthakker identified several dimensions to quality effects. Black hypothesized that price/income relationships are caused by differences in marketing services purchased; higher income households purchase more marketing services and, hence, pay higher average prices for commodities. Larger families generally pay lower average prices because of economies of size in purchasing and in household production/consumption activities (Prais and Houthakker).

Goldman and Grossman attribute sources of variation in cross-sectional prices to the opportunity costs of time and the marginal costs/benefits of information search. Cowling and Raynor attribute residual variation in prices, after controlling for the physical characteristics of commodities, to costs of information, brand loyalty, brand ties through distribution networks, etc. Assuming that households purchase a "market basket" of goods for their food-at-home expenditures, retail food merchandising strategy (e.g., promotion items, loss leaders, etc.) can induce additional price differences in household food consumption data. When this type of price variation results from changing supply conditions, a range of prices for similar commodities would be generated, allowing estimation of cross-sectional demand functions.

Cramer makes a distinction between commodities and goods to account for the heterogeneity problem.[1] He demonstrates the importance of adjusting for quality effects by drawing a distinction between income and expenditure elasticities. Note that the definition of expenditures (x) as price (p) times quantity (q) implies that

$$(1)\qquad \log x = \log q + \log p.$$

In conventional Engel analysis of household data, expenditures are hypothesized to depend on real income (y), family size and other household characteristics (c). The Engel relationship and (1) imply that

[1] Goods are defined as specific varieties or brands sold at a single price. If the quantities of these goods can be added together from the consumer's viewpoint, then these goods belong to the same commodity.

$$\frac{\partial \log x}{\partial \log y} = \frac{\partial \log q}{\partial \log y} + \frac{\partial \log p}{\partial \log y}, \quad (2)$$

where the second term on the right-hand side of (2) is the quality elasticity (Prais and Houthakker). For goods (as opposed to commodities), this quality elasticity is zero since homogenous goods only have one price. However, Engel analysis generally deals with commodities rather than goods, so that this quality elasticity is nonzero for commodities which are heterogenous (Prais and Houthakker, chap. 8; Cramer).

Commodity prices depend both on income and household characteristics, c_p, which can induce quality effects (Theil, Cramer). With this specification, cross-sectional, quantity-dependent demand functions have the form

$$q = h[p(y, c_p), p_s(y, c_p), c, y], \quad (3)$$

where p_s is a vector of prices for substitute and/or complementary commodities and c is the vector of household characteristics which shares some elements with c_p. Denote these shared elements as c_q. Partial differentiation of (3) with respect to c_q yields

$$\frac{\partial q}{\partial c_q} = \frac{\partial h}{\partial p}\frac{\partial p}{\partial c_q} + \frac{\partial h}{\partial p_s}\frac{\partial p_s}{\partial c_q} + \frac{\partial h}{\partial c_q}. \quad (4)$$

This partial derivative cannot be obtained directly through estimation of equation (3) using average commodity prices (without assuming $\partial p/\partial c_q = \partial p_s/\partial c_q = 1$). However, it can be obtained if the prices are first adjusted for the impact of quality effects induced by household characteristics. Therefore, quality-adjusted prices should be used to estimate commodity demand functions from cross-sectional data. The potential distortions from not adjusting cross-sectional prices for quality effects will increase with heterogeneity of the commodity aggregate (Cramer).

This approach to modeling demand is likely to be difficult for differentiated or generic goods because of the potential for multiple corner solutions. Given that households typically do not consume all component goods of a commodity aggregate, corner solutions due to nonconsumption yield multiple first-order conditions for quantity demands in a utility maximization framework. Simultaneous solution of these multiple first-order conditions and the associated first-order conditions for quality demand is generally intractable at branded or generic goods level without strong a priori assumptions (Hanemann 1981). At the commodity level of analysis, however, quality choice, as reflected in the quantity shares of the component goods, can be viewed as occurring prior to the quantity choice decisions. One can assume that the household first determines commodity quality through the selection of component goods and then the quantity of the composite commodity. This means the household quality decision (as reflected in the quality/price function) can be modeled independently of the quantity decision at the commodity level.

Theoretical Framework

Hanemann (1981) discusses two alternative approaches to specifying quality/quantity substitution in demand analysis. He also shows that these approaches can be viewed as generalizations of the traditional utility maximization approach. In the Houthakker-Theil framework, demand is modeled as quality/quantity choice among generic commodities which possess a scalar (Houthakker) or vector (Theil) of characteristics. Within this framework, commodities with different characteristics are treated as the same commodity with a continuum of quality levels. Theil assumes that the qualitative nature of an aggregate commodity is reflected in its average price and that this quality component can be described completely by a vector of quantitative or qualitative characteristics. In contrast, Houthakker assumes that the average commodity price reflects two components—the quality and quantity components of price.

The maximization problem in the Houthakker-Theil model is

$$\max_{q \geqq 0, b \geqq 0} U(q, b, z, c) \quad \text{subject to } \Sigma_i p_i(b_i) q_i + z = y, \quad (5)$$

where q is defined as a vector of commodities $(q_i, \ldots, q_n)$, b is a vector of commodity-specific characteristics $b_i = (b_{i,1}, \ldots, b_{i,ki})$, z represents all other goods taken as the *numéraire*, c is a vector of household characteristics hypothesized to influence preferences, and y is total household (real) income. The function U is assumed to possess the usual neoclassical properties. The hedonic (real) price functions, $p_i(b_i)$, reflect price/quality trade-offs. Following Houthakker, these functions can be written

$$p_i = \alpha_i + \Sigma_j \gamma_{ij} b_{ij}, \quad (6)$$

where α_i is interpreted as the quantity price and $\Sigma_j \gamma_{ij} b_{ij}$ reflects the sum of component quality prices per unit of q_i.

