Prices and Unit Values in Poverty Measurement and Tax Reform Analysis

John Gibson*
University of Waikato

Scott Rozelle
University of California, Davis

Abstract

Researchers often use *unit values* (household expenditures on a commodity divided by the quantity purchased) as proxies for market prices when calculating poverty lines and estimating consumer demand equations. Such proxies are often needed because community price surveys in developing countries are either absent or suffer quality problems. However, biases may result from using unit values, due to measurement error and quality effects. In this paper, we report evidence on a household survey experiment where information on prices was obtained in three ways: from unit values, from a market price survey, and from the opinions of householders who were shown pictures of various items and asked to report the local price. The three sets of price data are used to calculate poverty lines and to estimate price elasticities and analyse marginal tax reforms. There are substantial biases when unit values are used as a proxy for market price, even when sophisticated correction methods are applied. The performance of the price opinions obtained from members of the survey households on the basis of the pictures was better. The results highlight the importance of price collection methods and the need to consider the wider costs of having potentially unreliable community-level price data.

JEL: C81, D12, H21, I32

Keywords: Demand, Measurement error, Poverty, Prices, Tax reform, Unit values

Acknowledgements
We are grateful for assistance and helpful comments from Chris Hector, Tim Maloney, Susan Olivia, Berk Özler, Steven Stillman, and three anonymous referees, and seminar audiences at Canterbury University and the Northeast Universities Development Consortium Conference. Data used in this paper were originally collected as part of a World Bank poverty assessment for Papua New Guinea, for which financial support from the governments of Australia (TF-032753), Japan (TF-029460), and New Zealand (TF-033936) is gratefully acknowledged.

*Address for correspondence: Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand. Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz

Prices and Unit Values in Poverty Measurement and Tax Reform Analysis

I. Introduction

Prices are important. Economists need good measures of prices to conduct studies for many applications in developing countries. For example, when constructing computable general equilibrium models for trade policy analysis, researchers need to have matrices of own- and cross price elasticities of demand (Minot and Goletti, 2000). Similarly, the effective reform of indirect taxation and subsidy regimes requires accurately estimated price elasticities in order to predict changes in either public expenditure or tax revenues as demand changes following subsidy or tax rates shifts (Lariki, 1989; Ahmad and Stern, 1991). The elasticities also are needed for a proper accounting of the welfare effects of economic crises because first-order approximations that ignore consumer substitution can greatly overstate welfare losses (Friedman and Levinsohn, 2002). Finally, poverty analysts need accurate and timely price data to ensure that poverty lines correspond to the actual change in the cost of living for poor people; this issue has affected recent debates about poverty reduction in India (Deaton, 2003).

Surprisingly, despite being important for so many analyses, few studies systematically collect price data. State statistical bureaus in countries such as China, Indonesia and Pakistan do not collect market price data that can be matched to their rural household income and expenditure surveys. Consequently, in many countries, Poverty Reduction Strategy Papers (PRSPs) use poverty estimates that are based on *assumed* levels of rural prices.[1] Even research-driven surveys suffer from a lack of price data. For example, the Indonesia Family Life Survey (IFLS2) collected a tremendous amount of data from households and communities, including expenditures on 37 food items, but market price surveys were carried out for only nine foods. This incomplete information on prices makes it difficult to reliably measure the inflation rate that

Indonesian households faced during the Asian economic crisis, and may contribute to the large discrepancy between the poverty increases implied by the IFLS price data and those implied by the official (urban) inflation rates (Beegle, *et. al.* 1999). Even in the well-funded and comprehensive Living Standards Measurement Study (LSMS) surveys, there have been problems in gathering prices:

> "In most previous LSMS surveys, interviewers have collected price data by visiting markets and vendors and asking the price of particular goods. … Another possible way to collect prices would be to ask community informants or a sub-sample of household informants about prices. *Given how little is known about how to collect data on community-level prices and how many problems there have been in past LSMS studies* [emphasis ours], it is recommended that both methods be used (Frankenberg, 2000, p. 329)."

Community-level prices, of the type collected in most LSMS surveys, may be unreliable either because they are gathered from the wrong market, or are for the wrong specification of goods, or the prices quoted are not the prices actually paid by local residents (Deaton and Grosh, 2000). Indeed, in some LSMS surveys, the market price data have either never been released because of quality problems (e.g. Tajikistan), or analysts have been forced to discard some of the prices.[2]

This poor track record for collecting price data may not be surprising. In the rural areas of many developing countries it is hard for outsiders to find, understand and study markets. Markets may meet intermittently, at different places on different days and often at very early hours. Perhaps because managing the traditional part of the data collection effort (household expenditures) is already logistically difficult, adding another part of the survey (for collecting prices) with its own set of complications may cause a decline in overall survey quality. The problems are likely to be most apparent in countries with poor infrastructure and low population densities, which are the very places where price policy can be an important tool for government because of the high per capita administrative cost of delivering income interventions.

Without good price data, economists have had to turn to imperfect proxy measures, such as *unit values* (the ratio of household expenditure on a particular good to the quantity consumed).[3] The range of applications where unit values have recently been used include the calculation of purchasing-power-parity exchange rates (Deaton *et al*., 2004), the calculation and updating of poverty lines (Deaton, 2003), the assessment of household welfare changes from a trade liberalisation (Nicita, 2004) and an economic crisis (Friedman and Levinsohn, 2002), the analysis of indirect tax and subsidy reforms (Deaton and Grimard, 1992; Nicita, 2004a), and assessments of the distributional and nutritional impacts of devaluation (Minot, 1998). However, in some applications, such as demand studies, the use of unit values is believed to give biased results (Deaton, 1997). The problem with unit values is that, in contrast to market prices, they reflect household-specific quality and reporting error effects, and are subject to sample selection effects because they are unavailable for non-purchasing households. Even procedures developed by Deaton (1990) to correct these biases have been shown to produce inaccurate and imprecise results (Gibson and Rozelle, 2002). Alternative strategies, such as using more readily available *urban* price series as proxies for the prices faced by rural households, also may cause bias (Alderman, 1988).

Recognizing that these types of problems with the gathering of price data in household surveys appear to be pervasive, we devised an experiment during a survey in Papua New Guinea (PNG) to test alternative ways of collecting price data. We use three ways to obtain information on prices: from the unit values implicit in household expenditure data; from a market price survey (which we conducted by making repeated trips to the market and surveying traders); and from the ‘opinions’ of household respondents that were shown pictures of various items during the survey and asked to report the local price for the product in the picture. The picture-based methodology has several potential advantages over traditional, unit value-based approaches. Since it is easy to show pictures to all households and ask for their estimates of the price, there are likely to be fewer missing observations. More importantly, any measurement error in these price opinions should not be correlated with actual demands. Finally, biases due to quality effects should be less, since everyone sees and is responding to the same picture.

We use the prices from the market price survey as the standard against which we judge the two alternative price proxies. Although somewhat innocuous, such a preference for relying on market price surveys is not always apparent in the literature (Deaton and Grosh, 2000). In this paper, we explicitly assume that prices for well-defined items collected from market surveys using certain sampling rules are the appropriate standard. In some cases, there may be reasons to worry about the quality of market prices themselves. In the case of our study, however, three features of the case study country increase the reliability of the market price surveys. First, villages are small and in almost every village that the survey team visited, the market that serves the village is well-defined. Second, for whatever reason, haggling is uncommon in markets in PNG. Moreover, several of our products are sold only in trade stores and supermarkets, where transactions always take place at the listed prices. Both of these features mean that the prices observed by enumerators are likely to be the prices actually faced by households in the survey. Third, there is very little quality variation in many of the foods consumed in PNG, which are often branded products with well-defined package sizes. Often a single brand supplies the whole market either because local monopolies are protected by quotas and high tariffs or because a dominant importer controls port and distribution facilities. Even for the foods that are produced and marketed by the informal sector, there is little intra-market quality variation (as we show below) so significant variation in quality between markets seems unlikely.

Although our experiment relates to just a single country, we believe that there may be wider interest in our findings. We appear to have made the only systematic attempt to test an idea that was proposed early in the development of the LSMS surveys, which was to obtain price data by interviewing groups of housewives (Saunders and Grootaert, 1980).[4] This strategy was never implemented in part because subsequent LSMS papers were critical of this ‘novel but risky’ idea (Wood and Knight, 1985). The main criticisms were that these price opinions could be biased by differences in bargaining skill, by uncertainty about the reference period (which matters in inflationary environments), and by the lack of a representative sample. We believe that our development of the idea, based on a representative sample of households each shown a defined specification (in the form of a photograph) and asked to report the current price, overcomes several of these biases. We also find that these price opinions do not vary with observable household characteristics, so the concern about bias due to differences in bargaining skill may be misplaced.

Also, ours is one of the only papers to empirically demonstrate the magnitude of the bias from using unit values as proxies for market prices. Surprisingly, despite the widespread reliance on unit values and despite the plea by Deaton (1990), there has never been a ‘crucial experiment’ in which results calculated from market price data are compared with the results from either naïve or corrected unit value procedures. We are aware of only one paper in the literature that compares poverty estimates when poverty lines are priced with either unit values or market prices (Capéau and Dercon, 1998).[5] In comparison to that paper, we go further by having three types of prices, and by also looking at the effect on estimated demand elasticities and marginal tax reform calculations.

The rest of the paper is organized as follows. The next section describes our three methods of collecting price data. Section III reports the basic descriptive statistics and comparisons of the various price measures. The results of using the three sets of price data to calculate and compare poverty lines, poverty indices and price elasticities are reported in Section IV. Sections V and VI discuss the practicalities of the various types of price surveys and offers some conclusions.

