Quality, Quantity, and Spatial Variation of Price

By ANGUS DEATON*

In many household surveys, geographically clustered households report unit values of foods, which when corrected for quality effects and for measurement error, indicate the underlying spatial variation in prices, and can be matched to variation in demand patterns so as to estimate price elasticities. A 1979 household survey from the Côte d'Ivoire is used to estimate price elasticities for beef, meat, fish, cereals, and starches.

In the United States, as well as in many Western European countries, income taxes are a major source of government revenue, and it is appropriate that economists should have devoted a great deal of both theoretical and empirical effort to calculating the effects of income taxes on labor supply and on government revenue. By contrast, few developing countries possess the administrative machinery that would permit a comprehensive income tax or Social Security system, so that correspondingly greater emphasis is placed upon indirect taxation and subsidies. In such circumstances, intelligent policy design requires knowledge of price elasticities for taxable commodities. Such knowledge would normally be obtained by the analysis of time-series data on aggregate demands, prices, and incomes. Unfortunately, few developing countries have time-series of a length that is adequate to estimate even own-price elasticities, let alone the cross-price elasticities that are also required. However, many developing countries regularly collect high-quality household survey data on expenditures and quantities of a wide range of commodities. To my knowledge, such data currently exist for such diverse LDC's as Brazil, India, Sri Lanka, and Côte d'Ivoire, the Sudan, Morocco, and Indonesia, as well as for a number of developed countries, for example, the United States and the United Kingdom; a systematic search would likely reveal many more. *In principle*, these household surveys contain information on the *spatial* distribution of prices, so that if this information could be recovered in usable form, there is great potential for estimating the demand responses that are required for making policy. This paper is concerned with the development and implementation of an appropriate methodology to estimate price elasticities of demand using such cross-sectional household survey data.

In surveys where households report both expenditures and physical quantities, it is possible to divide one by the other to obtain unit values. These unit values, which depend on actual market prices, suggest that there is substantial spatial variation in prices in many developing countries, a finding that makes good sense in the presence of high transport costs. However, it is not possible to use unit values as direct substitutes for true market prices in the analysis of demand patterns. Consumers choose the *quality* of their purchases, and unit values reflect this choice. Moreover, quality choice may itself reflect the influence of prices as consumers respond to price changes by altering both quantity and quality. Measured unit values are also contaminated by errors of measurement in

*Research Program in Development Studies, Woodrow Wilson School, Princeton University, Princeton, NJ, 08544. This paper was written while I was visiting the Welfare and Human Resources Division of the World Bank. Dwayne Benjamin provided exemplary research assistance and many helpful suggestions. Guy Laroque helped clarify conceptual issues at the beginning of the research. Helpful comments were provided by Anne Case, and by the referees of this journal. John DiNardo and Tom Lemieux pointed out an error in the calculation of the standard errors. The World Bank does not accept responsibility for the views expressed herein, which are those of the author, and should not be attributed to the World Bank or to its affiliated organizations.

expenditures and in quantities and are likely to be spuriously negatively correlated with measured quantities. In the technique developed here, market prices are treated as unobservable variables that affect quantities purchased, and that determine observed unit values with both measurement error and quality effects. Since household surveys typically collect data on *clusters* of households that live together in the same village and are surveyed at the same time, there should be no genuine variation in market prices within each cluster. Within-cluster variation in purchases and unit values can therefore be used to estimate the influence of incomes and household characteristics on quantities and qualities, and can do so without data on prices. Variation in unit values within the clusters can also tell us a good deal about the importance of measurement error. By contrast, variation in behavior *between* clusters is at least partly due to cluster-to-cluster variation in prices, and this effect can be isolated by allowing for the quality effects and measurement errors that are estimated at the first, within-cluster stage.

The plan of the paper is as follows. In Section I, I discuss previous attempts to estimate price elasticities from cross-sectional data by relating quantities purchased to their "prices," where the latter are obtained by dividing recorded expenditures by recorded quantities. The problems associated with quality choice and with measurement error are identified. I also propose a simple model of quality choice under weak separability that generates a relationship between the price and income elasticities of quality and quantity. This relationship later plays a crucial role in removing the effects of quality choice from the estimate of the price elasticity. Section II presents a model of quantity and quality choice that can be estimated and Section III contains the results for a household survey from the Côte d'Ivoire.

I. Conceptual Background: Temptation and Its Consequences

Given the importance of estimating price elasticities in developing countries, as well as the difficulty of doing so, it is not surprising that researchers have been prepared to make use of whatever data are available. A majority of developing countries has at some time or another conducted a household expenditure survey, usually so as to obtain weights for the calculation of a consumer price index. By definition, such surveys collect data on household expenditures, but they often go further and collect data on *quantities* purchased, particularly for foodstuffs, where quantities are well-defined, and where the consumption levels are themselves important for assessing the nutritional status of the respondents. Armed with expenditures and with quantities, the temptation to divide one by the other is irresistible. In the early classic studies by Hendrik Houthakker and Sigbert Prais, 1952, and Prais and Houthakker, 1955, the authors thoroughly analyzed the behavior of the unit values obtained by such division, but the authors were (presumably) cautious enough to resist the further temptation to use the calculated "price" to estimate price elasticities. More recently, more valorous researchers have taken up the challenge, and there have been a series of papers, by Peter Timmer and Harold Alderman, 1979; Timmer, 1981; Dov Chernichovsky and Oey Astra Meesook, 1982; and Mark Pitt, 1983, using data from Indonesia (all save Pitt) and Bangladesh (Pitt), all of which have regressed quantities on unit values, and all of which have obtained sensible and pleasing results. The dangers were perhaps more apparent than real.

