Journal of Econometrics 36 (1987) 7–30. North-Holland

ESTIMATION OF OWN- AND CROSS-PRICE ELASTICITIES FROM HOUSEHOLD SURVEY DATA*

Angus DEATON

Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA
Living Standards Unit, The World Bank, Washington, DC 20433, USA

A methodology is proposed that allows the use of expenditure and quantity data from household surveys to estimate a system of demand equations including freely estimated own- and cross-price elasticities. Unit values of purchases by individual households are used to indicate spatial variation in prices, and the fact that households in such surveys are geographically clustered is used to separate out the effects of measurement error from genuine price variation. The method also allows for variation in quality with household incomes and with prices. Results for a five good system for the Cote d'Ivoire are presented and discussed.

1. Introduction

The estimation of price elasticities of demand, especially cross-price elasticities, is an undertaking of great difficulty. Even in developed countries, where data, particularly time series data, are relatively plentiful, many decades of intellectual activity have not succeeded in estimating a set of parameters or elasticities that can be used with real confidence. In developing countries, time series data are a good deal less plentiful, and even less is known about the way in which consumers respond to changes in prices. Yet a knowledge of price responses is required for any intelligent analysis of tax or subsidy reform, a topic of central importance for policy in developing countries. In this paper, I explore a technique that exploits price variation over space to estimate a system of price elasticities using household survey data. The data used here come from a 1979 household survey of the Cote d'Ivoire, but similar data are available for a wide range of developing countries, so that the technique should have wide applicability.

That there is considerable spatial variation in prices in most developing countries should not be doubted. Transport is frequently difficult and expensive, so that even efficient and fast-moving arbitrageurs cannot bring prices

*Dwayne Benjamin provided exemplary research assistance and many helpful suggestions. Anne Case and Rick Monteverde detected errors in an earlier draft. I am grateful to them and to a referee of this journal. 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.

0304-4076/87/$3.50

into uniformity. However, household surveys rarely collect data on market prices. Instead, households are asked to report expenditures on each good over a specified period, together with the quantities purchased, so that, for example, two kilograms of rice might have been purchased at a cost of 220 francs. The ratio of expenditure to quantity, the unit value of the purchase, gives an *indication* of the market price, though it will also be affected by the type or quality of rice that is bought by the household.

The basic model estimated here is one in which market prices are treated as unobservable variables, which directly determine quantities purchased and are 'indicated' by unit values. There are, however, a number of complicating factors that need to be taken into account. Firstly, because unit values reflect quality choice and because quality choice will generally be affected by prices, all prices affect the unit values of all goods, while each unit value is likely to vary less than proportionately with its own price. The paper develops a model of this process that at a later stage permits the identification of price elasticities from the data. The second complication is even more important. The fact that both expenditures and quantities are measured with error means that unit values, which are the ratios of expenditures to quantities, are also error-ridden, with errors that are likely to be negatively correlated with quantities. If this spurious correlation is not dealt with, little confidence can be placed in the estimates.

The paper follows its companion, Deaton (1986), in using the structure of the household survey to identify the separate effects of measurement error and variation in market prices. Although the sample is a national one, households belong to 'clusters,' usually of about twelve units, all of which are in close geographical proximity. This is typical of most such survey designs; if the respondents are grouped together with a dozen or so in the same village, enumerators can visit all of them at the same time without incurring the travel costs that would be involved if each household were randomly located on the map. The key assumption is that there is no variation in market prices within each cluster, so that within-cluster estimators of unit value and quantity equations can identify Engel and quality effects without contamination by the (unobservable) variation in market price. These within estimators can also be used to calculate the effects of the measurement error, since only the spurious variances and covariances will be present within clusters. The demand system itself is then estimated from the inter-cluster variation in corrected quantities and unit values, with due allowance for the predicted effects of the measurement errors as well as for the influence of price on quality.

In the companion paper, Deaton (1986), I developed a similar model in which each good was considered in isolation, with quantity and unit value responding only to own prices. In this paper, the main contribution is to extend the analysis to allow for cross-price effects. This turns out to be a much less straightforward exercise than I had initially supposed. The analysis of quality effects has to be extended to allow the price of each good to influence the qualities chosen for every other good, and this greatly complicates the correction of the estimated price elasticities for the effects of prices on qualities. The estimation of the model is also more complex, and the method is a far from trivial extension of the one-good one-price model. The exposition of this paper is self-contained, although I have avoided duplication as much as possible by limiting discussion of matters that are essentially unchanged from the simpler analysis.

The paper is organized as follows. Section 2 is concerned with the effects of price on quality and on unit values. Quality is modelled as the choice of commodities within a group, and a weakly separable structure is proposed wherein the effects of price on quality are related to the effects of total expenditure or income on quality. Section 3 outlines the econometric model and the procedures for estimation and calculation of standard errors. Section 4 describes the data and the empirical results.

2. Quantity, quality and price

In the empirical work below, I shall be working with such commodities as meat, fish and cereals. None of these is a homogeneous good but consists of a large number of components, each of which is itself available in many varieties. For example, in the data it is possible to separate fresh fish, dried fish and smoked fish, but even a category such as 'fresh fish' encompasses many different kinds of fish, some of which are more expensive than others. Since better-off households will tend to buy the better and more expensive varieties, the unit value of fish purchases will tend to be positively correlated with household income, a phenomenon studied by Prais and Houthakker (1955). Quality choice may also depend on prices, with both own- and cross-price effects possible. Consumers may well respond to high fish prices by buying cheaper varieties, but the same thing may happen in response to increases in the price of other staples such as meat and cereals.

I denote a groups of goods (fish) by the letter G, so that p_G is the price vector of goods in the group. Prices vary over space, and the location where the price is observed is denoted by (cluster) c. When there is no risk of confusion, I shall drop the cluster suffix. The corresponding quantity vector is q_G, so that $E_G = p_G \cdot q_G$ is expenditure on group G. It is necessary to be able to change the *level* of prices in each cluster while holding constant the structure of relative prices. To this end, define the price of the group as a whole, λ_G, as some linear homogeneous function of p_G, for example as a constant weight Laspeyres price index. I can then write

$$p_G = \lambda_G \cdot p_G^*, \tag{1}$$

where p_G^* represents the relative price structure. Corresponding to the group price λ_G is a group quantity Q_G defined as $k_G^0 \cdot q_G$, where k_G is a vector of constants defining the dimension of quantity in which we are interested. In the empirical analysis, I take k to be the vector of units, so that Q is simply the *weight* of purchases. However, k_G could also be the vector of calories per kilogram of each of the component goods in G, so that Q_G would be aggregate calories from fish or whatever group is being considered. Other schemes for adding apples and oranges could also be handled in the same way.

The unit value for G is denoted V_G and is given by

$$V_G = \frac{E_G}{Q_G} = \lambda_G \frac{p_G^* \cdot q_G}{k_G^0 \cdot q_G}. \tag{2}$$

Clearly, if better-off households purchase bundles that contain a larger share of high price per kilo items, i.e., items for which the ratio p_{Gi}^*/k_{Gi} is high, then V_G will rise with income. Indeed, the income elasticity of V_G is precisely Prais and Houthakker's 'quality elasticity'. Taking logarithms of eq. (2) gives

$$\ln V_G = \ln \lambda_G + \ln \rho_G, \tag{3}$$

where ρ_G, corresponding to the ratio term in (2), is the measure of quality. By (3), and if everything is measured in logarithms, unit value is the sum of quality and price.

