NBER WORKING PAPER SERIES

PARAMETRIC AND NON-PARAMETRIC APPROACHES TO PRICE AND TAX REFORM

Angus Deaton
Serena Ng

Working Paper 5564

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1996

We are grateful to Jim Powell for helpful discussions. This paper is part of NBER's research program in Public Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper 5564
May 1996

PARAMETRIC AND NON-PARAMETRIC APPROACHES TO PRICE AND TAX REFORM

ABSTRACT

In the analysis of tax reform, when equity is traded off against efficiency, the measurement of the latter requires us to know how tax-induced price changes affect quantities supplied and demanded. In this paper, we present various econometric procedures for estimating how taxes affect demand. We examine advantages and disadvantages of parametric methods of tax reform analysis and suggest that the nonparametric "average derivate estimator" is a useful alternative. We apply both parametric and nonparametric methods to analyze possible price reform for foods in rural Pakistan, and discuss the issues that remain to be dealt with in empirical welfare analyses.

Angus Deaton
221 Bendheim Hall
Research Program in Development Studies
Princeton University
Princeton, NJ 08544-1022
and NBER

Serena Ng
Department of Economics
University of Montreal
CP 6128 Succ Centre Ville
Montreal, Quebec, H3C 3J7
CANADA

0. Introduction: the basic problem

Most developing countries collect a large proportion of their revenue from indirect taxation, especially imposts on trade and on manufactured (modern-sector) goods. In Pakistan, which is the country examined in this paper, more than 80 percent of federal government revenue in the 1970s and 1980s came from indirect taxes, about half of which came from customs duties, Ahmed and Stern (1991, Table 2.5.) Many of these taxes and subsidies have grown up haphazardly over the years, and it is implausible that they are optimal by any coherent criterion. The issue with which we are concerned is the standard price and tax reform problem, in which we examine the equity and efficiency effects of "small" changes in the prices of goods and services. In this paper, we are concerned not with the development of the standard rules, but with their implementation in practice, and with the associated data and econometric requirements. We assume that there exist household survey data detailing the consumption of the relevant goods and services by households at different levels of living, and we shall be concerned with how to use those data to implement the theory. For the case of Pakistan, we use the rural data from the 1984–85 Household Income and Expenditure Survey, which covered over 9,000 rural households in the four provinces of the country during a twelve month period.

We review a number of the techniques that have been used to implement price reform analysis in previous studies, including the earlier work for Pakistan in Deaton and Grimard (1992). The contribution of the current paper is to consider the use of semi-parametric econometric methods in the shape of average derivative estimators, Härdle and Stoker (1989), Stoker (1991), to estimate the price responses that are required in assessing the effects of price changes on government revenues. These techniques have a number of very considerable advantages. They provide direct estimates of the quantities of interest; the theory asks for average derivatives, and the methods deliver estimates of average derivatives, thus bypassing the usual detour through utility theory, specification of functional form, and various econometric problems, perhaps the most serious of which is censoring at zero. However, nothing comes for free, and there are certain issues that can be dealt with by standard parametric methods and that do not seem to be readily incorporated into the non-parametric approaches. We hope that our calculations will also be of some methodological interest; our use of average derivative estimators provides a non-artificial example of the use of the technique, and gets us into some of the problems that are encountered in real applications using many thousands of observations, rather than in the small sample examples that are typically used to illustrate the econometric theory.

Price and tax reform formulas

The formulas given here are the standard ones as developed, for example, in Newbery and Stern (1987) or in the survey paper by Drèze and Stern (1987). We suppose that there are n goods, labeled by i. The price of good i is p_i, for the moment assumed to be the same for everybody, and the quantity consumed by household h is q_i^h. The price contains a tax or subsidy element, so that we can write

$$p_i = v_i + t_i \tag{1}$$

where v_i is the price before tax, and t_i is the amount of the tax or subsidy. In the standard analysis, it is supposed that t_i can be varied holding v_i constant, which although typically nonsensical for actual taxes in LDCs, is nevertheless a good place from which to start. We will return to the issue below.

There is supposed to be a social welfare function that encapsulates the equity trade-offs by the policy maker. (This is much disliked by political economists, but is no more than a convenient tool for summarizing the equity and efficiency effects of any policy change, and should not be treated as a descriptive device of how policy changes actually come about.) In any case, we can increase the tax (decrease the subsidy) and ignoring all but the direct welfare effects on prices, we can write

$$\frac{\partial W}{\partial t_i} = \sum_h \xi^h q_i^h \tag{2}$$

where W is social welfare, and ξ^h is the marginal social utility (welfare weight) of household h. The change in tax also affects government revenue R, and we can write

$$\frac{\partial R}{\partial t_i} = \sum_h q_i^h + \sum_h \sum_k t_k \frac{\partial q_k^h}{\partial p_i} \tag{3}$$

The ratio of (2) to (3) is the welfare cost of raising one unit of revenue by raising the price of good i, and is typically denoted λ_i, *viz.*

$$\lambda_i = \frac{\sum_h \xi^h q_i^h}{\sum_h q_i^h + \sum_h \sum_k t_k \frac{\partial q_k^h}{\partial p_i}} \tag{4}$$

Goods with large λ-ratios should have their prices raised, and those with small ratios should have their prices lowered; when all λ-ratios are the same, taxes and subsidies are optimal. The empirical challenge is to calculate these ratios in a convincing way.

Extensions and problems

Drèze and Stern (1987) show that equation (4) can be rewritten so as to yield a substantial generalization over the derivation given here. In particular, if we write s_i for the *shadow price* of good i, with shadow prices defined as social opportunity costs, then we can write

$$\lambda_i = \frac{\sum_h \xi^h q_i^h}{-\sum_h \sum_k s_k \frac{\partial q_k^h}{\partial p_i}} \tag{5}$$

It is straightforward to show that (4) and (5) are equivalent if the v's in (1) are proportional to the shadow prices, as would be the case, for example, when domestic prices are world prices plus tariffs. More generally however, if shadow prices can be properly calculated, (5) will capture the general equilibrium effects that will typically accompany a tax change.