Assuming an interior solution to (8) (that is, $q > 0$, $b > 0$), the ordinary demand functions for quantity and quality take the form

$$(7) \quad q_i = h_i^q(\alpha, \gamma, c, y) \quad i = 1, \ldots, N$$

$$(8) \quad b_{ij} = h_{ij}^b(\alpha, \gamma, c, y) \quad i = 1, \ldots, N; \; j = 1, \ldots, ki$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ is the vector of quantity prices and $\gamma = [\gamma_1 (\gamma_{1,1}, \ldots, \gamma_{1,k1}), \ldots, \gamma_n(\gamma_{n,1}, \ldots, \gamma_{n,kn})]$ is the vector of quality prices. These functions are continuously differentiable and homogenous of degree zero in (α, γ, y). Hanemann (1981) summarizes the properties of these quantity and quality demand functions. Since these results are not explicitly required for the purposes at hand, they are not presented here.

Note that (7) and (8) represent a potentially large system of simultaneous equations complete with a full set of restrictions implied by utility maximization. For most applications, this system is likely to be untractable, particularly for disaggregated commodities where corner solutions (in quantities or qualities) are likely to occur. Usually additional a priori structure is required to make this framework tractable. Hanemann (1981, 1984) identifies several special cases of this general model of quality/quantity choice and summarizes the numerous econometric difficulties of these approaches.

Goldman and Grossman argue that simultaneity of quantity and quality choice is likely to be important where the opportunity costs of search time are significant components of "full" commodity prices.[2] They treat quality-adjusted prices as endogenous and estimate the demand for quality and quantity with a two-stage least squares procedure. The feasibility of this procedure here, however, is constrained by data on time components of search and by estimation difficulties (e.g., simultaneous quantity/quality choice with multiple corner solutions). Since search costs for commodities with small budget shares (e.g., disaggregated food commodities) are negligible, simultaneity bias is unimportant. Moreover, if quality effects are due largely to heterogeneity of composite commodities, then the assumption of recursive quality/quantity choice is appropriate.

Hanemann (1984) shows how a random utility, discrete/continuous choice model can be used to analyze quality/quantity decisions. In this context, household characteristics are viewed as proxies for household preferences over unobservable quality characteristics. As a result, the simultaneous discrete/continuous choice of quantity and quality by households can be viewed as a switching regression model (Hanemann 1984). Heckman (1979) shows that the tobit model is a special case of the switching regression model for censored dependent variables.

Tractability of the Houthakker-Theil model is considerably enhanced by limiting the potential number of corner solutions. Given the estimation difficulties of multiple-equation, multiple-corner point-switching models (Hanemann 1981, 1984) separability assumptions are employed to reduce this general framework to a single-equation tobit model. In this approach, the household is assumed to decide whether or not to consume a commodity aggregate whose quality has been determined by prior decisions concerning the quantity shares of component goods. Conditional on this choice, a decision as to the quantity of this composite commodity is then made. In essence, the composite commodity is assumed to be weakly separable from all other commodities in order to reduce corner solutions and estimation difficulties.

Data and Empirical Specification

Data from the western region of the 1977–78 NFCS are used to estimate commodity demand functions and to determine the bias in parameter estimates from quality effects in cross-sectional prices. This regional focus reduces computational burdens, reduces concerns about heterogeneity in demand across regions, and supports the feasibility of using these types of data for estimating regional demand functions. The Cramer model indicates that the bias from failure to quality adjust cross-sectional prices increases with the heterogeneity of a commodity aggregate. Therefore, commodities exhibiting a range of heterogeneity were sought. Additional interest in substitution among product forms of disaggregated food commodities led to the selection

[2] Goldman and Grossman extend the Houthakker-Theil model through household production theory and define "full" commodity prices as commodity prices plus the opportunity costs of time spent in commodity production/consumption.

of the following three commodities: fresh, canned, and frozen vegetables excluding potatoes. These commodities are comprised of the following components (goods): dark green, deep yellow, light green, and other vegetables. The fresh vegetable aggregate includes all forms of home-produced vegetables.[3] The canned vegetable aggregate consists mainly of commercially canned vegetables and includes baby and junior foods as well as foods labeled as dietetic. The frozen vegetable aggregate consists of commercially frozen vegetables. The degree of truncation (i.e., percentage of nonconsuming households) is 6%, 30%, and 58% for the fresh, canned, and frozen commodities, respectively.