So far, everything is definitional. To derive useful results about the effect of price on quality, I assume that each of the goods, fish, meat, cereals, and so on, form a separable branch of preferences. The utility function is then written

$$u = v\{v_1(q_1), v_2(q_2), \dots, v_G(q_G), \dots, v_M(q_M)\}, \tag{4}$$

so that, for each group G, there exists a set of group demand functions

$$q_G = g_G(E_G, p_G) = g_G(E_G/\lambda_G, p_G^*), \tag{5}$$

where the last step follows by zero degree homogeneity of the demands. Since quality ρ_G is a function of q_G and since both p_G^* and k_G^0 are being held constant, the effects on ρ_G of changes in any price λ_H, say, operate entirely through the term E_G/λ_G in (5). The derivatives are therefore

$$\frac{\partial \ln \rho_G}{\partial \ln \lambda_H} = -\frac{\partial \ln \rho_G}{\partial \ln E_G}\left(\delta_{GH} - \frac{\partial \ln E_G}{\partial \ln \lambda_H}\right), \tag{6}$$

where δ_{GH} is the Kronecker delta. But expenditure is the product of quantity, quality and price, so that

$$\frac{\partial \ln E_G}{\partial \ln \lambda_H} = \delta_{GH} + \frac{\partial \ln \rho_G}{\partial \ln \lambda_H} + \varepsilon_{GH}, \tag{7}$$

where ε_{GH} is the cross- (or own-) price derivative of G with respect to H. Corresponding to the price derivative (6) there is also a derivative with respect to total expenditure (or income) x; this derivative $\partial \ln \rho_G / \partial \ln x$ is the Prais–Houthakker quality elasticity η_G. Because of the separability, the effects of x operate through E_G, so that

$$\frac{\partial \ln \rho_G}{\partial \ln x} = \frac{\partial \ln \rho_G}{\partial \ln E_G} \cdot \frac{\partial \ln E_G}{\partial \ln x} = \frac{\partial \ln \rho_G}{\partial \ln E_G}(\varepsilon_G + \eta_G), \tag{8}$$

where ε_G and η_G are the total expenditure (or income) elasticities of quantity and quality, respectively. Eq. (8) shows that the derivative of subgroup quality with respect to subgroup expenditure, $\partial \ln \rho_G / \partial \ln E_G$, is equal to $\eta_G/(\varepsilon_G + \eta_G)$. Using this result after substituting (7) into (6) yields

$$\frac{\partial \ln \rho_G}{\partial \ln \lambda_H} = \frac{\varepsilon_{GH} \partial \ln \rho_G / \partial \ln E_G}{1 - \partial \ln \rho_G / \partial \ln E_G} = \frac{\varepsilon_{GH} \eta_G}{\varepsilon_G}. \tag{9}$$

(If ε_G is zero, then so will be η_G, and the first equality in (9) remains valid.) For the unit value V_G, which is quality times prices,

$$\frac{\partial \ln V_G}{\partial \ln \lambda_H} = \delta_{GH} + \frac{\eta_G \varepsilon_{GH}}{\varepsilon_G}. \tag{10}$$

Note that, if the income quality elasticity η_G is zero, the unit value of G moves one for one with the price of G and is independent of prices outside the group. If the quality elasticity is non-zero, unit values will typically respond less than one for one to own price, as the quality is adjusted downward through the income effect. Similarly, if there are significant cross-price effects between groups of goods, then non-zero quality elasticities will generate non-zero cross effects between prices and qualities.

Eq. (10) will be used in the econometric work below. What is observed is a relationship between unit values and quantities. Part of this is due to the common influence of income and household characteristics on both, and this can be allowed for directly since income and household characteristics are observed. Market prices also exert an influence on both unit value and on quantity, but since market prices are not observed, the data are not informative about the separate effect on each. Eq. (10), by linking the effects of prices on quality to the effects of total expenditure on quality, gives the additional information that is required for identification.

3. Econometric specification and estimation

For each household i in cluster c, there are data on purchases of a range of goods, with both expenditure and weight provided. I shall work with quantities and with the generated variables, unit values, and postulate the following model:

$$\ln q_{Gic} = \alpha_G^0 + \beta_G^0 \ln x_{ic} + \gamma_G^0 \cdot z_{ic} + \sum_{H=1}^{5} \theta_{GH} \ln p_{Hc} + \left(f_{Gc} + u_{Gic}^0\right), \tag{11}$$

$$\ln v_{Gic} = \alpha_G^1 + \beta_G^1 \ln x_{ic} + \gamma_G^1 \cdot z_{ic} + \sum_{H=1}^{5} \psi_{GH} \ln p_{Hc} + u_{Gic}^1. \tag{12}$$

These two equations provide models for the quantity of good G consumed by household i in cluster c, q_{Gic}, and the associated unit value v_{Gic}. Eq. (11) is a standard double-logarithmic demand function, in which the logarithm of the quantity demanded is linked to total expenditures (in this case total *food* expenditure per capita), $\ln x_{ic}$, a vector of household demographic characteristics, z_{ic}, and the prices of each of the goods in the system. Here, I shall work with the five commodities, meat, fresh fish, dried and smoked fish, cereals and starches, so that there are five prices in each equation. The error term in the quantity equation has two components. The first, f_{Gc}, is a cluster-specific fixed (or random) effect, to be interpreted as the cluster-specific residual in the demand function for good G. I shall treat f_{Gc} as a fixed effect, but no difficulties arise if it is thought of as being random. As is usual with fixed effects, I can allow f_{Gc} to be correlated with the observable explanatory variables, $\ln x$ and z, but I must assume it to be uncorrelated with the unobservable prices p_{Hc}. Clearly, if arbitrary inter-cluster variation in tastes is allowed and if prices vary only between clusters, there would be no possibility of measuring price elasticities. The household specific error component u_{Gic}^0 has an expectation of zero within the cluster and is uncorrelated with all other right-hand-side variables, including the fixed effects. Its existence models the usual inexactness of econometric models as well as the presence of measurement error in quantities.

The unit value equation, eq. (12), shows unit value as the sum of quality and price, with price allowed to affect quality choice. The x and z variables reflect the influence of the household's living standard and demographic composition on the choice of quality, while all of the prices appear in line with the analysis of the previous section, particularly eq. (10) above. There are no fixed effects in this equation; apart from quality effects, unit value is a direct indication of price. Again, the assumption is central to the possibility of using spatial variation in unit values to yield the information about prices that will allow estimation of price responses. As for the quantity equation, there is an idiosyncratic error, u^1_{Gic}, reflecting, among other things, measurement error. It is of the greatest importance that u^0 and u^1 are allowed to be correlated. Since the logarithm of unit value is the difference between the logarithm of expenditure and the logarithm of quantity, measurement errors in the latter must be correlated with errors in the former, unless price is recalled perfectly and used by the respondent to calculate either quantity from expenditure or expenditure from quantity. While this is possible, it is unlikely that all households calculate in this way or that those who do are capable of recalling prices without error. Much of the rest of this paper is concerned with allowing for these correlated measurement errors.