Table 1 lists similar estimates for rural households from the Côte d'Ivoire for two commodities, beef and meat, the latter a broad category that includes the former. The survey and the data will be discussed in Section III below. The results given here were obtained by regressing (by ordinary least squares) the logarithm of annual quantity purchased, measured in kilos, on the logarithm of the unit value, obtained by dividing the total annual expenditure on the good by the total annual quantity purchased. In addition to the logarithm of the unit price, per capita total household expenditure on food was included as an explanatory vari-

TABLE 1—"PRICE" ELASTICITIES, BEEF AND MEAT: RURAL CÔTE D'IVOIRE, 1979

	Beef	Meat
No Cluster Effects	−0.56 (5.0)	−0.30 (4.3)
With Cluster Effects	−0.80 (4.6)	−0.39 (4.6)

able together with a range of household demographic indicators. There are 429 observations for beef, and 631 for meat; this is only a fraction of the total number of observations, but the (temporary) use of logarithms means that we are limited to those households that purchased positive quantities. The regression results in the first row of Table 1 are obtained by ignoring the cluster structure of the sample, while the second row involves deviations from cluster means of all variables. (Identical results would be obtained by including in the regressions dummy variables for each of the clusters.) The price elasticities are all reasonably well-determined and taken separately, either the first or second row would appear to yield satisfactory estimates. The fact that meat is less price elastic than is beef is what would be expected, given that there are substitutes for beef within the meat group as a whole. However, taken together, the results in the two rows are more disturbing. Since the model is supposedly one of spatial price variation, and since price variation within clusters should be absent, the subtraction of the cluster means should make estimation of a price elasticity impossible. In fact, the estimated price elasticity *increases*. The resolution of these (and other) problems involves a good deal of analysis and it is to this that the remainder of this section is devoted.

A. *Quality*

The unit value of "meat" is clearly not a price. Meat is not a homogeneous commodity, but a collection of commodities, in this case, agouti (a large rat), palm squirrel, venison, other game animals, game birds, chicken, guinea fowl, beef, pork, mutton, goat, and canned meat. Not all of these have the same income elasticity, so that richer households will consume not only more than poorer households, but also in different proportions. What is generally to be expected is that the price per kilo will be higher for the goods more heavily consumed by the rich so that there will be a positive relationship between the unit value of meat and the level of household income. Even for a more narrowly defined commodity such as beef, there are more and less expensive cuts, and there are lean, scrawny (and cheap) agoutis as well as fat, sleek, and tasty ones. Houthakker and Prais (1952) give several estimates of what they call the "elasticity of quality," defined as the elasticity of unit value with respect to total household expenditure (or income); see also J. S. Cramer (1973).

One immediate consequence of this analysis is that, insofar as unit values reflect quality as well as genuine price variation, they are *chosen* by consumers just as are quantities. The regression of quantity on unit value is therefore a regression of one choice variable on another, and runs all the usual risks of possible lack of identification, simultaneity bias, and interpretational ambiguity. But there is another, and possibly more important issue. Prices will themselves affect the choice of quality. If market prices rise, consumers can not only alter the quantity that they buy, but also the quality, or more precisely the composition of their purchases within the group. If protein and calories are of greater primary concern than is "taste," then a sensible reaction to bad times is to move to less expensive cuts with little sacrifice of nutritional levels. The effect of this sort of substitution will be that an increase in price will generate a less than proportionate increase in unit value. To fix ideas, suppose that market prices for all types of meat are higher in village A than in village B, and that the per capita weight of meat consumed in A is correspondingly lower. If price were directly observed, and other variables adequately taken into account, the price elasticity could be directly estimated from the relationship between price and quantity. But if unit values are used, the same quantity difference will be ascribed to a smaller unit value difference, because of the quality effect, so that the "price elasticity" will be exaggerated. Note that this would be true even if the econometric equations were to fit the data perfectly; the problems associated with the simultaneity of quantity and quality introduce additional complications.

Without information on all three of quantity, quality, and price, it is generally impossible to estimate everything that we want to know. Some theoretical restriction is required, and I obtain it from a simple model of the way in which quality is influenced by price. The basic assumption is that meat forms a separable branch of preferences, so that the demand for individual meats depends on the total meat budget and on the prices of the individual meats within the branch. In consequence, changes in the level of market prices of all meats together affect the demands for individual meats in exactly the same way as do changes in the total budget devoted to meat. But the "quality" of meat depends on the composition of demand within the meat group. In consequence, if we know how the quality of meat changes with changes in income, we can predict the effects of changes in absolute prices on the unit value index. As I shall show below, if the quality elasticity of meat is zero, the unit value index moves proportionately with the market price of meat. If the quality elasticity is positive, as would normally be the case, unit value will move less than proportionally with prices, the shortfall depending on the size of the quality elasticity as well as on the overall price elasticity of meat. The remainder of this subsection is devoted to a model that produces this result. The result itself will be required in the later sections to estimate price elasticities.

Suppose that at cluster (village, or site) c, the price vector for meats is p_c, where the components of p_c are the prices of the individual goods, agoutis, venison, beef, and so on. I need to assume that there exists a positive, linearly homogeneous, function of p_c, $\lambda_c(p_c)$, where the value λ_c is to be thought of as the *level* of meat prices in cluster c. For example, λ_c could be a fixed-weight Laspeyres index, but there are many other possibilities. Given λ_c, I write

$$(1) \qquad p_c = \lambda_c p_c^*.$$

For some purposes, it is useful to think of p_c^* as being the same for all clusters c. However, in practice the relative prices of different meats are clearly not the same in all locations; indeed not all varieties of meat will even be available in each of the clusters.