Total household income from wage and nonwage sources (*RINCOME*) is obtained by aggregating over individual household members. Households not reporting income were deleted from the sample (about 11%). Commodity prices are implicit prices, obtained by dividing expenditures on quantity consumed by the quantity consumed. In accordance with Prais and Houthakker (p. 112), observations with prices more than five standard deviations from the average observed price were deleted (about 2%). All prices and income are deflated using a region-specific, monthly consumer price index (CPI) for all items (unadjusted) for urban wage earners and clerical workers.[4]

Family size is linearly decomposed into the following age/sex categories: children of age 5 or less (*CHILD*0–5), children ages 6–10 (*CHILD* 6–10), children ages 11–15 (*CHILD*11–15), adult males (*ADULT-M*), and adult females (*ADULT-F*). While further decomposition of these categories and functional form explorations are possible, this linear specification was preferred given estimation constraints of the maximum likelihood (ML) tobit procedures. Additional household characteristics included two continuous variables: the proportion of meals consumed at home (*PROPHOME*) and age of the household head (*AGEHEAD*). Discrete zero-one variables are used to model the consumption impacts due to education of the meal planner (*ELEMED* for meal planners with elementary school education; *COLLED* for college-educated meal planners), sex of the food purchaser (*MALEPUR* for male food purchaser), urbanization (*SUBURBAN*, *NONMETRO*), region (*PACIFIC*), and quarter (*SPRING*, *SUMMER*, *FALL*).[5]

The Price/Quality Functions

The (real) deviations from regional/quarterly mean prices (*RDMP*) for the consuming households are assumed to reflect the "quality" effects induced by household characteristics and nonsystematic supply-related factors. Therefore, the price/quality functions (6) are specified as

$$(9) \qquad RDMP_i = p_i - \alpha_i = \Sigma_j \gamma_{ij} b_{ij} + e_i,$$

where α_i is the regional/quarterly mean price, e_i is the regression residual, and b_{ij} are household characteristics as proxies for household preferences for unobserved quality characteristics. Quadratic specifications of family size and income are used to capture economies of size in purchasing and home production/consumption as well as increases in quality which are hypothesized to increase with income. Thus, family size (income) is expected to have negative (positive) linear and positive (negative) quadratic impacts on the *RDMP* functions.

All household characteristic variables of the final demand model are included in (9) with the exceptions of the age/sex categories and the regional and quarterly effects.[6] Additional household characteristics controlling for shopping location (*SUPERMKT*, *NONSUPERMKT*), Spanish origin (*HISPANIC*, *NON-HISPANIC*), race (*BLACK*, *NON-BLACK*), and employment of the female household head (*FHEADEMP*, *FHEADNOT*) were also included. The shopping categories are hypothesized to capture the impacts of additional services rendered by supermarkets (i.e., selection, carryout, convenience, etc., see Black). Therefore, positive effects on

[3] The incorporation of home production into the household consumption categories follows the convention cited in Burk (p. 92). The procedure of allocating home production (regardless of form) to the fresh category is due mainly to convenience because the original data tapes are categorized in this manner.

[4] This monthly western region CPI for all items is calculated from the quarterly ratios of the western region to the U.S. city average CPI times the monthly U.S. city average.

[5] The discrete categories deleted for estimation purposes are high school-educated meal planners, female food purchasers, central city, mountain region, and winter quarter. The combined effects of these deleted variables are captured by the intercept.

[6] The age/sex categories, a linear decomposition of family size, are implicitly contained in the family size measure. Because the deviations from regional/quarterly means are modeled as dependent variables, region and quarter categories were dropped from the specification after conventional *F*-tests indicated their individual and combined effects were not statistically different from zero.

the *RDMP* are expected due to *SUPERMKT*. The ethnicity and race categories are hypothesized to reflect quality effects and potential price discrimination impacts as discussed by Black, and Prais and Houthakker. These impacts are likely offsetting, so the signs on these categories are indeterminate. Household production theory suggests substitution toward products with more convenience as the opportunity costs of time spent in food preparation increases. Therefore, positive effects on *RDMP* are expected from employment of the female household head (*FHEADEMP*). All sets of discrete categorical variables in the *RDMP* price/quality functions are constrained to sum to zero following Suits. Therefore, the coefficients on these variables are interpreted as the deviations from the intercept term (average *RDMP*) due to the particular category of the dummy variable.

Treatment of Missing Prices

Truncation difficulties generally increase as one disaggregates commodities in cross-sectional data. Aside from the limited dependent variables models which result, missing regressor difficulties are encountered where prices are not observed for nonconsuming households. Two popular and computationally simple solutions to this problem are (*a*) to discard all incomplete observations and estimate population parameters using the remaining observations; and (*b*) to use zero-order methods which substitute "appropriate" sample means for the missing values. Less simple solutions include various first-order methods (Afifi and Elashoff, Dagenais, Gourieroux and Monfort) and Heckman's (1974) sample selection procedure.

If regional/quarterly price differences reflect mainly commodity supply conditions, then regional/quarterly average prices are appropriate zero-order solutions for the missing price information.[7] This procedure was used to determine unadjusted missing pieces. Hence these zero-order missing prices imply that nonconsuming households faced average quality commodity prices (i.e., there is no deviation from the regional/quarterly average price).

First-order missing price procedures use the characteristics of the household with missing prices to determine an appropriate solution. A first-order missing regressor procedure which takes the price/quality structure estimated on the consuming households to infer additional information about the prices of the nonconsuming households is used to determine quality-adjusted missing price information. Thus, the parameters of the estimated price/quality function are used to predict household-specific, quality-adjusted prices for the nonconsuming households. Following Cowling and Raynor, and Goldman and Grossman, the quality adjusted prices ${p^*}_i$ and calculated as

$$p_i^* = \alpha_i + e_i, \qquad (10)$$

where α_i are the regional/quarterly average commodity prices and e_i is the residual from the *RDMP* price/quality regression in (9).[8]

