

Prices and Unit Values in Poverty Measurement and Tax Reform Analysis

John Gibson and Scott Rozelle

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 economies are either absent or suffer quality problems. However, using unit values may result in biases due to measurement error and quality effects. In a household survey experiment, 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 items and asked to report the local price. The three sets of price data are used to calculate poverty lines, estimate price elasticities, and analyze 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. Performance was better for the price opinions of household members. 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.

Prices are important. Economists need good measures of prices to conduct studies for many applications in developing economies. For example, they need matrices of own- and cross-price elasticities of demand for constructing computable general equilibrium models for trade policy analysis (Minot and Goletti 2000). Effective reform of indirect taxation and subsidy regimes requires accurately estimated price elasticities to predict changes in public expenditure and tax revenues as demand changes following subsidy or tax rates shifts (Ahmad and Stern 1991; Laraki 1989). Elasticities also are needed to account for the welfare effects of economic crises because first-order approximations

John Gibson is a professor in the Department of Economics at the University of Canterbury, New Zealand; his email address is john.gibson@canterbury.ac.nz. Scott Rozelle is a professor in the Department of Agricultural and Resource Economics at the University of California, Davis; his email address is rozelle@primal.ucdavis.edu. The authors are grateful for assistance and helpful comments from Chris Hector, Tim Maloney, Susan Olivia, Berk Ozler, Steven Stillman, three anonymous referees, and seminar audiences at Canterbury University and the Northeast Universities Development Consortium Conference. Data for this study were originally collected as part of a World Bank poverty assessment for Papua New Guinea, for which financial support from the governments of Australia, Japan, and New Zealand is gratefully acknowledged.

THE WORLD BANK ECONOMIC REVIEW, VOL. 19, NO. 1, pp. 69–97

doi:10.1093/wber/lhi002

© The Author 2005. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development/THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oupjournals.org.

that ignore consumer substitution can greatly overstate welfare losses (Friedman and Levinsohn 2002). 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 few studies systematically collect price data, despite their widespread importance. 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 some countries, such as Lao People's Democratic Republic and Pakistan, Poverty Reduction Strategy Papers use poverty estimates based on *assumed* levels of rural prices. Even research-driven surveys suffer from a lack of price data. The Indonesia Family Life Survey 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 9 foods. This incomplete information on prices makes it difficult to reliably measure the inflation rate that Indonesian households faced during the Asian economic crisis of the late 1990s and may contribute to the large discrepancy between the poverty increases implied by the survey price data and the increases implied by the official (urban) inflation rates (Beegle and others 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*, it is recommended that both methods be used. (Frankenberg 2000, p. 329; emphasis added)

Community-level prices of the type collected in most LSMS surveys may be unreliable because they are gathered from the wrong market or for the wrong specification of goods or because 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 (for example, Tajikistan) or analysts have been forced to discard some of the prices.¹

This poor track record for collecting price data may not be surprising. In the rural areas of many developing economies, it is hard for outsiders to find, understand, and study markets. Markets may assemble 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 to the survey (for

1. In Côte d'Ivoire, the price of canned tomato paste had to be used as a substitute for all nonfood prices, which were poorly measured (Glewwe 1991).

collecting prices) with its own complications may cause overall survey quality to decline. The problems are likely to be most apparent in countries with poor infrastructure and low population densities—the very places where price policy can be an important tool for government because of the high per capita administrative cost of 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).² Unit values have recently been used in calculating purchasing power parity exchange rates (Deaton and others 2004), calculating and updating poverty lines (Deaton 2003), assessing household welfare changes from trade liberalization (Nicita 2004b) and economic crises (Friedman and Levinsohn 2002), analyzing indirect tax and subsidy reforms (Deaton and Grimard 1992; Nicita 2004a), and assessing the distributional and nutritional impacts of devaluation (Minot 1998). In some applications, however, such as demand studies, the use of unit values is believed to give biased results (Deaton 1997). In contrast to market prices, unit values reflect household-specific quality and reporting error effects and are subject to sample selection effects because they are unavailable for nonpurchasing 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).

I. HOUSEHOLD SURVEY EXPERIMENT

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

2. 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 using any prices. 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).

The prices from the market price survey are used to assess the two 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). This article explicitly assumes that prices for well-defined items collected from market surveys using certain sampling rules are the appropriate standard.

Although in some cases there may be reasons to worry about the quality of market prices, three features of the case study country used here increase the reliability of the market price surveys. First, villages are small and almost every village visited had a well-defined market. Second, haggling is uncommon in markets in Papua New Guinea. Moreover, several of the selected products are sold only in trade stores and supermarkets, where transactions always take place at the listed prices. Thus 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 Papua New Guinea, which are often branded products with well-defined package sizes. Often a single brand supplies the whole market either because of local monopolies 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 quality variation within markets (as will be shown), so significant variation in quality between markets seems unlikely.

Although the experiment relates to a single country, the findings may be of wider interest. This appears to be the only systematic attempt to test an idea that was proposed early in the development of the LSMS surveys: to obtain price data by interviewing groups of housewives (Saunders and Grootaert 1980).³ This strategy was never implemented, in part because subsequent LSMS reports were critical of this “novel but risky” idea (Wood and Knight 1985). The main concerns were that such price opinions could be biased by differences in bargaining skill, uncertainty about the reference period (which matters in inflationary environments), and the lack of a representative sample. The experiment reported on here overcomes several of these shortcomings. It is based on a representative sample of households each shown a defined specification (a photograph) and asked to report the current price. 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, this is one of only two studies to demonstrate empirically 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 naive or corrected unit value procedures. The literature seems to include only one article that compares poverty estimates with poverty lines priced with unit values and

3. 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 prices collected from market surveys do not seem to be available.

market prices (Capéau and Dercon 1998).⁴ The current study goes further by having three types of prices and by looking at the effect on estimated demand elasticities and marginal tax reform calculations as well.