II. Data Collection

Data used in this paper come from the Papua New Guinea Household Survey (PNGHS), which was designed and supervised by the authors in 1995 and 1996. The survey covered a random sample of 1200 households, residing in 73 rural clusters (each providing 12 households to the sample), 40 clusters from the capital city (providing six households each) and seven clusters from smaller urban areas (12 households each). The survey fieldwork was spread over a 12-month period. The key feature of this survey is that it collected information on prices in three ways: market price surveys, unit values and price opinions of householders who were shown pictures of various items.

Market prices were collected in each cluster using two different surveys. The prices of 14 commercially produced food items (e.g., rice, sugar, beer) and nine non-food items (e.g., soap, kerosene) were collected from the two main trade stores or supermarkets used by the households in the cluster. These prices typically were for a finely defined specification (e.g., a 1kg bag of "Trukai" brand rice). For four of the foods and one of the non-foods, the prices covered two different specifications of the same commodity (e.g., a bottle of beer and a carton of beer). In these cases, the analyses use a simple average of the prices of the two specifications of the same commodity. The second market survey collected the prices of 11 locally produced foods from the nearest local market, with one food (banana) having prices collected for two different varieties. Enumerators recorded the price and weight of up to six different lots of each commodity (drawing the sample from different sellers). The market price survey was carried out on two different days in each cluster; potentially, up to 12 observations are available on the price of each of these foods for a given market.

The unit values were obtained from a closed interval consumption recall. After an initial interview to signal the start of the consumption recall period, enumerators revisited the households approximately two weeks later and asked respondents to recall the value and quantity of all purchases, gifts, and own-production made since the initial interview. This recall covered 36 categories of food and 20 categories of other frequent expenses.[6] The unit values are calculated as the ratio of purchase values to purchase quantities.

The data collection methods affect the unit values in the PNG survey in two important ways. First, the unit values are for the same period that the market price survey was carried out. In contrast, in some LSMS surveys (e.g., Vietnam) the unit values cover a 12 month period, which would weaken any comparison with current market prices. Second, the survey team used a method for reducing the problem of respondents not understanding metric quantities. At the time of the first interview, all households were given an empty 25 kg sack, with graduations of "¼", "½", and "¾" marked on the outside, to use for recording food volumes. This was the recommended unit for the bulky root crop staples and was used in over 90 percent of cases.[7] The other main unit used was a simple count of the number of items, which was recommended for items like coconut and betelnut, and for animals. Average volume-to-weight (and count-to-weight) conversion factors were established from weighing trials that were carried out in all regions of the country. While these procedures are undoubtedly crude, compared with the ideal of weighing all items consumed, they avoid the problem of enumerators and respondents using idiosyncratic conversion factors and so they reduce the relevance of the Capéau and Dercon (1998) procedure.

The “picture method” data come from price opinions that were gathered from each household for 15 food items (including beverages) and for three tobacco products. Since six of the food items were alternate specifications of a particular food (e.g., a bottle and a can of soft drink), the pictures refer to nine categories of food. On average, these nine foods comprise 30 percent of the household’s *total* consumption expenditure, with individual budget shares ranging from 11 percent (sweet potato) to one percent (flour, biscuits, and soft drinks). Central to the enumeration process, respondents were shown a series of 18 high-quality colour photographs (in A4 format). These photographs had been taken by professionals and showed each of the food items, presented in the typical bundle, pile, or package that is found in markets in PNG. For foods where there could be some confusion about the size of the items shown, a box of matches was included in the photograph so that respondents could put the item into perspective. Examples of these photographs are shown in Figure 1, for the four items with the largest budget shares – sweet potato, banana, betelnut (a mild narcotic, like *pan*), and rice. These photos were shown at the conclusion of the second visit to the household. Interviewers were instructed to ask the following question when showing the photograph:

> “How much does it currently cost to buy a *(Item)* like this, in the main market or store in this village/town?”

The questions about food were directed to the person in the household who typically buys most of the food, and the questions about drinks, betelnut and tobacco to the person who makes most of these purchases.

Respondents reported their opinion about the price of what they saw in the picture and no attempt was made to force them to report a price in kilograms, which are not widely used in markets in PNG.[8] Instead, the reported prices were transformed to kilogram prices at the analysis stage, using the actual weights of the items that were in the photographs. What was required of respondents was an ability to map a two-dimensional picture into volumes and weights and to form an assessment of quality based on the picture.[9] Pre-testing showed that respondents were quite good at this because reported prices based on pictures were quite close to the prices reported when respondents were shown the actual items that were in the picture. It was not practical to use actual products in the main survey because (a) interview teams would be burdened by carrying bulky products; and (b) the fact that the same product could not be used simultaneously in different survey locations means that quality variations could be introduced into the price opinions.[10]

III. Unit Values, Prices and Pictures

Our data collection effort provides us with three different measures of price (market prices, picture prices and unit values) for nine foods (sweet potato, banana, rice, betelnut, flour, biscuits, canned fish, soft drink and beer). In contrast, many surveys have just one measure of price, so analysts are often forced to use unit values even though they know that they are not a direct substitute for genuine price data.[11] We therefore use our survey to answer the following two questions: First, are the problems in using unit values as a measure of price sufficiently large to justify the expense of collecting additional information on prices? Second, if this additional information is needed, do picture prices have smaller problems than unit values?[12] Negative answers to both questions would suggest that current procedures using unit values are appropriate. Positive answers to both would suggest that some innovation in data collection methods is needed along the lines of our picture prices. Finally, if additional information on prices is needed but picture prices perform poorly, it would suggest that greater investment in properly carrying out community price surveys is needed.

In this section we report some simple descriptive analyses that may help to answer these two questions. To guard against outliers affecting the analyses, the original survey forms were re-examined and cases of data entry errors and obvious miscoding (e.g., kilograms entered as grams) were removed or rectified. As a further defence against the effect of outliers, we followed the rule of Cox and Wohlgenant (1986) and trimmed the sample by removing unit values and price opinions more than five standard deviations from their respective means. This procedure removed 23 unit values and 25 price opinions, which amounted to proportionately trimming more of the unit values because there were only 4550 of them, compared with 9100 observations on price opinions.

Even after trimming outliers, the unit values appear to be fairly noisy and biased measures of market prices. The correlations between household-specific unit values and market prices range between 0.38 and 0.59 for sweet potatoes, bananas and rice, the three foods with the largest budget shares.[13] Examining deviations from the 45-degree line in price plots also demonstrates the low correlations for the major food commodities (Figure 2). The correlations for the major food commodities, however, are even higher than those for the six other, more minor food commodities ($\bar{r} = 0.37$ -- results not shown).[14] Unit values also appear to be biased measures of mean market prices, according to the ratio, $\bar{x}_{uv} / \bar{x}_p$. The average unit value overstates the average market price by about 30 percent for sweet potato and banana, the two most common locally produced foods.

Picture prices appear to provide a better measure of market prices. When using the same households as the unit value analysis, the scatter plots of market prices and picture prices are distributed more symmetrically around the 45-degree line and the ratio of means of the two price series, $\bar{x}_{pp}/\bar{x}_p$, is closer to one, ranging from 0.94 to 1.01 (Figure 2). The correlations with market prices range from 0.48 to 0.79 for the three major foods. The average correlation coefficient between picture prices and market prices for the six minor food commodities is also higher, $\bar{r} = 0.64$ (compared with $\bar{r} = 0.37$ for the unit values).

There are several reasons why picture prices and, especially, unit values may be imperfect measures of market prices. Both may contain quality effects, although these tend to be small in our data, particularly for the picture prices (see below). Second, the specification of items may differ between the pictures, the market price surveys and the unit values. But for the foods where a clear comparison is possible, there is no evidence that a discrepancy between what the unit value refers to and the specification used in the market price surveys contributes to the low correlation.[15] Finally, both picture prices and unit values are subject to reporting error, and it could be that for unit values the errors are greater.

In fact if we treat all three series (market prices, unit values and picture prices) as error-ridden measures of true but unknown community prices, the intra-cluster correlations amongst each measure can provide an estimate of the "reliability ratio" – the proportion of measurement error in the variance of the observed price series. Following this logic, we find that the intra-cluster correlation of unit values (after the effects of quality have been purged) is systematically lower than those of market prices and picture prices. The average value of the correlation coefficients across the nine foods is only 0.379 for unit values, compared with 0.782 for market prices and 0.647 for picture prices. Hence, according to this analysis, unit values are the least reliably measured although there is imperfect reliability for all of the price measures.

By examining Figure 2, it seems possible that a few households disproportionately generate much of the reporting error bias in both picture prices and unit values. To see how important this source of bias is, we follow a common practice of forming cluster averages. When we use these averages the correlation between unit values and market prices improves, although the unit values still tend to be noisier measures than the picture prices (Table 1, columns 6 and 7). The average correlation of cluster-level unit values and market prices is 0.63, while the average correlation for picture prices is 0.77.[16]

Averaging by cluster, however, does not remove the bias that occurs when unit values are used to calculate average market prices (Table 1). On average, the mean price for each food and the mean of the cluster-level unit values for the same food differs by 14 percent (this is calculated for each food as: $\left|\bar{x}_{uv} - \bar{x}_p\right| / \bar{x}_p$). In contrast, the average error is only 6 percent for the picture prices. Hence, the conclusion that unit values are more biased measures of average market prices holds even for the cluster-level estimates.