Quantities are thought of as the purchases of single individuals located at different sites, but otherwise identical. The aggregate quantity of meat purchased is Q_c, where

$$(2) \qquad Q_c = k^0 \cdot q_c,$$

and q_c is the vector of meat purchases at location c. The vector k^0 will be a vector of ones if it is appropriate to aggregate by weight, in which case Q_c will simply be the number of kilos of meat. I allow the more general formulation so that aggregation could be done with respect to other characteristics, for example, calories. The element-by-element ratios p_i^*/k_i^0 are the prices per kilo of each of the meats, and I take these to be indicators of quality, with higher quality items costing more per kilo. Clearly, the variation in relative prices between clusters must be sufficiently limited for this interpretation to be justified.

Expenditure on the commodity group is denoted by E_c, which is $p_c \cdot q_c$, so that the unit value, V_c, is given by

$$(3) \qquad V_c = E_c/Q_c = \lambda_c\left(p_c^* \cdot q_c/k^0 \cdot q_c\right).$$

The term in brackets, which I denote by ν_c, is the measure of quality; it is the average cost per kilo at location c once price-level differences across clusters have been taken into account. It is this interpretation that limits the degree of allowable relative price dispersion across clusters. Equation (3) can be written

$$(4) \qquad \ln V_c = \ln \lambda_c + \ln \nu_c,$$

so that, measured in logarithms, unit value is the sum of price and quality.

The next step is to use the assumption that meat is separable in preferences to derive an expression for the effect of price changes on quality. The conceptual experiment here is to vary the level of prices, λ_c, while holding fixed the within-cluster relative price structure p_c^*. By weak separability, the vector of meat demands q_c is a function of total meat expenditure and the vector of meat prices, so that

$$(5)\quad q_c = g_c(E_c, p_c) = g_c(E_c/\lambda_c, p_c^*),$$

where the second equality follows from the fact that the (group) demand functions are zero-degree homogeneous in total expenditure and prices. The quality indicator ν_c is $p_c^* \cdot q_c / k^0 \cdot q_c$, which at fixed p_c^* and p^0 is simply a function of q_c, and therefore, by (5) of the ratio E_c/λ_c. In consequence,

$$(6)\quad \frac{\partial \ln \nu_c}{\partial \ln \lambda_c} = -\frac{\partial \ln \nu_c}{\partial \ln E_c} + \frac{\partial \ln \nu_c}{\partial \ln E_c} \cdot \frac{\partial \ln E_c}{\partial \ln \lambda_c},$$

since E_c, total expenditure on meat, will itself be affected by the price change. Expenditure E_c is the product of quality ν_c, price λ_c, and quantity Q_c, so that taking logarithms and differentiating with respect to λ_c gives

$$(7)\quad \frac{\partial \ln E_c}{\partial \ln \lambda_c} = 1 + \frac{\partial \ln \nu_c}{\partial \ln \lambda_c} + \varepsilon_p,$$

where ε_p is the price elasticity of the group with respect to the group price λ_c. Hence, from equation (6), using (7) to substitute for the last term,

$$(8)\quad \frac{\partial \ln \nu_c}{\partial \ln \lambda_c} = \frac{\varepsilon_p \partial \ln \nu_c / \partial \ln E_c}{1 - \partial \ln \nu_c / \partial \ln E_c}.$$

Equation (8) shows that the effects of price on quality operate as income effects; an increase in the group price depresses group demand through the group price elasticity, and it is this fall in demand that generates the change in quality. The result can be expressed in the Prais-Houthakker notation by noting that, since ν_c is a function of E_c,

$$(9)\quad \frac{\partial \ln \nu_c}{\partial \ln x} = \frac{\partial \ln \nu_c}{\partial \ln E_c} \cdot \frac{\partial \ln E_c}{\partial \ln x},$$

where x is total expenditure (or income, or total food expenditure). The left-hand side is η, the quality elasticity as defined by Prais and Houthakker, while the last term on the right-hand side is the sum of the quality elasticity η and the usual quantity elasticity ε_x. Making the substitutions in (9) and then in (8) gives $\partial \ln \nu_c / \partial \ln \lambda_c = \eta \varepsilon_p / \varepsilon_x$, so that, for unit value V_c, by equation (4),

$$(10)\quad \partial \ln V_c / \partial \ln \lambda_c = 1 + \eta \varepsilon_p / \varepsilon_x.$$

In consequence, if the group price changes, and we mistakenly measure the price elasticity by the relationship between the change in quantity and the change in unit value, we measure not ε_p but $\tilde{\varepsilon}_p$, where $\tilde{\varepsilon}_p = d \ln Q_c / d \ln V_c = \varepsilon_p / (1 + \eta \varepsilon_p / \varepsilon_x)$. This expression shows that, econometric issues apart, the comparison of quantities and unit values will tend to overstate the price elasticity in absolute magnitude, at least if, as would normally be the case, the price elasticity is negative, and the product of the price and quality elasticities is smaller in absolute value than the total expenditure elasticity. The equation also gives a means of repairing the bias, since the quality and quantity elasticities can be estimated. In Section III, I shall use the result for exactly this purpose.