Estimation proceeds in several stages. At the first, eqs. (11) and (12) are estimated by 'within-cluster' ordinary least squares. Cluster means are subtracted from all variables, thus annihilating the prices and fixed effects, but allowing consistent estimation of the income and demographic effects in both equations. Removing cluster means from (11) and (12) gives

$$(\ln q_{Gic} - \ln q_{G\cdot c}) = \beta^0_G(\ln x_{ic} - \ln x_{\cdot c}) + \gamma^0_G \cdot (z_{ic} - z_{\cdot c}) + (u^0_{Gic} - u^0_{G\cdot c}), \tag{13}$$

$$(\ln v_{Gic} - \ln v_{G\cdot c}) = \beta^1_G(\ln x_{ic} - \ln x_{\cdot c}) + \gamma^1_G \cdot (z_{ic} - z_{\cdot c}) + (u^1_{Gic} - u^1_{G\cdot c}), \tag{14}$$

where the '$\cdot$' notation indicates means over all households in cluster c, so that, for example, $\ln x_{\cdot c}$ is the mean of the logarithms of household food expenditure per capita in cluster c. Eqs. (13) and (14) comprise a set of ten classical multivariate regressions with identical right-hand-side variables so that ordinary least squares on each equation is fully efficient. The residuals from each equation can also be used to estimate the variances and covariances of the measurement errors, so that

$$\tilde{\sigma}^{rs}_{GH} = (n - C - k)^{-1} \sum_c \sum_i e^r_{Gic} e^s_{Hic}, \qquad r, s = 0, 1, \quad G, H = 1, 5, \tag{15}$$

where n is the sample size, i.e., the total number of households, C is the total number of clusters, k is the number of variables on the right-hand side of (13) and (14), and e^r_{Gic} is the residual for good G from household i in cluster c for the quantity equation if $r=0$ and the unit value equation if $r=1$.

The estimates of the β's and γ's from the within estimators are the final estimates of these parameters. Write these as $\tilde{\beta}^0_G, \tilde{\beta}^1_G, \tilde{\gamma}^0_G, \tilde{\gamma}^1_G$. To set up the second stage of estimation, define the 'corrected' quantities and unit values by

$$\tilde{y}_{G\cdot c} = \ln q_{G\cdot c} - \tilde{\beta}^0_G \ln x_{\cdot c} - \tilde{\gamma}^0_G \cdot z_{1\cdot c}, \tag{16}$$

$$\tilde{w}_{G\cdot c} = \ln v_{G\cdot c} - \tilde{\beta}^1_G \ln x_{\cdot c} - \tilde{\gamma}^1_G \cdot z_{1\cdot c}. \tag{17}$$

The population counterparts to (16) and (17) are, from (11) and (12),

$$y_{G\cdot c} = \tilde{y}_{G\cdot c} + (y_{G\cdot c} - \tilde{y}_{G\cdot c}) = \alpha^0_G + \sum_{H=1}^{5} \theta_{GH} \ln p_{Hc} + (f_{Gc} + u^0_{G\cdot c}), \tag{18}$$

$$w_{G\cdot c} = \tilde{w}_{G\cdot c} + (w_{G\cdot c} - \tilde{w}_{G\cdot c}) = \alpha^1_G + \sum_{H=1}^{5} \psi_{GH} \ln p_{Hc} + u^1_{G\cdot c}. \tag{19}$$

Define the matrix S as the variance–covariance matrix of the w's and R as the covariance matrix of the w's with the y's, i.e.,

$$\{S\}_{GH} = \text{cov}(w_{G\cdot c}, w_{H\cdot c}), \qquad \{R\}_{GH} = \text{cov}(w_{G\cdot c}, y_{H\cdot c}), \tag{20}$$

while the matrices Ω and Γ are the variance–covariance matrices for the u's, so that, comparing with (15),

$$\{\Omega\}_{GH} = \text{cov}(u^1_{Gic}, u^1_{Hic}) = \sigma^{11}_{GH}, \qquad \{\Gamma\}_{GH} = \text{cov}(u^1_{Gic}, u^0_{Hic}) = \sigma^{10}_{GH}. \tag{21}$$

Then, from (18) and (19), if M is the variance–covariance matrix of the unobservable logarithms of prices, and n_c is the number of households in cluster c,

$$S = \Psi M \Psi' + n_c^{-1}\Omega, \tag{22}$$

$$R = \Psi M \Theta' + n_c^{-1}\Gamma. \tag{23}$$

The sample quantities $\tilde{y}_{G\cdot c}$ and $\tilde{w}_{G\cdot c}$ are used to provide consistent estimators of S and R. It is important to be clear about what quantities are being allowed to tend to infinity when consistency is being discussed. Cluster sizes in household surveys are typically much the same no matter what the sample size, so that when the total number of households increases, it is sensible to assume that the number of clusters increases with the number of households per cluster remaining constant. The data to be used are also of this form; cluster sizes are typically small, but there are a large number of clusters.

Denote by $\tilde{S}$ and $\tilde{R}$ the empirical variances and covariance matrices from the estimates $\tilde{y}_{G\cdot c}$ and $\tilde{w}_{G\cdot c}$ and $\tilde{\Omega}$ and $\tilde{\Gamma}$ the estimates of Ω and Γ from (15). Since the first pairs are derived by averaging over clusters of possibly different sizes, $\tilde{S}-\nu^{-1}\tilde{\Omega}$ with $\nu^{-1}=C^{-1}\sum(n_c^{-1})$ is a consistent estimator of $\Psi M \Psi'$ and $\tilde{R}-\nu^{-1}\tilde{\Gamma}$ of $\Psi M \Theta'$. Hence, defining $\tilde{B}$ as

$$\tilde{B}=(\tilde{S}-\nu^{-1}\tilde{\Omega})^{-1}(\tilde{R}-\nu^{-1}\tilde{\Gamma}), \tag{24}$$

$$\operatorname*{plim}_{C\to\infty} \tilde{B}=(\Psi')^{-1}\Theta'. \tag{25}$$

Note that $\tilde{B}$ is essentially an errors-in-variables version of the standard classical multivariate regression estimator in which the variance–covariance matrix of the explanatory variables and the covariance matrix of the explanatory variables with the variables to be explained are each corrected for the influence of measurement error. The fact that what is estimated is not the matrix of own- and cross-price elasticities Θ, but the 'mixed' matrix $(\Psi')^{-1}\Theta'$, reflects the fact that unit values, even when purged of the effects of income and of demographics, are still contaminated by the influence of price on quality. These effects can be disentangled by the use of the price–quality model of section 2, and in particular from eq. (10). In current notation, (10) may be written

$$\Psi = I + D\Theta, \tag{26}$$

where D is a diagonal matrix with ratios of quality to quantity elasticities on the diagonal, i.e., $d_{GH}=\delta_{GH}\beta_G^1/\beta_G^0$, so that

$$\tilde{\Theta}=(I-\tilde{B}'\tilde{D})^{-1}\tilde{B}' \tag{27}$$

is a consistent estimator of Θ, the matrix of own- and cross-price elasticities.

The remainder of this section is concerned with the derivation of estimates of variances and covariances for the matrix Θ. A simplifying feature of the estimator (27) is that the estimates of B and of D are asymptotically independent since the former uses between-cluster variation in the data, while the latter uses the within-cluster variation. Each column of $\tilde{B}$ is a standard errors in variable estimator, for which variance–covariance matrices can be obtained following, for example, Fuller (1975). However, $\tilde{\Theta}$ involves all of the elements of $\tilde{B}$, so that a full variance–covariance matrix for $\tilde{B}$ is required. If the 'delta' method is applied to (27) and terms of smaller order in probability are ignored,

$$\begin{aligned}(\tilde{B}-B) &= (S-\nu^{-1}\Omega)^{-1}\{(\tilde{R}-R)-\nu^{-1}(\tilde{\Gamma}-\Gamma)\}-(S-\nu^{-1}\Omega)^{-1}\\ &\quad\times\{(\tilde{S}-S)-\nu^{-1}(\tilde{\Omega}-\Omega)\}(S-\nu^{-1}\Omega)^{-1}(R-\nu^{-1}\Gamma)\\ &= (S-\nu^{-1}\Omega)^{-1}\{[(\tilde{R}-R)-(\tilde{S}-S)B]\\ &\quad-\nu^{-1}[(\tilde{\Gamma}-\Gamma)-(\tilde{\Omega}-\Omega)B]\}. \end{aligned} \tag{28}$$