In addition to being a biased and noisy measure of market prices, there is a further statistical problem with unit values which becomes apparent when the cluster means are formed. A cluster mean unit value is available only when at least one household in that cluster made a purchase during the recall period. When there are no households making such a purchase, a sample selection problem occurs. In the case of some commodities, this can be a fairly serious problem. For example, in our sample, rather than the expected sample of 120 clusters, there are only 63 clusters with an average unit value for beer and 92 clusters with one for banana.[17] How serious this sample selection problem would be elsewhere is likely to depend on the length of the survey recall period, with longer recalls allowing more households to record a purchase.[18] In contrast to the unit values, the picture prices are much more widely available, with the most for any food being four clusters having all households with missing prices opinions. Thus, the method of obtaining opinions about prices rather than just relying on purchase behaviour can, potentially, capture the full range of spatial price variation in a sample.

IV. The Effects of the Alternative Price Collection Methods

In this section, we seek to measure the impact of using the alternative prices series as proxies for market prices. To do so, in the next subsection, we examine how using unit values (compared to using picture prices) will affect estimates of the poverty line and a number of different aggregate measures of poverty. In the following subsection, we do the same for price elasticity estimates and assess the implications for tax policy analysis.

Price Collection Methods and Poverty

Existing poverty lines for PNG are based on the market prices collected by the survey (World Bank, 1999). Specifically, the cost of buying a basket of foods that provides 2200 calories per day was calculated for five regions: the National Capital District (NCD), the South Coast, the Highlands, the North Coast, and the New Guinea Islands. Rural and urban areas within each region are combined because the sample usually had only one urban cluster per region and there are no rural clusters in the NCD.[19] The regional average prices used to calculate the cost of the poverty line basket of foods were themselves calculated from the cluster-level averages of the market prices, which have been described in Table 1.[20]

In this section of the paper we follow the above procedures used to calculate the food poverty line in PNG, but work instead with the unit values and price opinions. The aim of this replication is to construct alternative poverty lines, to see what impact the use of a different source of price information would have on measured poverty. In all cases, the unit values and price opinions are first averaged by cluster before forming the regional averages. This ensures that clusters with more purchasing households do not receive undue weight in the calculations. The cluster averages will also tend to dampen some of the measurement error. The unit values are also 'purged' of quality effects by running within-cluster regressions on a set of household characteristics (see Equation (1) below, for the particular characteristics included).

One constraint with these exercises is that while the poverty line contains 35 foods, there are only nine foods with data on both price opinions and unit values. These foods, however, contribute almost one-half of the value of the food poverty line. Thus, our experiments are, effectively, varying only one half of the value of the food poverty line, so the measured effect of different price collection methods on estimated poverty may be, if anything, understated.

The regional food poverty lines that result from using the market prices, unit values and price opinions are illustrated in Figure 3. When market prices are used, the food poverty lines range from K235 per year in the North Coast region to K626 in the NCD, and have a population-weighted average of K330.[21] While the existing poverty lines for PNG include a non-food allowance, which is equivalent to between one-third and one-half of the value of the food poverty line, we ignore that here because our different price information is only for foods.

The food poverty line is consistently overstated when unit values are used as the measure of price (Figure 3). In the NCD, South Coast and Islands regions, unit value overstate the poverty line by a relatively modest margin, of only 4-10 percent. However, in the other two regions, areas containing 70 percent of the population, unit value-based analysis overstates the food poverty line by 16 to 20 percent. In contrast, the use of picture prices creates a smaller bias in poverty line estimates. In two regions, the NCD and South Coast, the use of picture prices causes the food poverty line to be understated by about 10 percent. In the other three regions, it is overstated by 4 to 11 percent. On average, the food poverty line has a proportionate error, $\left|z_i - z_p\right| / z_p$ (where z is the food poverty line and p=market prices, and i=unit values or picture prices) of 14 percent with the unit values, and 9 percent with picture prices.[22]

When data collection methods create biased estimates of the poverty line, they also affect measures of poverty rates (Table 2). In particular, the overstatement of the food poverty line when unit values are used causes an upward bias in measured poverty rates. For example, the head-count index is estimated to be 28 percent (with a standard error of 2.6 percent) rather than the actual figure (based on market prices) of 22 percent (rows 1 and 2).[23] The poverty gap index is estimated as 8.0 percent rather than 5.9 percent. The differences between these estimates and those based on the market prices are statistically significant (the *t*-statistics for the null of no difference range from 4.8 to 6.8). This evidence that using unit values causes higher poverty measures is consistent with a result from Capéau and Dercon (1998) that headcount poverty in rural Ethiopia would be overstated by one-fifth if unit values are used rather than other price data.

In contrast, although there is also an upward bias associated with the use of the picture prices, the discrepancy is significantly smaller (Table 2, rows 1 and 3). Picture price-based estimates overstate the headcount poverty measure by only eight percent (the *t*-statistic for the null of no difference is 2.2). This overstatement is significantly less than when the unit values are used (the *t*-statistic is 4.5 for the test that the overstatement is the same for unit values and picture prices). Clearly, in this respect, the picture price series provide more accurate measures of poverty in PNG, although even the smaller overstatement from picture prices may be enough to justify the expense of collecting better price data from local stores and markets.

Price Collection Methods, Price Elasticity Estimates, and Indirect Tax Analysis

In this section we report the results of using the three different price measures to calculate own- and cross-price elasticities of demand. In developing countries, pricing policy plays the same central role in fiscal policy that income tax and social security plays in developed countries (Deaton, 1989). The matrix of price elasticities, which is needed to estimate the revenue effects of price reforms, can therefore provide fundamental information to governments.[24] Hence, it is important to establish what bias might occur when elasticities are calculated from either unit values or picture prices if estimates from prices based on the preferred data collection method (that is, market price surveys) are not available.

Although we have the three measures of price for nine different foods, we focus attention on the three major staples; sweet potato, banana, and rice.[25] These three foods comprise over one-fifth of total household consumption expenditures and supply about 45 percent of calories to households. In addition to their consumption and nutritional importance, these three foods have some policy significance because until recently rice was imported duty free, whereas all other food imports were subject to tariffs. But following a switch to a Value-Added Tax (VAT), rice is now taxed at the same ten percent rate as other imported goods. In contrast, sweet potato and banana effectively fall outside of the tax net because the farmers and traders who sell them in informal markets are not registered for the VAT.

There are 11 clusters with no market price survey data for either sweet potato or banana, so the demand system is estimated on the remaining 109 clusters (containing 1018 households). This reduced sample highlights one advantage of the picture method, because there would be only two clusters with missing data if only the picture prices were used. It is also notable that of the 109 clusters, only 86 have at least one household making purchases of either sweet potato or banana (the total number of purchasing households is ca. 350). Thus, we are forced to rely on methods of imputing unit values for those households and clusters that do not have any available.

The base model uses market prices and a "share-log" functional form (Deaton, 1989):

$$w_i = \alpha_i + \beta_i \ln x + \sum \theta_{ij} \ln p_j + \boldsymbol{\gamma}'\mathbf{z} + u_i \tag{1}$$

where w_i is the share of the budget devoted to good i, x is total expenditure, p_j are the prices and $\mathbf{z}$ is a vector of other household characteristics: (log) household size, the share of the household in seven demographic groups: males and females 0-6 years, 7-14 years, 15-50 years, and over 50 years (males excluded), dummy variables for whether the household head was either female or employed in the formal sector, and regional and quarterly dummy variables. An advantage of the functional form in equation (1) is that it is able to treat zero and non-zero consumption in the same way. While there is a literature on censored demand systems, this is not needed here; the analysis of tax and subsidy reform relies on unconditional demand functions because the revenue effect of a tax increase does not depend on whether demand changes take place at the extensive or intensive margins (Deaton, 1990). The price elasticities for equation (1) are given by:

$$\varepsilon_{ij} = \left(\theta_{ij} / w_i\right) - \delta_{ij}, \tag{2}$$

where δ_{ij} is the Kronecker delta (=1 if i=j, 0 otherwise) and budget shares are evaluated at their mean values.

The most common empirical strategy for using unit values is to simply replace the prices in equation (1) with unit values. Most of the variation concerns how analysts deal with the missing unit values and with the choice of leaving unit values at household level or aggregating them to cluster level. We use the following two methods:

UV1: using household-specific unit values, with missing unit values replaced by the mean unit value calculated across other households in the same region and season (following Minot, 1998);

UV2: using cluster median unit values, in place of both household-specific and missing unit values. This follows several studies that use averages, but with the median chosen for its robustness to outliers.

We also apply these same two methods to the picture prices, denoting them PP1 and PP2.

In addition to replacing unobserved prices with some form of unit value (as in UV1 and UV2) and estimating equation (1) and then getting elasticities from equation (2), we also use the procedures developed in Deaton (1990). The Deaton procedure uses a two-equation system of budget shares (w_{Gic}) and unit values (v_{Gic}) that are both functions of the *unobserved* prices, (p_{Hc}):

$$w_{Gic} = \alpha_G^0 = \beta_G^0 \ln x_{ic} + \gamma_G^0 \cdot z_{ic} + \sum_{H=1}^{N} \theta_{GH} \ln p_{Hc} + (f_{Gc} + u_{Gic}^0) \tag{3}$$

$$\ln v_{Gic} = \alpha_G^1 = \beta_G^1 \ln x_{ic} + \gamma_G^1 \cdot z_{ic} + \sum_{H=1}^{N} \psi_{GH} \ln p_{Hc} + u_{Gic}^1 \tag{4}$$

In addition to the variables previously defined, f_{Gc} is a cluster fixed-effect in the budget share for good G, u_{Gic}^0 and u_{Gic}^1 are idiosyncratic errors, and the i indexes households, the G and H index goods, and the c indexes clusters.