II. DATA COLLECTION

Data for this study come from the Papua New Guinea Household Survey, which was designed and supervised by the authors in 1995 and 1996, with fieldwork taking place over 12 months. The survey covered a random sample of 1,200 households residing in 73 rural clusters (each providing 12 households to the sample), 40 clusters from the capital city (6 households each), and 7 clusters from smaller urban areas (12 households each).

Market prices were collected in each cluster using two different surveys. The prices of 14 commercially produced food items (such as rice, sugar, and beer) and 9 nonfood items (such as soap and kerosene) were collected from the two main trade stores or supermarkets used by households in the cluster. These prices typically were for a finely defined specification (for example, a 1 kg bag of Trukai brand rice). For four of the foods and one of the nonfood items, the prices covered two different specifications of the same commodity (for example, a bottle of beer and a carton of beer); simple averages of the prices of the two specifications were used. The second market survey collected the prices of 11 locally produced foods from the nearest local market; for 1 food (bananas), prices were 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 took place on two different days in each cluster; potentially, up to 12 observations are available on the price of each food 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 about two weeks later and asked respondents to recall the value and quantity of all purchases, gifts, and own-production since the initial interview. This recall covered 36 categories of food and 20 categories of other frequent expenses.⁵ The unit values are calculated as the ratio of purchase values to purchase quantities.

The data-collection methods affect the unit values in the survey in two important ways. First, the unit values are for the same period as the market price survey. In contrast, in some LSMS surveys (for example, in Vietnam) the

4. In fact, the main point that Capéau and Dercon discuss is not the comparison of market prices and unit values but rather 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 Papua New Guinea survey make this conversion issue less of a problem.

5. 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.

unit values cover a 12-month period, which would weaken any comparison with current market prices. Second, to reduce problems stemming from a failure to understand metric quantities, all households received during the first interview an empty 25 kg sack, with graduations of 1/4, 1/2, and 3/4 marked on the outside to use in recording food volumes. This unit was recommended for bulky root crop staples and was used in more than 90 percent of cases.⁶ The other main unit used was a simple count of the number of items, recommended for items like coconuts and betelnuts and for livestock. Average volume to weight (and count to weight) conversion factors were established from weighing trials conducted in all regions of the country. Although crude compared with the ideal of weighing all items consumed, these procedures avoid the problem of enumerators and respondents using idiosyncratic conversion factors and so reduce the relevance of the Capéau and Dercon (1998) procedure.

The picture method data come from price opinions gathered from each household for 15 food items (including beverages) and 3 tobacco products. Because six of the food items were alternate specifications of a particular food (for example, a bottle and a can of soft drink), the pictures refer to nine categories of food. On average, these nine foods constitute 30 percent of the household's total consumption expenditure, with individual budget shares ranging from 11 percent (sweet potato) to 1 percent (flour, biscuits, and soft drinks).

Central to the enumeration process, respondents were shown a series of 18 high-quality color photographs taken by professionals. The photos showed each food item in the typical bundle, pile, or package found in markets. For foods where scale was important, a box of matches was included in the photograph (see examples in figure 1 for the four items with the largest budget shares—sweet potato, banana, betelnut, and rice).⁷ Interviewers were instructed to ask the following question when showing the photographs, which was done at the conclusion of the second visit: "How much does it currently cost to buy a [item] like this in the main market or store in this village or 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 photographs, and reported prices were transformed to kilogram prices at the analysis stage, using the actual weights of the items in the photographs.⁸ Respondents were expected to map a two-dimensional picture into volumes and

6. The average Papua New Guinea household consumes almost 100 kg of root crops every two weeks, so the sacks were filled several times during the recall period. This should reduce errors due to the relatively coarse graduations used.

7. Full color versions of the pictures in figure 1 can be viewed on the publisher's Web site.

8. No attempt was made to force respondents to report a price in kilograms, which are not widely used in markets in Papua New Guinea. This does not mean that people are unaware of size differences—they just use different terminology. For example, canned fish comes in three sizes and the smallest size (155 g) is known in the local vernacular as "battery" because its shape resembles that of a D-size battery.

FIGURE 1. Examples of Photographs Used for Eliciting Price Opinions



Source: Papua New Guinea Household Survey 1995/96.

weights and to form an assessment of quality based on the photograph.⁹ Pretesting showed that respondents were good at this: Reported prices based on pictures were close to the prices reported when respondents were shown the actual items instead of the picture. Actual products were not used in the main survey because the interview teams would be burdened by carrying bulky products and the same product could not be used simultaneously in different survey locations, so quality variations could be introduced into the price opinions.¹⁰

III. UNIT VALUES, PRICES, AND PRICE OPINIONS

The data-collection effort provided three different measures of price (market prices, price opinions, and unit values) for nine foods (sweet potatoes, bananas, rice, betelnuts, flour, biscuits, canned fish, soft drinks, and beer). With surveys that

9. This was fairly straightforward for trade store products because quantities are indicated on the packaging and were visible in the photographs, and 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 color and size of the individual tubers in which region of Papua New Guinea the sweet potato had been grown.

10. However, analysis suggests that there is little quality variation in the goods used and in the valuations that respondents placed on those goods.

have just one measure of price, analysts are often forced to use unit values even though they are not a direct substitute for genuine price data.¹¹ The survey data were used to answer two questions: Are the problems in using unit values as a measure of price large enough to justify the expense of collecting additional information on prices? If this additional information is needed, do price opinions 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 the photo-guided price opinions. Finally, if additional information on prices is needed but price opinions perform poorly, greater investment may be needed in properly carrying out community price surveys.