The two terms in square brackets on the right-hand side of (28) are asymptotically independent of one another and so can be dealt with one at a time. Since R and S are submatrices of the full variance–covariance matrix of the y's and w's, and similarly for Ω and Γ, it is convenient to define the full matrices as follows:

$$H=\begin{pmatrix} Q & R' \\ R & S\end{pmatrix}, \qquad \Lambda=\begin{pmatrix} \Sigma & \Gamma' \\ \Gamma & \Omega\end{pmatrix}, \tag{29}$$

where Q is the five-by-five covariance matrix of the 'corrected' quantities $y_{G\cdot c}$ [compare eq. (20) above] and Σ is the corresponding covariance matrix of the errors in the quantity equation, so that Σ has typical element σ^{00}_{GH}, [see eq. (21)]. If $J=(0|I)$, a five-by-ten matrix made up of a five-by-five matrix of zeros and a five-by-five identity matrix, and $P'=(I|-B')$, then (28) can be conveniently rewritten as

$$(\tilde{B}-B)=(S-\nu^{-1}\Omega)^{-1}J[(\tilde{H}-H)-(\tilde{\Lambda}-\Lambda)]P, \tag{30}$$

or in 'vec' notation,

$$\begin{aligned}\operatorname{vec}(\tilde{B}-B) &= \{I\otimes(S-\nu^{-1}\Omega)^{-1}J\}(P'\otimes I)\\ &\quad\times\{\operatorname{vec}(\tilde{H}-H)+\nu^{-1}\operatorname{vec}(\tilde{\Lambda}-\Lambda)\} \end{aligned}\tag{31}$$

$$\begin{aligned}&= \{P'\otimes(S-\nu^{-1}\Omega)^{-1}J\}\\ &\quad\times\{\operatorname{vec}(\tilde{H}-H)+\nu^{-1}\operatorname{vec}(\tilde{\Lambda}-\Lambda)\}.\end{aligned}\tag{32}$$

Given normality, $\tilde{H}$ and $\tilde{\Lambda}$ are Wishart matrices, so that the variance–covariance matrices of $\operatorname{vec}(\tilde{H})$ and $\operatorname{vec}(\tilde{\Lambda})$ are $(H \otimes H)(I+K)$ and $(\Lambda \otimes \Lambda)(I+K)$, respectively, where K is the (100×100) 'commutation' matrix, the matrix of ones and zeros that is defined by $K \operatorname{vec}(A) = \operatorname{vec}(A')$; see Magnus and Neudecker (1986) for an excellent survey of this and related results. Writing N for $(I+K)$, the variance–covariance matrix of $\operatorname{vec}(\tilde{B})$, V_b, say, is given by

$$V_b = \left\{ P' \otimes (S - \nu^{-1}\Omega)^{-1} J \right\}$$
$$\times \left\{ (C-1)^{-1}(H \otimes H) + (n - C - k)^{-1}\nu^{-2}(\Lambda \otimes \Lambda) \right\}$$
$$\times N \left\{ P \otimes J'(S - \nu^{-1}\Omega)^{-1} \right\}. \tag{33}$$

The commutation matrix K has the property that $K(A \otimes B) = (B \otimes A)K$, so that if the Kronecker products in (33) are multiplied out, we reach

$$V_b = (C-1)^{-1}\left\{ P'HP \otimes (S - \nu^{-1}\Omega)^{-1} JHJ'(S - \nu^{-1}\Omega)^{-1} \right\}$$
$$+ (C-1)^{-1}\left\{ P'HJ'(S - \nu^{-1}\Omega)^{-1} \otimes (S - \nu^{-1}\Omega)^{-1} JHP \right\} K$$
$$+ (n - C - k)^{-1}\nu^{-2}\left\{ P'\Lambda P \otimes (S - \nu^{-1}\Omega)^{-1} J\Lambda J'(S - \nu^{-1}\Omega)^{-1} \right\}$$
$$+ (n - C - k)^{-1}\nu^{-2}\left\{ P'\Lambda J'(S - \nu^{-1}\Omega)^{-1} \right.$$
$$\left. \otimes (S - \nu^{-1}\Omega)^{-1} J\Lambda P \right\} K. \tag{34}$$

The matrices J and P can be removed from this expression and the whole somewhat simplified by noting that $P'HP = Q - B'R - R'B - B'SB$, $JHJ' = S$ and $JHP = R - SB$, with similar equations for the expressions involving Λ.

The variance–covariance matrix of $\tilde{\Theta}$ can now be obtained by expanding (27) around its true value; hence

$$\tilde{\Theta} - \Theta \approx (I - B'D)^{-1}(\tilde{B}' - B')(I + D\Theta) + \Theta(\tilde{D} - D)\Theta, \tag{35}$$

or, stacking the matrices,

$$\operatorname{vec}(\tilde{\Theta} - \Theta) = \left\{ (I + \Theta'D) \otimes (I - B'D)^{-1} \right\} K \operatorname{vec}(\tilde{B} - B)$$
$$+ (\Theta' \otimes \Theta)\operatorname{vec}(\tilde{D} - D). \tag{36}$$

Since $\tilde{B}$ and $\tilde{D}$ are asymptotically independent, the variance–covariance matrix of $\tilde{\Theta}$ may readily be calculated from (36), viz.

$$V\{\mathrm{vec}(\tilde{\Theta})\} = \{(I+\Theta' D)\otimes(I-B'D)^{-1}\}KV_bK \times\{(I+D\Theta)\otimes(I-DB)^{-1}\} + (\Theta'\otimes\Theta)V_d(\Theta\otimes\Theta'), \quad (37)$$

where V_b and V_d are the variance–covariance matrices of vec(B) and vec(D), respectively. V_b is given by eq. (34). The diagonal matrix D, defined above by eq. (27), involves first-stage parameters only, so that the variance–covariance matrix V_d can be calculated using the usual rules for variances of SUR parameters. If d_i is the ith element of the diagonal of D and if the variance of β_G^0 estimated from (13) is $\xi\sigma_{GG}$, i.e., ξ is the diagonal element of $(X'X)^{-1}$ corresponding to the logarithm of total food expenditure, then it is easily shown that

$$\mathrm{cov}(\tilde{d}_i, \tilde{d}_j) = \xi\{\omega_{ij} + d_i d_j\sigma_{ij} - (d_i + d_j)\gamma_{ij}\}/(\beta_i^0\beta_j^0), \quad (38)$$

where ω_{ij}, σ_{ij} and γ_{ij} are elements of the matrices Ω, Σ and Γ defined in (29) and (21) above. Given (38), the variance–covariance matrix of vec(D) can be constructed.