Deaton's method recognises that the data are collected on *clusters* of households that are presumed to face the same market prices. The intra-cluster variation in budget shares and unit values is used to identify the effect of income and other household characteristics on both the quantity and quality demanded. For example, the coefficient β_G^1 is the elasticity of the unit value with respect to total expenditure (henceforth, called the *quality elasticity*) while the elasticity of quantity demanded with respect to total expenditure is derived from β_G^0.

The first-stage, within-cluster regressions are consistent even in the absence of market prices, which are treated as fixed effects. Any residual variation in unit values (and covariance with budget share residuals) is assumed to reflect measurement error, and the first-stage regression residuals give an empirical estimate of these errors. More specifically, the error terms, e_{Gic}^0 and e_{Gic}^1, from equations (3) and (4) contain all the variability around the cluster means of w_{Gc} and ln v_{Gc} that is not explained by household characteristics, so this residual variability is assumed to reflect measurement error.

Results from the first stage of the Deaton procedure are reported in Table 3. In order to compare the quality effects and measurement error properties of unit values and picture prices, equation (4) is estimated with both types of data. The quality elasticities are universally small, ranging from -0.07 to 0.06 for unit values and from -0.04 to 0.01 for picture prices. These small values are consistent with the evidence from Deaton (1990) and Gibson, Rozelle and Le (2002) that the quality problems with unit values are rather less important than the measurement error problems. Moreover, notwithstanding their small size, the quality elasticities are larger for unit values than for picture prices. On average, the absolute value of the quality elasticities is almost four times as large for unit values as for the picture prices.

The unit values also have higher measurement error variance (as measured by the variability around the cluster means of ln v_{Gc} that is not explained by household characteristics in equation (4)) than the picture prices for all nine foods. On average the measurement error was four times as great (almost ten times for soft drink and biscuits) for the unit values compared with the picture prices. Finally, the covariance between the errors in the unit value equation and the errors in the budget share equation also were higher for unit values than for picture prices for seven of the nine foods. On average, the covariance in the errors was almost 10 times as great for the unit values, suggesting that the errors in the picture prices are less correlated with actual demands than are the errors in the unit values.

In the second stage of the Deaton procedure, a between-clusters errors-in-variables regression is applied to the (adjusted) average budget shares and unit values, which have been purged of household characteristics at the first stage. If it were not for the effect of prices on cluster-wide quality variation, the parameters estimated at the second stage would be sufficient for calculating price elasticities. Instead, a separability theory of quality (Deaton 1988) has to be used to identify the price effects at the third and final stage. An important feature of the procedure is that it depends on a large number of clusters (rather than a large number of households) for its consistency properties.

When comparing the own-price elasticity estimates from the five price proxy series and methods (UV1, UV2, PP1, PP2, and the Deaton method) with those that are based on market prices, both picture price series (PP1 and PP2) create the estimates with the least bias (Figure 4). The point estimates of the elasticities estimated from picture price methods (particularly those using the cluster-medians--PP2) are close to those of the market price-based estimates. Also, the confidence intervals have a high degree of overlap.

There is less overlap for the two simple unit value procedures, UV1 and UV2, and for that of the Deaton method (Figure 4). For example, in the case of the estimates of the own-price elasticity of demand for sweet potato, the market price-based estimate is -1.33±0.09. When household-level unit values are used, however, the estimated elasticity is much lower in an absolute value sense (-1.00±0.08). When cluster median unit values are used (UV2), the absolute value of the estimated elasticities are even lower (-0.77±0.10). Moreover, while the Deaton procedure calculates point estimates of the own-price elasticities for sweet potato and rice that are relatively consistent with the estimates from market prices, it does a poor job of estimating the own-price elasticity for banana (giving a point estimate of -2.2 rather than -1.0). There is also considerable imprecision in the Deaton estimates. The imprecision, however, is not surprising because Deaton's method essentially reduces to a between-clusters regression, and, in our sample there are not many clusters.

Estimates of cross price elasticities, also important in indirect taxation analysis, are likewise affected adversely by the use of unit values. Although there are too many cross-price elasticity estimates to display individually, the aggregate bias (AB) can summarize the performance of each method. Let $\boldsymbol{\varepsilon}$ be the vector of elasticities calculated from the market price data and $\hat{\boldsymbol{\varepsilon}}$ the corresponding elasticity vector from unit values or picture prices, so that the bias is $\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}$, and $AB = (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})'(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})$, which is the sum of squared biases. The aggregate bias is calculated for the own-price elasticities alone (AB1) and for the full system of own- and cross-price elasticities (AB2).[26] With the exception of the Deaton method, where bootstrapping is used, standard errors for AB1 and AB2 are obtained from the delta method.

According to our results, the aggregate bias in the own-price elasticities is lowest (AB1=0.048) when the estimation uses cluster medians of the picture prices (Table 3, column 1). When the cross-price elasticities are included in the aggregate bias calculation (AB2), the use of household-specific picture prices performs best (AB2=0.904—column 2). It is notable that the bias estimates for either procedure using picture prices are less than 35 percent of those for the similar procedure using unit values. Moreover, while neither AB1 nor AB2 are statistically significant when using the picture prices, AB2 is statistically significant (at $p<0.03$ or smaller) for all three of the unit value procedures. Similarly, the correlation of the picture price elasticities (PP1 and PP2) with the market price elasticities is higher (0.94-0.96) than is the correlation for UV1 and UV2 (0.67-0.80—column 3). The Deaton procedure does worst in the aggregate bias calculations, although the standard errors for AB1 and AB2 are also widest with this procedure.[27]

The bias in the elasticities calculated from naïve unit value procedures could affect public policy decisions. One obvious use of the price elasticities is for deciding on the direction of marginal tax reform (Deaton and Grimard, 1992). The last three columns of Table 4 contain estimates of the social cost-benefit ratios, λ_i of a marginal increase in tax on each of the three foods, calculated from:

$$\lambda_i = \frac{w_i^{\varepsilon} / \tilde{w}_i}{1 + \frac{\tau_i}{1+\tau_i}\left(\frac{\theta_{ii}}{\tilde{w}_i} - 1\right) + \sum_{k \neq i} \frac{\tau_k}{1+\tau_k}\frac{\theta_{ki}}{\tilde{w}_i}} \tag{5}$$

where τ_i is the tax rate on good i (0.1 for rice and 0 for the others), θ_{ki} is the log price derivative of the budget share (from equation (1) or (3)), and the average budget shares w_i^{ε} and $\tilde{w}_i$ are:

$$w_i^{\varepsilon} = \left[\sum_{m=1}^{M} (x_m / n_m)^{-\varepsilon} x_m w_{im}\right] \Big/ \sum_{m=1}^{M} x_m \tag{6a}$$

$$\tilde{w}_i = \sum_{m=1}^{M} x_m w_{im} \Big/ \sum_{m=1}^{M} x_m \tag{6b}$$

where x_m and n_m are the total expenditure and size of household m, and ε is the coefficient of inequality aversion.[28] According to the calculations in Table 4, when market prices are used to estimate θ_{ki}, the highest ratio of social costs to benefits occurs when there is a marginal increase in the tax on sweet potato (λ=1.47±0.01), followed by a tax on rice (λ=1.44±0.05), while banana looks like the best candidate for a tax increase (λ=1.39±0.02). But this ranking is preserved by only two of the other estimation methods: picture prices with missing values replaced by regional and quarterly means (PP1) and the Deaton procedure applied to unit values.[29] The other two unit value procedures rank rice as the best candidate for tax increases. Hence, using unit values as proxies for market prices in an optimal tax reform exercise might lead policy makers in PNG to increase a tax which is not the socially least-cost source of revenue.

Part of the poor performance of the methods that rely on unit values may reflect the sample selection problem of several clusters having no unit value available. While this is an intrinsic disadvantage of unit value methods, in some settings there might be a wider availability of unit values either because households are more reliant on purchased food or because the consumption recall period is longer. In Table 5 we explore the performance of the cluster-median and Deaton estimators on the sub-sample of 86 clusters that have unit values available for all three foods. This change in the sample coverage does, in fact, improve the relative performance of the cluster-median unit values, although the aggregate bias (AB2) is still almost twice as large for unit value-based measures when compared to those using picture prices (but the difference is no longer statistically significant). The Deaton method also appears to do better on this sub-sample, in terms of the aggregate bias now being statistically insignificant and a higher correlation with the market price elasticities. Thus, unit value methods may not fail as badly as indicated in Table 4 and Figure 4, if the unit values are available for a wider range of clusters than they are in PNG.

However, a trend in the literature is to artificially reduce the number of clusters by redefining them at a broader geographic level. Starting with Gracia and Albisu (1998), several users of the Deaton method have treated *regions* as a cluster. For example, Nicita (2004) uses each of the 32 states in Mexico as a cluster, even though there are hundreds of lower level *municipios*. Likewise, Kedir (2001) groups households from an unclustered urban survey in

Ethiopia into 'clusters' of varying aggregation, even treating Addis Ababa as a single cluster. It is doubtful that the Deaton method can provide reliable elasticity estimates in these circumstances, because it assumes that "households in a single cluster live near one another" (Deaton, 1997, p. 73) and it needs a large number of clusters for its consistency properties. Intra-regional variation in unit values due to spatial price variation will wrongly be treated as measurement error when clusters are artificially aggregated. Using a single unit value for an entire region will overstate price in villages where market prices are low and understate it in high-price villages, so demand differences will be explained by attenuated price differences, usually causing the magnitude of elasticities to be overstated.

