

Household Survey Data and Pricing Policies in Developing Countries

Angus Deaton

In recent years, household survey data from developing countries have increasingly become available and have been increasingly used to cast light on important questions of policy. The reform of prices, whether agricultural prices, consumer taxes, subsidies, or tariffs, has consequences for individual welfare and for government revenues, and these can be investigated empirically with household survey data. The gainers and losers from price changes can be identified, and the magnitudes of their gains and losses measured. Nonparametric estimation techniques provide a straightforward and convenient way of displaying this information. The procedure is illustrated for the effects of rice pricing in Thailand using data from more than five thousand rural households. Estimates of the revenue effects of price reforms are harder to obtain, because they require estimates of supply and demand elasticities, estimates that are not easily obtained for many developing countries. A procedure is presented for estimating price elasticities of demand from spatial price variation as recorded in household survey data. The main innovations lie in the appropriate treatment of quality variations and measurement error. Applications of the procedure in Côte d'Ivoire, Indonesia, and Morocco are reviewed.

Although household surveys have been used to describe living standards for a very long time, there has recently been a considerable increase in interest in the use of such data, particularly for developing countries. This interest has been motivated in part by methodological concerns, in part by data availability, and in part by advances in computing. However, there has also been an increased awareness of the fact that household surveys are by far the richest source of

The author is a professor of economics at the Woodrow Wilson School of Public and International Affairs, Princeton University.

This article is a revision of "New Approaches to Household Survey Data from Developing Countries," presented to the Australian Economics Congress, Canberra, August 1988. The material in section III is a summary of a World Bank research project on estimating price elasticities carried out within the Welfare and Human Resources Development Division, Population and Human Resources Department. The work is reported in more detail in Deaton (1987a,b, 1988, forthcoming). Section II is adapted from Deaton (1989), work which was jointly funded by the Asian Development Bank and the McDonnell Foundation. Section I provides a framework for both areas of research and updates the conclusions reported in Deaton (1987a) in the World Bank's volume on taxation and development edited by Newbery and Stern (1987). The author is grateful to two referees for comments on the previous version.

©1989 The International Bank for Reconstruction and Development / THE WORLD BANK.

information about developing economies, and that they can be used to investigate a wide range of theoretical and practical questions. In economies in which perhaps 80 percent of the population is engaged in agriculture, and where household level activity is a large share of total national product, household surveys capture not only levels of living, but the processes of production, consumption, and labor use, and the determinants of health and nutrition that lie beneath.

This article explores the link between pricing policy and the econometric analysis of survey data. Pricing policies include agricultural procurement policy, taxation, and food subsidy schemes, and in many developing countries they play the same central role in fiscal policy as do income tax and social security schemes in most developed countries. Pricing policies influence the distribution of real income, levels of production and consumption, and government revenues. In section I below, I present a simple framework in which pricing policies can be analyzed; the model is the standard one but my emphasis is on the magnitudes that need to be measured so that the theory can be used in practice. I draw attention to, first, the distribution of income over households, second, the average levels of production and consumption conditional on income, and third, the elasticities of net supply with respect to price changes. In section II, I argue that recent statistical advances in nonparametric estimation of densities and regression functions provide a particularly simple method of using household survey data to calculate and display relations between income, production, and consumption and thus the effect of price changes on the distribution of real income. My examples are drawn from a rice price policy in Thailand. In section III, I argue that household survey data can also be used to measure price elasticities by using spatial variation in price. I review some recent experience with this technique, with examples from Côte d'Ivoire, Indonesia, Morocco, and by way of contrast, the United States.

I. MODELING PRICING POLICY AND ITS CONSEQUENCES

The model that I use here is the standard one for price reform and tax analysis, discussed in detail in the recent volume by Newbery and Stern (1987), but dating back at least to Diamond and Mirrlees (1971). Although not everyone is comfortable with the optimal tax framework in which these models are often set, the basic issues are simply those of equity and efficiency: price changes affect the welfare of consumers, they affect different consumers differently, and they alter government revenues and expenditures. As a result, an analysis of price reform is useful even for those situations in which one could not accurately characterize the political economy of price setting as being a benevolent central authority maximizing a social welfare function.

In this article, I consider only the question of price reform, and I attempt to trace the consequences of a small change in some price. The price change affects consumers and producers in proportion to the amount of the commodity

that they consume or produce, with net consumers losing from a price increase and net producers gaining. Government revenues and expenditures are also affected, directly if the price change is a change in a tax or subsidy and indirectly as consumers and producers react to the price change by altering their purchases and sales of items that carry taxes and subsidies. On the production and consumption side, since different individuals consume and produce different amounts, some rule is required for aggregating welfare changes across individuals. I shall follow the practice of postulating a social welfare function, but an important aim of the analysis is to demonstrate how the empirical work and the valuation of welfare effects can be separated, so that disagreements about welfare judgments need not spill over into disagreements about what needs to be calculated.

For reasons that will become clearer as I proceed, it is convenient to work with a continuous distribution of welfare over individuals. If u is used to denote the welfare level of an individual, then I write

$$(1) \quad u = \psi(m, p, \epsilon)$$

where m is the individual's budget (income or total expenditure), p is a vector of prices of commodities bought or sold by the individual, and ϵ is a vector of taste differences that vary across individuals. The indirect utility function, ψ , converts money into real income; it is the same for all individuals. The welfare levels of members of society are aggregated by means of a social welfare function, which I write in the form

$$(2) \quad W = \int \phi(u) dF(u) = \iint \phi[\psi(m, p, \epsilon)] f(m, \epsilon) dm d\epsilon$$

where $F(u)$ is the distribution function of u in the population, and $f(m, \epsilon)$ is the joint density function of m and ϵ .

Direct Effects on Households' Real Income

The effect of a price change on social welfare depends on the direct effect through equation 2 and the indirect effect through the social value of the change in government revenue. The utility of an individual is affected according to

$$(3) \quad \partial u / \partial p_i = -q_i(m, p, \epsilon) \cdot \partial u / \partial m$$

where q_i is the net consumption of good i , so that equation 3 is positive for net producers, who have negative net consumption and who benefit from a price increase, and negative for consumers, who lose from the higher price. The derivative of social welfare is given by substituting equation 3 in equation 2 to give

$$(4) \quad \begin{aligned} \partial W / \partial p_i &= \int \phi'(u) \partial u / \partial p_i dF(u) \\ &= - \iint \{ \partial \phi(m, p, \epsilon) / \partial m \} q_i(m, p, \epsilon) f(m, \epsilon) dm d\epsilon \end{aligned}$$

The expression $\partial\phi/\partial m$ is the social marginal utility of money to an individual with budget m and characteristics ϵ who faces prices p .

For practical purposes, it is useful to make two modifications to equation 4. First, define the "net benefit ratio," $\sigma_i(m, p, \epsilon)$ by

$$(5) \quad \sigma_i(m, p, \epsilon) = -p_i q_i / m$$

The net benefit ratio expresses the marginal effect of a price change as a fraction of income; it is positive for individuals who are net producers of the commodity and negative for net consumers. It can be thought of as the elasticity of (the money value of) welfare with respect to a price change. Multiply through equation 4 by p_i so that it can be rewritten as

$$(6) \quad \partial W / \partial \ln p_i = \iint \{ \partial \phi(m, p, \epsilon) / \partial \ln m \} \sigma_i(m, p, \epsilon) f(m, \epsilon) dm d\epsilon$$

If the net benefit ratio, σ_i , were the same for everyone, and could be taken out from the integral in equation 6, then $\partial W / \partial \ln p_i$ would be σ_i multiplied by the rest of the integral, which is the effect on welfare of a proportional increase in everyone's income (see equation 2). In consequence, the mean values of σ_i within a region or income group tell us the net income transfer to that group that will result from price changes. The deviations of σ_i from its mean tell us what effects the price change has on the distribution of income over different households.

The second modification to equation 4 concerns the social marginal utility of money, $\partial\phi/\partial m$, or $\partial\phi/\partial \ln m$ in equation 6. In the form given, the social marginal utility of money is tailored to the characteristics of each individual, but it is hard to imagine that any real authority would wish to discriminate to this degree, or that it would be capable of doing so, even if it wished. Instead, I shall make the assumption that there exists some single household characteristic, $x(m, \epsilon)$, upon which, in addition to price, the social marginal utility of money depends. The example I have in mind, and which I shall use below, is the one in which m is total household expenditure and $x(m, \epsilon)$ is total household expenditure per capita, or better, if we only knew how to measure it, total household expenditure per equivalent adult. Note that it would be easy enough to extend this case to allow the marginal social utility of money to depend on a small list of household characteristics, including perhaps region of residence, but I keep to the simplest case here.

Write the social marginal utility of money multiplied by income as $\theta(x, p)$, so that, after the appropriate substitutions, equation 6 becomes

$$\begin{aligned} \partial W / \partial \ln p_i &= \iint \theta(x, p) \sigma_i(x, p, \epsilon) f(x, \epsilon) dx d\epsilon \\ (7) \quad &= \iint \theta(x, p) \{ \sigma_i(x, p, \epsilon) f_\epsilon(\epsilon | x) d\epsilon \} g(x) dx \\ &= \iint \theta(x, p) E(\sigma_i | x) g(x) dx \end{aligned}$$

where, in only a slight abuse of notation, $f(x, \epsilon)$ is the joint density of x and ϵ , $f_\epsilon(\epsilon | x)$ is the conditional density of ϵ given x , $g(x)$ is the marginal density of x , and $E(\sigma_i | x)$ is the conditional expectation of σ_i given x .