This more general formulation is useful for thinking about actual taxes in the context of Pakistan. Although a good deal of revenue is raised through tariffs, a situation that fits well with (4), there are also important export taxes, to which (4) does not readily apply. Pakistan exports rice and maintains a domestic price that is lower than the world price by taxing exports; similar schemes can be found in many LDCs. Such taxes have the attraction for many governments of raising revenue while simultaneously lowering the price of rice to domestic (urban) constituents. Of course, the direct costs are born by the producers, whose welfare and behavior is not explicitly captured by the consumer oriented version of (4). While it is possible to generalize the model according to (5), and to build in the effects on producers, this would not be useful for the current empirical exercise, which uses household survey data which has no information on production. We are therefore constrained to look at only the consumption side of the reform process. This involves the conceptual experiment of supposing (counterfactually) that the government of Pakistan could separate the price paid for rice (for example) for those who eat it from the price paid to those who grow it. (Many are the self-same people, which is the heart of the problem.) An increase in the price of rice as considered here should not therefore be thought of as being brought about by a decrease in the export tax, but through a revenue raising tax on domestic consumption, with the export tax left in place. While this is an artificial construct, it captures part of what we have to know, even if a final decision on the desirability of reform also has to take into account producer responses and producer welfare.

Data sources and measurement

Given a national household survey, two of the three terms in (4) can readily be evaluated. The social weights in the numerator are typically replaced by some function of household income (or total expenditure), most commonly by writing

$$\xi^h = (x^h/n^h)^{-\epsilon} \tag{6}$$

where n^h is the number of people in household h, and ϵ is an index of inequality aversion, with larger values of ϵ leading to greater relative weight being placed on the poor. If top and bottom of (4) are divided by H, the total number of households, then the numerator is an average of quantities that can be observed in the survey, and so can be estimated directly by the corresponding sample average, using sampling weights as necessary. After division by H, the first term in the denominator is the average quantity consumed, and again can be estimated from the survey, or from administrative records—customs records and the like. The difficulties lie entirely with the last term, which captures the effects on government revenue of the behavioral consequences of the policy change. To estimate this, we need to know something about behavior, for which we need some sort of model of demand.

2. *Parametric approaches*

Ignoring cross-price effects

The most primitive approach to calculating the last term in (4) is to guess, using likely numbers for price elasticities, converted to the appropriate form. This might be done by assuming that cross-price elasticities are small enough to ignore, and then "calibrating" the own-price effects using estimates from other countries or from *ad hoc* time series models in which the aggregate demand for the good is regressed on aggregate real income and its own price relative to some price index. Since time series data are scarce in many LDCs, such methods make a sensible choice for estimating what is probably the most important and readily measured of the behavioral effects. However, there are obvious dangers. In particular, it is one thing to assume that an individual cross-price response is small, but quite another to assume that the sum of the cross-price effects on revenue is small, especially given that substitutes are more common than complements. We also know from optimal tax theory that if cross-price elasticities are zero, we should have inverse elasticity pricing, but we also know that the result is not robust, and that uniform tax rates are potential solutions to even the simplest Ramsey problems.

Additive preferences and the linear expenditure system

Frisch (1959) showed that if utility functions are additive across goods, then own and cross-price elasticities are linked by the formula

$$e_{ij} = \delta_{ij}\phi e_i - e_i w_j(1 + \phi e_j) \tag{7}$$

where e_{ij} and e_i are cross-price and total expenditure (income) elasticities, w_i is the budget share, and ϕ is a scalar. The budget shares are known, so that if the total expenditure elasticities can be estimated from an Engel curve analysis of the budget survey, we only need ϕ to calculate all the price elasticities, and thus to fill in our formula. Frisch thought that ϕ should be around –0.5, although he expected it to vary with the level of income, and he proposed constructing a world atlas of values to be used in studies such as the present one.

The Frisch scheme is still used from time to time, especially in the food policy literature, but it is more frequently implemented via the linear expenditure system, Stone (1954), which has the form

$$p_i q_i^h = p_i \gamma_i + \beta_i (x^h - \sum_{k=1}^{n} \gamma_k p_k) \tag{8}$$

for parameters β and γ. Although this model cannot (quite) be estimated on a single cross-section, it is economical enough in parameters to be estimated on even the most meager time series.

The linear expenditure system—as above or with more complicated Engel curves—has been extensively used in empirical tax reform exercises in LDCs, see for example Ahmad and Stern, and Ahmad, Ludlow, and Stern (1988) for Pakistan. Once again, use of additive demand systems is a pragmatic response to a difficult situation. There are not enough data to estimate a general model, so it is necessary to pick something that is tightly parametrized, like the linear expenditure system or some other additive specification. Unfortunately, and as with neglecting cross-price elasticities, the approach can have severe consequences. When preferences are additive and such as to generate linear Engel curves, a class characterized by

Pollak (1971), uniform taxes are optimal, and the tax reform prescriptions are extremely simple: increase the prices of those goods whose tax rates are below average, and decrease the prices of those goods whose tax rates are above average, see Deaton (1987). Under these assumptions, empirical analysis is unnecessary, since the answer is predetermined. Indeed, if we believe the assumptions, the problem is done, and there is no point in empirical analysis. But the assumptions are very implausible. Since different foods are subsidized and taxed at different rates, the nature of substitution or complementarity between them matters for tax policy, and there is no reason to suppose that it will always conform to the simple pattern of universal and uniform substitutability that is required by additive preferences.

Using spatial price variation

Although LDCs are relatively poorly endowed with long time-series data on consumption, they are often well provided with household survey data. Usually—although not universally—consumption data are collected on both value and volume terms, at least for food. For each household that purchases rice, say, we observe both the number of kilos bought in the last month and the number of rupees paid for those kilos. If the latter is divided by the former, we get a unit value of rice for that household. If we were to make the further leap and suppose that the unit value is a price—or at least a good indicator of price—we would have a plentiful source of quantity and price information, enough to estimate a very general demand system, allowing for quite unrestricted patterns of substitution and complementarity between the various goods. Since it has often been observed that markets are not necessarily well integrated across space in LDCs, it is entirely plausible that there is variation in price from one household to another, and there seems no reason why the use of spatial price variation should in any way be inferior to the use of temporal price variation. Early versions of this technique, for example Timmer and Alderman (1979) and Timmer (1981) run double logarithmic regressions with log quantities on the left hand side, and log unit values, together with log total expenditure and household socioeconomic and demographic characteristics on the right hand side. These regressions yielded sensible and well-determined price elasticities.