To see how large is the impact of aggregating clusters, the Deaton method was re-run with each of the 19 provinces in PNG treated as a cluster. For both sweet potato and rice, the estimated own-price elasticities move further from the values estimated when market prices are used, so aggregating seems to impair the Deaton estimator. The effect on the elasticity for rice is especially large; the own-price elasticity is -4.1 when provinces are used as clusters but only -2.3 when the original 109 clusters are used.[30] It is not surprising that the elasticity for rice is affected most, because an ANOVA shows that rice has the highest proportion of within-province variation in market price (0.77) and in quality-purged unit values (0.56). Thus, if most price variation is typically within regions, the strategy of applying the Deaton method to artificially aggregated clusters is likely to bias elasticity estimates and mislead subsequent analyses.

V. Practicalities of Price Surveys

Is the improvement in data quality from using picture prices, as compared to using unit values, worth the additional effort? According to our experience in PNG, price opinions can be collected with the aid of a picture in a fairly efficient, timely way. The enumerators in our survey collected picture prices for 18 products at the same time that the rest of the household survey was being done. Hence, picture papers required no additional logistical effort (besides having to remind enumerators to bring their picture price albums with them to the field). On average, the typical household only spent about ten minutes on this block of the survey. Since each cluster included 12 households, the survey team, on average, spent about two hours per cluster collecting picture prices. Moreover, enumerators and respondents intrinsically liked this part of the survey, because it provided a break from the normal questioning.

In the PNG survey more effort was needed surveying local stores and markets than was needed to obtain the price opinions. The typical community price survey required visits to the local fresh produce market and to two trade stores. It usually took less than 15 minutes to obtain the required prices from each trade store. The survey in fresh produce markets, however, was more time consuming. It typically took an enumerators up to an hour to weigh and record the prices of up to six lots of 11 different items. In sparsely populated areas even more time was spent in getting to the market, which was commonly located by the nearest Community School. In some cases, the nearest market was over an hour's walk from the village. Because the survey in the fresh produce market was repeated when the team returned to each community for the consumption recall, the travel and survey times were doubled. In addition, in many communities there were prohibitions on betelnut being sold in the main market and on beer being sold in trade stores. In such cases, enumerators had to spend additional time traveling to the roadside betelnut markets and to the nearest beer sellers. On average, the total time spent surveying local stores and markets collecting market price data was about four hours per community, which was twice the time spent on the picture prices.

Another problem with the market surveys was that in some areas the markets would either meet infrequently (typically one day per week) or else convene at daybreak and only last for one or two hours. When the market day was missed, due to logistical failures or due to timing problems, survey teams often had to spend a lot of time and resources to leave a member in the village until the market convened or to send back a team at a later date. None of these timing problems affected the collection of picture prices.

Based on our experience, it would seem feasible to collect picture prices for 30-40 items.[31] This would imply a time commitment of 15-20 minutes per household or 3-4 hours per cluster. If more prices were required, market price surveys might become more attractive because the fixed cost of finding and getting to the market could be spread over the larger number of items whose prices were surveyed. However, the success of market price surveys would depend on whether it is possible to find every item in a rural marketplace; more detailed surveys might run into the problem of a large number of missing prices.[32] Thus, some consideration of how markets operate may be needed when choosing whether to rely on price opinions rather than market price surveys.

An increase in the detail of a survey also can undermine methods that rely on unit values because of the greater likelihood that entire clusters have no households reporting the purchase of a narrowly specified item. For example, 21 percent of the clusters in the PNG survey did not have a unit value for flour whereas the broader category of “cereals” had purchasers (and hence unit values) in most clusters. Price opinions are less dependent on actual purchases so a survey that sought details on many items rather than a few commonly consumed ones would still have information available. For example, 77 percent of the PNG sample offered opinions on the price of flour, even though only 30 percent purchased it during the recall period. However, this raises the question of how reliable are the opinions of households who do not purchase a particular good. In the PNG survey there was only a small gap, of 0.06, in the average correlations between market prices and price opinions, when the sample was divided according to whether households purchased each item ($\bar{r} = 0.62$ for purchasers versus 0.56 for non-purchasers). We speculate that the reason that non-purchasers seem relatively well informed about the prices of things they do not consume is because they still observe those prices in stores and the market when they are shopping for other goods. Consistent with this, the item where the discrepancy was largest was beer (r=0.89 for purchasers versus 0.75 for non-purchasers) which is usually sold in hotels, clubs and licensed outlets. Hence, non consumers have less chance to observe the price of such a commodity because they never see it while shopping for other goods. Thus, the usefulness of price opinions may also depend on how segmented are markets, which affects the ease of observing prices for items that the household does not usually consume.

VI. Conclusions

This paper has presented evidence on the accuracy of poverty lines, poverty rates and price elasticities of demand estimated from household budget surveys. Three different measures of price have been used: average market prices as established from a market price survey, unit values and the price opinions of household members shown pictures of specified foods. The sort of cross-sectional household survey data studied here are increasingly being used as economists try to exploit one of the few data sources in developing countries that can help provide estimates of the demand responses that are needed for evaluating tax and subsidy reforms.

Our findings suggest that unit values, whether used in naïve or improved estimation procedures, lead to biased estimates of poverty rates and biased estimates of price elasticities. In contrast, the price opinions perform better, with both poverty estimates and demand elasticities being closer to the values established from market price surveys. There are good reasons to expect this better performance from the price opinions. The picture-based method can provide price estimates for a much wider ranger of households than unit values can, the errors in the estimates are unlikely to be correlated with demands and the price opinions should have less quality variation because everyone sees the same picture. Thus, based on these findings it may be worthwhile to pursue the approach of directly asking households about prices, rather than indirectly obtaining price information from unit values.

Whether relying on price opinions would be better than collecting good measures of prices by surveying local stores and markets will depend somewhat on the nature of each survey and on the nature of rural markets in a given country. What is clear is that in many developing countries, for whatever reason, the logistics of collecting market prices appear to be so difficult that many surveys do not attempt this, and of those that do, some end up not using the data. Consequently, many important analyses of poverty and price and tax policies rely on very imperfect price information.

The findings of this paper also should provide an incentive for others to experiment with methods of gathering price data in rural areas of developing countries. For example, instead of pictures, three-dimensional models might be used. In fact, a broader experiment could be designed: using price opinions from an informed respondent without an aid (as was done in the Indonesia Family Life Survey); using pictures (as we did); and using three-dimensional models to elicit price opinions. It would also be of interest to learn whether certain types of respondents within the household have more informative opinions that others; such comparisons are precluded by our design which asked only for the opinion from 'the most informed' person.

Using pictures or other aids to help gather data from households on their beliefs about existing prices could also lead to questions about hypothetical prices. For example, one use of price data is to econometrically estimate price elasticities of demand. Another approach would be to directly ask households how they would change their demand for a pictured item following some hypothetical price changes. It would be an interesting experiment to find out how well the direct approach approximated the econometric estimates. These willingness-to-pay type questions might usefully be applied to more than just food, with medicines and other health interventions being plausible candidates for this approach.[33]

References

Ahmad, E., and Stern, N. (1984) The theory of reform and Indian indirect taxes. *Journal of Public Economics* 25(3): 259-298.

Ahmad, E., and Stern, N. (1991) *The theory and practice of tax reform in developing countries* Cambridge University Press

Alderman, H. (1988) Estimates of consumer price response in Pakistan using market prices as data. *Pakistan Development Review* 27(2): 89-107.

Beegle, K., Frankenberg, E. and Thomas, D. (1999) Measuring change in Indonesia. *Labour and Population Program Working Paper Series* 99-07. Santa Monica, CA: RAND.

Capéau, B. and Dercon, S. (1998) Prices, local measurement units and subsistence consumption in rural surveys: an econometric approach with an application to Ethiopia. *Working Paper* 98-10, Centre for the Study of African Economies, Oxford University.

Cox, T. and Wohlgenant, M. (1986) Prices and quality effects in cross-sectional demand analysis. *American Journal of Agricultural Economics* 68(4): 908-919.

Deaton, A. (1988) Quality, quantity, and spatial variation of price. *American Economic Review* 78(3): 418-430.

Deaton, A. (1989) Household survey data and pricing policies in developing countries. *The World Bank Economic Review* 3(2): 183-210.

Deaton, A. (1990) Price elasticities from survey data: extensions and Indonesian results. *Journal of Econometrics* 44(3): 281-309.

Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* Johns Hopkins, Baltimore.

Deaton, A. (2003) Prices and poverty in India, 1987-2000. *Economic and Political Weekly* January 25, pp.362-368.

Deaton, A. and Grimard, F. (1992) Demand analysis for tax reform in Pakistan. *LSMS Working Paper* No. 85, Washington D.C., World Bank.

Deaton, A. and Grosh, M. (2000) Consumption. In M. Grosh and P. Glewwe (eds) *Designing Household Survey Questionnaires for Developing Countries* The World Bank, Washington, pp. 91-133.

Deaton, A. and Tarozzi, A. (2000) Prices and poverty in India. *Mimeo* Research Program in Development Studies, Princeton University.