This section reports some simple descriptive analyses that may help answer these two questions. To guard against outliers affecting the analysis, the survey forms were reexamined and data entry errors and obvious miscoding (such as kg entered as g) were rectified or removed. Following the rule of Cox and Wohlgemant (1986), unit values and price opinions more than five standard deviations from their respective means were also removed to further reduce outlier effects. This procedure removed 23 unit values (of 4,550) and 25 price opinions (of 9,100), a proportionately greater trimming of the unit values.

Even after outliers were trimmed, 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 still higher than those for the six other food commodities ($\bar{r}=0.37$; results not shown).¹⁴ Unit values also appear

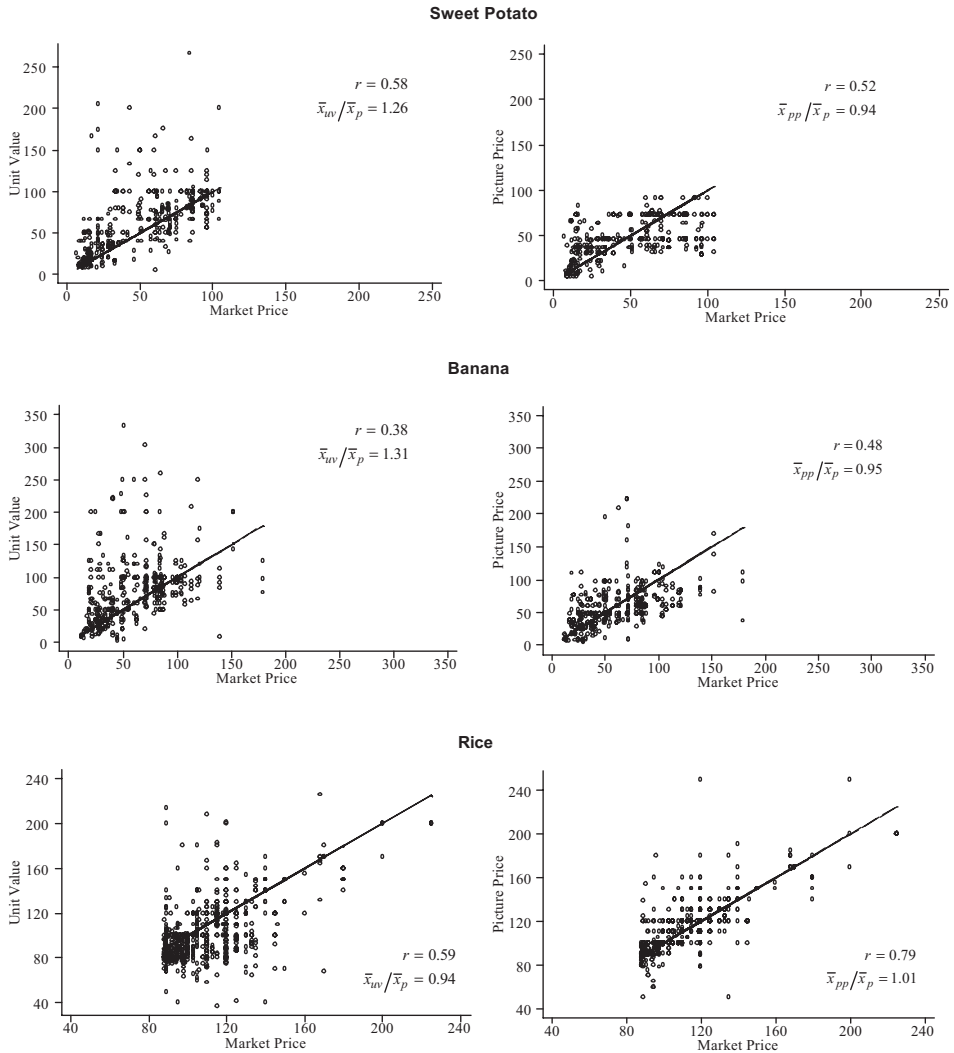
11. An exception is Minot and Goletti (2000), who estimate a demand system for 14 foods (in the context of a study of trade liberalization in Vietnam), where unit values are used for 7 of the foods and market prices for the other 7. This use in the same demand system implies a direct substitutability 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 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 as reflective of the unusual conditions in Papua New Guinea. 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 and others 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, whereas 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 among the different types of unit values: For households that both purchased and produced either sweet potatoes, bananas, or betelnut, the average correlation between the two types of unit values is only 0.26. For those that both purchased and received gifts, the average correlation is 0.43.

FIGURE 2. Comparisons of Market Prices and Household-Specific Unit Values and Price Opinions



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).

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96.

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 potatoes and bananas, the two most common locally produced foods.

Photo-guided price opinions appear to provide a better measure of market prices. For the same households as for the unit value analysis, the scatter plots of market prices and price opinions 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 1, ranging from 0.94 to 1.01 (see figure 2). The correlations with market prices range from 0.48 to 0.79 for the three major foods. The average correlation coefficient between price opinions 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 price opinions and especially unit values may be imperfect measures of market prices. Both may contain quality effects, although these tend to be small in the data, particularly for the price opinions (see later discussion). The specification of items may differ for 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 such a discrepancy between the specifications for the unit value and the market price surveys contributes to the low correlation.¹⁵ Finally, both price opinions and unit values are subject to reporting error, and it could be that the errors are greater for unit values.