There are a number of problems with such specifications. First, the logarithmic specification makes it difficult to deal with zero purchases, which are common for some goods. Indeed, in Pakistan only slightly more than a half of rural households buy rice, which is one of the goods with which we are most concerned. As a result, the logarithmic regressions have to discard a large fraction of the data, with the usual selectivity and efficiency consequences. More generically, we are frequently interested for fiscal purposes in a good that is only consumed by a small fraction of the population, so that techniques that cannot deal with zero purchases appropriately are not well suited to the problem at hand. Second, unit values are not prices, although they are likely to be spatially correlated with prices. Even within categories like rice or wheat, different households can choose different varieties, and rich households will typically choose more expensive varieties than poor households. As always, there is some discomfort in using one endogenous variable to explain another, but there is a more specific problem here. If price increases are offset by quality shading and as well as quantity reductions, unit values will rise less than the true prices, so that quantity responses to prices will be overestimated by regressing quantities on unit values. Third, any measurement error in either reported expenditures or reported quantities will be transmitted into calculated unit values, and it is entirely plausible that there will be a spurious negative correlation between quantity and unit value. Someone who bought 1 kilo for 2 rupees, but is recorded as purchasing 10 kilos for 2 rupees, will also be recorded as having found a great bargain, and there is an obvious danger of "explaining" the former by the latter.

In a series of papers, Deaton (1988, 1990), Deaton and Grimard (1992), one of us has been grappling with these various difficulties, and has evolved a set of parametric procedures that circumvents at least some of them. In this paper, our concern is to give the briefest outline that is required to compare with the non-parametric analysis that follows, and to give some baseline results for Pakistan.

Allowing for quality effects, village effects, and measurement error

The model used in Deaton and Grimard (1992) and Deaton (1996, Ch. 5) can be written as a system of paired equations, one for each of the budget shares, and one for each of the unit values. The budget share equations are written

$$w_i^{ch} = \alpha_i^0 + \beta_i^0 \ln x^{ch} + \gamma_i^0 . z^{ch} + \sum_{j=1}^{n} \theta_{ij} \ln p_j^c + (f_i^c + u_{0i}^{ch}) \tag{9}$$

and the unit value equations as

$$\ln v_i^{ch} = \alpha_i^1 + \beta_i^1 \ln x^{ch} + \gamma_i^1 . z^{ch} + \sum_{j=1}^{n} \psi_{ij} \ln p_j^c + u_i^{ch}. \tag{10}$$

As before, w is a budget share, and v denotes a unit value. The three levels of subscripts and superscripts refer to the good, i, the household, h, and the village (or sample cluster) in which it lives, c. The quantity x is total household expenditure, z is socio-demographics, p is (unobservable) price, assumed to be constant within each cluster, f is a cluster fixed effect, and the u's are error terms that are allowed to be correlated across goods and between the share and unit value terms. The estimation of (9) and (10) is done in two stages. At the first, village effects are swept out by working with deviations from village means. This removes not only the fixed effects, but also the unobservable prices, so that the β and γ parameters can be consistently estimated. The correlations between the paired residuals in share and unit value equations are calculated and interpreted as indicating the variance and covariance of measurement error in the two equations. At the second stage, the estimated β's and γ's are used to subtract out the demographic effects from (9) and (10), and the "purged" budget shares and log unit values are averaged, village by village. The between village regression of "purged" budget shares on "purged" log unit values gives an estimate of price effects, which is corrected for the measurement error calculated at the first step. As is apparent from (9), the matrices Θ and Ψ are not separately identified, and in fact, the above procedure yields an estimate only of $\Theta \Psi^{-1}$, so that the estimation of Θ requires more information. This is provided by noting that it makes sense to link quality shading in response to price changes to quality changes in response to income, and is formally done using appropriate separability structures that permit the two matrices to be separately identified.

These are complicated operations, and it is worth trying to assess which parts of it really seem to matter in practice, and which might usefully be short-circuited. Start with the quality issue. As first noticed by Prais and Houthakker (1955), these quality effects really do exist, and there is nearly always a significant positive relationship between unit values and total expenditure. However, the effects are not very large. The β-coefficients in (10) are estimated to be 0.10 for both rice and wheat in rural Pakistan, they are a little higher for dairy produce (0.14) and meat (0.15), but essentially zero for edible oils and fats, and for sugar. Similar results were obtained for Indonesia, Deaton (1990), Côte d'Ivoire, Deaton (1988), and the state of Maharashtra in India, Deaton, Parikh, and Subramanian (1995). Numbers of this size make sense. It is hard to imagine a rich person paying much more than twice as much per kilo than a poor person for any broad aggregate of goods, and if rich people spend about six times as much as poor people—a useful rule of thumb—the elasticities will be of the size we have estimated. Given that the income elasticities of quality are so low, it is implausible that price elasticities of quality are high, which implies that the matrix Ψ is close to the identity matrix, and the final correction of the previous paragraph is not very important.

Even if quality effects can be ignored, measurement error remains a real hazard. In particular, evidence in Deaton (1988) suggested that regressing logarithms of quantities on logarithms of unit values is not a good idea. However, in (9) and (10), the transformation to budget shares and log unit values seems to remove the worst of the measurement error, at least in the sense that the correlations between the residuals of the first-stage within-village regressions are typically not very large, so that the correction for measurement error at the second stage also has little impact. There is no reason that this has to happen, although it is not implausible that while there are correlated measurement errors between reported quantities and reported unit values, there should be relatively little correlation in the reporting errors of expenditures and unit values. Note also that the second stage regressions are run using village averages, so that the effects of measurement error, if not eliminated, are likely to be reduced.

The importance of the village fixed effects is harder to judge. While it is often the case that sweeping out the village means does not affect the estimates of β's and γ's—something that is true in Pakistan—the inclusion of village fixed effects seems like good practice. Villages often differ a great deal from one to another, and are frequently internally homogeneous, so that intravillage correlations are to be expected. Furthermore, the village specific factors, such as prices, are quite likely to be correlated with included characteristics such as income. Indeed, in earlier work with Côte d'Ivoire, there were several cases where inclusion or exclusion of village fixed effects had a marked effect on the first stage estimates.

A much more difficult issue is the treatment of censoring, the fact that not all households are recorded as purchasing each commodity. Home-production complicates the picture further; food grown for domestic consumption, although added into total consumption (and income) has a price imputed to it by the enumerators, but does not produce a genuine "market price" observation. As in the original work on selectivity in wage equations, there are two separate problems. First, we have no observations on unit values for households that do not make purchases in the market, so that such households cannot be included in (9) and (10), at least not without some imputation procedure. Second, we have the "Tobit problem," that the presence of zero purchases will typically make it inappropriate to estimate linear regressions by OLS. Note that zero purchases that arise because of "frequency of purchase" considerations are no problem. While some households happen to buy nothing in the observation period, others are stocking up, so that expectations—including conditional expectations—are not affected. But there are typically many "genuine" zeros, where households are not purchasing because they do not consume the good. Note also that, whatever econometric procedure is adopted to handle the censoring, what we want for tax purposes is *all* households, whether they purchase or not. The welfare and revenue derivatives in (2) and (3) hold for households who do not purchase the good, just as for households who do, and the λ-ratios must be computed including all households. If zero-purchase households are excluded, the average welfare costs of a price rise will be understated, and the average revenue consequences will be incorrect.