Equation 7 is in a form which can be empirically implemented and which provides a useful separation of the three elements that determine the effects of prices on welfare. The first, $\theta(x, p)$, concerns the valuation placed by the observer or the evaluating authority on an extra unit of currency to an individual with characteristic (per capita expenditure) x . The second, the conditional expectation, $E(\sigma_i | x)$, tells us how net production and consumption relate to per capita expenditure x . For consumers who do not produce good i , $E(\sigma_i | x)$ is (the negative of) an Engel curve in budget share form. More generally for a randomly selected household with per capita expenditure x , it is the expected elasticity of real income with respect to a price change. In statistical terms $E(\sigma_i | x)$ is the regression function of σ_i on x . The third and final term, $g(x)$, is simply the distribution of the characteristic x in the population. In my usage here, x represents real income or welfare, so that $g(x)$ is the distribution of real income in the population. The first term is the product of value judgments by the policymaker or analyst, and so must be given exogenously. But empirical estimates of the other two terms can be readily combined with any given set of value judgments.

In the next section, I shall show how household survey data can be used to provide direct estimates of the second and third of these terms, the price elasticity of income and the distribution of income. These estimates can be constructed without arbitrary parametric assumptions about the shape of the income distribution or the relevant Engel curves. The standard approach requires specifying production and utility functions, estimating demands and supplies, and allowing for households who have either zero production or consumption. This turns out to be unnecessary here; a much simpler approach is readily available.

The Effects on the Government Budget

An expression like equation 7 captures only the direct effects of a price change on individual welfare. The other side to the story is that price changes influence government revenues and expenditures, and that these changes, weighted by the appropriate shadow price of government funds, should be added to the direct effects on individuals so as to evaluate the total effects of the price reform. Suppose that from all tariffs, taxes and subsidies the government collects a net surplus or deficit R , so that we can write

$$(8) \quad R = \sum_{k \in S} (p_k - \pi_k) \int q_k(x, p, \epsilon) f(x, \epsilon) dx d\epsilon$$

where p_k and π_k are, respectively, consumer and producer prices for some subset of goods, S , for which it is possible for the government to set separate prices. In the simplest case, π_k is a world price, p_k is a domestic price, and the difference is the tariff collected. Alternatively, q_k is the level of exports of good k , and $(p_k - \pi_k)$ is the export tax, so that $(p_k - \pi_k)$ multiplied by the integral of q_k is the total revenue from the export tax. If we denote by Q_k the mean of the individual q_{kS} , and if we make the simplifying (but not innocuous) assumption

tion that producer prices (π) remain constant when consumer prices (p) are changed, we can calculate the derivative of government revenue with respect to a change in p_i as

$$(9) \quad \partial R / \partial p_i = Q_i + \sum_{k \in S} (p_k - \pi_k) \partial Q_k / \partial p_i$$

on the assumption that good i is in the taxable set S . Note that the fixity of the π s makes sense if the goods are traded so that π is the world price. Lack of response of producer prices also follows from the assumptions of the nonsubstitution theorem, that there are constant returns to scale, one nonproduced factor, and no joint products. When these assumptions are not valid, and they are unlikely to be except for traded goods, the market clearing conditions for the commodity have to be explicitly taken into account. For the purposes of this article, we ignore this complication.

The information needed to estimate equation 9, while not insubstantial, is relatively limited when compared with the detailed information on individual supply and demand required for equation 7. Provided we can obtain information on the tax rates themselves (often a far from simple matter), the evaluation of equation 9 requires only the own and cross derivatives, $\partial Q_k / \partial p_i$, of the aggregate Q_k . Although it should be recognized that the estimation of any price elasticity is always a matter of difficulty, and that a century or more of research has failed to reach an agreed set of figures on either the demand or supply side, the requirements here are about as minimal as it is possible to imagine. We do not require estimates of individual supply and demand elasticities, and we do not need estimates of preferences or technology. It does not matter whether the additional demand or supply comes from changes in behavior of existing consumers and producers, or from the entry into the market of new agents. All that is required, though it is still a great deal, is the response of aggregate net supply/demand to a change in price. Partial evidence about such a response can be obtained from many different sources. Section III below reviews recent work that uses spatial price variation to measure these price elasticities. Taken together, sections II and III show how to use household survey data to measure both the direct and indirect effects of changes in pricing policy.

II. NONPARAMETRIC ESTIMATION OF WELFARE EFFECTS: AN EXAMPLE FROM THAILAND

In Thailand, the government has raised revenue for at least a century through an export tax on rice (see Bock 1884 and Siamwalla and Setboonsarng 1989). In the last few years, when the world price of rice has been historically low and the economy has become more diversified and less dependent on rice, the tax has been removed, or at least set to zero. However, revenue from the tax had been important in the government budget and reached close to 8 percent of government revenue at the peak of the rice price boom in 1974. The operation of the export tax is described in some detail by Siamwalla and Setboonsarng

(1989); in the simplest terms, the existence of the "premium" on exports ensures that the domestic price is lower than the world price, so that households that are net consumers of rice benefit from the tax, as does the government, while the losers are those rice farmers who produce more than they consume. Clearly, the tax redistributes income from rural to urban areas, and since rural households in Thailand are substantially poorer than urban households, the tax would seem to be regressive in effect. However, there are also many rural households that either do not produce rice or who consume more rice than they produce, and they too must be counted among the beneficiaries. Furthermore, to the extent that a proportion of sales of rice are made by farmers who are relatively well-off, the tax may be more progressive than appears by looking only at the rural-urban distinction. It is on these distributional grounds that Trairatvorakul (1984) has argued in favor of keeping rice prices low.

Thailand has a regular system of socioeconomic surveys which cover the whole country and which take place every five or six years. The surveys collect data on household income and expenditure patterns, on household characteristics, and for farm households, on cropping patterns and on the amounts sold of each crop. From such data, measures of living standards can be constructed, as well as estimates of the net sales or purchases of various commodities. The data used here come from the 1981-82 Socioeconomic Survey, which collected information on nearly 12,000 households at various times from February 1981 until January 1982. The survey is stratified by three levels of urbanization: "municipal areas," which are urban, "sanitary districts," which are rural agglomerations of villages or small towns, and "villages," which represent the rural areas. In this article, I shall concern myself with the villages; there are 5,836 village households, themselves a random sample of all Thai village households.

For each household, I calculate two numbers. The first, the x variable described in section I, is the logarithm of household per capita total expenditure on all goods and services, ($\ln xpc$). Total consumption is a better measure of living standards than is household income, and the deflation by household size is a crude but straightforward method for correcting for different numbers of household members sharing the same budget. I also make the conventional assumption that prices are the same for all households in the survey, so that it is unnecessary to make any explicit correction to convert money into real expenditures. This assumption seems reasonable in Thailand, which has a highly developed transport system and a very efficient trade and distribution network, so that spatial price variation is likely to be small. Even so, a complete absence of spatial price variation is not to be expected and would rule out use of the methods to be discussed in section III.

The second number calculated for each household is the net benefit ratio, that is, the net sales of rice as a fraction of total household expenditure. Although farmers typically sell paddy and consumers purchase rice, provided the two prices move together the value of the difference between sales of one

and purchases of the other is still the correct measure of the money value of the welfare change resulting from the alteration in price.

Having calculated 5,836 values for $(\ln xpc)$ and for the net benefit ratio, how should the theory of section I be implemented? Perhaps the most obvious procedure is to estimate equation 7 directly. Given a formula relating the social marginal utility of money to the logarithm of total household per capita expenditure, $(\ln xpc)$ —that is, an explicit function $\theta(x, p)$ —the value of $\theta(x, p)\sigma_i$ can be calculated for each household, and an average taken over all households in the sample. If there is no doubt about the appropriate choice for $\theta(x, p)$, this is undoubtedly the right thing to do. But if there are several possible weighting schemes, or if we want to leave the choice to someone else, it is useful to have a means of displaying the consequences of different weighting schemes. To do so given equation 7, we need not the final number $\partial W / \partial \ln p_i$, but the two intermediate expressions $E(\sigma_i | x)$ and $g(x)$. The first is a regression, the second a density function. Standard procedures for estimating each one would involve the selection of a parametric form, followed by some appropriate estimation procedure, such as least squares, maximum likelihood, or minimum χ^2 . For the regression function, the relevant literature is that on Engel curves, at least for the households that are net consumers, and there are many different functional forms from which to choose (see, for example, Prais and Houthakker 1955, or Deaton and Muellbauer 1980). For the density function, there is also a large literature on the various functional forms that mimic the distribution of income (see, for example, Cramer 1969, or the appendix to Cowell 1977).

All these approaches leave something to be desired. The literature on demand and supply is much concerned with deriving functional forms that are consistent with choice and production theory. Although such a methodology is necessary for many purposes, its theoretical elegance comes at a real price in terms of the ability to model important aspects of real data. In particular, it has proved extremely difficult to construct tractable empirical models that can deal with producers or consumers who are sometimes optimally located at “corners” where they produce or consume nothing of the good in question (see, for example, the attempts in Wales and Woodland 1983, and Lee and Pitt 1986). But the analysis of pricing questions does not require any distinction between households at internal and households at corner solutions; the quantity that is required is the expectation $E(\sigma_i | x)$, an expectation that is taken over all households, whether they purchase or not and whether they consume or not.