If all three series (market prices, unit values, and price opinions) are treated as error-ridden measures of true but unknown community prices, the intraclass correlations among each measure can provide an estimate of the “reliability ratio”—the proportion of measurement error in the variance of the observed price series. The intraclass correlation is systematically lower for unit values (after the effects of quality have been purged) than for market prices and price opinions. The average value of the correlation coefficients across the nine foods is only 0.38 for unit values, compared with 0.78 for market prices and 0.65 for price opinions. By this analysis then, unit values are the least reliably measured, although there is imperfect reliability for all the price measures.

Figure 2 suggests that it is possible that a few households disproportionately generate much of the reporting error bias in both price opinions and unit values. The use of cluster averages can reduce this source of bias and indeed results in improved correlation between unit values and market prices, although the unit values still tend to be noisier measures than the price opinions (table 1, columns 6 and 7). The average correlation of cluster-level unit values and market prices is 0.63, and the average correlation for price opinions is 0.77.¹⁶

15. For example, the brand of rice used for the market price survey (Trukai) accounted for 86 percent of rice sales in Papua New Guinea in 1996, and most of those sales were for the specified 1 kg pack size, according to Neville Whitecross of Trukai Industries, Port Moresby. The correlation between unit values and market prices is almost the same for households that report purchasing only 1 kg 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 corresponds to that of the market price survey, there is a low correlation between unit values and market prices, suggesting that reporting errors are important. The intraclass 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. Moreover, this variation in market prices within a cluster would also affect the calculated reliability of the price opinions, 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.

TABLE 1. Descriptive Statistics for Cluster-Level Market Prices, Unit Values, and Price Opinions

Product	Mean Market Price ^a	Mean Unit Value ^a	Mean Price Opinion ^a	No. Clusters with Data On ^b		Correlation with Market Prices	
				Unit Values	Price Opinions	Unit Values	Price Opinions
Sweet potatoes	43.9	59.0	42.5	93	118	0.74	0.74
Bananas	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

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96.

^aToea per kg, as calculated from cluster-level averages; 130 toea = US\$1 in 1996.

^bOf a possible 120 clusters.

Averaging by cluster, however, does not remove the bias that occurs when unit values are used to calculate average market prices (see table 1). On average, the mean price for each food and the mean of the cluster-level unit values for the same food differ by 14 percent (this is calculated for each food as: $|\bar{x}_{uv} - \bar{x}_p|/\bar{x}_p$). In contrast, the average error is only 6 percent for the price opinions. 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, unit values exhibit a further statistical problem that becomes apparent when the cluster means are formed. A cluster mean unit value is available only when at least one household in a cluster makes a purchase during the recall period. When no households make such a purchase, a sample selection problem occurs. This can be a serious problem for some commodities. For example, in the sample there are only 63 clusters with an average unit value for beer and 92 clusters with one for bananas rather than the expected sample of 120 clusters.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 price opinions are much more widely available. The most for any food was four clusters with missing price opinions for all households. Thus, the method of obtaining opinions about prices rather than just relying on purchase behavior can potentially capture the full range of spatial price variation in a sample.

IV. THE EFFECTS OF THE ALTERNATIVE PRICE COLLECTION METHODS

This section measures the impact of using the alternative prices series as proxies for market prices. First, it examines how using unit values compares with using photo-guided price opinions in estimating the poverty line and various aggregate measures of poverty. Next, the same comparison is made for price elasticity estimates, and implications are drawn for tax policy analysis.

Effects on Poverty Measures

Poverty lines for Papua New Guinea are based on the market prices collected by the survey (World Bank 1999). Specifically, the cost of buying a basket of food that provides 2,200 calories a day was calculated for five regions: the National

17. This lack of unit values affects rural areas particularly. 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 used 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 potatoes is available, the average market price is 46.8 toea per kg (slightly above the average across all clusters), which is still 20 percent below the mean unit value for those clusters.

Capital District, 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 National Capital District.¹⁹ The regional average prices used to calculate the cost of the poverty line basket of foods were calculated from the cluster-level averages of the market prices (see table 1).²⁰

This section follows the same procedures to calculate the food poverty line, substituting the unit values and price opinions in place of the market prices. The unit values and price opinions are first averaged by cluster before the regional averages are calculated. This ensures that clusters with more purchasing households do not receive undue weight in the calculations. The cluster averages also tend to dampen 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 for the characteristics included).

One constraint with these exercises is that the poverty line food basket contains 35 foods, but there are only 9 foods with data from both price opinions and unit values. Although these foods contribute almost half the value of the poverty line food basket, the experiments are effectively varying only half of the value of the food poverty line. Thus 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 market price, unit value, and price opinion data are illustrated in figure 3. When market prices are used, the food poverty line ranges from 235 kina (K) a year in the North Coast region to K626 in the National Capital District, with a population-weighted average of K330.²¹ Although the existing poverty lines for Papua New Guinea include a nonfood allowance, which is equivalent to between a third and a half of the value of the food poverty line, that is ignored here because price information was gathered only for foods.