B. *Spurious Correlations*

The contamination of unit values by quality effects is not the only problem that lies in the way of using unit values to indicate prices, and it may not even be the most serious. Additional problems are generated by errors of measurement. Unit values are calculated by dividing expenditures by quantities, so that errors of measurement in either will not only cause the unit value to be measured with error, but will also most likely generate a spurious negative correlation between quantity and unit value. Suppose that regressions are run in logarithms, so that the point can be illustrated with a simple bivariate regression. Suppose that for observation i (either a cluster or a household), expenditure E_i and quantity Q_i are measured, each with error. Write this

$$(11)\qquad \ln E_i = \ln E_i^* + e_{1i}$$

$$\ln Q_i = \ln Q_i^* + e_{2i},$$

where true values are marked with asterisks, so that, for the unit value V_i, we have

$$(12)\qquad \ln V_i = \ln V_i^* + e_{1i} - e_{2i}.$$

The errors e_1 and e_2 have variances ω_1^2 and ω_2^2 and covariance ω_{12}. It is important that this covariance be taken into account; although the data *recorded* are expenditures and quantities, it does not follow that the measurement errors of these two magnitudes should be independent. Respondents may recall the latter by dividing the former by the price, or reconstruct expenditures from price multiplied by quantity.

Suppose that the true regression function for Q conditional on V is $E(\ln Q_i^* | \ln V_i^*) = \alpha + \beta \ln V_i^*$, although note that β is not likely to be the price elasticity, if only for the reasons in the previous subsection. If the errors are normal, the regression function of the observed quantities on observed unit values is also linear and has slope not β, but $\tilde{\beta}$, where

$$(13)\qquad \tilde{\beta} = \beta(\omega^2_*/\omega_v^2) - \{(\omega_2^2 - \omega_{12})/\omega_v^2\},$$

where ω^2_* is the true and ω_v^2 is the measured variance of V. The first term is the standard "attenuation bias" whereby β is multiplied by a positive factor less than unity, while the second term arises from the fact that unit values are derived by division. If unit values are correctly recalled, $e_{1i} = e_{2i}$, so that $\omega_2^2 = \omega_{12}$, and $\tilde{\beta} = \beta$. More generally, it seems reasonable to expect the second term to be negative, as it must be if the measurement errors in expenditures and quantities are independent. In this case, and if, as expected, β is negative, the two terms in (13) act in opposite directions so that the biased estimate could be either larger or smaller than the true value.

II. Specification and Estimation

The data to be analyzed come from surveys in which the unit of observation is the household. However, such surveys are invariably geographically clustered, with clusters of a dozen or so households living in the same place and surveyed at the same time. Such a design minimizes the transport costs for the enumerators, who can spend some time in a given location, instead of constantly having to move between units that are widely separated in space. My basic assumption is that all households in the same cluster face the same market price; this price is not itself observed, but makes its presence felt in quantities purchased and in their unit values, both of which are observed. Denoting the household by i and the cluster by c, I propose two basic equations:

$$(14)\qquad w_{ic} = \alpha_1 + \beta_1 \ln x_{ic} + \gamma_1 \cdot z_{ic} + \theta_1 \ln p_c + f_c + u_{1ic},$$

$$(15)\qquad \ln v_{ic} = \alpha_2 + \beta_2 \ln x_{ic} + \gamma_2 \cdot z_{ic} + \theta_2 \ln p_c + u_{2ic},$$

where w_{ic} is the share of the budget devoted to the good (including both actual purchases and imputed expenditures), x_{ic} is the total budget, v_{ic} is the calculated unit value, and z_{ic} is a vector of household characteristics, all of which are observed. The logarithm of the cluster price, p_c, is not observed, nor is the cluster fixed-effect f_c, nor the two error terms u_{1ic} and u_{2ic}. The demand equation (14) is simply the regression function of the budget share conditional on the right-hand side variables, and should not be regarded as a structural demand equation at the level of the individual household. In particular, households that do not consume the commodity, for whom $w_{ic} = 0$, are included in the equation. This is the correct concept for policy analysis; if the government can impose a tax that increases all spatially dispersed prices for the good by the same amount, then the effect on revenue depends on total demand and on the response of total demand to price, including purchasers and nonpurchasers alike. Given this, the equation is a standard Engel curve specification, linking expenditure to total outlay, price, and household characteristics. The unit value equation (15), which is observed only for households that record positive market purchases, follows the analysis of quality choice in Section I by relating unit value to the budget, to household characteristics, and to the price. The coefficient θ_2 is the response of unit value to price, and is related to the other elasticities by equation (10) above.

The share equation (14) contains a set of cluster fixed- (or random) effects f_c that represent unobservable taste variation from cluster to cluster. These are taken to be orthogonal to the unobservable price term, but they need not be orthogonal to the $\ln x$ and z variables; they can be thought of as "residuals" in a cross-cluster explanation of purchases. Alternatively, $f_c + u_{1ic}$ can be thought of as an error term with both cluster and idiosyncratic components. Note that the fixed effects are excluded from the unit value equation; fixed effects would preclude any inference about price from unit values, and the model would not be identified. The error term u_1 has variance σ_{11}, and, conditional on the household making purchases in the market, u_2 has variance σ_{22} and covariances σ_{12} with u_1. (Note that, because of home production, not making market purchases is not the same as having a zero-budget share.) These variances and covariances allow the model to capture the spurious relationships between quantity and price that do not come from genuine price responses; note that, in the case where the measurement errors of expenditures and unit values are independent, σ_{12} would be zero. More complex versions of (14) and (15) can be proposed, for example, so as to include cross-price effects, see Angus Deaton (1987), or to allow dummies to pick up broad regional taste differences that are not generated by price differences, see my working version of this paper (1986).