Dealing with the missing prices is somewhat easier than dealing with the censoring of purchases. The model rests on the supposition that actual prices—although not unit values—are constant within villages, so that prices should be well predicted by a regression of unit values on a system of village dummies. Such regressions have high predictive power, and in those cases where it has been possible to check—for example in the Maharashtran case in Deaton, Parikh and Subramanian—the estimates prices match well with independently collected price data. Recall once again that the price elasticities are estimated from (9) using village averages, so that prices are only truly missing when no one in the village buys the commodity, a much less serious censoring problem than when we have to exclude all households that do not purchase.

The censoring has essentially been ignored in the work so far, and regressions run on all households, with zeros treated the same as positive purchases. If OLS were to estimate the true regression function, this would give the right answer, but this is a poor defense because, in the presence of extensive censoring, linear functional forms are likely to fit poorly. Note that the use of Tobit procedures is not recommended. Tobits are inconsistent in the presence of heteroskedasticity and/or non-normality, and there is no reason to suppose that Tobit estimates will be in any way superior to OLS. Non-parametric Tobit estimators, such as Powell's (1984) censored LAD estimator are robust against alternative distributional assumptions about the errors (including heteroskedasticity), but maintain the linearity of the "structural" part of the regression, something that seems just as arbitrary as the assumption that the residuals are normal and homoskedastic. Even then, if we want to make the demand functions coherent with some underlying utility function—which might appear to be desirable for a welfare calculation such as the current one—equation by equation Tobit cannot be the correct specification. For example, if some good is not purchased when total expenditure is below some critical threshold, then the marginal propensities to consume other goods have to adjust so as to satisfy the budget constraint, so that in general, the functional forms of all the equations have to change at censoring points for any of them. Handling the resulting multiple regime model is *extremely* complex even in the simplest possible cases, see Lee and Pitt (1986). As a result, while it is impossible to defend the practice of ignoring the censoring, there seems to be no better methodology, at least within the range of structural, parametric models.

Rewriting the formulas

The methodology described in the previous subsection has been applied to Pakistan in Deaton and Grimard (1992). Before summarizing some results, it is convenient to rewrite the formulas for the λ-ratios in a way that fits better with (9) and (10) and also with the new results in the next section. Rearrangement of (4) yields the following expression:

$$\lambda_i = \frac{w_i^e/\bar{w}_i}{1 + \frac{\tau_i}{1+\tau_i}\left((p_i\bar{q}_i)^{-1}\frac{\partial(p_i\bar{q}_i)}{\partial \ln p_i} - 1\right) + \sum_{k \neq i} \frac{\tau_k}{1+\tau_k}(p_i\bar{q}_i)^{-1}\frac{\partial(p_k\bar{q}_k)}{\partial \ln p_i}} \tag{11}$$

where τ is the ratio of tax to the pre-tax price, and $\bar{w}_i$ and w_i^e are the average and socially weighted budget shares of good i, the latter defined by

$$w_i^e = \frac{\sum_h \xi^h x^h w_i^h}{\sum_h x^h} \tag{12}$$

and where superimposed bars denote means. (For clarity, we have here retained the fiction that all prices are the same, but in practice they are not, and (11) remains valid if price multiplied by mean quantity is replaced by mean expenditure, and taxes are *ad valorem*, so that a change in tax rates changes all prices proportionately.) Since the ξ-weights can be arbitrarily scaled—see (2) above—we can think of (12) as a weighted average, and the ratio in the numerator of (11) as a commodity specific indicator of inequality. For goods that are relatively heavily consumed by the rich, the ratio will be low, and *vice versa* for goods more heavily consumed by the poor. Note also that the second term in the denominator of (11) captures the behavioral effects on revenue through changes in the demand for the good whose price is being changed—the own-price effects—while the last term captures the cross-price effects through other goods and the taxes and subsidies they bear. The term in large brackets in the denominator is the aggregate own-price elasticity for good i.

Equation (11) is a rewritten version of the original (4), and so will be valid independently of how we specify demand. In the special case of (9), (11) becomes

$$\lambda_i = \frac{w_i^e/\bar{w}_i}{1 + \frac{\tau_i}{1+\tau_i}\left(\frac{\theta_{ii}}{\bar{w}_i} - 1\right) + \bar{w}_i^{-1}\sum_{k \neq i}\frac{\tau_k}{1+\tau_k}\theta_{ki}} \tag{13}$$

which can be evaluated once the parameters have been estimated. Since there are a large number of parameters, we focus in this paper on presentating the estimated own-price effects, not because cross-price effects are unimportant, but in order to keep the number of comparisons within manageable bounds.

Results from the parametric approach

In Grimard and Deaton (1992), we give a set of (shadow) tax rates that has been roughly characteristic of tax policy in Pakistan over the 1970s and 1980s. There are substantial shadow subsidies on wheat and rice, with domestic prices about 40 percent below the world prices. Sugar bears a shadow tax of about 50 percent, with tariffs set to protect an inefficient domestic refining industry. Wheat is the basic staple for most households in Pakistan, and the average consumption of rice takes place at much higher levels of living than the average consumption of wheat. Equity factors therefore tend to favor a low price of wheat, but cannot readily be used to justify low prices of other goods.

Rather than present the λ-ratios, which depend on the particular configuration of tax rates, we look at the own-price elasticities, defined as the term in large brackets in the denominator of (13), and as determined by the average budget shares and by θ_{ii} in (13). Table 1 presents two sets of own-price elasticities for seven broad food groups, taken from Deaton and Grimard (1992). The wheat, rice, sugar, and edible oils groupings choose themselves because they are the groups whose prices most obviously deviate from shadow prices; the other groups are chosen arbitrarily but conform to the usual decomposition of foods in Pakistani surveys. The first column shows the budget shares for the seven groups; these are the averages over households of the individual budget shares. The second column shows unrestricted own price elasticities, which are calculated according to the methods outlined above. The order of consumption of cereals in Pakistan is wheat and then rice, with the latter consumed at significantly higher incomes, and this pattern is reflected in the price elasticities, with wheat inelastic, and rice—the relative luxury good—much more price elastic. (It is worth noting that similar estimates for the state of Maharashtra in India, where wheat is the relative luxury and rice the more basic good, show a much higher price elasticity for wheat than for rice.) Conditional on the parametric specification, the standard errors are small, except for the two groups, oils & fats and sugar, whose prices are to some extent controlled, and which therefore display much less spatial price variation. This relative lack of variation also results in quite imprecise estimates of cross-price responses with respect to the prices of sugar and edible oils, and in an attempt to overcome this, a version of the model was estimated that imposed the Slutsky symmetry restrictions. The own-price elasticities from the symmetry constrained model are shown in the third column, but do not differ by much from the unconstrained estimates.