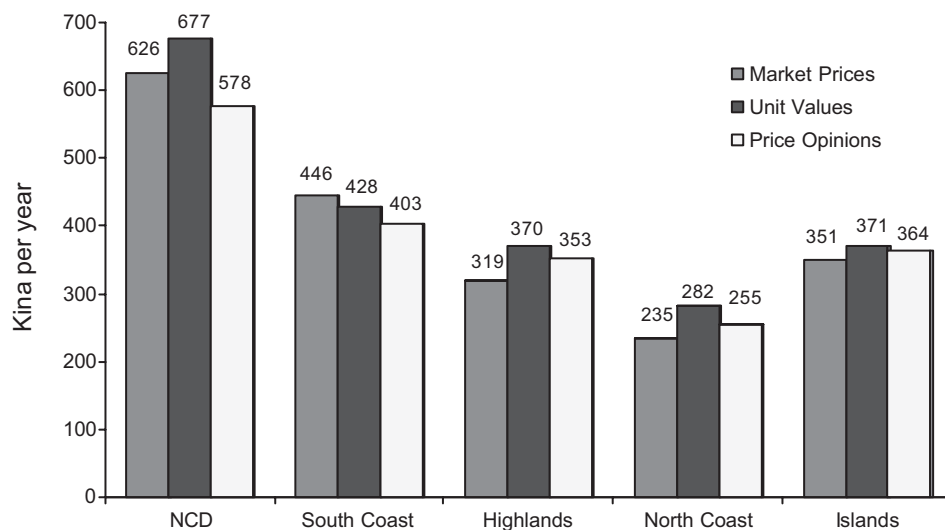
The food poverty line is consistently overstated when unit values are used as the measure of price (see figure 3). In the National Capital District, South Coast, and Islands regions unit values overstate the poverty line by a slight margin of 4–10 percent. However, in the other two regions, which contain 70 percent of the population, unit value-based analysis overstates the food poverty line by 16–20 percent. In contrast, the use of photo-guided price opinions creates a smaller bias. The use of price opinions causes the food poverty line to be understated by

19. An analysis of covariance also showed that urban–rural price differentials within regions were less important than interregional price variations (World Bank 1999).

20. The National Capital District is an exception, with the average price formed directly from the raw prices rather than from the cluster-level prices. This reflects 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.

FIGURE 3. Regional Food Poverty Lines



Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96.

about 10 percent in the National Capital District and South Coast and to be overstated by 4–11 percent in the other three districts. On average the food poverty line has a proportionate error, $|\bar{z}_i - \bar{z}_p|/\bar{z}_p$ (where z is the food poverty line, p is market prices, and i is unit values or price opinions), of 14 percent with the unit values and 9 percent with price opinions.²²

When data-collection methods create biased estimates of the poverty line, they also affect measures of poverty rates (table 2). The overstatement of the food poverty line when unit values are used causes an upward bias in poverty measures. Thus, for example, the headcount index is estimated as 28 percent (with a standard error of 2.6 percent) rather than the 22 percent based on market prices,²³ and the poverty gap index is estimated as 8.0 percent rather than as 5.9 percent. The differences between unit value estimates and those based on market prices are statistically significant (the t -statistics for the null hypothesis of no difference range from 4.8 to 6.8). This finding that using unit values results in higher poverty measures is consistent with Capéau and Dercon (1998), who conclude that headcount poverty in rural Ethiopia would be overstated by one-fifth if unit values were used instead of other price data.

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).

TABLE 2. Aggregate Food Poverty Measures for Papua New Guinea, 1996

Poverty Line Food Basket Calculated From	Headcount Index	Poverty Gap Index	Poverty Severity Index
Market prices	22.0 (2.4)	5.9 (0.9)	2.4 (0.4)
Unit values	28.0 (2.6)	8.0 (1.0)	3.4 (0.6)
Price opinions	23.8 (2.5)	6.8 (1.0)	2.8 (0.5)

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96.

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. Numbers in parentheses are SEs corrected for the effect of clustering, sampling weights, and stratification.

There is also an upward bias associated with the use of price opinions, but the discrepancy is significantly smaller (see table 2). Estimates based on price opinions overstate the headcount poverty measure by only 8 percent (the *t*-statistic for the null hypothesis of no difference is 2.2). This overstatement is significantly less than when unit values are used (the *t*-statistic is 4.5 for the test that the overstatement is the same for unit values and price opinions). Clearly, the price opinions provide more accurate measures of poverty in Papua New Guinea, although even the smaller overstatement may be enough to justify the expense of collecting better price data from local stores and markets.

Effects on Price Elasticity Estimates and Indirect Tax Analysis

In developing economies, pricing policy plays the same central role in fiscal policy that income tax and social security policies play in industrial countries (Deaton 1989). The matrix of price elasticities needed to estimate the revenue effects of price reforms can therefore provide fundamental information to governments.²⁴ That makes it important to establish what bias might occur when elasticities are calculated from either unit values or price opinions if estimates from market price surveys are not available.

Attention is focused here on the three major staples—sweet potatoes, bananas, and rice²⁵—which account for more than one-fifth of household consumption expenditures and supply about 45 percent of calories to households. These three foods have some policy significance as well as consumption and nutritional importance, 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),

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, provide sufficient information when price changes are small (Ahmad and Stern 1984).

25. All of the other foods and nonfoods are aggregated into a composite fourth commodity in the demand system, and leisure is assumed to be separable from goods demand (an assumption forced by the fact that the survey did not gather data on wage rates).

rice is now taxed at the same 10 percent rate as other imported goods. Sweet potatoes and bananas effectively fall outside of the tax net because the farmers and traders who sell them in informal markets are not registered for the VAT.

Eleven clusters have no market price survey data for either sweet potatoes or bananas, so the demand system is estimated on the remaining 109 clusters (containing 1,018 households). This reduced sample highlights one advantage of price opinions—there would be only two clusters with missing data if only the price opinions were used. Of the 109 clusters, only 86 have at least one household purchasing either sweet potatoes or bananas (the total number of purchasing households is around 350). Thus, imputed unit values must be used for the other clusters.