Results

Table 1 summarizes the estimated *RDMP* price/quality functions. The results are consistent with a priori expectations concerning family size and income for all coefficients which are statistically significant. While the linear family size coefficient is significant in all models, only fresh vegetable *RDMP* exhibit a significant income influence. Significant impacts from urbanization are indicated for fresh and canned vegetables; suburban (nonmetro) households pay about 0.6¢ less (more) than the average *RDMP* for fresh vegetables. Significant impacts due to shopping location indicate higher *RDMP* for canned vegetables in supermarkets (1.1¢ above average). Age of household head (*AGEHEAD*) and proportion of meals at home (*PROPHOME*) generally indicate significant and negative impacts on *RDMP* (exceptions: *PROPHOME* for the canned and frozen vegetables). Blacks pay significantly lower-than-average prices for fresh (0.8¢) and canned (1.3¢) vegetables. With the exception of college-educated meal planners (canned vegetables, 1.0¢), ethnicity,

[7] It should be noted that the assignment of regional, monthly components of a CPI as the appropriate cross-sectional prices (Benus, Kmenta, and Shapiro) is very similar to the practice of assigning regional/quarterly average prices for the missing information.

[8] Since the literature is not clear concerning the superiority of alternative missing regressor procedures in the present context, several alternatives were evaluated (see Cox). Zero versus first-order quality-adjusted missing prices were generally found to yield insignificant differences in estimated slope and elasticity response measures. While the first-order missing price procedure used has considerable intuitive appeal, evaluation of alternatives such as Heckman (1974) should be considered in further research.

Table 1. Estimated *RDMP* Price/Quality Functions for the Vegetable Commodities

Household Characteristic	Fresh Vegetables	Canned Vegetables	Frozen Vegetables
INTERCEPT	0.0530***[a]	0.0283	0.0699**
FAMILY SIZE	−0.0072***	−0.0125***	−0.0158**
FAMILY SIZE SQUARED	−0.00009	0.0005	0.0009
INCOME	0.0013***	0.0009	0.0001
INCOME SQUARED	−0.000001**	0.0000	0.00002
URBAN	−0.0004	0.0011	0.0050
SUBURBAN	−0.0059***	−0.0068*	0.0088
NON-METRO	0.0064**	0.0056	−0.0137
SUPERMKT	0.0043	0.0105**	0.0173
NONSUPERMKT	−0.0043	−0.0105**	−0.0173
AGEHEAD	−0.0005***	−0.0004**	−0.0007**
PROPHOME	−0.0363***	−0.0079	−0.0278
NON-HISPANIC	−0.0025	−0.0073	−0.0065
HISPANIC	−0.0025	−0.0073	−0.0065
NON-BLACK	0.0084**	0.0127*	−0.0137
BLACK	−0.0084**	−0.0127*	0.0137
FEMALEPUR	−0.0008	−0.0037	0.0094
MALEPUR	0.0008	0.0037	−0.0094
ELEMED	−0.0016	−0.0080	0.0018
HIGHSCH	0.0008	−0.0017	0.0043
COLLED	0.0007	0.0097**	−0.0061
FHEADEMP	−0.0004	0.0022	0.0011
FHEADNOT	0.0004	−0.0022	−0.0011
Summary statistics			
Number of observations	2,047	1,524	908
Mean-squared error	0.0099	0.0110	0.0156
R-square	0.0532	0.0348	0.0384

Source: Computations by the authors.
Note: The dependent variables are (real) deviations from regional/quarterly average prices ($/lb.). The zero-one dummy variables for each category are constrained to sum to zero and hence reflect deviations from the intercept term due to that factor.
[a] Significance levels of 0.01, 0.05, and 0.1 are indicated by ***, **, and *, respectively.

sex of the meal purchaser, education of the meal planner, and employment of the female household head were not found to be significant.

The low R-squared values for these *RDMP* price/quality functions indicate that considerable variation remains unexplained after adjusting for quality effects. Because data on physical characteristics reflecting commodity quality were not available, this residual variation is assumed to reflect nonsystematic, supply-related factors. In addition, the large residual variation suggests that these aggregates can be considered fairly homogenous in the context of Cramer's analysis. Hence, quality correction likely will not have much impact on the price variables nor imply much parameter bias resulting from unadjusted prices.

Given these results, unadjusted and quality-adjusted prices for both the consuming and, with the missing price procedures, the total samples were computed. For comparison of these results, not presented here because of space limitations, see Cox. Not surprisingly, the quality price components from equation (9) were found to be quite small for all commodities. The net impact of quality-adjusting the observed prices is slightly less price variability (about 2% less for all commodities), little or no impact on the average price level, and an overall decrease in the minimum and maximum prices paid.

A comparison of observed prices with the zero-order missing prices indicated that price variability decreased by 3%, 20%, and 53% for the fresh, canned, and frozen vegetable prices, respectively. Similar results were found when the zero-order quality-adjusted prices are compared with the first-order quality-adjusted prices where variability decreased by 2%, 17%, and 49% for the fresh, canned, and frozen vegetable prices, respectively. These reductions in variability from different missing regressor procedures roughly parallel the degree of commodity truncation (6%, 30%, and 58% for the fresh, canned, and frozen vegetables, respectively). What is most striking about these results is that considerable price variability remained after the quality correction and missing regressor procedures even for the heavily truncated, frozen vegetable prices. Whether this variability is sufficient to estimate price responses in cross-sectional demand functions, however, is an empirical issue which is considered next.

Table 2 summarizes the ML tobit parameter estimates of the various cross-sectional commodity demand functions.[9] These results indicate significant own-price effects with expected signs for all commodities. Significant and positive cross-price effects in the frozen vegetables models indicate canned vegetables are substitutes for the frozen form. Significant income coefficients for all commodities indicate the fresh and frozen forms of vegetables are normal goods, while canned vegetables are inferior goods. The proportion of meals consumed at home has a positive, significant impact on consumption of all commodities, while age of the household head has a significant positive (negative) impact on fresh (canned) vegetables. College-educated meal planners, Pacific region, and summer quarter have significant impacts on the consumption of all commodities.