In the next section, I shall require one further covariance matrix, that of vec(Θ) and β^0. The estimate of Θ in (27) depends on B and on D, and it is only the estimate of the latter that is asymptotically correlated with the first-stage estimates β^0. By a similar calculation to that in the previous paragraph, the covariance of d_i and β_j^0 is

$$\mathrm{cov}(\tilde{d}_i, \tilde{\beta}_j^0) = \xi(\gamma_{ij} - d_i\sigma_{ij})/\beta_i^0. \quad (39)$$

From this, the (25×5) covariance matrix of vec(D) and β^0, Υ, say, can be constructed, so that, finally,

$$\mathrm{cov}\{\mathrm{vec}(\Theta), \beta^{0\prime}\} = (\Theta'\otimes\Theta)\Upsilon. \quad (40)$$

4. Data and results

4.1. Data and empirical procedures

The data used are taken from the 1979 *Enquete Budget Consommation* of the Cote d'Ivoire, and relate to 1920 households, 1200 of which are located in urban areas and 720 in rural areas. 522 of the urban households are located in the principal city, Abidjan, and 678 in other towns. The rural households are distributed between the northern Savannah (264) and the two coastal areas West Forest (144) and East Forest (312). The survey was carried out throughout the calendar year of 1979, so that there is price variation between households according to the time at which they were interviewed. Each rural household was visited four times, once during each calendar quarter, while urban households were visited only once. All households in each cluster were interviewed at (approximately) the same time so that the assumption of no price variation within the cluster is a reasonable one, although for rural clusters it is necessary to keep separate each of the repeat visits. Each of the rural clusters contains 12 households, so that the rural sample comprises 60 clusters – 22 in the Savannah, 12 in the West Forest and 26 in the East Forest. Because each of these is visited on four occasions and because I require that there be no price variation within the cluster, I treat the 4×60 clusters as 240 separate 'clusters'. I make no attempt to follow individual rural households from one visit to the next; as discussed below, I am limited to households that made at least one purchase of *one* of the goods and do not wish to impose the further limitation that households should make purchases in each of the four quarters. There are 13 clusters in Abidjan, comprising 522 households in total, and 21 clusters in the other towns, containing between them a total of 678 households. Each of the urban households was visited on only one occasion, but some of the clusters were visited more then once, so that, once again, the same cluster on a different occasion is treated as if it were distinct.

The survey collected data on 99 different foods, including expenditures and weights for each purchase during the previous week. From these, data on four groups of commodities were constructed. These are as follows: (i) meat, comprising beef, agouti (bush rat), palm squirrel, venison, game animals, game birds, chicken, guinea fowl, pork, mutton and goat (from which list beef is by far the most important), (ii) fresh fish, (iii) other fish, comprising dried and smoked fish, (iv) starches, comprising early yams, late yams, taro, sweet potatoes, potatoes, fresh plantain, plantain meal, fresh cassava, cassava meal and attieke, (v) cereals, comprising bleached rice, rice flour, corn on the cob, corn meal, millet meal, sorghum, grain fonio, wheat bread, macaroni and other dried products. Data are available for each of the components listed, and the groupings were chosen partly because they conform to common usage and partly to make groups large enough so that a reasonable number of purchases of each were actually recorded. Together, the five commodity groups account for 50–60% of all purchases, with meat and fish expenditures rather more important than cereals or starches. The data used below come from aggregating *purchase* data for each household to reach a total weight and expenditure of meat, fresh fish, other fish, starches and cereals over the previous week.

Apart from food purchases, I have access to a rather limited selection of other data. The total expenditure (or income) variable is taken to be the total value of annual food consumption divided by the number of people in the household; in the rural areas this measure contains a large component of imputed expenditures. Valuation of non-market consumption is always difficult, and since there is little such consumption in the urban areas, I allow for possible inconsistencies between the two sectors by dealing with each separately. The z variables in the within cluster regressions are taken to be household demographics; there are fourteen such variables, one each for the number of males and females in the seven age groups, 65 and over, 55–64, 35–54, 18–34, 12–17, 6–11 and 0–5.

Perhaps the major unsolved difficulty here, as in any study using household level purchase data, is the question of how to deal with households that record zero purchases. For meat, which is a very important part of the budget and diet in Cote d'Ivoire, there are on average 6.5 purchases per cluster in the Savannah, so that during the survey week there is approximately one purchase for every second household. This figure is low because rural households obtain a good deal of their meat from their own livestock or from occasional hunting, and the data relate to market purchases. In Abidjan, where there is less livestock and wild game, the typical household makes 3.25 purchases per week, and 70% of households report some purchase of meat during the survey. However, in the East and West Forest regions, where fish tends to be favored over meat, only a quarter of households record meat purchases in a given week. The zero problem is therefore pervasive in these data. It is important to note that for important and largely necessary categories such as those used here, most zero purchases do not reflect zero consumption; it just happens that the household does not purchase the commodity during the survey week. Other surveys that use longer recall periods typically find very many fewer zero records. Hence, although there has been some recent progress in modelling zero consumption [see, in particular, Lee and Pitt (1986)], such models cannot describe the current data. In this study, I simply confine attention to households that record positive purchases. As argued in Deaton (1986), this is correct if all households consume the good, But purchases are randomly distributed over time with a distribution that is unaffected by prices or other variables that determine purchases. This is not likely to be true, if only because rural households are likely to substitute between own and market consumption in response to price fluctuations, but I do not know of any better way of dealing with the problem. The estimates cannot therefore be regarded as consistent for the parameters of household preferences; the regressions are estimates of the expectations of amounts purchased conditional on purchases being positive.

All of the estimates that are presented below are calculated separately for the rural and urban sectors. This is necessary not only because of the differences in imputation procedures between the two sectors, but also because the cluster structures are not the same. Urban clusters are larger than rural clusters, and because large towns have several clusters, they are not all widely separated. One possibility would be to confine attention to rural areas where the assumption of separate markets makes sense. However, even clusters in the same city are geographically separated, so that it is at least worth trying out the assumption that all households in the cluster face the same prices. Even so, it is clearly wise to place greater confidence in the results from the countryside.

The econometric procedures discussed in section 3 are implemented as follows. The within-cluster regressions (13) and (14) are run, two equations for each of the four goods, using only those households that record positive purchases of the good in question. Note that households are not required to purchase all four goods, so that, for example, the sample of households included in the meat regressions will not be the same as the sample in the fish regressions. These regressions yield estimates of the vectors β^0, the total food expenditure elasticities, and β^1, the Prais–Houthakker quality elasticities with respect to total outlay on food. The within regressions are also used to calculate the variances and covariances of the two vectors u^0 and u^1 as per eq. (15). Given the prevalence of zero purchases, there is little point in attempting to calculate the cross-commodity covariances from the within estimator. Each such covariance could only be based on those households that recorded positive purchases of both goods, and is therefore likely to be poorly estimated. I therefore assume that u^0_G and u^1_G, although not independent of one another, are independent of u^0_H and u^1_H for $G \neq H$. The important issue here is the covariance between u^0_H and u^1_H that results from measurement error, and this is estimated directly according to (15). Note that the cross-commodity independence assumption means that the matrices Σ, Γ and Ω that make up Λ in (29) are each diagonal.

Given the within-cluster estimates, cluster means of 'corrected' quantities and unit values are calculated according to eqs. (16) and (17), and the results used to give the full matrix of inter-cluster variances and covariances; see the definitions of S and R in (20) and of Q and H in (29). Although most clusters have at least one household recording a purchase of each commodity, there are a few exceptions. In consequence, each covariance was calculated over the subset of clusters that contained means for that pair alone, so that the meat–fish covariance is not calculated over the same clusters as the cereals–fish covariance or the cereals–starches covariance. The number of clusters C is taken to be the average of the numbers in each of the pairwise comparisons; while such a choice is arbitrary, it clearly makes no difference to the asymptotic properties of the estimators.

The estimator of B in (24) uses the inter-cluster variances and covariances as well as the estimates of Ω and Γ from the within-cluster regression errors. It also requires the quantity ν, the 'average' cluster size, defined by $\nu^{-1} = C^{-1}\sum n_c^{-1}$. Since the within-cluster regressions use only those households recording positive purchases, the number of households per cluster actually used is different for each good, sometimes substantially so. I therefore replace ν^{-1} by a four-element vector containing the appropriate 'average' cluster size for each of the four goods, and define T^{-1} to be a diagonal matrix with this vector as its diagonal. T^{-1} is then used to replace ν^{-1} in eq. (24), and it is a straightforward exercise to make the corresponding changes in the formulae for standard errors.