Deaton, A., Friedman, J., and Alatas, V. (2004) Purchasing power parity exchange rates from household survey data: India and Indonesia. *Mimeo* Princeton University.

Frankenberg, E. (2000) Community and price data. In M. Grosh and P. Glewwe (eds) *Designing Household Survey Questionnaires for Developing Countries* The World Bank, Washington, pp. 315-338.

Friedman, J. and Levinsohn, J. (2002) The distributional impact of Indonesia's financial crisis on household welfare: a 'rapid response' methodology. *World Bank Economic Review* 16(3): 397-423.

Gibson, J. (2000) A Poverty Profile of Cambodia, 1999. A Report to the World Bank and the Ministry of Planning, Phnom Penh, Cambodia.

Gibson, J. and Rozelle, S. (2002) Demand systems with unit values: comparisons with elasticities from market prices. *Mimeo* Department of Economics, University of Waikato.

Gibson, J., Rozelle, S., and Le, T. (2002) Evaluating the use of unit values and community prices in demand analysis. *Mimeo* Department of Economics, University of Waikato.

Gracia, A., and Albisu, L. (1998) The demand for meat and fish in Spain: urban and rural areas. *Agricultural Economics* 19(3): 359-366.

Glewwe, P. (1991) Investigating the determinants of household welfare in Côte d'Ivoire. *Journal of Development Studies* 35(2): 307-337.

Jolliffe, D. and Semykina, A. (1999) Robust standard errors for the Foster-Greer-Thorbecke class of poverty indices: SEPOV. *Stata Technical Bulletin*, STB-51.

Kedir, A. (2001) Some issues in using unit values as prices in the estimation of own-price elasticities: evidence from urban Ethiopia. *CREDIT Research Paper* 01/11, University of Nottingham.

Laraki, K. (1989) Ending food subsidies: nutritional, welfare, and budgetary effects. *World Bank Economic Review* 3(3): 395-408.

Minot, N. (1998) Distributional and nutritional impact of devaluation in Rwanda. *Economic Development and Cultural Change* 46(2): 379-402.

Minot, N., and Goletti, F. (2000) *Rice Market Liberalization and Poverty in Viet Nam* Research Report 114, International Food Policy Research Institute, Washington.

Nicita, A. (2004) Who benefited from trade liberalization in Mexico? Measuring the effects on household welfare. *World Bank Policy Working Paper* No. 3265.

Nicita, A. (2004a) Efficiency and equity of a marginal tax reform: income, quality and price elasticities for Mexico. *World Bank Policy Working Paper* No. 3266.

Ravallion, M., and Dearden, L. (1988) Social security in a "moral economy": An analysis for Java. *Review of Economics and Statistics* 70(1): 36-44.

Saunders, C, and Grootaert, C. (1980) Reflections on the LSMS group meeting. *Living Standards Measurement Study Working Paper* No. 10, The World Bank.

Wood, D., and Knight, J. (1985) The collection of price data for the measurement of living standards. *Living Standards Measurement Study Working Paper* No. 21, The World Bank.

World Bank. (1999) *Papua New Guinea: Poverty and Access to Public Services* Washington, DC: World Bank.

Table 1: Descriptive statistics for cluster-level market prices, unit values and price opinions

	Mean market price[a]	Mean unit value[a]	Mean price opinion[a]	No. of clusters with data on[b]		Correlation with market prices	
				Unit values	Price opinions	Unit values	Price opinions
Sweet potato	43.9	59.0	42.5	93	118	0.74	0.74
Banana	54.2	75.9	51.3	92	118	0.65	0.71
Rice	114.7	107.3	115.5	114	118	0.75	0.93
Flour	143.6	114.9	158.3	95	116	0.43	0.72
Biscuits	444.4	450.0	452.4	112	118	0.50	0.83
Canned fish	432.7	437.0	422.7	115	118	0.42	0.56
Betelnut	510.8	566.0	419.9	107	117	0.63	0.64
Soft drink	272.8	263.3	287.9	100	118	0.73	0.91
Beer	558.3	507.0	586.8	63	116	0.86	0.93

[a] Toea per kilogram, as calculated from cluster-level averages. 130 toea=US$1 in 1996.
[b] Out of a possible n=120.

Table 2: Aggregate food poverty measures for Papua New Guinea, 1996

Cost of poverty line food basket calculated from:	Headcount index	Poverty gap index	Poverty severity index
Market prices	22.0	5.9	2.4
	(2.4)	(0.9)	(0.4)
Unit values	28.0	8.0	3.4
	(2.6)	(1.0)	(0.6)
Price opinions	23.8	6.8	2.8
	(2.5)	(1.0)	(0.5)

Note: Based on the food poverty lines in Figure 3. The poverty estimates are in terms of adult-equivalents. The unit values have been purged of quality effects using a regression. Standard errors in ( ) are corrected for the effect of clustering, sampling weights and stratification.

Table 3: Quality and Measurement Error Indicators for Unit Values and Picture Prices from the First Stage Regressions of the Deaton Procedure

	Quality Elasticity[a]		Residual variance[b]		Residual covariance[c]	
	Unit values	Price opinions	Unit values	Price opinions	Unit values	Price opinions
Sweet potato	-0.016 (0.039)	-0.040 (0.027)	0.152	0.151	-0.042	0.466
Banana	0.059 (0.055)	-0.005 (0.025)	0.334	0.166	7.255	-0.131
Rice	-0.019 (0.011)*	-0.005 (0.007)	0.031	0.011	-0.803	0.149
Flour	-0.045 (0.037)	0.010 (0.020)	0.121	0.064	-0.865	0.233
Biscuits	0.035 (0.026)	0.009 (0.006)	0.111	0.011	0.276	0.018
Canned fish	0.018 (0.020)	0.005 (0.008)	0.074	0.019	-0.076	-0.022
Betelnut	-0.012 (0.038)	-0.003 (0.017)	0.260	0.079	0.854	-0.381
Soft drink	0.028 (0.021)	-0.006 (0.007)	0.071	0.007	0.221	0.027
Beer	-0.074 (0.056)	0.001 (0.012)	0.058	0.011	-0.331	0.530

[a] Standard errors in ( ); * statistically significant at 10%.
The quality elasticity is the coefficient β_G^1 in equation (4).

[b] Calculated from e_{Gic}^1 in equation (4).

[C] The residual covariance is x1000 and is from equations (3) and (4).

Table 4: Summary Comparisons of Estimates Using Market Prices, Picture Prices and Unit Values

Data source and estimation method	AB1	AB2	Corr	Cost-benefit ratio (λ_i) of tax rise for: Sweet potato	Banana	Rice
Market prices				1.47 [3] (0.01)	1.39 [1] (0.02)	1.44 [2] (0.05)
PP1 (missing=reg/qtr mean)	0.089 (0.133)	0.904 (0.503)	0.958	1.46 [3] (0.01)	1.40 [1] (0.01)	1.41 [2] (0.04)
PP2 (cluster medians)	0.048 (0.147)	1.448 (0.874)	0.938	1.45 [2] (0.01)	1.40 [1] (0.02)	1.47 [3] (0.07)
UV1 (missing=reg/qtr mean)	0.369 (0.356)	3.323 (1.444)	0.804	1.49 [3] (0.01)	1.40 [2] (0.02)	1.35 [1] (0.03)
UV2 (cluster medians)	0.653 (0.408)	4.844 (1.553)	0.669	1.48 [3] (0.01)	1.42 [2] (0.02)	1.34 [1] (0.03)
Unit Values (Deaton method)	1.415 (0.943)	7.775 (3.582)	0.737	1.53 [3] (0.04)	1.34 [1] (0.06)	1.43 [2] (0.08)

Note: AB1 is the aggregate bias on the own-price elasticities, AB2 is the aggregate bias on own- and cross-price elasticities, "Corr" is the correlation between the elements of the elasticity matrix and the market price elasticities. The calculations exclude the elasticities for "other goods" derived from the adding-up and homogeneity restrictions.

PP refers to "picture prices" and UV to "unit values". The cost-benefit ratio, λ_i is calculated from equation (5), using an inequality aversion parameter, ε=0.5. The values in [ ] are the good's rank in terms of λ_i, where "1" denotes the good with the lowest cost-benefit ratio from a marginal tax increase.

Standard errors in ( ) are derived from the delta method, except for those for unit values estimated using the Deaton method, which are bootstrapped from the second stage regression using the approach outlined in Deaton (1997).