The parameters β_1, β_2, γ_1, and γ_2 can be estimated by ordinary least squares applied to the data with cluster means removed; denote these within cluster estimates by $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\gamma}_1$, and $\tilde{\gamma}_2$. Note that there is no selectivity problem involved in estimating the unit value equation using only those households that make market purchases; all households in the cluster face the same price, and there is no reason to link measurement error with the amount consumed. Let n be the total number of households in the C clusters, and let n_1 be the number of households that record market purchases. Then if e_1 and e_2 are the OLS residuals from the within-cluster regressions of (14) and (15), then the variances and covariance can be estimated from $\tilde{\sigma}_{11} = (n-k-C)^{-1}e_1'e_1$, $\tilde{\sigma}_{22} = (n_1 - C - k)^{-1}e_2'e_2$, and $\tilde{\sigma}_{12} = (n_1 - C - k)^{-1}e_2'e_1^+$, where e_1^+ are the elements of e_1 corresponding to the households that make purchases in the market. To estimate the other parameters, use the OLS estimates to "correct" the shares and unit values by calculating the two variables $\tilde{y}_{1ic} = w_{ic} - \tilde{\beta}_1 \ln x_{ic} - \tilde{\gamma}_1 \cdot z_{1ic}$, and $\tilde{y}_{2ic} = \ln v_{ic} - \tilde{\beta}_2 \ln x_{ic} - \tilde{\gamma}_2 \cdot z_{1ic}$. We are interested in the between-cluster variation in these magnitudes, so consider their cluster means, $\tilde{y}_{1.c}$ and $\tilde{y}_{2.c}$; from (14) and (15), the population counterparts of these magnitudes $y_{1.c}$ and $y_{2.c}$ satisfy

$$(16) \qquad y_{1.c} = \alpha_1 + \theta_1 \ln p_c + f_c + u_{1.c},$$

$$(17) \qquad y_{2.c} = \alpha_2 + \theta_2 \ln p_c + u_{2.c}.$$

It is then easy to show that, over C clusters, where cluster c has n_c households, n_c^+ of which show positive consumption, we have

$$(18) \quad \operatorname{cov}(y_{1.c}, y_{2.c}) = \theta_1 \theta_2 m_p + \sigma_{12}/n_c,$$

$$(19) \qquad \operatorname{var}(y_{2.c}) = \theta_2^2 m_p + \sigma_{22}/n_c^+,$$

where m_p is the intercluster variance of $\ln p_c$. Hence, if we define τ as $C/(\Sigma n_c^{-1})$, and τ^+ as $C/(\Sigma n_c^{+-1})$, which are the appropriate measures of average cluster size, then we take covariances over the clusters, the empirical ratio

$$\tilde{\phi} = \frac{\operatorname{cov}(\tilde{y}_{1.}, \tilde{y}_{2.}) - \tilde{\sigma}_{12}/\tau}{\operatorname{var}(\tilde{y}_{2.}) - \tilde{\sigma}_{22}/\tau^{+}} \tag{20}$$

will consistently estimate the ratio θ_1/θ_2, consistency referring to the situation where the number of clusters becomes large but the number of households per cluster remains fixed.

To understand the intuition behind this estimator, note that, if there were no "corrections" to the numerator and denominator in (20), it would be the ratio of a covariance to a variance, that is, an ordinary least squares estimator, in this case of cluster-budget share on cluster price, at least once both variables have had the effects of household characteristics netted out. There are two problems with this OLS estimator. First, there is measurement error in both share and unit value. As discussed in Section I, Part B, there is likely to be a spurious negative correlation between quantity and price, and this will result both in an inflated variance for the logarithm of the unit value, as well as in a spurious correlation between the share and the log unit value. Averaging over clusters will reduce, but not eliminate the bias induced by these effects and both the covariance and the variance have to be corrected using the error covariance and variance σ_{12} and σ_{11} estimated at the first stage. Second, since unit value responds to price with an elasticity of θ_2, which is not in general equal to unity, budget shares respond to unit values with a coefficient θ_1/θ_2, rather than the coefficient θ_1 that would be obtained if θ_2 were unity. This second problem cannot be corrected from the data, but requires application of the quality model of Section I. In particular, (10) gives θ_2 as $1+\beta_2\varepsilon_p/\varepsilon_x$. In the model defined by (14) and (15), neither the price elasticity ε_p nor the expenditure elasticity ε_x are constant, but since w is the product of the unit value and the quantity divided by the budget, we have $\partial \ln w/\partial \ln p = \theta_2 + \varepsilon_p = \theta_1/w$, and $\partial \ln w/\partial \ln x = \beta_2 + \varepsilon_x - 1 = \beta_1/w$, so that, first substituting these expressions in the formula for θ_2, and then replacing θ_2 by θ_1/ϕ, we get, after some rearrangement

$$\theta_1 = \phi\{\beta_1 + w(1-\beta_2)\}/(\beta_1 + w - \phi\beta_2). \tag{21}$$

An estimate of θ_1 can be obtained by replacing the quantities on the right-hand side by their estimates, although notice that θ_1 will generally vary with the budget share. Our estimates below will be calculated at the sample mean of the w's. Note that, if $\beta_2 = 0$, so that there are no income effects on quality, then $\theta_1 = \phi$, and the estimator (20) requires no further correction.

In summary, there are three stages to the estimation. At the first, OLS applied to the within-cluster data yields estimates of the effects of total expenditure and household characteristics, as well as of error measurement variances and covariances. At the second state, the effects of the budget and characteristics are netted out, and cluster averages of the "corrected" budget shares and unit values calculated. A regression of shares on unit values, corrected for measurement error, yields an estimate of the ratio of the responses to price of the share and the unit value. At the third and final stage, the effect of price on the budget share is extracted from the ratio by use of the theory linking quality and quantity elasticities. It seems not to be the case that there exists any obvious instrumental variable estimator that will short-circuit the first two stages. For example, if (15) is used to express price in terms of unit value and the result substituted into (14), we obtain a relation between share and unit value in which the unit value term is correlated with the error term. The cluster dummies would seem to be likely instruments, but it is easy to see that this is equivalent to replacing individual unit values by the cluster means of unit values, and these, like the individual observations, are still contaminated by measurement error. Indeed, it is precisely this contamination that requires the use of the errors-in-variable estimator in (20) rather than OLS.