Perhaps an even more fundamental problem with conventional approaches is that it is very hard to do justice to the richness of a data set that contains several thousand observations. Economists who use time series data used to be content with twenty or thirty observations to estimate half a dozen parameters, although with quarterly data, they now enjoy the luxury of ten or twenty observations per parameter estimated. By this sort of criterion, and with 5,836 observations, we should be estimating Engel curves or income distribution functions with five or six hundred parameters. But the traditional thrust in the

literature on production, demand, and income distribution has been rather toward economy of parametrization, on finding two or three parameter Engel curves, or three and four parameter distribution functions to fit income distribution data. The problems with which I am concerned here provide no reward for economy of parametrization. Much more important is to use the plentiful data to provide an accurate description of the two functions that we need, $E(\sigma_i|x)$ and $g(x)$.

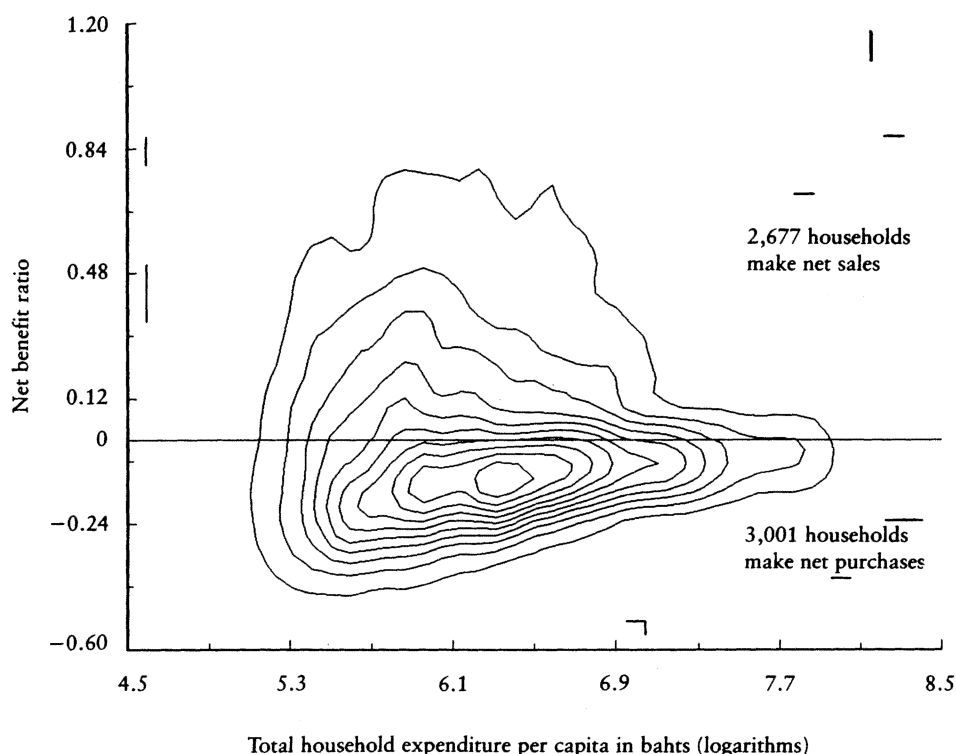
With so many observations, it seems that only very minimal assumptions about functional form ought to be required, and such is the case. There has been a great deal of recent work in both statistics and economics on the nonparametric estimation of both densities (see particularly Silverman 1986), and of regressions (see Härdle 1988). Many of the techniques are both simple to understand and to use, and they provide a natural way to handle the problems posed in this article.

Figure 1 provides the first illustration. It shows contours of the estimated joint density function of the net benefit ratio, reflecting welfare effects, and the logarithm of total household expenditure per capita for all village households in Thailand. The vertical axis shows the net benefit ratio running from 1.20, for producers whose net sales are 120 percent of their total household expenditure, to -0.60 , for households who spend 60 percent of their budget on rice. The horizontal line through zero separates those who gain from incremental increases in rice prices, below the line, from those who lose, above it. Each of the ten contour lines in figure 1 connects points at which the estimated joint density is the same, and the contours are equally spaced in that the vertical distances between any two adjacent contours are equal.

To understand how such a "map" is constructed, imagine first drawing a scatter diagram of the two variables. In principle, such a diagram would contain the same information, but with so many observations, scatter diagrams are rarely informative; there are too many superimposed points, and the eye finds it difficult to interpret the results. The contour map provides a "smoothed" version of the scatterplot. The idea is that, for each point on the map, a "height" is calculated by counting the number of scatter points that are in the neighborhood. A point is in the neighborhood if it is within a given radius, although it is a good idea to first transform the data so that the scatter is approximately circular. The degree to which the scatter is smoothed depends on the radius of the neighborhood, or "bandwidth," and the idea is to shrink the bandwidth as the sample size becomes larger. A bandwidth that is too small will produce estimates that are less biased but unnecessarily variable, whereas too large a bandwidth will sacrifice bias for smoothness. Techniques exist for optimizing this tradeoff, but in the current examples, it seems sufficient to select the bandwidth by judging the appearance of the resulting plots.

One final modification to the simple procedure outlined above is required. As we move over the map calculating heights, simple counting of occurrences within a fixed radius will produce jumps whenever a point comes into or moves

Figure 1. *Contour Plot: Thai Village Households Distributed by Net Income Effect of Rice Price Changes and Logarithm of Total Household Expenditure Per Capita*



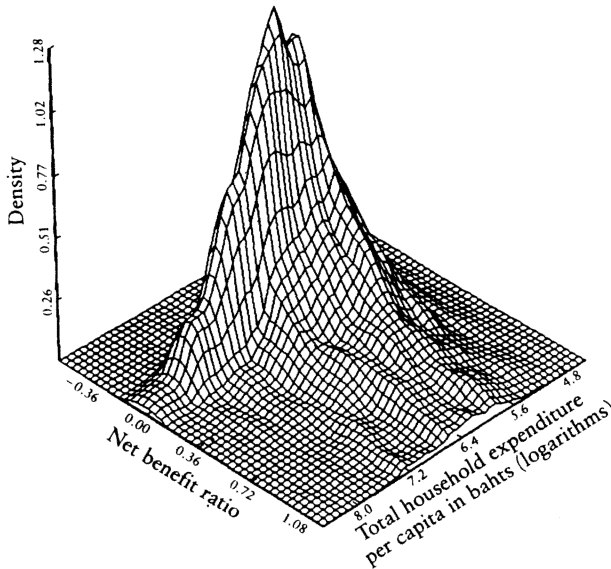
Note: The net benefit ratio $\sigma(m, p, \epsilon) = -pq/m$, where p is the price of rice, q is the net consumption of rice for the household, and m is total household expenditure. The isolated fragments of contour lines reflect the presence of outliers.

Source: Calculations based on data from the National Statistical Office, Government of Thailand.

out of the moving circle. These discontinuities are avoided by taking a weighted count, with weights summing to one, but declining as we move outward from the center of the circle. The weighting function, or kernel, is what gives this type of nonparametric estimation its name. The choice of kernel is less important than the choice of bandwidth (the calculations here use the Epanechnikov kernel, as described, for example, in Silverman 1986). The statistical appendix to this article provides the formulas used in the calculations; these are simply implemented with minimal programming.

Figure 2 provides a direct three-dimensional representation of the joint distribution of the two variables, whereas figure 1 shows the same data through a contour map. Figure 2 corresponds to figure 1 seen from above the top right-hand corner. Figure 1 shows that the mode of the distribution is located among net consumers, with a budget share for rice of about 10 percent (-0.10 on the

Figure 2. *Surface Plot: Thai Village Households Distributed by Net Income Effect of Rice Price Changes and Logarithm of Total Household Expenditure Per Capita*