The pattern of price elasticities in the Table tends to reinforce the reform recommendations based on equity issues alone. In particular, the low price of rice, which is undesirable on equity grounds, is also highly distortionary, because of the high price elasticity. Because wheat is estimated to be a substitute for rice—not shown here—the beneficial effects of raising the price of rice would be moderated somewhat by the switch to wheat, which also carries a shadow subsidy, but the effect is small relative to the equity and own-price effects.

3. Non-parametric approaches

What is to be estimated?

If we refer back to (11), and we wish to avoid specifying a parametric set of demand functions, then we need to find a way to estimate the quantities in the denominator, which requires us to evaluate, for all k and i,

$$\frac{\partial(p_k\bar{q}_k)}{\partial \ln p_i} = \frac{1}{H}\sum_h \frac{\partial(p_k q_k^h)}{\partial \ln p_i}. \tag{14}$$

By itself, (14) is not well-defined; we must specify the variables on which we are conditioning. In principle, we are concerned with the partial derivative with respect to each price, so that we should condition on all other variables that effect expenditures on each good. In this sense, there is no non-parametric escape route; just as in the parametric case, we must specify the right hand side variables. We denote the vector of conditioning variables by z; we use a subset of the variables in the parametric analysis, namely the logarithms of total household expenditure, total household size, and the seven unit values or prices. In the parametric analysis, these were supplemented by a set of family composition variables; these had some explanatory powe, but were less important than the other variables. As we shall see, there are strong incentives to limit the number of conditioning variables in the analysis, and our choice of seven variables in a data set with over 9,000 observations is a compromise that is designed to give a realistic example without stretching our computational resources beyond reasonable bounds.

For this subsection only, we denote the dependent variable by y, which will be the expenditure on each of the seven goods in turn. We can then write, for each i in turn

$$E(p_i q_i | z) = E(y|z) = m(z) \tag{15}$$

and the quantity that we wish to measure is

$$b_j = E_z\left(\frac{\partial E(y|z)}{\partial z_j}\right) = E_z\left(\frac{\partial m(z)}{\partial z_j}\right). \tag{16}$$

The expression (16) is an *average derivative*, and to estimate it, we follow Härdle and Stoker (1989) and Stoker (1991) and use *average derivative estimators*.

Estimation: the theory and implementation

Average derivative estimation is sometimes thought of as a semi-parametric technique, perhaps because it can be used to estimate index regression models where the conditional expectation of y is given by $\phi(x'\beta)$ for unknown function ϕ. In the current context, however, we are not restricted to such models, and the method produces estimates of average derivatives without having to restrict the functional form. Even so, the rate of convergence of the estimators is much faster than is typically the case for fully non-parametric treatments; like OLS, the estimates converge at rate $n^{0.5}$, and not at $n^{0.2}$ as would be the case if we were estimating a regression function or its derivatives. The more rapid convergence here is because we are estimating only the *average* of the derivatives, so that although derivatives at any given point will typically be estimated much more imprecisely, their average can be relatively accurate.

The theory of average derivative estimation is straightforwardly developed as follows. Suppose that the joint density function of the conditioning variables z is $f(z)$. Denote the "scores" by

$$w_j = -\partial \ln f(z)/\partial z_j \tag{17}$$

a quantity that, in principle, could be evaluated at each z^h in the sample, generating an $H \times K$ matrix, where K is the number of conditioning variables, here nine. Consider now the unconditional expectation

$$E(w_j y) = \int\int w_j(z) y f(y,z)\, dy\, dz = -\int\int f_j(z) y f(y|z)\, dy\, dz \tag{18}$$

where the last expression comes from substituting (17) into (18), and then splitting the joint density function into the product of a conditional and a marginal. If this last term is integrated by parts, and if we assume that $f(z)$ is zero on the boundary, then (18) becomes

$$E(w_j y) = \int\int f(z) y f_j(y|z) dy dz = E_z(\partial m(z)/\partial z_j). \tag{19}$$

Hence, if we knew the w's in (17), and formed the $H \times K$ matrix W, we could calculate

$$\hat{b}_1 = H^{-1} W' y \tag{20}$$

which, by (19) would converge to the vector of average derivatives (16). Note from (19) that $E(w_j z_k)$ is the derivative with respect to z_j of the expectation of z_k conditional on z, which is δ_{jk}, the Kronecker delta, so that the probability limit of $H^{-1}W'Z$ is the identity matrix. In consequence, another consistent estimator of the average derivatives is provided by the "instrumental variable" estimator

$$\hat{b}_2 = (W'Z)^{-1} W' y \tag{21}$$

According to Stoker (1991), (21) is to be preferred to (20) because common biases in the denominator and numerator offset one another in a way that does not occur with (20).

To make either of these estimators feasible, we require a method of estimating the scores (17). This first-stage estimation is based on a kernel estimate of the joint density $f(z)$ given by

$$\hat{f}(z) = \frac{1}{\tau^k H} \sum_{h=1}^{H} K\left(\frac{z - z^h}{\tau}\right), \tag{22}$$

where τ is a bandwidth and $K(.)$ is some suitable kernel function. In this paper, we use the quartic product kernel

$$K(z) = \prod_{i=1}^{K} \kappa(z_i), \quad \kappa(z) = \frac{15}{16}(1 - |z|^2)^2 \, I(|z| \le 1), \tag{23}$$

where $I(.)$ is the indicator function that is 1 if the statement in brackets is true, and otherwise is zero. The bandwidth τ controls the degree of smoothing. Once a bandwidth is chosen—on which more below—the logarithmic derivatives of (22), which are the estimates of the scores (16), are calculated by differentiating (22) and using the data to calculate the resulting formula. This is straightforward but time consuming; for each of the $H \times K$, here 9119 by 9 elements of the matrix W, the evaluation requires a complete pass through the sample of 9119 points. In practice, we follow the standard recommendations in Silverman (1986, p. 77–8), and first transform and scale the Z matrix so that it has a unit variance covariance matrix, after which the calculated scores are transformed back to restore the original dimensions. The bandwidth is chosen according to the recommendations in Powell and Stoker (1993); in this case, we set τ to be unity for the transformed data.