The estimator $\tilde{\phi}$ given by (20) can be compared with the estimator $\hat{\phi}$, where $\hat{\phi}$ is the same estimator with the tildes removed, that is, the estimate with the first-stage parameters known. $\hat{\phi}$ is a standard errors in variables estimator (see Wayne Fuller, 1987, p. 108), and its asymptotic variance is given by

$$(22)\quad v(\hat{\phi}) = (m_{22} - \sigma_{22}/\tau)^{-2} \times \{\pi^0 m_{22} + (m_{12} - \phi m_{22})^2\} (C-1)^{-1},$$

$$(23)\quad \pi^0 = m_{11} - 2m_{12}\phi + m_{22}\phi^2,$$

where m_{11} is the variance of $y_{1.c}$ and m_{12} and m_{22} are, respectively, the covariance and variance in equation (20). This formula understates the variance of $\tilde{\phi}$ because it ignores the estimation uncertainty associated with the β's and σ's. The correct formula is given in the Appendix; in practice, and for these data, (22) is an extremely accurate approximation. The variance of the estimate of θ_1 calculated from (21) can be obtained by application of the "delta" method.

III. Empirical Results

The data are taken from the *Enquête Budget Consommation*,[1] which collected data from a random national sample of households in the Côte d'Ivoire during the calendar year 1979. There are 1200 urban households in the sample, 522 in Abidjan, and 678 in other towns. The remaining 720 households are divided between the northern Savannah (264 households), and the coastal East and West Forest regions, with 312 and 144 households, respectively. Although the survey was a very ambitious one, not all of the data collected are now usable, and we are limited to only a fraction of them. Data are available on the value and weight of each purchase during "last week" for 100 different foodstuffs, as well as on the volume and imputed value of "autoconsommation," foods grown, manufactured, or captured for own consumption. There are also data on a limited number of household characteristics including the ages and sexes of family members and the household location.

Each of the rural clusters contains twelve households, so that there are 26 East Forest clusters, 12 West Forest clusters, and 22 Savannah clusters. Since each of these clusters was visited during four different seasons of the year, there will generally be genuine (seasonal) price variation within each cluster. To deal with this, I treat clusters in different quarters as different clusters, so that there are effectively 240 rural "clusters" or more accurately cluster-quarters. The urban clusters are naturally much less dispersed than those in the countryside, so that urban households might easily buy commodities in clusters different from those in which they live. In consequence, I confine the presentation here to the rural results; even so, and although the results are not the same, the methodology appears to work just as well for urban households. Cluster-quarters in which no households purchase the good in the market have to be excluded since no unit value can be calculated; there are a total of 49 such clusters out of 240. Although such exclusions may generate some selectivity bias, note that the situation is much better than would be the case if we had to exclude all households that made no market purchases.

The variables in the specifications (14) and (15) are defined as follows. The income or expenditure variable is total annualized household expenditure on food divided by the number of people in the household. Expenditure on food rather than total expenditure is necessitated by absence of data on the latter, and will be theoretically acceptable if food is separable in preferences. Annualized expenditure is thirteen times the total value of purchases and imputed consumption observed over the total of the four weekly visits. The z variables are household demographics; we have data on the numbers of household members by sex in each of seven age-groups. The first thirteen z variables are

[1]The underlying data from this survey, as well as those from the *Living Standards Survey* used later in the paper are the property of the government of the Côte d'Ivoire. The tapes are lodged with the Welfare and Human Resources Division of the World Bank, to whom requests for access should be directed.

the ratios of each of these numbers to total household size (the effect of the fourteenth can be inferred from the intercept), and I also enter the logarithm of household size to allow for the possibility that demands are not linearly homogeneous in household numbers and total expenditure taken together. Expenditures and quantities were recorded at the purchase level, but are here aggregated so that unit values were derived by dividing total expenditures for the household in the relevant week by total quantities in kilograms for the same week. Nonpurchased quantities (home-produced or hunted goods) and the corresponding imputed expenditures were then added to market expenditures to give the total from which the budget shares are formed.

The detailed analysis here is confined to meat; summary results will be given for four other categories of expenditure, fresh fish, other fish (dried and smoked), cereals (rice, maize, millet, etc.), and starches (cassava, yams, potatoes, and plantains). Table 2 shows a preliminary analysis of the unit values that tests for regional and temporal price differences, as well as for quality effects. The vertical panels refer to regressions in which the number of included variables increases from left to right. The fourteen demographic variables as defined above are included in all of the regressions, as is $\ln(PCX)$, the logarithm of per capita total food expenditure. The quality elasticity of meat is small or zero, although for the urban households (not shown), there is a statistically significant but still small (0.10) effect. Greater variety is typically available in urban markets, so there is presumably more scope for a quality effect than in the countryside. The second, third, and fourth panels explore the effect of adding locational and time variables. The second pair of regressions includes dummies for the regions and for the quarters. Meat prices are very much higher in the East Forest than in either West Forest or in the omitted Savannah region. In the third panel the regions are replaced by a set of cluster dummies for the 59 rural clusters. The F-statistics in the bottom panel relate to these regressions and test for the significance of the cluster, quarter, and demographic effects. The demographic effects are jointly insignificant, while the seasonal effects generate F-statistics that are significant at conventional levels. The cluster effects generate a very large F-value, one large enough to pass even the (very stringent) test proposed by Gideon Schwartz (1978), which in this context is that F should be larger than the logarithm of the sample size. The test itself is of no great moment, but the strength and significance of the cluster effects in the countryside is very important because it indicates the existence of the cross-cluster price variation that is necessary to make possible the estimation of the price elasticities. The fourth and final panel in Table 2 shows the quality elasticity from the within-cluster estimator; in this regression the same cluster in two different quarters is treated as two separate clusters, so that geographical and temporal dummies are fully interacted. It is this estimate that is taken forward to the calculation of the price elasticities. Estimation of the within-cluster Engel curve for meat yields an estimate of β_1 of .052 with a standard error of .01; it is worth noting that without allowing for the cluster effects, though with regional dummies, the estimate of 0.026 is only half the size, so that the estimated expenditure elasticity from the within-cluster regression is twice as far from unity as that from the whole sample regression, 1.37 versus 1.19. Cluster effects are also important in the share equation and it is at least plausible that some of this importance comes from prices. The within-cluster share and unit value regressions have an estimated covariance, σ_{12}, of 0.00845, corresponding to a correlation coefficient of 0.07, so that, in this instance, the reporting errors in expenditures and unit values are close to being independent. Such a finding implies a negative covariance between the reporting errors in quantities and unit values, an implication which is confirmed if we estimate not a share log unit value pair of equations, but a log quantity/log unit value pair; see my paper (1986).