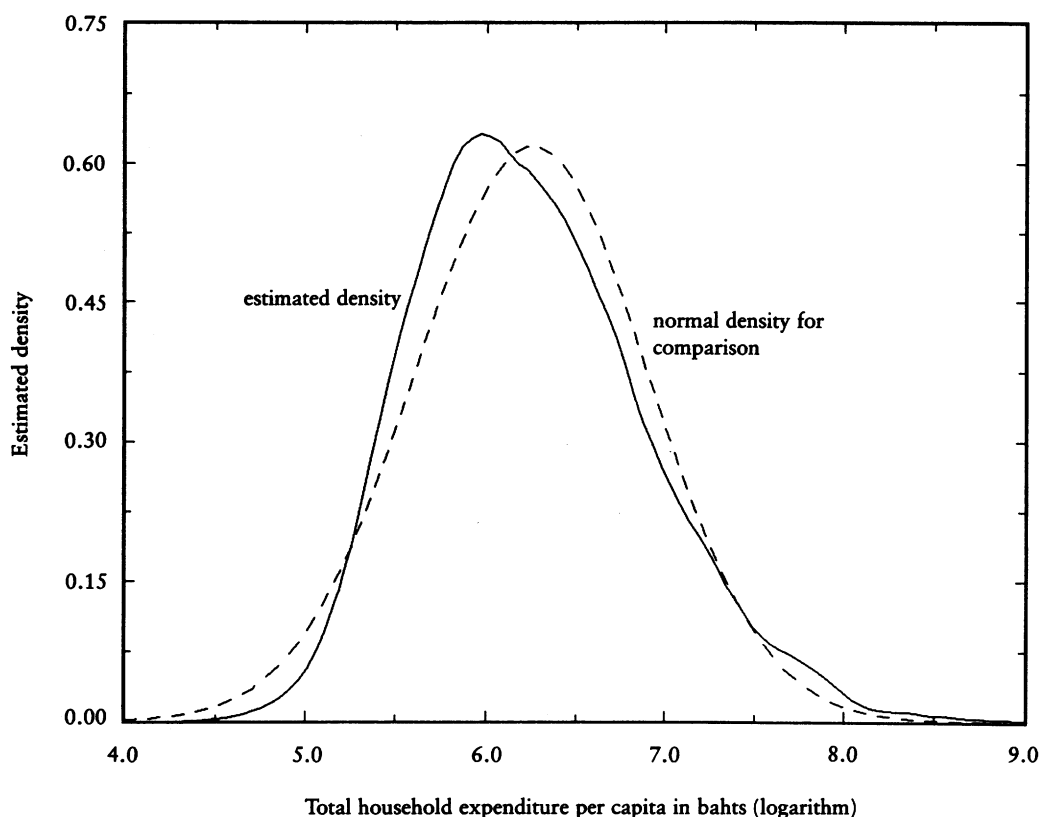
Note: The net benefit ratio $\sigma(m, p, \epsilon) = -pq/m$, where p is the price of rice, q is the net consumption of rice for the household, and m is total household expenditure.

Source: Calculations based on data from the National Statistical Office, Government of Thailand.

diagram); figure 2 suggests the possibility that there may actually be two closely adjacent modes.

The most marked feature of the distribution is the location in the real income distribution of the net sellers of rice. The big gainers from increases in the rice price are not located either among the poor or the rich but in the middle of the distribution. This outcome is the result of two effects working in opposite directions. On the one hand, the fraction of rice farmers who are net sellers increases with living standards. Wealthier farmers are much more likely to sell some of their crop than are poor farmers. On the other hand, however, although nearly 80 percent of the poorest rural households grow rice, the fraction declines among better-off households, so that less than 20 percent of the best-off households grow any rice at all. In consequence, and as figures 1 and 2 show, there are few households among the poor that gain much from higher rice prices, because poor households tend to be subsistence farmers, and there are also few gainers among the rich, because rich households tend not to grow rice at all.

Figure 3. Distribution of Household Expenditure Per Capita across Thai Village Households, 1981–82

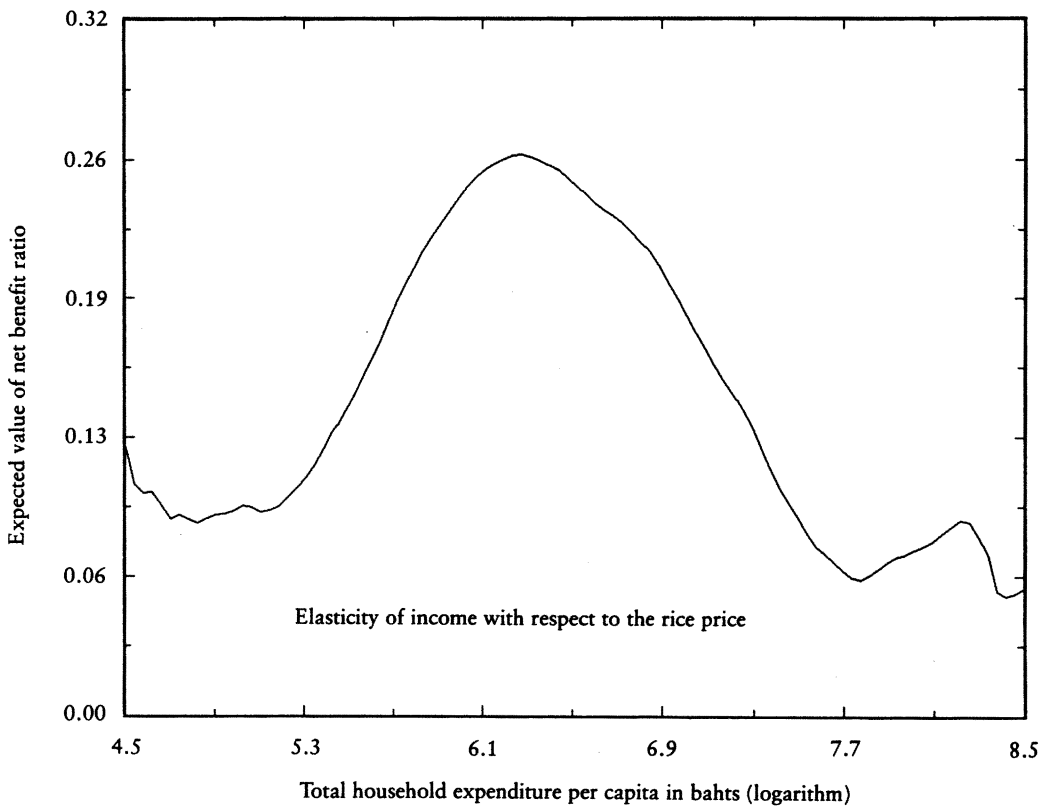


Source: Calculations based on data from the National Statistical Office, Government of Thailand.

The two quantities from equation 7 that we require, the regression function of rice sales on the level of living and the distribution of living standards, are both derivable from the joint distribution in figures 1 and 2. The latter is the marginal distribution of $(\ln xpc)$, whereas the former is the expectation of sales with respect to the distribution of sales conditional on $(\ln xpc)$. In practice, both the regression and the marginal distribution are estimated directly rather than indirectly through the estimated joint distribution (see the statistical appendix for the formulas that were used). Estimation of the distribution of $(\ln xpc)$ is simply the one-dimensional analog of the two-dimensional estimation described above—instead of counting points in a circle around each point, points are counted within a band around each point on the real line. The regression function is calculated at the same time as the density. For each point along a grid on the x -axis, a weighted mean is calculated of all y -values within the band, the weights declining with distance from the center of the band.

Figure 3 shows the resulting estimate of the distribution of $(\ln xpc)$ in village

Figure 4. *Expected Value of Welfare Effects of Rice Price Increase across Thai Village Households, 1981-82*



Note: Expected value of the net benefit ratio, σ , is calculated as $E(\sigma|x)$, where the net benefit ratio $\sigma(m,p,\epsilon) = -pq/m$, where p is the price of rice, q is the net consumption of rice for the household, and m is total household expenditure.

Source: Calculations based on data from the National Statistical Office, Government of Thailand.

Thailand, and a reference normal distribution with the same mean and standard deviation. Although the two figures are quite close, the actual distribution is more positively skewed than the normal, so that its mode lies to the left of the mean of the comparison distribution. Despite generally very low levels of living in rural Thailand, there are clearly some very well-off individual households.

Figure 4 summarizes the evidence in the joint distribution that is relevant for the policy discussion: it indicates the average change in household income due to a rise in rice prices, across different income levels. It not only can identify the gainers and losers, but also show by how much their welfare will change. If figure 4 showed a horizontal line, then households at all levels of living would benefit equally (in percentage terms) from a price increase in rice. As it is, the graph suggests that on average rural households in all income groups

can expect to benefit, but that the major beneficiaries are those in the middle of the distribution, with the poorest and the richest gaining relatively little, as was indicated in figures 1 and 2. Note that the shape revealed in figure 4 is one that would have been quite difficult to replicate with a simple parametric regression, at least if we restricted ourselves to the usual polynomial functional forms.

Armed with figures 3 and 4, we are now in a position to examine the social effects of rice price changes, at least within the rural sector. Since the gain from a price increase accrues to households at all income levels, no social welfare function is required to aggregate and weight the benefits and losses by income within the sector. When the function in figure 4 is multiplied by the density in figure 3 (see the last two terms in the integral in equation 7), the benefits of a price increase are even more heavily concentrated in the middle of the distribution. The distributional question then becomes the tradeoff between urban consumers and these middle-income rural producers. Urban households in Thailand are generally a good deal better off than are rural households, so that the rural beneficiaries of increases in the rice price are quite poor by urban standards. In consequence, the distributional case for an export tax seems to be a weak one.

I conclude this section with a number of important caveats. The discussion has been concerned only with a few of the issues that enter a question as complex as the setting of rice prices. In particular, the focus has been entirely on the effects of a "small" change in prices on households. Real price reforms sometimes involve quite large price changes, and these will generate important additional effects on production and consumption. I also have not discussed the revenue consequences of tax reform, or how the export tax fits into the general fiscal picture in Thailand. But perhaps the greatest difficulty of all is my neglect of what may be first-order effects of price changes on other prices, and particularly on wages. Higher rice prices may or may not generate higher rural wages, and if they do, the distributional effects could be quite different from those shown above. This problem is a general one in price policy analysis (see, for example, Sah and Stiglitz 1987), and there is at least one attempt to come up with plausible answers in the Thai context (Trairatvorakul 1984). In Thailand, as in most other developing countries, we lack the empirical knowledge of rural labor markets that would enable us to go much further than the sort of elementary analysis given here.