Although it would be possible to estimate (21) using the estimate of the score matrix as described, there will be problems where the estimated density is small, since the evaluation of the logarithmic derivative requires division by the density. For the same reason, Powell, Stock and Stoker (1989) evaluate a density weighted average derivative estimator, which corresponds to (16) weighted by $f(z)$. But Powell, Stock and Stoker are only interested in estimating the average derivatives up to scale, so that they can weight with impunity, whereas in the current case, we need the average derivatives themselves. We therefore adopt an alternative approach, "trimming" the data by deleting the 5 percent of the observations for which the estimated density is smallest. If the cutoff on the density to achieve this 5 percent is α, we can define the $H \times H$ diagonal matrix Ω by its typical element

$$\omega^{hh'} = \delta_{hh'} \mathrm{I}(\hat{f}(z^h) \geq \alpha). \tag{24}$$

Then the estimator that we actually use can be thought of as the weighted IV estimator

$$\hat{b}_3 = (\hat{W}'\Omega Z)^{-1}(\hat{W}'\Omega y). \tag{25}$$

The estimates of the average derivatives are $\sqrt{H}$ consistent and asymptotically normal, with asymptotic standard errors given in (3.5) and (3.6) of Härdle and Stoker (1989) under the assumption that the observations are independent and identically distributed. Let the estimated "residuals" corresponding to household h for good i be given by:

$$\hat{r}_{hi} = \hat{w}_{hi} y_h \zeta_h + H^{-1}\tau^{-k} \sum_{h'=1}^{H} [\tau^{-1} K'(\frac{z_h - z_{h'}}{\tau}) - K(\frac{z_h - z_h'}{\tau})\hat{w}_{h'i}] \frac{y_{h'}\zeta_{h'}}{\hat{f}_{h'}} \tag{26}$$

Then the sample covariance matrix for the average derivative estimates is given by:

$$\hat{V}_0 = H^{-1} \sum_{h=1}^{H} \hat{r}_h \hat{r}_h \zeta_h - \bar{r}\bar{r}' \tag{27}$$

Although this formula will be robust to heteroskedasticity in the residuals, it does not account for the likely more serious bias to standard errors that comes from ignoring intracluster correlations, a particularly inappropriate omission in a context where the cluster structure is such an important part of the analysis. It is also known that the bias to the standard errors tends to be largest when the regressors vary little within the clusters, as is the case for the prices here, see Kloek (1981) and Pfefferman and Smith (1985). The standard errors of the OLS estimates can readily be corrected using a generalization of the Huber-White procedures as implemented in STATA, and the results suggest that the problem is non-trivial, with the robust standard errors typically more than twice the size of those reported. To adapt the standard errors of the ADEs to allow for similar effects, we consider a modified sample variance-covariance matrix, defined as:

$$\hat{V}_1 = H^{-1}\sum_{h=1}^{H}\sum_{h'=1}^{H}(\hat{r}_h - \bar{r})(\hat{r}_{h'} - \bar{r})\zeta_h\zeta_{hh'} \tag{28}$$

where $\zeta_{hh'} = 1$ if household h and h' belong to the same cluster.

Some practical considerations

The non-parametric estimation procedure, like the parametric one, requires some way of dealing with those cases where there is no unit value recorded because the household in question did not buy the good. In the parametric estimation, price elasticities were estimated from the cross-cluster variation in unit values, effectively using average within-cluster unit values to represent each cluster. In the same spirit, we have filled in the missing values for each household with the average unit value for those households in the cluster who do make purchases. In the few cases where no one in the village records a purchase, we use the average for the province as a whole.

The calculations as described above are straightforward in principle; there are no implicit equations to be solved and no iterative techniques are required. The computation is burdensome only because there is a large number of observations H and a large number of conditioning variables, K, and because the number of evaluations is proportional to KH^2. The baseline calculations, using all observations, took over one day (over 200,000 seconds) to compute using a 8 processor (MIPS R4400) Unix workstation. This has prompted us to experiment with alternative methods to speed up the computations, all of which involve discarding some of the data. The first method is the simplest, and works by sampling a subset of the data randomly. The second and third methods make a more deliberate attempt to preserve the structure of the data. Both start by running a preliminary OLS regression of expenditure on the nine conditioning variables, and then ordering the observations by the size of the predicted values. This is done because we need a unidimensional quantity on which to sort the data, but want to avoid selectivity bias induced by selecting on the dependent variable. Method two divides the sorted sample into bins (quintiles, deciles, for example), and then randomly select from each. Method three selects the fifth, tenth, or twentieth observation from the sorted data, starting from some observation other than the first.

Our experience suggests that the computation time is essentially determined by the effective sample size, and not by the method of sample reduction. Estimates from a reduced sample of 2000 observations took one-third of the time required to obtain estimates from the full sample of 9119 observations. The point estimates are quite robust to sample reduction. As might be expected, reducing the sample size raises the standard errors of the estimates, so that borderline significant estimates in the full sample tend to become insignificant in the smaller sample. However, estimates that are statistically significant in the full sample remain so in the smaller sample. There was no clear winner as to which of the three methods dominate in the sense of giving smaller standard errors. As these sample reduction methods have weak theoretical foundations, we are not particularly confident with these results, which we omit from this paper. Suffice it to mention that more sophisticated time saving methods based upon the idea of "discretization" and "convolution" are now being considered in the literature, Härdle and Linton (1994). We are, however, somewhat skeptical if such methods are readily suited for high dimensional problems as in ours since the discretization procedure could itself be very time consuming .

We have nevertheless found a way of computing the average derivatives that results in a tenfold reduction in computing time. This comes from replacing the quartic kernel in (23) by the Gaussian kernel,

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{29}$$

Admittedly, we have not taken extra steps to optimize the execution speed of the programs, as this was not the intent of our paper. Indeed, we have been computing the average derivatives by "brute force." However, the average derivatives can be obtained in a mere 20,000 seconds (instead of 200,000) upon replacing the quartic kernel by the Gaussian kernel while leaving the rest of the program unchanged. Our cursory investigation into this difference suggests that the product Gaussian kernel involves one exponentiation and k addition operations. This kernel apparently entails drastically few computer instructions than the quartic kernel, which requires raising variables to higher power k times. Although not reported in published papers, this time-saving phenomenon was apparently noted by others (Professor T. Stoker in personal communication). In Table 2, we report the full sample estimates using the Gaussian kernel and the quartic kernel. When evaluated at the same bandwidth of 1.0, the Gaussian estimates are practically identical to the quartic estimates We report results for the Gaussian kernel for $\tau=.75$ and $\tau=1.25$, which the readers can verify are sufficiently close to the quartic kernel with $\tau=1.0$. Clearly, for the application to hand, the estimates are robust to the choice of the kernel and the bandwidth.