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

$$(1) \quad w_i = \alpha_i + \beta_i \ln x + \sum \theta_{ij} \ln p_j + \gamma' \mathbf{z} + \mathbf{u}_i,$$

where w_i is the share of the budget devoted to good i , x is total expenditure, p_j is 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 old, 7–14 years old, and 15–50 years old; women over 50 years old), dummy variables for whether the household head is 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 nonzero consumption in the same way. 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). Thus the literature on censored demand systems is not needed here. The price elasticities for equation 1 are given by:

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

where δ_{ij} is the Kronecker delta (equal to 1 if $i=j$ or to 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 in the literature concerns how to deal with the missing unit values and whether to leave unit values at the household level or aggregate them to the cluster level. Two methods are used here:

- UV1, which uses 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, which uses 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.

These same two methods are also applied to the photo-guided price opinions, denoted as PP1 and PP2.

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

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

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

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, the i indexes households, the G and H index goods, and the c indexes clusters.

Deaton's method recognizes that the data are collected on clusters of households that are presumed to face the same market prices. The variation in budget shares and unit values within clusters is used to identify the effect of income and other household characteristics on 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. To compare the quality effects and measurement error properties of unit values and price opinions, 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 price opinions. These small values are consistent with the evidence from Deaton (1990) and Gibson and others (2002) that the quality problems with unit values are less important than the measurement error problems. Moreover, although the quality elasticities are small, they are larger for unit values than for price opinions. On average, the absolute value of the quality elasticities is almost four times as large for unit values as for the price opinions.

The unit values also have higher measurement error variance (variability around the cluster means of $\ln v_{Gc}$ that is not explained by household characteristics in equation 4) than the price opinions for all nine foods. On average, the measurement error was 4 times greater (almost 10 times for soft drink and

TABLE 3. Quality and Measurement Error Indicators for Unit Values and Price Opinions from the First-Stage Regressions of the Deaton Procedure

Product	Quality Elasticity ^a		Residual Variance ^b		Residual Covariance ^c	
	Unit Values	Price Opinions	Unit Values	Price Opinions	Unit Values	Price Opinions
Sweet potatoes	-0.016 (0.039)	-0.040 (0.027)	0.152	0.151	-0.042	0.466
Bananas	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

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96. and the procedures developed in Deaton (1990).

Note: Numbers in parentheses are standard errors. * Significant at the 10 percent level.

^aThe coefficient β^1_G in equation 4.

^bCalculated from e^1_{Gic} in equation 4.

^cFrom equations 3 and 4, $\times 1,000$.

biscuits) for the unit values as for the price opinions. 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 price opinions for seven of the nine foods. On average, the covariance in the errors was almost 10 times greater for the unit values, suggesting that the errors in the price opinions are less correlated with actual demands than are the errors in the unit values.

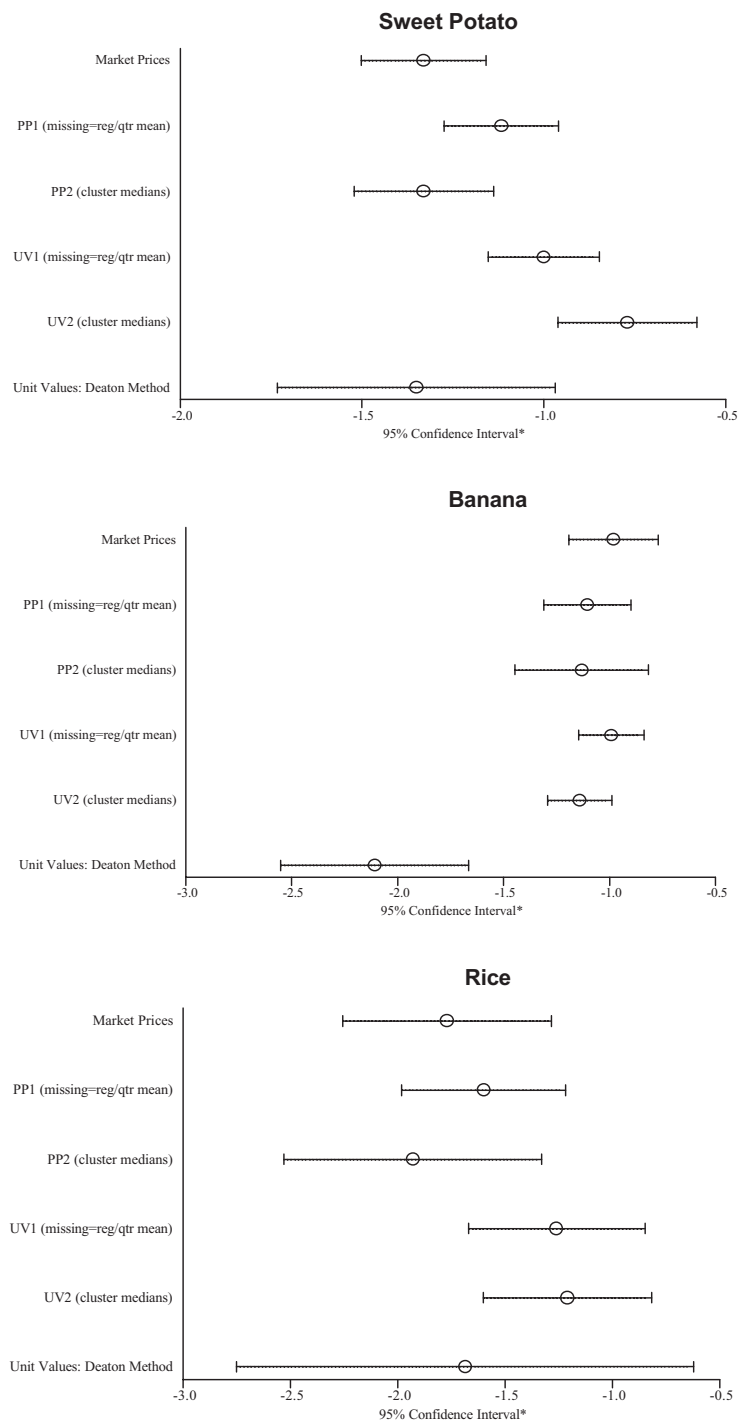
In the second stage of the Deaton procedure, a between-clusters errors-in-variables regression is applied to the (adjusted) average budget shares and unit values, which have been purged of household characteristics at the first stage. If it were not for the effect of prices on clusterwide quality variation, 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 elasticity estimates based on market prices, both price opinions series (PP1 and PP2) create the estimates with the least bias (figure 4). The point estimates of the elasticities estimated from photo-guided price opinions (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 (see figure 4). For example, in 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, although the Deaton procedure calculates point estimates of the own-price elasticities for sweet potatoes and rice that are relatively consistent with the estimates from market prices, it does a poor job of estimating the own-price elasticity for bananas (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 the sample used here does not have many clusters.