III. THE ESTIMATION OF PRICE RESPONSES FROM SPATIAL PRICE VARIATION

Assessment of price reform requires the ability to predict the budgetary consequences for the government. When changes in taxes or subsidies alter consumer and producer prices, there will be income and substitution effects that will alter demand and supply for many or all of the goods on which the

government budget depends. These effects can be of great practical importance. For example, when there are two or more crops that compete for the same land, as with coffee and cocoa in Côte d'Ivoire, tax rates for both crops have to be set jointly. As it is, coffee has been taxed at a much higher implicit rate than cocoa. There has been a long-term switch out of coffee, a switch which, in the long run, has probably lowered government revenue beyond what it might have been. Consumer subsidy schemes are another example. The subsidization of a particular staple, bread or wheat, often becomes increasingly expensive over time as consumers adjust their consumption patterns toward the subsidized good and away from higher-priced alternatives. As equation 9 shows, the desirability of a price change depends on the price elasticities of supply and demand for the item being considered, and on the cross-price elasticities between it and other commodities.

The Method

There is a substantial literature on agricultural supply and on the estimation of supply elasticities (see, for example, Bond 1983 for a collection of some of the estimates). In developed countries, demand analysis is also a well-developed branch of applied econometrics, and there exist many estimates of price elasticities, mostly derived from time series data. Much less is known about demand elasticities in developing countries. The obvious problem has been lack of data. Lluch, Powell, and Williams (1977) have provided estimates of the linear expenditure system for a large number of developing countries, and although this work is the best that can be done with the available data, it is hard to be comfortable with the very strong restrictions that are embodied in the linear expenditure system, and with the use of price elasticities that are determined as much by assumption as by measurement. Such restrictions are particularly dangerous in considering tax and price reform, where assumptions about preferences have very direct consequences for policy. For example, Deaton (1987a) shows that, with the linear expenditure system, and with a standard model of indirect taxation, it is always desirable to increase a tax rate that is below the mean and to decrease one that is above it. Such rules have their attractions, but it would be better to base such recommendations on measurement than on assumption.

In a series of recent papers (Deaton 1987b, 1988, forthcoming), I have been exploring the possibility of using household survey data as a basis for estimating price elasticities. The basic ideas have been known for some time. In their famous 1955 monograph, Prais and Houthakker noted that many household surveys collected data on both quantity and expenditure, particularly for foodstuffs, where the former is well-defined in terms of weight. It is therefore possible to calculate, for each household, a unit value for each of the foods purchased. For example, the category "meat" will cover a range of commodities, so that if the vector of meats is written q , the i th element being the number of kilos of meat type i , and the price per kilo of type i is p , then the unit value index for meat

would be defined for a group F as value divided by volume or weight, or

$$(10) \quad V_F = \sum_{k \in F} p_k q_k / \sum q_k$$

These unit values clearly have something to do with price. But as Prais and Houthakker noted, the balance between high-quality, expensive items and lower-quality, cheaper ones is something that is chosen by the consumer, and we might reasonably expect better-off consumers to choose more expensive items within the same general category. Prais and Houthakker confirmed this supposition by calculating "quality" elasticities from regressions of the logarithm of the unit value on the logarithm of household income and found strongly significant effects, some of which were reasonably large, 0.3 for example.

After Prais and Houthakker, a number of authors, mostly in the agricultural economics and food policy literatures, estimated price elasticities from household survey data by the simple expedient of regressing the logarithm of quantity ($\ln \Sigma q_k$) on the logarithm of unit value ($\ln V$), together with the usual range of other variables included in Engel curve estimation—income, household composition, and other characteristics. Such a procedure, if it can be justified, is a very attractive one. The size of household surveys for many developing countries is a demand analyst's dream—the Indian National Sample Survey now has a sample size of around a quarter of a million households. With so many observations, the precise estimation of large numbers of own- and cross-price elasticities seems to be within reach. Further, it is quite reasonable to suppose that much of the interhousehold variation in unit values is due to genuine price variation. Poor, largely rural economies often have poorly integrated markets, transport is expensive and unreliable, and large price differences between villages are frequently reported.

However, the simple regression of log quantity on log unit price has two immediate problems, both of which are likely to bias the procedure toward finding elasticities that are (absolutely) too large. The first is the quality issue, and the fact that quality is not an exogenous variable but one chosen by consumers. This becomes important if, as seems likely, consumers shade down both quality and quantity as prices rise. In places, or at times when prices are high, qualities will be low, and vice versa when prices are low. In consequence, unit values, which reflect both price and quality, will tend to vary by less than prices. The comparison of quantities with unit values, rather than with prices, will therefore overstate the extent to which quantities are affected by price.

The second problem, and it turns out to be the more important in practice, is measurement error. Since the logarithm of unit value is defined as the logarithm of expenditure less the logarithm of quantity, reporting errors in either expenditure or quantity will generally be transmitted to the unit value. Further, and unless either quantity or unit value is reported without any error at all, there will be a spurious negative correlation between reported quantity and reported unit value. An individual that remembers the value of a purchase but

understates its weight will record a spuriously high unit value, which can be used to "explain" the spuriously low weight.

In the final analysis, the quality issue can only be solved by additional direct measurement of quality or by independent recording of market price. Without such data, there is an identification problem that comes from working with three magnitudes, quality, quantity, and price, with data on only two of them.

The solution that I have adopted relies on an assumption that the quality downgrading that accompanies a rise in price is an income effect, so that an increase in all meat prices causes substitution toward lower quality cuts in the same way that would follow from a cut in money income. If it is supposed that meat (or whatever the group is) is a separable branch in consumers' preferences, price effects are automatically related to income effects, and it is possible to write

$$(11) \quad \partial \ln V_F / \partial \ln P_F = 1 + \eta \epsilon_p / \epsilon_x$$

where P_F is the price of meat, η is the income elasticity of the unit value, and ϵ_p and ϵ_x are the price and income elasticities of meat (see Deaton 1988b: p. 422).

The conceptual experiment here is one in which the price of all meats rises together, so that the first term on the right-hand side, unity, is the effect on the unit value if there were no quality shading. The second term is proportional to η , which is the elasticity of quality with respect to income, so that if quality does not upgrade in response to income, it will not downgrade in response to price. The size of this effect is controlled by ϵ_x , because the income effect of increasing the prices of all meat depends on the income elasticity of meat, and by ϵ_p , because the scope for downgrading depends on the extent to which the price change induces a reduction in quantity. Given estimates of the relation between quantities and unit values, equation 11 provides a means to separate out the true price elasticities since it shows how much of the relation results from quality downgrading in response to price. In practice, the correction turns out not to be very important, because in the work so far, the effects of income on quality, the parameter η has always been estimated to be small, so that, according to equation 11, unit values move approximately one-for-one with price.

The solution to the measurement error problem relies on a feature of sample design that is common to nearly all household surveys, in developed and developing countries alike. Sample households are never randomly scattered in space but are clustered in "blocks." For rural areas in developing countries, these blocks can nearly always be associated with villages. The team of survey enumerators comes to the village, selects a dozen or so households, and works through the questionnaires. Such a scheme helps keep down travel costs, and is almost inevitable if, as is usually the case, the survey design requires more than one visit to each household. In many developing countries, villages are also associated with markets, so that it will often be reasonable to assume that households in the same block make their purchases in the same market and

thus face the same prices. Of course, prices vary over time as well as over space, but the households in the same cluster are usually interviewed within a short span of time during which the survey team is in the village. What this gives us is a sample design in which the households break up into clusters, and within each cluster all households face the same prices for the goods that they buy.

In the actual data, the unit values vary within the cluster. One reason for this is the quality issue; better-off households may record higher unit values even if they purchase their goods in the same market, provided that there is sufficient variety from which to choose. The other reason is measurement error. In consequence, if we can control for quality, the intracluster variation in reported unit values can be used to estimate the part of the variance in unit values that is due to measurement error. If quality is not important, one possible procedure is to average unit values over each cluster (see Strauss 1982 for a similar treatment of prices in Sierra Leone). If, as seems reasonable, measurement error is uncorrelated across households, the measurement error in the average cluster unit value will have a lower variance, by a factor of the cluster size, than the measurement error in the individual unit values. If the cluster size is large enough, this technique could be adequate, and it certainly ensures the consistency of the parameter estimates as the cluster size tends to infinity. The problem is that cluster sizes are often quite small, sometimes as small as four or six households, so that, even after averaging, the measurement error may remain quite serious.

The procedure with which I have worked can be outlined as follows. At the first stage, estimate within-cluster regressions of the unit value on household incomes (or total expenditures) and possibly on other household characteristics. The within-cluster regression is computed by ordinary least squares using as data the household observations less their cluster means, exactly as for a within-estimator using panel data. The residual variance of this regression, which in many cases is close to the total variance, is an estimate of the part of the observed variance in unit values that is due to measurement error.