Results concerning the age/sex categories are quite reasonable. The relative magnitudes of the significant age/sex categories indicate that fresh vegetable consumption increases with age, while male and female adults have about equal impacts. The large impacts of the children categories on canned vegetable consumption partially reflect the junior and baby foods included in this category. While children have no significant effects on frozen vegetable consumption, adult females are indicated to have roughly twice the impact as adult males.

The summary statistics of table 2 suggest that the models without prices are inferior to both the unadjusted and quality-adjusted price models. Using the unconditional and conditional expectations as defined in McDonald and Moffitt, the root mean square errors (*RMSE*) evaluated as the unconditional expectation over the full sample [*RMSE*(*EY*)] and as the conditional expectation over the consuming subsample [*RMSE*(*EY**)] support this conclusion. The models with unadjusted prices appear marginally superior to those with quality-adjusted prices using the *RMSE*(*EY**) criterion for fresh and canned vegetables. The percentage of correct forecasts of discrete household decisions to consume or not [i.e., the % correct (probit) criterion] suggests that the models without prices are marginally superior for fresh and canned vegetables.

Log-likelihood ratio tests of the without prices model nested within either the unadjusted or quality-adjusted prices models were performed. The results of these tests showed that the null hypotheses of no price effects are rejected for all models and commodities. These results and the significance of region (*PACIFIC*) and quarter (*SUMMER*, all models and commodities; *WINTER*, all models, fresh and canned vegetables) in the models with prices indicate nonconstant price effects (i.e., price effects beyond those captured by spatial and temporal dummy variables) are identifiable in these cross-sectional demand functions.

Table 3 summarizes the total effect elasticities of prices, income, and age/sex categories for the alternative models and commodities. These elasticities reflect the unconditional response (i.e., the sum of the conditional and participation responses) to a change in an independent variable as defined in McDonald and Moffitt (p. 319) and are evaluated in elasticity terms at the full sample means. Approximate, asymptotic standard errors are computed using the general method described in Silvey. Two comparisons are of primary interest: between the unadjusted and quality-adjusted results and between the results with and without prices. Most of the differences between significant elasticity estimates in table 3 occur at the third decimal place and are statistically insignificant.

Own-price and income elasticities using quality-adjusted versus unadjusted prices are found to be slightly smaller in absolute value in all commodities except canned vegetables. Comparison of the significant age/sex elasticities across these alternative price models yields the opposite conclusion. These results suggest that price and income elasticities from the unadjusted price models reflect some of the impacts due to family composition. While the differences in magnitude differences are less systematic, comparisons of significant elasticities from the models with and without

[9] The computationally convenient procedures proposed by Fair were used to obtain these results. Results from Amemiya's ML tobit procedures were found to be essentially identical. Since parameter variance estimates evaluated at the negative inverse of the Hessian and the asymptotic covariance matrix of Amemiya (pp. 1007–10) yielded identical inferences concerning parameter significance, the Hessian method was used to compute all parameter standard errors.

Table 2. Comparison of ML Tobit Parameter Estimates for the Fresh, Canned, and Frozen Vegetable Models