Estimates of cross-price elasticities, also important in indirect taxation analysis, are likewise adversely affected 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 ϵ be the vector of elasticities calculated from the market price data and $\hat{\epsilon}$ the corresponding elasticity vector from unit values or price opinions, so that the bias is $\hat{\epsilon} - \epsilon$, and $AB = \hat{\epsilon} - \epsilon / (\hat{\epsilon} - \epsilon)$, 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

FIGURE 4. Own-Price Elasticity Comparisons for Market Prices, Price Opinions, and Unit Values



*68 percent confidence interval (± 1 STD error) for the elasticities from the Deaton method.
Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96.

own- and cross-price elasticities (AB2). For both AB1 and AB2 the calculation excludes the results for “other goods,” which are simply derived from the other elasticities. With the exception of the Deaton method, where bootstrapping is used, standard errors for AB1 and AB2 are obtained from the delta method.

The aggregate bias in the own-price elasticities is lowest (AB1 = 0.048) when the estimation uses cluster medians of the price opinions (table 4, column 1). When the cross-price elasticities are included in the aggregate bias calculation (AB2), household-specific price opinions perform best (AB2 = 0.904, see table 4, column 2). It is notable that the bias estimates for both procedures using price opinions are less than 35 percent of those for the similar procedure using unit values. Moreover, although neither AB1 nor AB2 is statistically significant when price opinions are used, AB2 is statistically significant (at $p < 0.03$ or smaller) for all three of the unit value procedures. Similarly, the correlation of the elasticities from price opinions (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, see table 4, 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.²⁶

The bias in the elasticities calculated from naive unit value procedures could affect public policy decisions. An obvious use of the price elasticities is in deciding on the direction of marginal tax reform (Deaton and Grimard 1992). Social cost-benefit ratios, λ_i , of a marginal increase in tax on each of the three foods are estimated from:

$$(5) \quad \lambda_i = (w_i^\epsilon / \tilde{w}_i) / (1 + [\tau_i / (1 + \tau_i)] [\theta_{ij} / \tilde{w}_i - 1] + \sum_{k \neq i} [\tau_k / (1 + \tau_k)] [\theta_{ki} / \tilde{w}_i]),$$

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:

$$(6a) \quad w_i^\epsilon = [\sum_{m=1}^M (x_m / n_m)^{-\epsilon} x_m w_{im}] / \sum_{m=1}^M x_m$$

$$(6b) \quad \tilde{w}_i = \sum_{m=1}^M x_m w_{im} / \sum_{m=1}^M x_m$$

where x_m and n_m are the total expenditure and size of household m , and ϵ is the coefficient of inequality aversion.²⁷ 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 ($\lambda = 1.47 \pm 0.01$), followed by a tax on rice

26. To verify that there was no 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.

27. 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.

TABLE 4. Summary Comparisons of Estimates Using Market Prices, Price Opinions, and Unit Values

Data Source and Estimation Method ^a	Cost-Benefit Ratio of Tax Rise For ^e				
	AB1 ^b	AB2 ^c	Corr ^d	Sweet Potatoes	Bananas
Market prices					
PP1 (missing = reg/qtr mean)	0.089 (0.133)	0.904 (0.503)	0.958	1.47 [3] (0.01)	1.39 [1] (0.02)
PP2 (cluster medians)	0.048 (0.147)	1.448 (0.874)	0.938	1.46 [3] (0.01)	1.40 [1] (0.01)
UV1 (missing = reg/qtr mean)	0.369 (0.356)	3.323 (1.444)	0.804	1.45 [2] (0.01)	1.40 [1] (0.02)
UV2 (cluster medians)	0.653 (0.408)	4.844 (1.553)	0.669	1.49 [3] (0.01)	1.40 [2] (0.02)
Unit values, Deaton method	1.415 (0.943)	7.775 (3.582)	0.737	1.48 [3] (0.01)	1.42 [2] (0.02)
				1.53 [3] (0.04)	1.34 [1] (0.06)

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96. and the procedures developed in Deaton (1990).

Note: Numbers in parentheses are SEs 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). Numbers in brackets are the good's rank in terms of the cost-benefit ratio, λ_i , where 1 denotes the good with the lowest cost-benefit ratio from a marginal tax increase.

^aPP refers to "photo-guided price opinions" and UV to "unit values."

^bAggregate bias on the own-price elasticities.

^cAggregate bias on own- and cross-price elasticities.

^dCorrelation between the elements of the elasticity matrix and the market price elasticities. Calculations exclude the elasticities for "other goods" derived from the adding-up and homogeneity restrictions.

^eCalculated from equation 5, using an inequality aversion parameter, $\epsilon = 0.5$.