At the second stage, a conventional demand system is estimated in which cluster averages of quantities (or expenditures, or expenditure shares, depending on the functional form) are related to cluster averages of income, household characteristics, and cluster average unit values. In the second-stage estimation, allowance has to be made for the variance of measurement error in the unit values. This is computed from the first-stage estimates, corrected for the numbers of households in the cluster, and then subtracted from the observed cross-cluster variance of the unit values. This is essentially an errors-in-variables procedure, in which the standard least squares estimator $(X'X)^{-1}(X'y)$ is replaced by $(X'X - Q)^{-1}(X'y - q)$ where Q and q are the components of measurement error in $X'X$ and $X'y$ respectively. For the few cases in which important quality effects are estimated at the first stage, the second-stage estimator of the price elasticity has to be further manipulated to eliminate the

upward bias from quality downgrading. The details for the simplest one-good, one-price case are given in the statistical appendix; the general system case is presented in Deaton (forthcoming), where there is also an account of the very tedious manipulations that are required to produce estimates of standard errors.

Before summarizing the results, it is worth clarifying some of the basic assumptions that are required for the analysis to be successful. The procedure relies on the existence of genuine variation in prices across space, and requires that such price variation be exogenous to the process that determines demand. The existence of spatial price variation is unlikely to be doubted seriously, but exogeneity is much more difficult. If local prices are determined by world prices, border taxes, and transport costs, the assumptions will be satisfied, because local demand has no effect on prices. But more generally, if village prices depend on demand in the village, the estimates will not be consistent, for all the usual simultaneity reasons. In principle, the procedures discussed here could be modified to handle this case, but the work has not yet been done.

The second crucial assumption, required for the treatment of measurement error, is that there is no genuine price variation within clusters, only between clusters. Here the assumption seems quite plausible, but there may be some clusters with more than one market, and some households may make purchases outside the cluster. There are also several assumptions that go into the story about quality, and these have the effect of limiting the range of relative price variation that can be allowed between different clusters over space. But these are a good deal less crucial, simply because the quality effects themselves are not very important and have little effect on the estimates. Note finally that, in general, I do not need to assume that each household buys each good, nor do I need to deal with the problems of corner solutions. Zero purchases are combined with nonzero purchases because, as before, I am interested in the effects of price changes on mean purchases over all households, whether or not they buy positive amounts.

Application of the Approach

The procedure has so far been applied to data from three developing countries, Côte d'Ivoire, Indonesia, and Morocco. In Côte d'Ivoire, which was the first country to be studied, data on 1,920 households from the 1979 *Enquête Budget Consommation* were used to estimate demand functions for five food categories: meat, cereals, starches, fresh fish, and dried and smoked fish.

In the first experiments, I worked with a double logarithmic demand function relating log quantities to the logarithms of the total budget and the logarithm of the price/unit-value. Such a system is simple to interpret and permits a particularly simple characterization of the effects of measurement error. However, logarithmic demand functions suffer from well-known theoretical drawbacks, and more seriously, they pose severe difficulties in handling households that do not purchase the good being analyzed. Such cases are very common, particularly in rural Côte d'Ivoire where many consumer goods are not ob-

tained through the market, and the exclusion of zero purchases not only introduces potentially serious selectivity bias but also involves the loss of a substantial fraction of the sample, a fraction that we wish to include to estimate the desired aggregate price responses.

A more satisfactory procedure is to use a functional form that permits zeros, and subsequent work has been based on the "share-log" form

$$(12) \quad w_i = \alpha_i + \beta_i \ln x + \sum_j \gamma_{ij} \ln p_j + \theta' z + u_i$$

where w_i is the share of the budget devoted to good i , x is total expenditure, (or for a food subsystem, total expenditure on food), p_j are the prices, and z is a vector of other relevant household characteristics. If the model is estimated for own-price elasticities alone, with cross-price elasticities being ignored, substantial price elasticities are estimated for all five goods, with figures of -0.8 for meat and starches, -1.1 for cereals, and -1.6 and -1.2 for fresh and dried/smoked fish respectively. The addition of cross-price effects makes relatively little difference in this case; there is a strong degree of substitutability between the two types of fish, but no other very marked effects. This result is somewhat surprising, particularly in view of the regional pattern of fish and meat consumption in Côte d'Ivoire. Fish is consumed along the coast and near major lakes and rivers, mostly in the south, while, in the north, beef is much more heavily consumed. Apparently this pattern can be adequately accounted for by assigning high own-price elasticities to each good, without reference to cross-price effects.

The second, more ambitious attempt used a sample of more than 14,000 households from rural Java, taken from the Indonesian National Socioeconomic Survey (SUSENAS) of 1981. Given the large sample size, eleven food commodities were included in the analysis: rice, wheat, maize, cassava, roots, vegetables, legumes, fruit, meat, fresh fish, and dried fish. From a policy perspective, rice is the major focus of interest. As in Thailand, it is the basic staple, and its price is influenced by a number of policy interventions. Rice also accounts for nearly 25 percent of the total budget; the highest among the other goods considered here is 5.8 percent for maize (table 1). Cassava is also important in discussions of food policy. It requires time-intensive preparation prior to consumption, and it is much more important in the budget of poor households than of rich. It is therefore an attractive candidate for subsidization, particularly if it can be shown to have a relatively low price elasticity.

Table 1 shows the average budget shares for the eleven goods, together with the estimated elasticities of quality and quantity with respect to total expenditure, and the elasticity of quantity with respect to price. The figures in parentheses are estimated asymptotic t -values. In this sort of model, the elasticity of expenditure for a particular good is the sum of the elasticity of quantity and the elasticity of quality. Thus, to take an example from the table, the elasticity of total expenditure on fresh fish with respect to total household expenditure is estimated to be 1.30, which decomposes into a quantity component of 1.08

Table 1. *Demand Elasticity Estimates for Java, 1981*

<i>Commodity</i>	<i>Share of total budget (percent)</i>	<i>Elasticity of quality with respect to total expenditure</i>	<i>Elasticity of quantity with respect to total expenditure</i>	<i>Elasticity of quantity with respect to total price</i>
Rice	24.53	0.03 (9.0)	0.49 (58)	-0.42 (5.1)
Wheat	0.52	0.10 (1.1)	1.57 (23)	-0.69 (14)
Maize	5.77	-0.00 (0.0)	0.09 (3.3)	-0.82 (7.5)
Cassava	1.39	0.02 (0.7)	0.14 (3.5)	-0.33 (2.8)
Roots	0.60	0.17 (2.8)	0.71 (14)	-0.95 (22)
Vegetables	5.57	-0.04 (1.8)	0.67 (25)	-1.11 (29)
Legumes	3.66	0.04 (4.6)	0.85 (42)	-0.95 (0.8)
Fruit	1.88	0.07 (2.7)	1.39 (40)	-0.95 (0.5)
Meat	2.07	0.09 (1.9)	2.30 (43)	-1.09 (0.7)
Fresh fish	2.95	0.22 (8.5)	1.08 (35)	-0.76 (3.8)
Dried fish	2.83	0.06 (5.7)	0.57 (26)	-0.24 (9.4)

Note: Figures in parentheses are *t*-statistics.

Source: Calculations based on data from the Central Bureau of Statistics, Government of Indonesia.

and a quality component of 0.22. This estimated quality elasticity for fresh fish is much the largest of those in the table. Apart from roots, with an elasticity of 0.17, all the others are less than 0.1, usually substantially so. And although the estimate for rice, at 0.03, is significantly different from zero, it is clear that quality upgrading is not an issue for this important commodity. The quantity elasticities show that both maize and cassava respond very little to extra income; likewise rice and dried fish are classed as necessities. Meat, wheat, fruit, and fresh fish are all classed as luxury foods.

When interpreting estimates of the own-price elasticities (table 2), it is important to bear in mind that the model (equation 12) has the feature that an estimated price coefficient γ_{ii} of zero yields not a price elasticity of zero but a price elasticity of -1. A number of the estimates, those for vegetables, legumes, fruit, and meat, are not significantly different from this "default" value. For the

Table 2. *Own- and Cross-price Demand Elasticities, Java, 1981*

	<i>Rice</i>	<i>Maize</i>	<i>Cassava</i>	<i>Vegetables</i>	<i>Legumes</i>	<i>Meat</i>	<i>Fresh fish</i>	<i>Dried fish</i>
Rice	-0.42*	-0.03	0.09*	-0.06*	0.06*	-0.06	0.05*	0.21*
Maize	1.25*	-0.82*	-0.26*	0.60*	0.01	0.17	-0.44*	-0.11
Cassava	0.15	0.19	-0.33*	0.12	0.01	0.18	0.05	-0.67*
Vegetables	-0.05	-0.10*	0.10*	-1.11*	-0.08	-0.06	0.05	-0.17*
Legumes	0.11	-0.20	-0.05	-0.12	-0.95	0.20	-0.08	-0.13
Meat	-0.19	-0.06	-0.17*	0.18*	0.03	-1.09	-0.04	0.20
Fresh fish	0.40*	0.24*	-0.02	0.32*	0.03	0.05	-0.76*	0.26*
Dried fish	0.39*	0.15	0.06	0.38*	-0.02	-0.06	0.67*	-0.24*

Note: The column identifies the commodity the price of which is changing and the row the commodity being affected. An asterisk indicates that the estimate is at least as large as twice its estimated standard error.