Results and evaluation

Table 2 lists the OLS regression coefficients for each good, together with the average derivative estimates for two kernels and different bandwidths. The most obvious feature of the last column in Table 2 is its close proximity to the first column in Table 2; although there are a few differences, the OLS and average derivative estimates are never very far apart, and when the estimates are converted into elasticities in Table 3, the only difference of any importance is in the estimate for oils and fats, where the OLS elasticity is shifted from −1.31 to −1.77. The much larger differences are between the first column, which replicates the first column of Table 1, and the second and third columns containing the OLS and ADE based estimates, not between the OLS and ADE elasticities themselves. Wheat, rice, and sugar are all more price elastic according to the new calculations than under the original method, so that if these new results are correct, the subsidies on wheat and rice and the tax on sugar are more distortionary than originally estimated. The closeness of the OLS and average derivative estimates is somewhat disappointing; the object of the paper was not to find the most expensive possible way of calculating least squares! It is also rather surprising. If the conditioning variables are jointly normally distributed, ADE and OLS will coincide, since under normality, the scores are linear functions of the variables themselves, see Stoker (1986). Perhaps this is approximately the case for the current example, although we have taken no steps to transform the variables to make it so, even approximately.

As far as choosing between the original method and the non-parametric procedure, we currently have no way of deciding which set of estimates should be preferred. The original method deals with issues that have to be ignored by the average derivative estimates, and *vice versa*, and it is perhaps a good idea to summarize these, if only because future work with average derivative estimation may allow more general assumptions.

Consider first the *advantages* of the average derivative method. First, it offers a direct measurement of the quantities that we need. The elasticities in the last column of Table 3 do not come from evaluating the derivatives of some arbitrarily specified functional form at some arbitrarily specified point. Rather they are estimates of the second term in the denominator of the tax reform formula (11), and while the estimates are interpretable as price elasticities, this is because it is convenient for the purposes of comparison to think of them in such a way, not because we have chosen to parametrize the elasticities. Second, the average derivative estimators do not require that a functional form be specified for the demand functions. We do not have to concern ourselves with utility theory, nor with the relationship between demand functions and utility functions. Third, we lose the multitude of problems associated with extensive and intensive margins of consumption. It simply does not matter whether some consumers buy some goods and not others and there is no need to treat zero demands any differently from positive demands. The welfare and tax reform formulae are correct no matter who consumes what, and the average derivative estimators estimate average derivatives, even if these derivatives are zero for some households because they do not buy the good. This is a major advantage, since there is no known feasible technique for correctly estimating large demand systems when different households consume different subsets of commodities.

There are also a number of important *disadvantages*. The parametric model allows for village fixed effects in demands, thus recognizing that there are likely to be common but unobservable features of behavior in each village, and that these effects may well be correlated with village observables. No such possibility is allowed by average derivative procedures. Indeed, the theory of average derivative estimation requires that the conditioning variables be continuous, thus prohibiting any dummy variables, and certainly the close to 1,000 dummy variables that are required for the village effects in the Pakistani data. Secondly, the parametric estimation method has a procedure for treating the measurement error. While the effectiveness of the treatment is unclear, and there are examples where the correction makes relatively little difference, the fact remains that within the parametric model there is a range of possibilities for dealing with measurement error, instrumental variables being the most obvious, and that none of these are available for average derivative estimators. Thirdly, it is unclear how the quality correction procedures in the parametric model can be adapted to the nonparametric case. However as we have seen, these effects are typically small, and are unlikely to be a major source of difference between the two methodologies. If this were thought not to be the case, average derivative estimators could certainly be applied as far as the estimation of the elasticities of quality with respect to total expenditure. It is the next step, where the price elasticities are corrected, that has no obvious counterpart in the non-parametric case.

There are also a number of difficulties that are common to both approaches, and that are not resolved by the nonparametric techniques. In both methods, it is necessary to specify a list of conditioning variables, and the results will typically depend on the choice. There is therefore no possibility of a fully non-parametric treatment, so that the estimation of the behavioral responses in the denominator of (11) is still on a very different footing from the estimation of the means in the numerator. Second, the problem of missing unit values is still largely unresolved. In the implementation of the ADEs we imputed prices to clusters on the basis of geographical information, a procedure that is obviously sensible and that is supported by the good fit that is obtained when observed unit values are regressed on cluster dummies. However, the parametric approach, by sweeping out cluster fixed effects, requires prices only at the cluster level, so that less imputation is required. Furthermore, it seems that different imputations schemes give different results, so that, for example, the OLS estimates in Table 2 are changed non-trivially if regressions are run on cluster means rather than on the individual data. It could also be argued that the need for a parametric imputation technique for missing values is a good deal less comfortable in a nonparametric setting than in a parametric model. It is also possible that "automatic" imputation techniques could be developed in the nonparametric context.

There is clearly a great deal more work to be done. Even so, and although there are still very real difficulties, we feel that the use of average derivative estimators is sufficiently promising to reward that work. ADEs solve some problems that appear to be completely intractable in the parametric approach, and their own problems seem to be addressable.

4. List of works cited:—

Ahmad, S. Ehtisham, David Coady, and Nicholas H. Stern, 1991, *The theory and practice of tax reform in developing countries*, Cambridge. Cambridge University Press.

—— Stephen Ludlow, and Nicholas H. Stern, 1988, "Demand responses in Pakistan: a modification of the linear expenditure system for 1976," *Pakistan Development Review*, **27**, 7–43.

Deaton, Angus S., 1987, "Econometric issues for tax design in developing countries," in Newbery and Stern, 92–113.

—— 1988, "Quality, quantity, and spatial variation of price," *American Economic Review*, **78**, 418–30.

—— 1989, "Household survey data and pricing policies in developing countries," *World Bank Economic Review*, **3**, 183–210.

—— 1990, "Price elasticities from survey data: extensions and Indonesian results," *Journal of Econometrics*, **44**, 281–309.

—— 1996, *The analysis of household surveys: a microeconometric approach to development policy*, Baltimore, Md. Johns Hopkins U. Press for the World Bank. (forthcoming.)

—— and Franque Grimard, 1992, "Demand analysis for tax reform in Pakistan," Princeton University Research Program in Development Studies, Princeton, NJ, processed. (Jan.)