Table 5: Results for the sub-sample with each cluster having a unit value available[a]

	Price Elasticities of Demand Calculated From:			
		Cluster Medians of		Deaton
	Market Prices	Picture Prices	Unit Values	Procedure
Own-Price Elasticity for				
Sweet potato	-1.19	-1.30	-0.90	-2.05
	(0.10)	(0.11)	(0.11)	(0.58)
Banana	-1.12	-0.70	-1.34	-2.16
	(0.14)	(0.16)	(0.10)	(0.91)
Rice	-1.59	-1.77	-1.95	-3.00
	(0.33)	(0.39)	(0.29)	(1.86)
Aggregate Bias		0.22	0.26	3.53
(own-price elasticities only)		(0.27)	(0.34)	(3.08)
Aggregate Bias (own- and		1.23	2.07	6.88
cross-price elasticities)		(1.32)	(1.44)	(4.31)
Correlation with elasticities from market prices		0.89	0.88	0.95

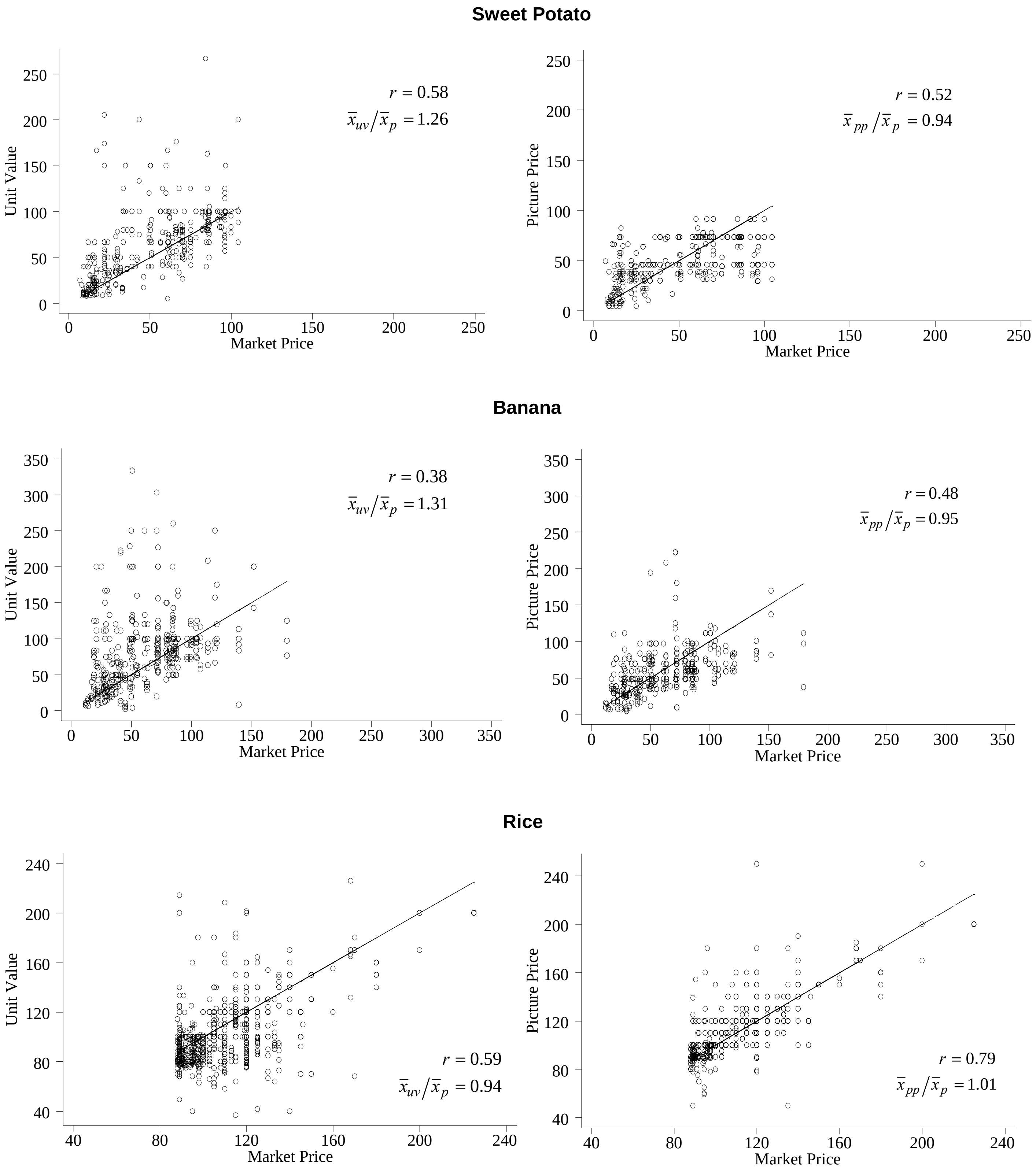
Notes

See Table 4.

[a] 86 clusters, containing 755 households.

Figure 1: Examples of Photographs Used When Eliciting Price Opinions

Figure 2: Comparisons of Market Prices and Household-Specific Unit Values and Picture Prices



Note: Prices are in toea per kilogram (130 toea=US$1 in 1996). The 45° line shows the points where market prices equal unit values (or picture prices).

Figure 3: Regional Food Poverty Lines

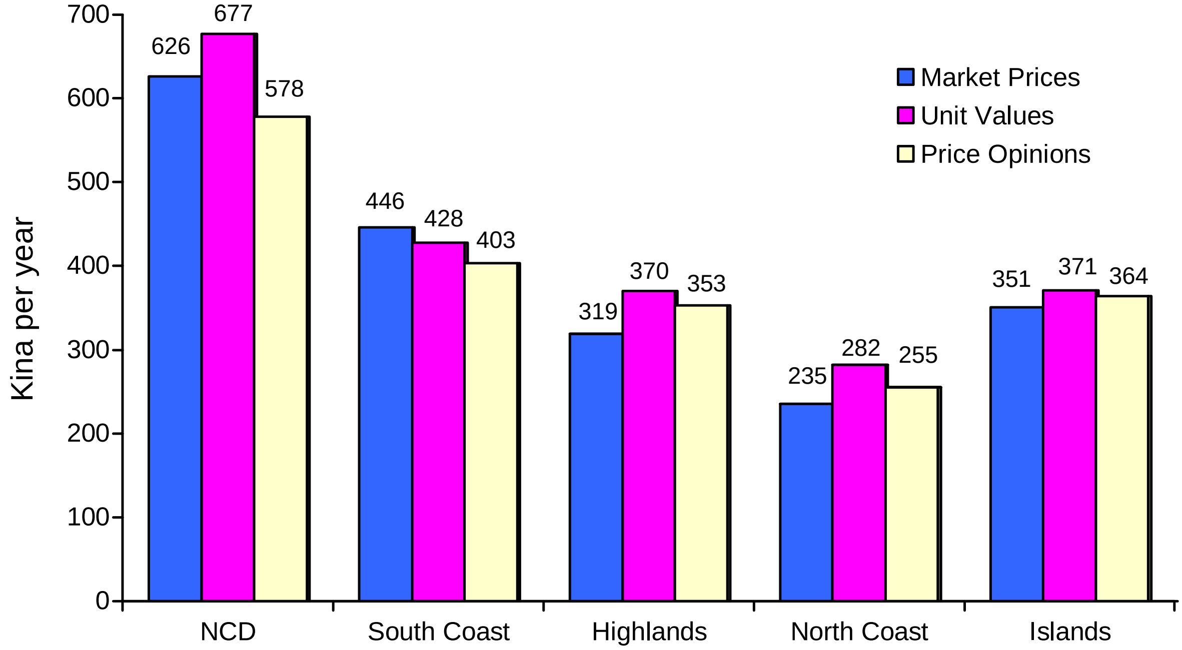
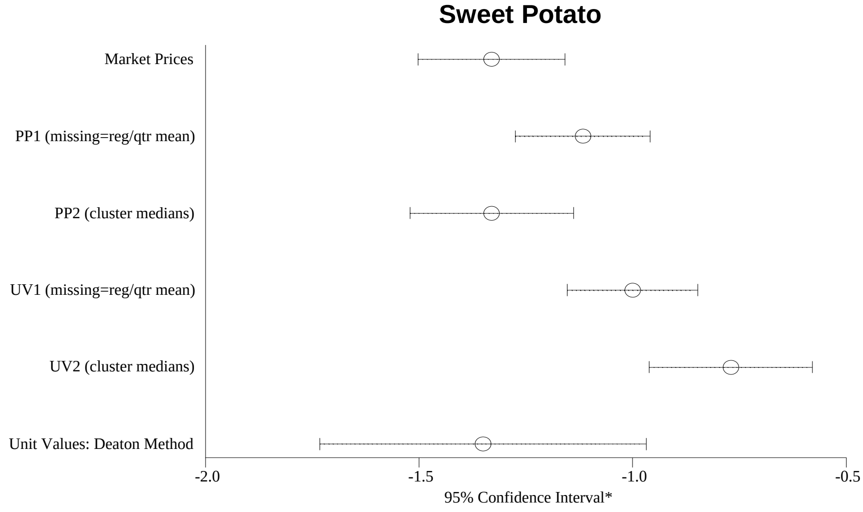
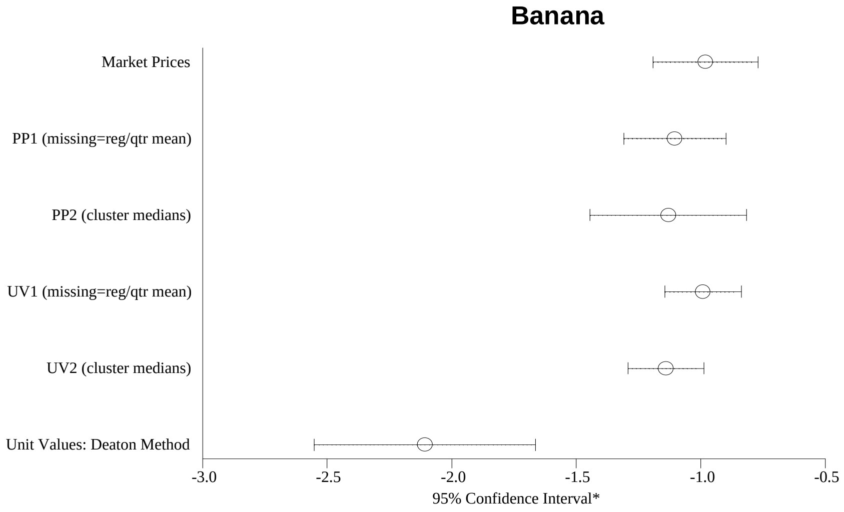
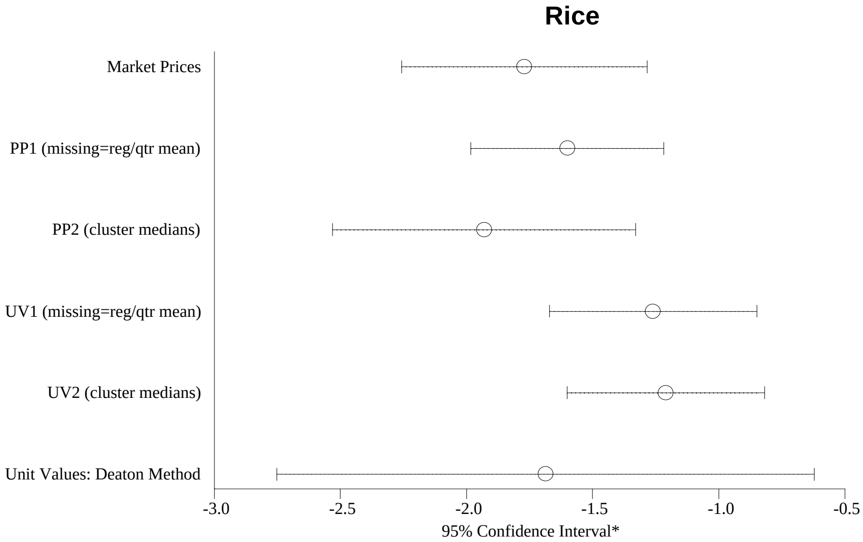