Source: Calculations based on data from the Central Bureau of Statistics, Government of Indonesia.

other commodities, however, the estimates of γ_{ii} are significantly positive, so that the estimated price elasticities are less than unity in absolute value. The estimated price elasticity of rice, at -0.42 , is well within the range of other studies, and it is also worth noting that two of the other necessities, cassava and dried fish, both have low estimated price elasticities, -0.33 and -0.24 respectively. This combination of low quantity elasticities and low price elasticities is a genuine finding of this study. Alternative procedures based on models such as the linear expenditure system that incorporate additive preferences will also produce the same result, but it is a consequence of prior assumption and will be "discovered" whatever are the true underlying patterns in the data.

A selection of the cross-price elasticities is shown in table 2. Judging these for plausibility is not easy, partly because there are few estimates with which they can be compared, and partly because it is too easy to provide an *ex post* rationalization for almost any pattern, however extraordinary. It is reassuring that most of the estimates are quite small, and that the prices of most goods do not have dramatic effects on the consumption of other goods. Some of the nonsmall figures are also very sensible. As was the case for Cote d'Ivoire, dried and fresh fish are substitutes for one another, although in other respects they behave quite differently. Looking at the effects of fish prices, fresh fish is a complement with maize, and dried fish with cassava. Vegetables appear to be substitutes for a number of other goods, including maize and both kinds of fish. The positive estimates in the first column do not have obvious explanations, particularly given the fairly large negative income effects that are to be expected when the price of rice increases. There are also some odd pairings; maize and vegetables, maize and fresh fish, and vegetables and dried fish are three cases in which the cross-price elasticities are each significantly different from zero but are of opposite signs in the two pairings, in apparent violation of Slutsky symmetry. A Wald test for symmetry produces a test statistic of 416, to be compared with a χ^2 -distribution with 55 degrees of freedom (under the null hypothesis, that symmetry is true). Clearly, such a test fails under any classical testing procedure, but it should be remembered that the price elasticities are estimated from a sample of 2,000 clusters, and the use of Schwartz's (1978) large sample Bayesian test criterion gives a critical value of 418, by which symmetry would (just) be accepted.

There is a third study using the same model, that of Laraki (1988) of food subsidies in Morocco. Over a long period of time, the government of Morocco has subsidized three important staples, soft wheat, vegetable oil, and sugar powder. In recent years, the cost of these subsidies has reached nearly 10 percent of government expenditure, but political pressures have made it extremely difficult to eliminate or reform the system. Laraki uses a 1984–85 household survey to estimate a demand system of seven commodities, the three subsidized goods together with possible substitutes and complements, hard wheat, barley, olive oil, and sugar loaf. His results are complex and not easily summarized. There are, for example, very different patterns of quantity and

price elasticities between the urban and rural sectors, with hard wheat sharply inferior in the urban areas and soft wheat inferior in the rural sector. Price elasticities also vary a great deal by sector, with barley estimated to be extremely price elastic in the urban sector but quite inelastic in the countryside. One surprising feature of these results is that they do not show the expected patterns of strong substitutability between the commodities in the three groups, cereals, oil, and sugar. Laraki uses his estimates to consider the budgetary implications of various price reforms and also goes beyond the theoretical framework in section I to examine the nutritional effects of price changes, effects that also depend on price elasticities.

I note finally a study of the demand for food in the United States by Nelson (1987). Nelson uses data on eleven foods from the 1985 U.S. Bureau of Labor Statistics (BLS) Consumer Expenditure Survey to estimate double-logarithmic demand equations without cross-price effects. What she finds is in very sharp contrast to any of the results from the three developing countries reported here. Quality elasticities are typically quite large and frequently dominate the total expenditure elasticity. For example, for sugar and other sweets, Nelson finds an expenditure elasticity of 0.022, which decomposes into a quantity elasticity of -0.231 and a quality elasticity of 0.254. But it is in the price elasticities where this study differs most sharply from those in developing countries. Nelson obtains a wide range of figures, sometimes positive and sometimes negative, but always with very large standard errors. The problem seems twofold. Because the BLS survey is mostly urban, the study has to include primary sampling units (PSUs) in urban areas, and some of these are large enough (for example, the Los Angeles metropolitan area) to violate the requirement that all households in the cluster (here PSU) face the same prices. Secondly, the U.S. consumer market is so well-integrated that it would be surprising if there were very sharp regional differences in prices. In Nelson's calculations, the measurement error accounts for most of the price variability, and there is not sufficient information remaining to provide sharp estimates of price elasticities. None of this, of course, suggests that the method is not appropriate for rural areas of developing countries.

IV. CONCLUSIONS

I have tried here to give an outline of the contributions that can be made by household budget analysis to the analysis of pricing policies in developing countries. I hope that I have conveyed some of the enthusiasm that I feel for this branch of research. Several years ago, I wrote about the econometric problems associated with tax design in developing countries, (Deaton 1987a), and at that time, I was very skeptical of whether econometric analysis had anything to offer that could reasonably be expected to improve the sort of recommendations that could be reached by purely deductive reasoning. I am a good deal more hopeful now, and I think that the techniques discussed in this

article have a good deal to contribute. However, it is only fair that I conclude with some warnings, and that I acknowledge a range of issues that have still not been adequately addressed.

The nonparametric display techniques of section II enable researchers and policymakers to "see" survey data in a way that is not otherwise easily done. It is easy to minimize the benefits of simple methods for displaying evidence. Techniques that give rapidly assimilable information on who gains and who loses from a price change can immediately cut through the sort of inflated (and often imaginary) claims that are sometimes made, for example, for subsidies helping the poor. Many subsidy schemes in developing countries have powerful constituencies that are by no means poor, and the survey data can rapidly reveal the discrepancies. But it is difficult to go a great deal further using these techniques. Most economic phenomena are not simply described in terms of relations between two or three variables, and econometricians have evolved a great number of techniques that (at least sometimes) permit us to disentangle phenomena of interest from a mass of uncontrolled evidence. Nonparametric analysis is only beginning to explore this territory. It is difficult to extend diagrams like figures 1 and 2 to more than two dimensions, partly because the results are very hard to display, and partly because of the "curse of dimensionality" which, as dimensions expand, quickly demands astronomical sample sizes. Nevertheless, the area is one of immensely active research by statisticians and econometricians, and there are undoubtedly great advances still to be made.

The techniques discussed in section III are much more conventional, and it is clear that the modeling in that area could benefit from some of the nonparametric techniques in section II. Although the price responses are probably too delicate to be modeled without strong assumptions about functional form, much could be done with the estimation of the Engel curves, perhaps by applying nonparametric regression to the within-cluster estimation. But there are other perhaps more fundamental questions both about the use of spatial price variation to identify price responses in general, and also about the particular techniques I have used to capture it. I conclude with a list of these questions, a sort of shopping list for further enquiry.

First, the contrasts between "spatial" demand analysis and "temporal" demand analysis force some thought about what is meant by a price elasticity. People who live in coastal Côte d'Ivoire, where fish is plentiful and cheap, consume a great deal of fish, while consumers in the Northern Savannah, where fish are nowhere to be seen, consume very little fish. In Europe, Norwegians consume a great deal more fish than do Swiss, and so on. The price differences associated with these different consumption patterns have been in place for a long time, perhaps thousands of years, and it is far from obvious that comparable changes in consumption patterns would come about in response to policy-induced price changes. The underlying issue is one of dynamics, and how short-run price responses relate to long-run responses, what are the associated changes

in complementary factors (including perhaps induced habits), and what is the time horizon with which policymakers are concerned.

Second, although the results of the various studies reported here are quite encouraging, a great deal more work has to be done before it can be asserted definitively that these procedures actually estimate price elasticities. The best confirmation would be from direct measurement of prices, and some work along these lines has been done in the World Bank's recent Living Standards Surveys in Côte d'Ivoire, Ghana, and Peru, where the household surveys were supported by community surveys which recorded prices in the markets used by the households in the survey. In some preliminary experiments reported in Deaton (1988), the Ivorian data from the Living Standards Survey were used to produce estimates that were not too far from the figures obtained by the methods reported here and using a quite different earlier survey. Unfortunately, the living standards surveys do not collect data on quantities consumed, so that it is not possible to make the crucial experiment, which is to compare the unit values from the household survey with the prices reported directly from the markets.

Third, a number of theoretical issues also need to be addressed. The quality model used in this work is a very simple one in which expenditures are simply the product of price, quality, and quantity. More work needs to be done on the choice-theoretic foundations of such models, particularly in relation to the welfare implications. Preferences are typically defined over quantities, but such a specification is inadequate when quality is also subject to choice. Until more work is done on these issues, the results reported here must be regarded as preliminary ones.