	Fresh Vegetables			Canned Vegetables			Frozen Vegetables		
	Unadjusted Prices	Quality-Adjusted Prices	Without Prices	Unadjusted Prices	Quality-Adjusted Prices	Without Prices	Unadjusted Prices	Quality-Adjusted Prices	Without Prices
INTERCEPT	−4.3016***[a]	−4.9911***	−6.7243***	−0.1407	−0.2798	−1.0933**	−2.3909**	−2.6542***	−3.3984
	(1.1905)[b]	(1.0774)	(0.8984)	(0.7144)	(0.6492)	(0.5450)	(0.5621)	(0.5167)	(0.4491)
PFRESH	−8.6389***	−8.5113***		0.1599	0.1225		0.0328	−0.1078	
	(1.9940)	(2.0012)		(1.1940)	(1.1984)		(0.9457)	(0.9508)	
PCANNED	2.2351	2.4007		−2.9979***	−3.0130***		2.3139***	2.3038***	
	(1.5388)	(1.5426)		(0.8677)	(0.8699)		(0.6865)	(0.6878)	
PFROZEN	−2.2958	−2.3505		−0.7866	−0.9064		−3.9829***	−3.7891***	
	(1.6547)	(1.6476)		(0.9696)	(0.9679)		(0.6730)	(0.6752)	
INCOME	0.0782***	0.0727***	0.0713***	−0.0299**	−0.0319***	−0.0315**	0.0455***	0.0432***	0.0435***
	(0.0202)	(0.0201)	(0.0202)	(0.0124)	(0.0124)	(0.0125)	(0.0089)	(0.0088)	(0.0091)
PROPHOME	6.5720***	6.9518***	6.8721***	3.8351***	3.9015***	3.8606***	1.8584***	1.9539***	1.9320***
	(0.8182)	(0.8172)	(0.8213)	(0.5046)	(0.5035)	(0.5061)	(0.3982)	(0.3978)	(0.4076)
AGEHEAD	0.0398***	0.0446***	0.0431***	−0.0256***	−0.0239***	−0.0246***	−0.0060	−0.0045	0.0057
	(0.0098)	(0.0098)	(0.0098)	(0.0058)	(0.0058)	(0.0058)	(0.0046)	(0.0046)	(0.0047)
ELEMED	−0.0939	−0.0818	0.0613	−0.0003	0.0222	−0.0157	−0.6155**	−0.6281**	0.6301**
	(0.5034)	(0.5033)	(0.5062)	(0.2966)	(0.2965)	(0.2981)	(0.2516)	(0.2520)	(0.2577)
COLLED	1.0794***	1.1157***	1.1022***	−0.8854***	−0.9230***	−0.9186***	0.3168***	0.3875***	0.3732***
	(0.2993)	(0.2989)	(0.3006)	(0.1774)	(0.1711)	(0.1781)	(0.1383)	(0.1384)	(0.1416)
MALEPUR	−1.6810***	−1.6448***	−1.6782***	−0.5169*	−0.5264*	−0.5304*	−0.0321	0.0391	−0.0013
	(0.4726)	(0.4728)	(0.4754)	(0.2837)	(0.2837)	(0.2853)	(0.2233)	(0.2236)	(0.2286)
SUBURBAN	−0.1221	−0.1050	−0.0908	−0.2781	−0.2613	−0.2657	−0.1563	−0.1841	−0.1783
	(0.3170)	(0.3168)	(0.3187)	(0.1884)	(0.1883)	(0.1894)	(0.1468)	(0.1470)	(0.1505)
NONMETRO	−0.0613	−0.0670	−0.0977	−0.0436	−0.0425	−0.0487	−0.5925***	−0.5012***	−0.5427***
	(0.3855)	(0.3854)	(0.3872)	(0.2264)	(0.2261)	(0.2272)	(0.1845)	(0.1844)	(0.1886)
PACIFIC	1.1297***	1.1248***	1.0723***	−0.7244***	−0.7246***	−0.7785***	0.4185***	0.4187***	0.4853***
	(0.3342)	(0.3342)	(0.3340)	(0.1958)	(0.1958)	(0.1957)	(0.1589)	(0.1591)	(0.1622)
SPRING	−0.5352	−0.5357	0.7021*	−0.1735	−0.1699	−0.2046	−0.0186	−0.0078	−0.0069
	(0.3885)	(0.3885)	(0.3883)	(0.2295)	(0.2295)	(0.2290)	(0.1800)	(0.1803)	(0.1835)
SUMMER	1.7977***	1.7983***	1.8590***	−0.6084***	−0.6053***	−0.6526***	−0.6162***	−0.6103***	−0.5726***
	(0.3843)	(0.3843)	(0.3858)	(0.2287)	(0.2286)	(0.2295)	(0.1820)	(0.1823)	(0.1861)
WINTER	−0.6504*	−0.6514*	−0.6486*	0.3940*	0.3931*	0.3981*	−0.2295	−0.2215	−0.1764
	(0.3733)	(0.3732)	(0.3746)	(0.2189)	(0.2188)	(0.2196)	(0.1735)	(0.1737)	(0.1776)
*CHILD*0–5	−0.1291	−0.0801	0.1219	0.6447***	0.6753***	0.6649***	0.1059	0.1175	0.1122
	(0.2414)	(0.2416)	(0.2427)	(0.1396)	(0.1396)	(0.1402)	(0.1117)	(0.1119)	(0.1144)
*CHILD*6–10	0.7825***	0.8339***	0.8018***	1.0624***	1.0912***	1.0876***	0.1689	0.1821	0.1689
	(0.2524)	(0.2524)	(0.2536)	(0.1452)	(0.1451)	(0.1458)	(0.1170)	(0.1171)	(0.1198)
*CHILD*11–15	1.0430***	1.0876***	1.0857***	0.7291***	0.7520***	0.7550***	0.1124	0.1194	0.1023
	(0.2328)	(0.2323)	(0.2336)	(0.1348)	(0.1344)	(0.1351)	(0.1069)	(0.1068)	(0.1094)
ADULT-M	1.8628***	1.9116***	1.9401***	1.1095***	1.1399***	1.1268***	0.2462**	0.2722***	0.2834***
	(0.2193)	(0.2187)	(0.2198)	(0.1296)	(0.1292)	(0.1299)	(0.1028)	(0.1026)	(0.1050)
ADULT-F	1.9030***	1.9497***	1.9568***	0.2304	0.2538	0.2373	0.5610***	0.5721***	0.5857***
	(0.2631)	(0.2629)	(0.2644)	(0.1554)	(0.1554)	(0.1562)	(0.1192)	(0.1192)	(0.1219)
VARIANCE	38.9854	38.9914	39.4518	12.4458	12.4410	12.5797	6.4635	6.4850	6.7957
LOGL	−4817.79	−4817.19	−4828.28	−3131.95	−3131.67	−3138.33	−2000.46	−1999.81	−2019.69
RMSE(Ey)	6.035	6.035	6.078	2.830	2.830	2.851	1.466	1.467	1.506
RMSE(Ey)*	6.046	6.048	6.087	3.069	3.071	3.093	2.102	2.100	2.172
% Correct (Probit)	93.6	93.7	94.3	71.3	71.1	71.6	65.0	64.9	62.1

Source: Computations by the authors.
[a] Significance levels of .01, .05, and .10 are indicated by ***, **, and *, respectively.
[b] Asymptotic standard errors are in parentheses.