It is perhaps worth concluding this section with a note on calculation. None of the estimators used here require iterative or search procedures and are from that point of view rather straightforward. In practice, there is a great deal of very tedious work in extracting the data from the tape by clusters and in removing the cluster means. That such work is tedious and conceptually straightforward should not disguise the meticulous attention that is required to get it right. Beyond that, the use of non-standard estimators, and particularly non-standard formulae for variances, raises serious difficulties for checking the correctness of both the algebra and the programming. In the current case, the calculations were programmed independently by the author, using GAUSS on an IBM-XT, and the research assistant, Dwayne Benjamin, using SAS on a mainframe. It became clear very quickly that without such cross-checks the results were worthless. Not all errors result in patently nonsensical results, and several iterations were required to produce identical results from both approaches. The computer programs for the standard errors were first written using an algebraic derivation that, unlike that given above, made no use of vec or Kronecker product operations, but worked with the basic matrix elements. At a final stage, the derivations given above were programmed in GAUSS and the comparison of results revealed a (minor) error in the original algebra. I believe that the results given below are now correct. Even so, it is clear from this experience (unless we are simply very poor programmers) that extraordinary efforts should be made to cross-check any type of non-routine calculations. The evidence in Dewald, Thursby and Anderson (1986) suggests that these problems are widespread among economists.

4.2. Results

Table 1 shows the results from the within-cluster estimation. I do not give the full set of demographic effects, but only those estimates that are carried forward to the second stage. The table is divided into two panels, the upper for the rural sector and the lower for the urban sector. The first two rows in each panel show the elasticities with respect to total food expenditure – β^0, the quantity elasticity, and β^1, the quality elasticity. Note that all the expenditure elasticities are quite precisely determined, particularly in the urban sector, and are all between zero and unity. They are larger in the urban than in the rural sector as is to be expected from the large amount of non-market consumption

Table 1

First-stage results.

	Meat	Fresh fish	Other fish	Starches	Cereals
			Rural		
$\beta^0(t)$	0.753 (7.5)	0.303 (4.3)	0.207 (5.0)	0.468 (4.4)	0.422 (5.2)
$\beta^1(t)$	0.059 (1.4)	0.016 (0.7)	0.027 (1.7)	0.009 (0.2)	0.028 (1.4)
σ^{00}	0.894	0.740	0.631	1.724	1.570
σ^{10}	−0.070	−0.068	−0.070	−0.392	−0.201
σ^{11}	0.151	0.081	0.091	0.359	0.102
ν	1.984	2.408	4.986	2.592	3.032
		$C = 195.4$	$n - C - k = 817.4$		
			Urban		
$\beta^0(t)$	0.930 (18.)	0.546 (9.1)	0.344 (6.4)	0.699 (11.)	0.804 (12.)
$\beta^1(t)$	0.111 (5.7)	0.045 (2.0)	0.046 (2.8)	0.028 (1.1)	0.016 (0.9)
σ^{00}	0.645	0.960	0.907	1.266	1.387
σ^{10}	−0.016	−0.149	−0.067	−0.261	−0.294
σ^{11}	0.093	0.132	0.085	0.193	0.089
ν	4.770	4.659	5.293	5.332	5.288
		$C = 86.4$	$n - C - k = 755.0$		

in the countryside. Meat is the closest to a luxury good, and fresh fish has a higher expenditure elasticity than does dired or smoked fish. Note that all these elasticities are computed conditional on market purchase; if recourse to the market at all is highly income-elastic, as it may well be in the rural areas, then the expenditure elasticities in the table are likely to be much lower than those for aggregate or average demand. The quality elasticities are not large, and in the rural sector they are not significantly different from zero; presumably there is a good deal less choice among varieties than there is in the cities. For neither sector are the quality elasticities for starches or cereals significantly different from zero.

The second set of numbers in table 1 are, first, the estimated residual variance from the within-cluster quantity regression, last, the corresponding figure for the unit value equation, and in the middle, the estimated covariance between the residuals in the two equations. The first row tells us something about taste variation, but it is the second and third that are of most interest here. The estimates of σ^{10} are uniformly negative, as would be expected if there is measurement error in either quantities or expenditures. They are also typically of the same magnitude as the estimates of σ^{11}, which measure the variance of measurement error in measured unit values. Since these measurement errors will also contaminate the between-cluster estimates, both must be allowed for at the second stage. Note finally the 'average' clusters sizes ν; since

Table 2
Inter-cluster variances and covariances.

	Meat	Fresh fish	Other fish	Starches	Cereals
			Rural		
Q matrix	0.7009	0.0631	0.0388	0.1432	0.1902
		1.2281	0.1646	0.3162	0.1070
			0.6581	0.1397	0.0459
				1.1352	0.4030
					1.1388
R matrix	−0.1161	−0.0463	0.1005	0.0300	0.0041
	−0.0380	−0.2478	−0.0678	−0.0292	−0.0705
	0.0060	−0.0648	−0.1534	0.0089	0.0393
	0.0359	−0.1031	0.0163	−0.2138	−0.0083
	−0.0007	−0.0174	0.0055	−0.0714	−0.1952
S matrix	0.3288	0.0136	−0.0032	0.0479	−0.0035
		0.1353	0.0529	0.0218	0.0112
			0.1285	−0.0032	0.0074
				0.2761	0.0128
					0.1022
			Urban		
Q matrix	0.2473	−0.0582	−0.0650	−0.0071	0.0027
		0.4201	−0.0736	0.0725	−0.0472
			0.5667	0.1101	−0.1051
				0.4213	−0.1086
					0.3785
R matrix	−0.0416	−0.0027	0.0156	−0.0123	−0.0234
	−0.0003	−0.1403	0.0472	−0.0063	−0.0102
	−0.0071	0.0009	−0.0570	−0.0105	0.0135
	0.0174	0.0029	0.0419	−0.0780	0.0010
	−0.0192	0.0148	−0.0030	−0.0006	−0.0479
S matrix	0.0538	0.0057	0.0158	−0.0014	0.0099
		0.1279	−0.0070	0.0075	0.0047
			0.0379	−0.0090	−0.0004
				0.0858	0.0032
					0.0371

rural clusters each contain 12 households, these give some idea of the prevalence of zero purchases.

Table 2 gives the inter-cluster variances and covariances for the five goods. These are the variance and covariance matrices of the vectors of quantities and unit values, each corrected for the effects of per capita food expenditure and household demographics. Corrected for measurement error and for quality effects, these are the basic data for estimating the price elasticities. Several features are notable. First, the diagonal elements of the R matrices are negative for all goods in both sectors, so that, once and demographic effects have been taken into account, there is a negative correlation between quantity and unit value. Some of this, of course, is measurement error, and the estimates of σ^{10} in table 1, scaled by the average cluster sizes, provide an estimate of the contribution of measurement error to the covariance. Second, note that the figures in the lower panel tend to be smaller than those in the upper panel. Dispersion in quantities and in prices is less in urban than in rural areas, and the covariance is correspondingly lower. The urban clusters are much less geographically dispersed than are the rural clusters, and indeed there are 13 urban clusters in Abidjan alone. Although these clusters are not adjacent, there is much less possibility for large price differences than is the case between the rural clusters which are widely dispersed across the country. Although other explanations are clearly possible, for example that tastes are more homogeneous in urban areas, it is certainly consistent with the model that a lack of price variation should accompany a corresponding lack of quantity variation. The additional fact that the R matrix is in proportion to the other two matrices across the sectors means that the estimates of price elasticities are unlikely to be implausibly different between urban and rural.