STATISTICAL APPENDIX

This brief appendix provides some of the formulas that are used to calculate the figures and tables presented in the main text. I start with the kernel estimation techniques used to produce figures 1 to 4. There are two variables in the analysis, the net benefit ratio and the logarithm of per capita household expenditure. So that I can use standard notation, relabel these as Y_i (the benefit ratio) and X_i ($\ln xpc$), where i refers to the observation (household) number, running from 1 through n (5,836). The figures show estimates of the joint density of X and Y (figures 1 and 2), the marginal density of X (figure 3), and the regression of Y on X (figure 4). First define a "kernel" or weighting function $K(u)$, which has the property that over the range of u it integrates to unity, has a single mode at $u = 0$, is symmetric around the origin, and is nonincreasing in the absolute value of u . In this paper, I use the Epanechnikov kernel, which takes the form

$$(A-1) \quad K(u) = 0.75(1 - u^2) \text{ if } |u| \leq 1 \\ = 0, \text{ otherwise.}$$

But the results would not be very sensitive to a different choice, for example, the standardized normal density function. The bandwidth, h , is incorporated by writing $K_h(u) = h^{-1}K(u/h)$, and the kernel estimate of the marginal density of X is given by

$$(A-2) \quad \tilde{f}_h(X) = n^{-1} \sum K_h(X_i - X).$$

This estimate can be thought of as a weighted "count" of points in the vicinity of X as a proportion of the total sample size. The regression function in figure 4 is computed using the kernels as weights, so that, if the regression function $E(Y|X)$ is denoted as $m(X)$, the estimate is

$$(A-3) \quad \tilde{m}(X) = \sum w_i(X, X_i) Y_i; w_i = K_h(X_i - X) / \sum K_h(X_j - X).$$

To calculate the two-dimensional density estimate shown in figures 1 and 2, define first the two-dimensional Epanechnikov kernel

$$(A-4) \quad K(d) = (2/\pi) (1 - d'd) \text{ if } d'd \leq 1 \\ = 0, \text{ otherwise}$$

where d is a two-element vector. Define $K_h(d)$ by $h^{-2}K(d/h)$, and write Z for the two-element vector of X and Y . The estimate of the bivariate density, $g(X, Y)$ or $g(Z)$, is, then,

$$(A-5) \quad g_h(Z) = n^{-1} (\det S)^{-1/2} \sum k\{(Z_i - Z)' S^{-1} (Z_i - Z)\}$$

where $k(d'd) = K(d)$, and S is the sample variance covariance matrix of Z .

For the estimation of price elasticities, I give details here for only the one-good, one-price case; the calculations are very much more complex for the system case, and are given in detail in Deaton (1988a). The basic equations are 12 in the text, and a corresponding equation for the unit value $\ln V$, that is:

$$(A-6) \quad w_{hc} = \alpha_1 + \beta_1 \ln x_{hc} + \gamma_1 \ln p_c + \theta'_1 z_{hc} + u_{1hc}$$

$$(A-7) \quad \ln V_{hc} = \alpha_2 + \beta_2 \ln x_{hc} + \gamma_2 \ln p_c + \theta'_2 z_{hc} + u_{2hc}$$

where the subscripts h and c refer to household h in cluster c . The error term u_1 contains a village fixed effect, while both u_1 and u_2 contain measurement error. The price terms are not observed, but there are data on the budget shares w , the unit values V , household per capita expenditure x , and household demographic variables z . The first stage of the estimation is accomplished by removing from each of the observables the corresponding cluster mean, and then running ordinary least squares (OLS) regressions corresponding to equations A-6 and A-7, without, of course, the prices which have been removed by the within cluster differencing. These OLS regressions yield the final estimates of the β and θ parameters. The residuals from these regressions, e_1 and e_2 , are used to estimate $\tilde{\sigma}_{11} = (n - k - C)^{-1} e'_1 e_1$, $\tilde{\sigma}_{22} = (n_1 - C - k)^{-1} e'_2 e_2$, 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, n_1 is the number of such

households, n is the total number of households, k is the number of explanatory variables, and C is the number of clusters. These σ parameters are estimates of the degree of measurement error.

At the second stage of estimation, the raw data on budget shares and log unit values (without cluster means removed) have the effects of total expenditure and demographics removed by calculating $\tilde{y}_{1hc} = w_{hc} - \tilde{\beta}_1 \ln x_{hc} - \tilde{\gamma}_1 z_{1hc}$, and $\tilde{y}_{2hc} = \ln V_{hc} - \tilde{\beta}_2 \ln x_{hc} - \tilde{\gamma}_2 z_{1hc}$. These are then used in a cross-cluster errors-in-variables regression to estimate $\phi = \theta_1/\theta_2$, by

$$(A-8) \quad \tilde{\phi} = \frac{\text{cov}(\tilde{y}_1, \tilde{y}_2) - \tilde{\sigma}_{12}/\tau}{\text{var}(\tilde{y}_2) - \tilde{\sigma}_{22}/\tau^+}$$

where $\tau = C/(\Sigma n_c^{-1})$, $\tau^+ = C/(\Sigma n_c^{+1})$, and n_c is the number of households in cluster c . The estimate of θ_1 is calculated from the estimate of ϕ using the formula $\theta_1 = \phi\{\beta_1 + w(1 - \beta_2)\}/(\beta_1 + w - \phi\beta_2)$, which can be deduced using the result on the response of unit value to price (equation 11). Given this estimate and the other parameters, price and total expenditure elasticities can be estimated in the usual way.

The evaluation of standard errors is complicated by the need to recognize that the magnitudes in A-8 are themselves estimated at the first stage. However, as is often the case, the correction is typically small, and if it is ignored, the variance can be computed by standard formulas, namely,

$$(A-9) \quad v(\tilde{\phi}) = (m_{22} - \sigma_{22}/\tau)^{-2} \{ \pi^0 m_{22} + (m_{12} - \phi m_{22})^2 \} (C-1)^{-1}$$

$$(A-10) \quad \pi^0 = m_{11} - 2m_{12}\phi + m_{22}\phi^2$$

where m_{11} is the variance of y_{1c} and m_{12} and m_{22} are, respectively, the covariance and variance in equation A-8. The corrections for first-stage estimation are described in Deaton (1988b: appendix). All calculations, for this and for the general case, were carried out using PROC MATRIX in SAS.

REFERENCES

- Bock, Carl. 1884. *Temples and Elephants: Travels in Siam 1881-1882*. London: Sampson Low, Marston, Searle, and Rivington.
- Bond, M. E. 1983. "Agricultural Responses to Prices in Sub-Saharan African Countries." *International Monetary Fund Staff Papers* 30: 703-26.
- Cowell, F. A. 1977. *Measuring Inequality*. Deddington: Philip Allan.
- Cramer, J. S. 1969. *Empirical Econometrics*. Amsterdam: North-Holland.
- Deaton, A. S. 1987a. "Econometric Issues for Tax Design in Developing Countries." In D. M. G. Newbery and N. H. Stern, eds., *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.
- . 1987b. "Estimation of Own and Cross Price Elasticities from Household Survey Data." *Journal of Econometrics* 36: 7-30.

- . 1988b. "Quality, Quantity, and Spatial Variation of Price." *American Economic Review* 78: 418–30.
- . 1989. "Rice Prices and Income Distribution in Thailand: a Nonparametric Analysis." *Economic Journal: Supplement* 99: 1–37.
- . Forthcoming. "Price Elasticities from Survey Data: Extensions and Indonesian Results." *Journal of Econometrics*.
- Deaton, A. S., and J. Muellbauer. 1980. *Economics and Consumer Behavior*. New York: Cambridge University Press.
- Diamond, P. A., and J. A. Mirrlees. 1971. "Optimal Taxation and Public Production I and II." *American Economic Review* 61: 8–27, 261–78.
- Härdle, Wolfgang. 1988. "Applied Nonparametric Regression." Fakultät Rechts-und Staatswissenschaften. University of Bonn.
- Laraki, Karim. 1988. *Food Subsidy Programs: A Case Study of Price Reform in Morocco*. World Bank Living Standards Measurement Studies. Working Paper 50. Washington, D.C.
- Lee, L. F., and M. M. Pitt. 1986. "Microeconomic Demand Systems with Binding Non-negativity Constraints: the Dual Approach." *Econometrica* 54: 1237–42.
- Lluch, Constantino, A. A. Powell, and R. A. Williams. 1977. *Patterns in Household Demand and Saving*. New York: Oxford University Press.
- Nelson, J. A. 1987. "Quantity Aggregation and Price Variation in U.S. Consumer Demand for Food." Bureau of Labor Statistics, Offices of Prices and Living Conditions, Working Paper 168 (July). Washington, D.C. Processed.
- Newbery, D. M. G., and N. H. Stern. 1987. *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.
- Prais, S. J., and H. S. Houthakker. 1955. *The Analysis of Family Budgets*. Cambridge: Cambridge University Press.
- Sah, R. K., and J. E. Stiglitz. 1987. "The Taxation and Pricing of Agricultural and Industrial Goods in Developing Economies." In David Newbery and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.
- Schwartz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461–64.
- Siamwalla, Ammar, and S. Setboonsarng. 1987. "Agricultural Pricing Policies in Thailand: 1960–85." World Bank Special Studies Division, Country Economics Department. Washington, D.C. Processed.
- Silverman, Bernard. 1986. "Density Estimation for Statistics and Data Analysis." London and New York: Methuen.
- Strauss, John. 1982. "Determinants of Food Consumption in Rural Sierra Leone." *Journal of Development Economics* 11: 327–53.
- Trairatvorakul, Prasarn. 1984. *The Effects on Income Distribution and Nutrition of Alternative Rice Price Policies in Thailand*. International Food Policy Research Institute Research Report 46. Washington, D.C.
- Wales, T. J., and A. D. Woodland. 1983. "Estimation of Consumer Demand Systems with Binding Non-negativity Constraints." *Journal of Econometrics* 21: 263–85.