Table 3. Comparison of Total Effect Elasticities for Selected Independent Variables across Models and Commodities

	Fresh Vegetables			Canned Vegetables			Frozen Vegetables		
	Unadjusted Prices	Quality-Adjusted Prices	Without Prices	Unadjusted Prices	Quality-Adjusted Prices	Without Prices	Unadjusted Prices	Quality-Adjusted Prices	Without Prices
PFRESH	−0.2079***[a]	−0.2036***		0.0095	0.0073		−0.0037	−0.0120	
	(0.0481)[b]	(0.0479)		(0.0712)	(0.0711)		(0.1059)	(0.1057)	
PCANNED	0.0613	0.0652		−0.2037***	−0.2028***		0.2951***	0.2904***	
	(0.0422)	(0.0419)		(0.0591)	(0.0587)		(0.0876)	(0.0870)	
PFROZEN	−0.0870	−0.0892		−0.0739	−0.0853		−0.7025***	−0.6680***	
	(0.0626)	(0.0625)		(0.0911)	(0.0911)		(0.1199)	(0.1202)	
RINCOME	0.0767***	0.0713***	0.0697***	−0.0727**	−0.0775**	−0.0762**	0.2074***	0.1966***	0.1942***
	(0.0198)	(0.0197)	(0.0198)	(0.0304)	(0.0303)	(0.0303)	(0.0405)	(0.0403)	(0.0405)
*CHILD*0–5	−0.0044	−0.0027	−0.0041	0.0540***	0.0566***	0.0555***	0.0167	0.0184	0.0173
	(0.0082)	(0.0082)	(0.0082)	(0.0117)	(0.0117)	(0.0117)	(0.0176)	(0.0176)	(0.0176)
*CHILD*6–10	0.0217***	0.0231***	0.0222***	0.0729***	0.0749***	0.0744***	0.0218	0.0234	0.0213
	(0.0070)	(0.0070)	(0.0070)	(0.0100)	(0.0100)	(0.0100)	(0.0151)	(0.0151)	(0.0151)
*CHILD*11–16	0.0333***	0.0347***	0.0346***	0.0577***	0.0595***	0.0596***	0.0167	0.0177	0.0149
	(0.0074)	(0.0074)	(0.0075)	(0.0107)	(0.0107)	(0.0107)	(0.0159)	(0.0159)	(0.0159)
ADULT-M	0.2132***	0.2188***	0.2215***	0.3149***	0.3235***	0.3187***	0.1311**	0.1447***	0.1477***
	(0.0253)	(0.0252)	(0.0253)	(0.0371)	(0.0370)	(0.0371)	(0.0548)	(0.0546)	(0.0548)
ADULT-F	0.2447***	0.2507***	0.2510***	0.0734	0.0809	0.0754	0.3355***	0.3416***	0.3428***
	(0.0340)	(0.0340)	(0.0341)	(0.0496)	(0.0495)	(0.0496)	(0.0715)	(0.0714)	(0.0716)

Source: Computations by the authors.
[a] Significance levels of .10, .05, and .01 are indicated by *, **, and ***, respectively.
[b] Approximate asymptotic standard errors are in parentheses.

prices suggest that the correlations of the omitted price variables with the remaining regressors do not induce much parameter bias for the commodities studied.

As a further test of the hypothesis that quality is a function of heterogeneity, the three vegetable commodity forms were aggregated into one commodity. Empirical procedures were replicated using this vegetable aggregate. Overall, the results were similar to those for disaggregated commodities, indicating that even this aggregated vegetable commodity is not sufficiently heterogenous to induce significant quality effects. Similarly, despite the significance of the nonconstant, aggregate vegetable price effect, a traditional Engel approach did not imply much bias in estimated income and family composition elasticities.

Conclusions

The conceptual framework for this analysis suggests that clarification of the sources and meaning of cross-sectional price variation is required to model appropriately these price effects. The hypothesis of quality effects due to commodity heterogeneity was developed. The Cramer commodity/goods distinction and a generalization of his quality elasticity decomposition were used to develop an analytical framework which is presented as a special case of the Houthakker-Theil model of quality/quantity choice. Two distinct, but related, hypotheses were generated by this discussion. First, are prices nonconstant in cross section, and if so, can they be adequately modeled by the traditional Engel function approach using dummy variables? Second, if prices are nonconstant in cross section, how important are quality effects, and does their importance vary with commodity heterogeneity as hypothesized by Houthakker and Cramer?

Analysis of the sources of variation in cross-sectional prices suggests that spatial and temporal price differences can be assumed to reflect primarily supply factors. While regional and quarterly price differences were found to be small, considerable variation remained after controlling for region and quarter. This remaining variation is assumed to reflect quality effects induced by household characteristics and nonsystematic supply-related factors such as retail-merchandising behavior.

The *RDMP* price/quality functions of table 1 lend support to the hypothesis of quality effects in cross-sectional prices. Negative and significant impacts from family size were found for all vegetable *RDMP* functions, as hypothesized. Positive impacts from income were found to be significant in the fresh vegetable *RDMP* function. The conceptual framework indicates that these effects are expected to increase with commodity heterogeneity and that a failure to adjust prices for quality effects could induce parameter bias. Table 2 suggests that the commodities analyzed are fairly homogenous using Cramer's criterion; hence, the impacts of quality effects were found to be quite small.