The second stage of estimation moves from within- to between-cluster analysis. Once the effects of expenditure and demographics have been removed from the shares and the unit values, the intercluster covariance between

TABLE 2—MEAN UNIT VALUES: ln V

Rural: $n=631$, $\ln(n)=6.45$	Regression 1	Regression 2	Regression 3	"Within" Regression 4
ln PCX	.088 (.037)	.030 (.038)	.056 (.045)	.065 (.042)
East Forest	–	.255 (.053)	–	–
West Forest	–	.007 (.070)	–	–
Clusters	–	–	×	×
Q2	–	−.173 (.061)	−.170 (.052)	×
Q3	–	−.142 (.059)	−.114 (.050)	×
Q4	–	−.085 (.061)	−.063 (.053)	×
R^2	.052	.109	.454	–

Note: Demos, $F=1.25$, $p=.24$; quarters, $F=3.97$, $p=.01$; clusters, $F=6.72$; $p<10^{-4}$.

TABLE 3—BUDGET SHARES AND ESTIMATES OF QUALITY AND QUANTITY ELASTICITIES

	w	β_2	ε_x	ε_p
Meat	0.139	0.065 (.04)	1.305 (.09)	−0.793 (.12)
Cereals	0.201	0.040 (.02)	1.091 (.07)	−1.076 (.30)
Starches	0.310	0.023 (.04)	0.840 (.06)	−0.847 (.10)
Fresh Fish	0.050	0.026 (.02)	0.682 (.11)	−1.575 (.26)
Other Fish	0.080	0.030 (.02)	0.536 (.05)	−1.189 (.14)

shares and unit values is 0.0074, and the intercluster unit value variance is 0.3250, so that the OLS estimate of the response of the share to the log price, uncorrected for measurement error or quality effects, is the ratio of these two numbers, or 0.0227. The "average" cluster sizes τ and τ^+ are 11.6 and 1.98, respectively, so that to correct for measurement error using (20), 0.0084 (the estimate of σ_{12}) divided by 11.6 is subtracted from 0.0074, and the result divided by 0.3267 less 0.1074 (the estimate of σ_{22}) divided by 1.98, giving a result of 0.0245 with an estimated standard error of 0.0173. If unit values moved one for one with price, so that $\theta_2=1$, this would be the final estimate of θ_1, the response of the budget share to price. As it is, the estimated quality elasticity of meat is small, 0.065, so that the application of (21) makes little difference, with a final quality-corrected estimate of θ_1 of 0.0235 with a standard error of 0.0166. The corresponding estimate of θ_2 is 0.9604, so that the final price and (total food) expenditure elasticities for the quantity of meat are −0.793 and 1.305. By contrast, a direct regression of the meat share on log unit value, log PCX, the demographics and regional dummies yield estimates of the same elasticities of −0.450 and 0.775. The differences lie in the treatment of quality, and of measurement error, as well as in the fact that the direct regression can include only those households that record market purchases of meat and for whom a unit value index is directly available.

Table 3 shows the full set of quality and quantity elasticities for the five goods; for each, the food budget share is given first, followed by the estimates of the expenditure elasticity of quality, the expenditure elasticity of quantity, and the price elasticity of quantity. While all of these numbers seem reasonable, there is little with which to compare them, and the model developed in this paper is essentially exactly identified, so that it is difficult to explore the validity of the assumptions that lie behind the estimates. However, there now exist new, and very different data from the Côte d'Ivoire that allow some limited comparisons to be made. In 1985, the World Bank, in conjunction with the government of the Côte d'Ivoire, conducted a *Living Standards Survey* which

TABLE 4—SOME COMPARISON TOTAL EXPENDITURE AND PRICE ELASTICITIES: CÔTE D'IVOIRE, 1985

	ε_x	ε_p		ε_x	ε_p
Beef	1.56	−1.91	**Maize**	0.52	−1.19
Fish	0.74	−1.31	**Yams**	1.00	−1.49
Imported Rice	0.73	−1.40	**Plantain**	0.95	−1.41
Domestic Rice	0.73	−1.02	**Cassava**	0.85	−0.91

collected a large volume of household survey data. Although household expenditures were surveyed, no data were collected on physical quantities. However, a complementary survey gathered data on prices, by direct observation in the markets used by the households in the survey. It is therefore possible to associate market prices with individual households, and to examine the relationship between budget shares, household total expenditures, and the market prices. Of course, this comparison is far from perfect; without quantity data, no allowance can be made for quality effects, the market price data are also subject to considerable measurement error, and the definitions of the goods in the two surveys are not the same. However, the analysis given above suggest that the quality effects may not be very large, so the comparisons are worth making.