—— Kirit Parikh and Shankar Subramanian, 1995, "Food demand patterns and pricing policy in Maharashtra: am analysis using household level survey data," Princeton University Research Program in Development Studies, Princeton, NJ, processed. (Jan.)

Drèze, Jean and Nicholas H. Stern, 1987, "The theory of cost-benefit analysis," in Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Amsterdam. North-Holland.

Frisch, Ragnar, 1959, "A complete scheme for computing all direct and cross demand elasticities in a model with many sectors," *Econometrica*, **27**, 177–96.

Härdle, Wolfgang, and Thomas M. Stoker, 1989, "Investigating smooth multiple regression by the method of average derivatives," *Journal of the American Statistical Association*, **84**, 986–95.

—— and Oliver Linton, 1994, "Applied Nonparametric Methods," in Engel, Robert F. and Daniel F. McFadden, eds., *Handbook of Econometrics*, Volume IV, Amsterdam. North Holland.

Kloek, Teun, 1981, "OLS estimation in a model where a microvariable is explained by aggregates and contemporaneous disturbances are equicorrelated," *Econometrica*, **49**, 205–7.

Lee, Lung-Fei and Mark M. Pitt, 1986, "Micoreconometric demand systems with binding non-negativity constraints: the dual approach," *Econometrica*, **54**, 1237–42.

Newbery, David M.G., and Nicholas H. Stern, 1987, *The theory of taxation for developing countries*, Oxford. Oxford University Press for the World Bank.

Pfefferman, D., and T. M. F. Smith, 1985, "Regression models for grouped populations in cross-section surveys," *International Statistical Review*, **53**, 37–59.

Pollak, Robert A., 1971, "Additive utility functions and linear Engel curves," *Review of Economic Studies*, **38**, 401–13.

Powell, James L., 1984, "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, **25**, 303–25.

—— James H. Stock, and Thomas M. Stoker, 1989, "Semiparametric estimation of index coefficients," *Econometrica*, **57**, 1403–30.

—— and Thomas M. Stoker, (1992) "Optimal bandwidth choice for density weight averages", mimeo, Princeton University.

Prais, Sigbert J., and Hendrik S. Houthakker, 1955, *The analysis of family budgets*, Cambridge. Cambridge University Press. (2nd ed. 1972)

Silverman, Bernard W., 1986, *Density estimation for statistics and data analysis*, London and New York. Chapman and Hall.

Stoker, Thomas M., 1986, "Consistent estimation of scaled coefficients," *Econometrica*, **54**, 1461–81.

Stoker, Thomas M., 1991, *Lectures on semi-parametric econometrics*, Louvain-la-Neuve, Belgium. CORE Foundation.

Stone, J. Richard N., 1954, "Linear expenditure systems and demand analysis: an application to the pattern of British demand," *Economic Journal*, **64**, 511–27.

Timmer, C. Peter, 1981, "Is there `curvature' in the Slutsky matrix?" *Review of Economics and Statistics*, **63**, 395–402.

Timmer, C. Peter and Harold Alderman, 1979, "Estimating consumption parameters for food-policy analysis," *American Journal of Agricultural Economics*, **61**, 982–4.

Table 1
Own-price elasticities for seven goods
Rural Pakistan 1984–85

	budget share	unrestricted		symmetric	
wheat	0.1276	−0.51	(0.09)	−0.51	(0.08)
rice	0.0267	−1.59	(0.13)	−1.53	(0.12)
dairy produce	0.1269	−0.87	(0.03)	−0.89	(0.03)
meat	0.0366	−0.57	(0.13)	−0.61	(0.13)
oils & fats	0.0414	−2.04	(0.44)	−1.59	(0.40)
sugar	0.0293	−0.07	(0.44)	0.03	(0.40)
other food	0.1219	−0.35	(0.08)	−0.40	(0.08)

Notes: From Tables 5 and 6 of Deaton and Grimard (1992). Unrestricted estimates of price elasticities are given in Column 2, and those with the Slusky symmetry restriction imposed are in Column 3.

Table 2
OLS and Average Derivative estimates of expenditures with respect to log prices
Rural Pakistan 1984–85

	OLS		ADE			ADE			ADE		
			Quartic kernel			Gaussian kernel			Gaussian kernel		
			τ=1.0			τ=1.25			τ=0.75		
wheat	11.71	(3.9)	14.33	(5.4)	[8.6]	14.39	(5.4)	[8.6]	14.16	(5.4)	[8.6]
rice	−46.12	(1.9)	−46.08	(2.6)	[5.9]	−46.07	(2.6)	[5.9]	−46.08	(2.6)	[5.9]
dairy produce	−8.30	(2.4)	−14.97	(3.8)	[7.5]	−14.26	(3.9)	[7.7]	−14.96	(3.8)	[7.5]
meat	38.30	(2.1)	38.72	(3.0)	[5.0]	38.79	(3.0)	[5.0]	38.72	(3.0)	[5.0]
oils & fats	−16.66	(7.4)	−42.06	(9.4)	[19.6]	−42.10	(9.4)	[19.6]	−42.06	(9.4)	[19.6]
sugar	26.76	(8.7)	28.44	(6.7)	[13.4]	28.41	(6.7)	[13.4]	28.44	(6.7)	[13.4]
other food	79.08	(4.4)	62.77	(6.5)	[11.5]	62.77	(6.5)	[11.5]	62.77	(6.5)	[11.5]

Notes: OLS is ordinary least squares using all 9,119 observations. ADE is average derivative estimation, τ is the bandwidth. Heteroskedastic consistent standard errors are in rounded brackets; standard errors corrected for cluster effects are in square brackets.

Table 3
Own-price elasticities from OLS and ADE estimates
Rural Pakistan 1984–85

	Parametric		OLS		ADE	
wheat	-0.51	(0.09)	–0.93	(0.02)	–0.91	(0.03)
rice	–1.59	(0.13)	–2.31	(0.06)	–2.31	(0.07)
dairy produce	–0.87	(0.03)	–1.05	(0.01)	–1.09	(0.02)
meat	–0.57	(0.13)	–0.20	(0.04)	–0.20	(0.06)
oils & fats	–2.04	(0.44)	–1.31	(0.14)	–1.77	(0.17)
sugar	–0.07	(0.44)	–0.30	(0.22)	–0.26	(0.17)
other food	–0.35	(0.08)	–0.51	(0.03)	–0.61	(0.04)

Notes: The first column is the second column of Table 1. The second and third columns are calculated from the first and secon columns of Table 2.