Assuming the structure of demand to be approximately constant, the conceptual framework of this analysis can be used to model quality-adjusted prices and to estimate commodity demand functions. Significant and negative own-price effects are found for all commodities (table 3). The significant and positive impact of canned vegetable price on frozen vegetable consumption indicates that this type of data can yield insight on substitution effects between disaggregated commodities. Furthermore, table 2 suggests that the use of dummy variables in Engel functions is not sufficient to model adequately cross-sectional price effects. However, assuming a correct model specification, the bias on income and family size elasticities because of a failure to model nonconstant price effects was found to be quite small (table 3). This result held even when all commodities were aggregated into a composite vegetable commodity. Therefore, while the conceptual structure clearly indicates the potential difficulties of not correctly modeling cross-sectional price effects, these difficulties appear to be minor for the disaggregated and relatively homogenous commodities considered. Whether these results will hold for more aggregated and/or less homogenous commodities is an empirical issue worthy of further research.

[*Received August 1985; no revision.*]

References

Afifi, A. A., and R. M. Elashoff. "Missing Observations in Multivariate Statistics: I. Review of the Literature." *J. Amer. Statist. Assoc.* 61(1966):595–604.

Alderman, Harold. *The Effect of Income and Food Price Changes on the Acquisition of Food by Low-Income Households*. Report submitted to the Agency for International Development, Office of Nutrition. Washington DC: International Food Policy Research Institute, Oct. 1984.

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

Amemiya, Takeshi. "Regression Analysis When the Dependent Variable Is Truncated Normal." *Econometrica* 41(1973):997–1016.

Benus, J., J. Kmenta, and H. Shapiro. "The Dynamics of Household Budget Allocation to Food Expenditures." *Rev. Econ. and Statist*. 58(1976):129–38.

Black, Guy. "Variations in Prices Paid for Food as Affected by Income Level." *J. Farm Econ*. 34(1952):52–66.

Burk, Marguerite C. *Consumption Economics: A Multidisciplinary Approach*. New York: John Wiley & Sons, 1968.

Capps, Oral, Jr., and Joseph Havlicek, Jr. "National and Regional Household Demand for Meat, Poultry and Seafood: A Complete Systems Approach." *Can. J. Agr. Econ*. 32(1984):93–101.

Cowling, Keith, and A. J. Raynor. "Price, Quality, and Market Share." *J. Polit. Econ*. 78(1970):1292–1309.

Cox, Thomas L. "An Analysis of Price Effects in Cross-Section Data." Ph.D. thesis, Texas A&M University, Aug. 1984.

Cox, Thomas L., Rod F. Ziemer, and Jean-Paul Chavas. "Household Demand for Fresh Potatoes: A Disaggregated Cross-Sectional Analysis." *West. J. Agr. Econ*. 9(1984):41–57.

Cramer, J. S. "Interaction of Income and Price in Consumer Demand." *Int. Econ. Rev*. 14(1973):351–63.

Dagenais, M. G. "The Use of Incomplete Observations in Multiple Regression Analysis." *J. Econometrics* 1(1973):317–28.

Fair, R. C. "A Note on the Computation of the Tobit Estimator." *Econometrica* 45(1977):1723–27.

Friedman, Milton. *Price Theory*. New York: Aldine Publishing Co., 1976.

George, P. S., and G. A. King. *Consumer Demand for Food Commodities in the United States with Projections for 1980*. Giannini Foundation Monograph No. 26, University of California, Berkeley, March 1971.

Goldman, Fred, and Michael Grossman. "The Demand for Pediatric Care: An Hedonic Approach." *J. Polit. Econ*. 86(1978):259–80.

Gourieroux, Christian, and Alain Monfort. "On the Problem of Missing Data in Linear Models." *Rev. Econ. Stud*. 48(1981):579–86.

Hanemann, W. Michael. "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52(1984):541–61.

———. "Quality and Demand Analysis." *New Directions in Econometric Modeling and Forecasting in U.S. Agriculture*, ed. Gordon C. Rausser. New York: Elsevier/North-Holland Publishing Co., 1981.

Heckman, J. J. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1979):153–61.

———. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42(1974):679–94.

Houthakker, Hendrick S. "Compensated Changes in Quantities and Qualities Consumed." *Rev. Econ. Stud*. 19(1952):155–64.

McDonald, John F., and Robert A. Moffitt. "The Uses of Tobit Analysis." *Rev. Econ. and Statist*. 62(1980):318–21.

Muellbauer, John. "Testing the Barten Model of Household Composition Effects and the Cost of Children." *Econ. J*. 87(1977):460–87.

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

Pollak, R. A., and T. J. Wales. "Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure System." *Amer. Econ. Rev*. 68(1978):348–59.

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

Purcell, Joseph C., and Robert Raunikar. "Price Elasticities from Panel Data: Meat, Poultry, and Fish." *Amer. J. Agr. Econ*. 51(1971):216–21.

Ray, Ranjan. "The Testing and Estimation of Complete Demand Systems on Household Budget Surveys: An Application of AIDS." *Eur. Econ. Rev*. 17(1982):349–69.

Silvey, S. D. *Statistical Inference*. London: Chapman and Hall, 1975.

Suits, Daniel B. "Dummy Variables: Mechanics versus Interpretation." *Rev. Econ. and Statist*. 66(1984):177–80.

Theil, Henri. "Qualities, Prices, and Budget Enquiries." *Rev. Econ. Stud*. 19(1952):129–47.

Thraen, C. S., J. W. Hammond, and B. M. Buxton. "Estimating Components of Demand Elasticities from Cross-Sectional Data." *Amer. J. Agr. Econ*. 60(1978):674–77.

Working, E. J. "What Do Statistical 'Demand Curves' Show?" *Quart. J. Econ*. 4(1927):212–35.