The third feature to note is that the two S matrices are close to being diagonal. While there is no reason to rule out any particular correlations between the unobservable prices, measurement errors that were correlated across goods would tend to show up in the off-diagonal terms of these two matrices. Table 2 does not therefore contradict the assumption that the measurement errors are not so correlated.

Simple calculations of elasticities can be made directly from tables 1 and 2. For example, if measurement error and cross-price effects are ignored, the own-price elasticities would simply be the ratio of the diagonal of R to the diagonal of S. Still ignoring cross-price effects, measurement error can be taken into account by subtracting from the diagonal of R the vector σ^{10} divided by the average numbers of (purchasing) households in each cluster, and dividing the result by the diagonal of S less σ^{11} again divided by the average cluster size. Such estimates are presented and discussed in Deaton (1986) where the cross-price effects are ignored. Tables 3 and 4 show the final estimates of own-price elasticities using the methodology of the previous study together with the new estimates of own- and cross-price elasticities. Estimates for the rural sector are presented in table 3 and those for the urban sector in table 4. For each sector, the first panel gives the estimate of B', evaluated according to eq. (24), the second the estimate of Θ, evaluated according to eq. (27), and the third the estimates ignoring the cross-price effects.

Perhaps the most obvious feature of tables 3 and 4 is that the estimates of Θ are not dramatically different from those of B'; this is a straight-forward consequence of the fact that the quality elasticities in table 1 are small, so that the D matrix in (27) is close to zero. The correction that *does* make a large difference is the correction for measurement errors; if $\tilde{B}$ is evaluated, not according to (24), but simply as $S^{-1}R$, quite different estimates are obtained.

Table 3

Estimates of own- and cross-price elasticities for the rural sector.[a]

	Meat	Fresh fish	Other fish	Starches	Cereals
		B' matrix: price elasticities without quality correction			
M	−0.379 (2.6)	−0.609 (2.2)	0.354 (1.5)	0.504 (1.9)	−0.062 (0.2)
FF	0.026 (0.2)	−2.378 (7.7)	0.534 (2.0)	−0.382 (1.4)	0.148 (0.5)
OF	0.403 (3.3)	−0.099 (0.4)	−1.225 (6.1)	−0.059 (0.3)	0.259 (1.1)
S	0.197 (1.1)	−0.265 (0.8)	0.265 (0.9)	−0.389 (1.2)	−0.942 (2.7)
C	0.003 (0.0)	−1.081 (3.6)	1.010 (3.9)	0.307 (1.1)	−1.867 (6.2)
		Θ matrix: price elasticities with quality correction			
M	−0.353 (2.5)	−0.529 (2.0)	0.283 (1.4)	0.493 (1.9)	−0.056 (0.2)
FF	0.043 (0.3)	−2.131 (4.7)	0.417 (1.9)	−0.336 (1.3)	0.135 (0.5)
OF	0.338 (3.2)	−0.103 (0.6)	−1.039 (5.9)	−0.031 (0.2)	0.197 (1.1)
S	0.199 (1.2)	−0.191 (0.6)	0.179 (0.7)	−0.393 (1.3)	−0.827 (2.6)
C	0.041 (0.3)	−0.867 (2.9)	0.755 (3.4)	0.284 (1.1)	−1.647 (5.7)
		Original own-price elasticities ignoring cross-price effects			
	−0.312 (2.4)	−1.944 (5.1)	−1.085 (7.2)	−0.452 (1.6)	−1.667 (6.1)

[a] The elasticities are presented so that the columns are the goods whose prices are changing and the rows the goods whose quantities are being affected. Hence, for example, the 0.043 in row 2 and column 1 of the rural Θ matrix is the estimate of the elasticity of the demand for fish with respect to the price of meat.

Table 4

Estimates of own- and cross-price elasticities for the urban sector.[a]

	Meat	Fresh fish	Other fish	Starches	Cereals
		B' matrix: price elasticities without quality correction			
M	−1.471 (1.9)	0.127 (0.6)	0.976 (1.0)	0.486 (1.3)	−0.318 (0.5)
FF	−0.085 (0.1)	−1.152 (5.3)	−0.196 (0.2)	0.128 (0.3)	1.014 (1.4)
OF	2.672 (1.9)	0.104 (0.3)	−3.839 (2.3)	0.310 (0.5)	−1.590 (1.3)
S	−0.065 (0.1)	−0.063 (0.3)	−0.752 (0.7)	−0.722 (1.7)	0.119 (0.1)
C	−2.186 (1.5)	0.101 (0.3)	2.374 (1.4)	0.281 (0.5)	1.420 (1.0)
		Θ matrix: price elasticities with quality correction			
M	−1.075 (2.5)	0.107 (0.7)	0.542 (1.1)	0.432 (1.4)	−0.392 (0.7)
FF	−0.133 (0.3)	−1.055 (5.5)	−0.101 (0.2)	0.108 (0.3)	0.981 (1.5)
OF	1.570 (2.4)	0.083 (0.4)	−2.463 (3.2)	0.286 (0.7)	−1.154 (1.5)
S	−0.210 (0.4)	−0.065 (0.3)	−0.492 (0.8)	−0.732 (1.8)	0.229 (0.3)
C	−1.454 (1.6)	0.092 (0.3)	1.491 (1.5)	0.258 (0.5)	1.202 (1.0)
		Original own-price elasticities ignoring cross-price effects			
	−0.982 (3.6)	−0.997 (5.6)	−1.604 (3.6)	−0.572 (1.5)	0.383 (0.5)

[a] See note to table 3 above.

Although several of the cross-elasticities are large and significantly different from zero, the rural own-price elasticities are very similar to those obtained earlier by ignoring the cross-price effects, compare the diagonal of the second panel with the third panel. All estimated own-price elasticities are negative and, apart from that for starches, are significantly different from zero. This is a somewhat surprising finding, at least for meat and fish. A great deal of fish is consumed in the Cote d'Ivoire, but there is a negative geographical association between fish consumption and meat consumption, the former being high along the coast and low in the Northern Savannah, and vice versa for meat. Not surprisingly, fish is also relatively cheap along the shore and meat relatively cheap in the Savannah. Even so, allowing for cross-price elasticities between meat and fish causes no dramatic alteration in the estimates of the own-price elasticities, and although there is a significant positive cross-price effect of the meat price on the demand for dried and smoked fish, there is an estimated *negative* cross-price effect of the fresh fish price on the demand for meat. Both types of fish have large estimated own-price elasticities, whether or not cross-price effects are taken into account; fresh fish is more elastic to total food expenditure and has a price elasticity of approximately -2, which is twice as large as that for other fish. The price of dried and smoked fish exerts a positive effect on the demand for fresh fish, although this substitutability does not appear in any strong corresponding effect of the price of fresh fish on the demand for other fish. Starches and meat are estimated to be much less price-elastic than either cereals or fish; note that cereals in the rural Cote d'Ivoire takes up less than 15% of the budget, and the dispersion of household budgets over a wide range of foods means that high price elasticities are less implausible than would be the case if demand were more concentrated.