Table 4 shows estimates for some of the relevant goods; the numbers are elasticities with respect to total expenditure, not just food, while the price elasticities are taken holding total expenditure constant as opposed to food expenditure. In consequence, the expenditure elasticities in Table 4 ought to be rather less than those in Table 3, while if food as a whole is not very price elastic, the same will be true of the price elasticities. Allowing for the usual degree of uncertainty, the expenditure elasticities correspond rather well between the tables, while the price elasticities in Table 4 are larger than would be expected from Table 3, perhaps because most of the goods in the former are more narrowly defined. Given the difficulties of estimating price elasticities in any context, I view these results as being encouraging, although a final verdict on the method will have to await further experiments in other countries.

APPENDIX: DERIVATION OF STANDARD ERRORS

This brief Appendix derives a formula for the asymptotic variance of $\tilde{\phi}$ in equation (20) that takes into account the fact that the y's and σ's are estimated, not known. Write $\tilde{m}_{12}$ and $\tilde{m}_{22}$ for the covariance and variance in (20) and $\hat{m}_{12}$ and $\hat{m}_{22}$ for the corresponding magnitudes using the unknown $y_{1.c}$ and $y_{2.c}$. Then, ignoring terms of higher order,

$$\text{(A1)}\quad \sqrt{C}(\tilde{m}_{12}-\hat{m}_{12}) = -\left(C^{-1}\Sigma y_{1.c}w'_{.c}\right)\left\{\sqrt{C}\left(\tilde{b}_2-b_2\right)\right\} - \left(C^{-1}\Sigma y_{2.c}w'_{.c}\right)\left\{\sqrt{C}\left(\tilde{b}_1-b_1\right)\right\},$$

where $w_{.c}$ are the cluster means of the variables included in the first-stage regressions, expressed as deviations around the grand mean. Using (A1) and the similar expression for $\tilde{m}_{22}$ to expand (20) around the true value ϕ,

$$\text{(A2)}\quad (\tilde{\phi}-\phi) = (\hat{\phi}-\phi) - \left\{(s_1-2\phi s_2)'\left(\tilde{b}_2-b_2\right)+s_2'\left(\tilde{b}_1-b_1\right)\right\} - \left(\tau^+ m^*_{22}\right)^{-1}\left\{(\tilde{\sigma}_{12}-\sigma_{12}) - \rho\phi(\tilde{\sigma}_{22}-\sigma_{22})\right\},$$

where ρ is the ratio τ^+/τ, m^*_{22} is $m_{22}-\sigma_{22}/\tau^+$, and s_1 and s_2 are the probability limits of $\Sigma y_{1.c}w'_{.c}/Cm^*_{22}$ and $\Sigma y_{2.c}w'_{.c}/Cm^*_{22}$, respectively. The estimates of the b's and the σ's are asymptotically independent, and both are independent of $\hat{\phi}$, so that the variance has three terms corresponding to the three terms on the right-hand side of

(A2). Hence,

$$(A3)\quad v(\tilde{\phi}) = v(\hat{\phi}) + \left\{\sigma_{22}(s_1' - 2\phi s_2')(X_2'X_2)^{-1} \times (s_1 - 2\phi s_2) + \sigma_{11}s_2'(X_1'X_1)^{-1}s_2 - 2\sigma_{12}(s_1' - 2\phi s_2')(X_1'X_1)^{-1}s_2\right\} + (\tau^+ m_{22}^*)^{-2}(n-C)^{-1} \times \left\{\pi^1\sigma_{22} + (\sigma_{12} - \rho\phi\sigma_{22})^2\right\},$$

where $(X_1'X_1)$ and $(X_2'X_2)$ are the two moment matrices from the within-cluster regressions, $\pi^1 = \sigma_{11} - 2\rho\phi\sigma_{12} + \sigma_{22}\rho^2\phi^2$, and $v(\hat{\phi})$ is as given by (22).

REFERENCES

Chernichovsky, Dov and Meesook, Oey Astra, "Patterns of Food Consumption and Nutrition in Indonesia," unpublished paper, Washington: The World Bank, 1982.

Cramer, J. S., "Interaction of Income and Price in Consumer Demand," *International Economic Review*, June 1973, *14*, 351–63.

Deaton, Angus, "Quality, Quantity, and Spatial Variation of Price," Research Program in Development Studies Working Paper No. 127, Woodrow Wilson School, Princeton, NJ, 1986.

______ "Estimation of Own and Cross Price Elasticities from Household Survey Data," *Journal of Econometrics*, September/October 1987, *36*, 7–30.

Fuller, Wayne A., *Measurement Error Models*, New York: Wiley & Sons, 1987.

Houthakker, Hendrik S. and Prais, Sigbert J., "Les Variations de Qualité dans les Budgets de Famille," *Économie Appliquée*, January-March 1952, *5*, 65–78.

Pitt, Mark M., "Food Preferences and Nutrition in Rural Bangladesh," *Review of Economics and Statistics*, February 1982, *65*, 105–14.

Prais, Sigbert J. and Houthakker, Hendrik S., *The Analysis of Family Budgets*, New York: Cambridge University Press, 1955.

Schwartz, Gideon, "Estimating the Dimension of a Model," *Annals of Statistics*, March 1978, *6*, 461–4.

Timmer, C. Peter, "Is There "Curvature" in the Slutsky Matrix?," *Review of Economic Statistics*, August 1981, *63*, 395–402.

______ **and Alderman, Harold,** "Estimating Consumption Parameters for Food Policy Analysis," *American Journal of Agricultural Economics*, December 1979, *61*, 982–87.