Figure 4: Own-Price Elasticity Comparisons for Market Prices, Picture Prices and Unit Values







Note: *68% Confidence Interval (± 1 std error) for the elasticities from the Deaton method

Notes

[1] Examples include Lao P.D.R. and Pakistan.

[2] In one example, the price of canned tomato paste had to be used as a substitute for all non-food prices (which were poorly measured) in Côte d'Ivoire (Glewwe, 1991).

[3] In some applications it is also possible to substitute assumptions for data. For example, researchers often use additivity assumptions, such as in the linear expenditure system, to get price elasticities from household budget data, without any prices required. But additive preferences imply that expenditure and own-price elasticities are roughly proportional, forcing a tradeoff between equity and efficiency, and leading to recommendations of uniform rates of commodity taxes regardless of the patterns in the data (Deaton, 1997).

[4] This data collection strategy has recently been used in the Indonesia Family Life Survey, with price opinions collected from 'key informants' (the Ibu PKK women's groups). However, comparisons of those prices with the prices collected from market surveys do not seem to be available.

[5] In fact, the main point that Capéau and Dercon discuss in their paper is not the comparison of market prices and unit values, but rather, the issue of how to collect data on crops for which households have difficulty converting from their traditional units of measure (what the farmer knows) to kilograms (what the economist needs). The data collection methods in the PNG survey make this conversion issue less of a problem.

[6] In addition to these short period measures of consumption, the estimate of household's total expenditure used an annual recall of 31 categories of infrequent expenses and an inventory of durable assets, which provides estimates of the flow of annual services from durables and dwellings.

[7] The average PNG household consumes almost 100 kg of root crops every fortnight, so the sacks were filled several times over during the recall period. This should reduce errors due to the relatively coarse graduations used.

[8] This does not mean that people are unaware of size differences, they just use different terminology. For example, canned fish comes in three can sizes and the smallest size (155 gram) is known in the local vernacular as "battery" because its shape resembles that of a D-size battery.

[9] This was fairly straightforward for trade store products because quantities are indicated on the packaging and were visible in the photographs, while quality is easily known from the brand name. But even for fresh produce, the picture conveys quality information. For example, people could tell from the colour and size of the individual tubers, in which region of PNG the sweet potato had been grown.

[10] However, the analysis reported below suggests that there is little quality variation in the goods we use and in the valuations that respondents placed on those goods.

[11] An exception is Minot and Goletti (2000) who estimate a demand system for 14 foods (in the context of a study of trade liberalisation in Vietnam), where unit values are used for seven of the foods and market prices for the other seven. This use in the same demand system implies a direct substitutability between the two types of price data.

[12] Unit values are likely to be collected in many surveys anyway, because of the interest in quantities (for example, for studies of nutrition) so picture prices might reduce problems either by substituting for unit values or complementing them by acting as an instrument.

[13] These correlations should not be seen as either atypically low or reflective of the unusual conditions in PNG. A comparison of market prices and unit values for 33 items in the 1997-98 Vietnam Living Standards Survey (VLSS) yields an average correlation of only 0.25 (Gibson, Rozelle and Le, 2002). Using a more restricted set of foods, and data from the 1992-93 VLSS, Deaton and Grosh (2000) report a median correlation of 0.34. A caveat to both comparisons is that the unit values in the VLSS are meant to refer to the previous 12 months while the market prices

are from the month when the household was actually surveyed.

[14] The correlations with market prices are even lower for the unit values applied to self-produced foods ($\bar{r}$ = 0.35) and for the unit values for gifts received ($\bar{r}$ = 0.36). There is also little agreement amongst the different types of unit values: for those households who both purchased and produced either sweet potato, banana or betelnut, the average correlation between the two types of unit values is only 0.26. For those who both purchased and received gifts, the average correlation is 0.43.

[15] For example, the brand of rice used for the market price survey ("Trukai") accounted for 86 percent of rice sales in PNG in 1996 and most of those sales were for the specified 1 kilogram pack size. (We are grateful to Neville Whitecross of Trukai Industries, Port Moresby, for these details.) The correlation between unit values and market prices is almost the same for the households who report purchasing only one kilogram of rice during the recall period (r=0.61) as it is for other households (r=0.57). Thus, even when the pack size for the unit value report corresponds to that used for the market price survey, there is a low correlation between unit values and market prices, suggesting that reporting errors are important. The intracluster correlation in the rice prices collected from the market survey is 0.82 so variation in the prices charged by different trade stores within each cluster is unlikely to account for the discrepancy with unit values. Moreover, this variation in market prices within a cluster would also affect the calculated reliability of the picture prices, so it cannot account for the relatively poor performance of the unit values.

[16] The average correlation is no higher (r=0.63) if a more broadly defined unit value is formed, based on the ratio of the combined value of purchases, net gifts received and own-production to the combined quantity.

[17] This lack of unit values particularly affects rural areas. For example, a unit value for beer is available for 35 of the 40 clusters in the capital city but in only 28 of the 80 clusters elsewhere. Hence, the spatial distribution of prices may not be measured in a reliable way when unit values are relied on as the proxy for market prices.

[18] However, even without this sample selection issue, there is still bias in the unit values. For example, in the 93 clusters where a unit value for sweet potato is available, the average market price is 46.8 toea per kilogram (slightly above the average across all clusters), which is still 20 percent below the mean unit value for those same clusters.

[19] An analysis of covariance also showed that urban/rural price differentials within regions were less important than inter-regional price variations (World Bank, 1999).

[20] The NCD is an exception, with the average price formed directly from the raw prices rather than from the cluster-level prices. This reflected the assumption that there is less need for the average to reflect the spatial distribution of prices within a city than there is in larger geographical regions (World Bank, 1999).

[21] This is equivalent to US$250 per year, and refers to adult-equivalents rather than per capita.

[22] The overstatement would be even higher, at 17 percent, if the unit values had not been purged of quality effects.

[23] These standard errors correct for weighting, clustering and stratification, using the program of Jolliffe and Semykina, (1999).

[24] The elasticities are not needed for evaluating the welfare effects of *marginal* tax and subsidy reforms. The existing demand structure, and some social weights for aggregating the effects across households, provides sufficient information when price changes are small (Ahmad and Stern, 1984).

[25] All of the other foods and non-foods are aggregated into a composite fourth commodity in the demand system and we assume that leisure is separable from goods demand (an assumption forced by the fact that the survey did not gather data on wage rates).

[26] For both AB1 and AB2 the calculation excludes the results for “other goods” which are simply derived from the other elasticities.

[27] To check that there was not some flaw in the programming, the market prices were passed through the STATA code for the Deaton procedure. The correlation between these elasticities and the market price elasticities reported in Figure 4 and Table 4 was 0.999.

[28] This expression for the cost-benefit ratio of a marginal tax increase is adapted by Deaton (1997) from the more usual one (see, for example, Ahmad and Stern (1984), equation (38)) and allows for both quantity and quality responses to tax-induced price changes.

[29] This finding is sensitive to the value of the inequality aversion parameter used. As ε increases, the equity effects of not taxing sweet potato and banana, which tend to be consumed by the poor, dominate the tax derivative effects and the rankings are not sensitive to differences in the price elasticities. However, attempts to econometrically estimate ε, using the approach of Ravallion and Dearden (1988), suggests that ε is likely to be close to zero in PNG.

[30] The point estimate of -2.3 differs from that in Figure 4 because the elasticities in the figure have controls for region and quarter, while these were not used in the experiment where provinces were treated as clusters.

[31] Thus, it would be most suitable for multi-topic living standards surveys that do not aim to get especially detailed measures of consumption but which do need prices for modeling household behavior.

[32] For example, a 1999 survey in Cambodia tried to obtain prices for 50 food items in 600 villages but data were obtained on less than half of the price-village combinations because of items missing from markets (Gibson, 2000).

[33] The 1993 LSMS survey in Tanzania used willingness-to-pay questions in the parts of the questionnaire on health and education facilities.