Given the smaller numbers of clusters in the urban areas, the urban price elasticities in table 4 are rather less precisely estimated than those from the rural clusters. Once again, both types of fish are strongly price-elastic, although here dried and smoked fish has the higher own-price elasticity of the two. Meat is estimated to be more price-elastic in the urban areas, while the estimates for starches are not very different from those in table 3. The urban data are not capable of establishing a good estimate for the cereals price elasticity; all three estimates in table 4 are *positive*, though none are significantly different from zero. Although only one is significantly different from zero, there are more *large* off-diagonal point estimates in the urban matrix; if these estimates are taken seriously, the dominant effects of price changes in the countryside are on the demand for the good whose price has changed, while in the urban areas there is more substitution between the goods as prices alter. Although there has been some discussion in the literature on the effects of urbanization on tastes, notably by Kuznets (1962), it is unclear why urbanization should be associated with greater substitutably.

Of the cross-effects, only five out of twenty are significantly different from zero in the rural matrix and only one of twenty in the urban matrix. And since there are 195 rural clusters and 86 urban clusters, conventional significance levels are perhaps somewhat generous, particularly for the rural estimates. Even so, there are some interesting patterns. I have already commented on the fish–meat pattern in the rural matrix. There are also reasonably large pairs of positive cross-price elasticities between meat and starches and between cereals and dried or smoked fish. Since the income effects are not very large for any of these goods, these pairs would be classified as substitutes. For the meat and fish trio, and giving greater weight to larger and more significant coefficients, meat appears to be substitutable for dried and smoked fish, but a complement to fresh fish, while fresh and dried fish would probably be classed as complements. The demand for cereals responds negatively to the price of fresh fish, indicating complementarity, but there is no corresponding effect of the cereals price on the demand for fresh fish. Note finally that while starches respond *negatively* to the price of cereals, cereals respond *positively* to the price of starches. Nevertheless, the overall picture in the rural areas is one of fairly large own-price elasticities, with the cross-price effects playing a secondary role.

The pattern of urban cross-price elasticities is rather different. There are four cross-price elasticities that have estimated absolute values greater than unity; the meat price positively influences the demand for other fish (again a substitutable pair as in the rural sector) and negatively affects the demand for cereals (a complementary pair). The other two large effects are between other fish and cereals, but, contrary to theory have opposite signs.

It is of some interest to construct formal tests for the symmetry of the substitution matrices underlying the uncompensated elasticities in table 3. Note that it is not possible to use the results to test for homogeneity of demands; that would require a complete catalog of goods that between them exhausted the food budget. Even so, note that homogeneity implies that the rows of the full Θ matrix should add to zero, so that it is possible to calculate the cross-price elasticities with the omitted 'other foods' category that is implied by the homogeneity of the food subgroup demands. The details are not of great interest, but it is clear from tables 3 and 4 that for several of the goods (fresh fish, starches and cereals in the rural sector and other fish and starches in the urban sector), there is insufficient substitutability to offset the large negative own-price elasticities without assuming an implausibly large degree of substitutability with the omitted category.

As to symmetry, if the uncompensated own- or cross-price elasticity is θ_{ij}, the compensated elasticity η_{ij}, say, is given by $\eta_{ij} = \theta_{ij} + \beta_i^0 w_j$, where w_j is the budget share of good j in total expenditure, in this case in total food expenditure. The substitution matrix will be symmetric if and only if $w_i \eta_{ij}$ is symmetric, i.e., if $w_i \theta_{ij} + \beta_i^0 w_i w_j$ is symmetric. The second term of this expression is small, hence the legitimacy of cataloging substitutes and complements in terms of the sign of θ_{ij}. Since the budget shares vary from observation to observation, I test the restrictions at the sample means of the budget shares; the relevant vectors are (0.1792, 0.0990, 0.1143, 0.0982 and 0.1417) for meat, fresh fish, dried fish, starches and cereals in the rural sector and (0.2248, 0.1212, 0.2111, 0.1474 and 0.1730) in the urban sector. Given that the w's are taken as fixed, the symmetry conditions yield a set of ten restrictions on each of the two Θ matrices and β^0 vectors, i.e.,

$$w_i\theta_{ij} + \beta_i^0 w_i w_j = w_j\theta_{ji} + \beta_j^0 w_j w_i, \qquad i, j = 1,5, \quad i \neq j. \tag{41}$$

I define a 30-element vector $\alpha' = [\text{vec}(\Theta)', \beta^{0\prime}]$ with a variance–covariance matrix V_α, which is composed of the variance–covariance matrix of vec(Θ), given in (37), the variances of the elements of β^0 from the first-stage regressions, and the covariance of vec(Θ) and β^0 from (40). The symmetry restrictions take the form $\Xi\alpha = 0$, where Ξ is a (10×30) matrix of restrictions. Symmetry can be tested by examining the vector $\Xi\tilde{\alpha}$ and using its variance–covariance matrix $\Xi V_\alpha \Xi'$ to construct standard errors and t-tests, while an overall Wald test is given by calculating $\tilde{\alpha}'\Xi'(\Xi V_\alpha \Xi')^{-1}\Xi\tilde{\alpha}$, which, under the null of symmetry, is asymptotically distributed as a χ^2 with ten degrees of freedom.

The Wald statistic for the rural sector is 18.2, while that for the urban sector is 11.5, neither of which is significant at conventional levels. Given that there are a number of significant cross-price elasticities, at least in the rural sector, this result would seem not to derive from the possibility of placing almost any restriction on an imprecisely estimated elasticity matrices. In the rural sector, the pairings of cereals with fresh fish, other fish and starches do not easily fit with symmetry, while in the urban area the pair that appears to violate symmetry is cereals and other fish. However, in the overall test, these problem pairings are more than offset by the fact that the other goods fit well with the pattern required by symmetry.

5. Conclusions

This paper has proposed a method for estimating own- and cross-price elasticities from the spatial variation in prices within a standard household survey. The methodology is applied to data from the Cote d'Ivoire and is successful in that estimates are generated that are not clearly absurd. With one (insignificant) exception, the own-price elasticities are negative, the cross-price elasticities are mostly plausible, and they pass a formal test of the symmetry restrictions of demand theory. Somewhat different patterns of demand behavior and responsiveness to prices are estimated as between urban and rural areas, differences that may well be real, possibly reflecting differences in food availability between cities and countryside, the prevalence of own-consumption over market purchases in the countryside or differences in tastes between modern city dwellers and the largely traditional inhabitants of the countryside. If such differences are indeed real, their existence has important implications for the design of pricing and tax systems and for the effects that such schemes exert on the allocation of resources and distribution of income between urban and rural sectors.

What cannot be done in this study is to provide any independent validation of the estimates. There are still unsolved econometric problems, particularly in dealing with households that report no expenditure on various goods. But perhaps most important will be future attempts to replicate this sort of analysis on other data sets, both in the Cote d'Ivoire and elsewhere. That work is currently in progress.

References

Deaton, A.S., 1986, Quality, quantity, and spatial variation of price (Woodrow Wilson School, Princeton University, Princeton, NJ) processed.

Dewald, W.G., J.G. Thursby and R.G. Anderson, 1986, Replication in empirical economics, American Economic Review 76, 587–603.

Fuller, W.A., 1975, Regression analysis for sample surveys, Sankhya: The Indian Journal of Statistics C 37, 117–132.

Kuznets, S., 1962, Quantitative aspects of the economic growth of nations, VII: The share and structure of consumption, Economic Development and Cultural Change 10, 1–92.

Lee, L.F. and M.M. Pitt, 1986, Microeconometric demand systems with binding nonnegativity constraints: The dual approach, Econometrica 54, 1237–1242.

Magnus, J.R. and H. Neudecker, 1986, Symmetry, 0–1 matrices and Jacobians: A review, Econometric Theory 2, 157–190.

Prais, S.J. and H.S. Houthakker, 1955, The analysis of family budgets (Cambridge University Press, Cambridge).