($\lambda = 1.44 \pm 0.05$), while banana looks like the best candidate for a tax increase ($\lambda = 1.39 \pm 0.02$) (see table 4). But this ranking is preserved by only two of the other estimation methods: price opinions with missing values replaced by regional and quarterly means (PP1) and the Deaton procedure applied to unit values.²⁸ 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 policymakers in Papua New Guinea to increase a tax that is not the socially least-cost source of revenue.

Some 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. Although this is an intrinsic disadvantage of unit value methods, unit values may be more widely available in some settings either because households are more reliant on purchased food or because the consumption recall period is longer. The performance of the cluster-median and Deaton estimators is explored for the subsample of 86 clusters that have unit values available for all three foods (table 5). This change in sample coverage does 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 as for those using price opinions (but the difference is no longer statistically significant). The Deaton method also appears to do better on this subsample, with the aggregate bias now statistically insignificant and with a higher correlation with 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 Papua New Guinea.

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 (2004a, b) 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. Intraregional variation in unit values because of 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 villages where

28. This finding is sensitive to the value of the inequality aversion parameter used. As ϵ increases, the equity effects of not taxing sweet potatoes and bananas, 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 Papua New Guinea.

TABLE 5. Results for the Subsample with Each Cluster Having a Unit Value Available

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

Source: Authors' computations based on data from Papua New Guinea Household Survey 1995/96, and the procedures developed in Deaton (1990).

Note: Total of 86 clusters, containing 755 households. See table 4 for details on aggregate bias and correlation with elasticities from market prices.

prices are high, so demand differences will be explained by attenuated price differences, usually causing elasticities to be overstated.

To see the impact of aggregating clusters, the Deaton method was rerun with each of the 19 provinces in Papua New Guinea treated as a cluster. For both sweet potatoes 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.²⁹ It is not surprising that the elasticity for rice is affected most, because an analysis of variance 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 price opinions instead of unit values worth the additional effort? In Papua New Guinea, price opinions were collected

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

with the aid of a picture in a fairly efficient, timely way. The enumerators collected price opinions for 18 products at the same time that they conducted the rest of the household survey. No additional logistical effort was required other than reminding enumerators to bring their picture albums with them. On average the typical household spent only about 10 minutes on this block of the survey. Because each cluster included 12 households, the survey team spent about two hours per cluster collecting price opinions. Moreover, enumerators and respondents liked this part of the survey because it provided a break from the normal questioning.

In the Papua New Guinea survey more effort was needed to survey local stores and markets than to obtain the price opinions. The typical community price survey required visits of 15 minutes or less to each of two trade stores. The survey in fresh produce markets, however, was more time consuming, typically taking the enumerator 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 getting to the market, which was commonly near the community school. In some cases, the nearest market was more than 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, many communities had prohibitions on selling betelnuts in the main market and selling beer in trade stores. 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 to collect market price data was about four hours per community or twice the time spent on the price opinions.

In some areas the markets would assemble infrequently (typically one day a week) or start at daybreak and last for only one or two hours. When the market day was missed, survey teams had to spend time and resources to leave a member in the village until the market convened again or send back a team at a later date. None of these timing problems affected the collection of price opinions.

The experience in Papua New Guinea suggests that with a time commitment of 15–20 minutes per household or 3–4 hours per cluster, it would be feasible to collect price opinions for 30–40 items. Thus, this method would be most suitable for multitopic living standards surveys that do not aim to get especially detailed measures of consumption but that do need prices for modeling household behavior. As more prices are required, market price surveys become more attractive because the fixed cost of finding and getting to the market can be spread over the larger number of items whose prices are surveyed. However, the success of market price surveys depends on the ability to find every item in a rural marketplace; more detailed surveys might run into the problem of a large number of missing prices.³⁰ Thus, some consideration of how markets operate

30. For example, a 1999 survey in Cambodia sought 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).

may be needed when choosing whether to rely on price opinions or market price surveys.

An increase in the detail of a survey can also 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 Papua New Guinea 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 Papua New Guinea sample offered opinions on the price of flour, even though only 30 percent purchased it during the recall period. This does raise questions about the reliability of the price opinions of households that do not purchase a particular good. In the Papua New Guinea 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 by households that purchased each item and households that did not ($\bar{r}=0.62$ for purchasers and 0.56 for nonpurchasers). Nonpurchasers may be relatively well informed about the prices of goods they do not consume because they still observe those prices in stores and the market when they are shopping for other goods. Consistent with this explanation, the item with the largest discrepancy was beer ($r=0.89$ for purchasers and 0.75 for nonpurchasers), which is usually sold in less commonly frequented hotels, clubs, and licensed outlets. Thus, the usefulness of price opinions may also depend on how segmented markets are, which affects the ease of observing prices for items that the household does not usually consume.

VI. CONCLUSION

Cross-sectional household survey data of the kind examined here are increasingly being used as economists try to exploit one of the few data sources in developing areas that can help provide estimates of the demand responses needed for evaluating tax and subsidy reforms. The findings suggest that unit values, whether used in naive or improved estimation procedures, lead to biased estimates of poverty rates and biased estimates of price elasticities. In contrast, 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 range of households than unit values can, the errors in the estimates are unlikely to be correlated with demand, and the price opinions should have less quality variation because everyone sees the same picture. It may thus 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 depends 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 economies, for a variety of reasons, the logistics of collecting market prices appear to be so difficult that many surveys do not attempt it, and of those that do, some end up rejecting the data. Consequently, many important analyses of poverty and price and tax policies rely on very imperfect price information.

The findings here should also provide an incentive for others to experiment with methods of gathering price data in rural areas of developing areas. For example, three-dimensional models might be used instead of pictures. A broader experiment could gather price data using price opinions from an informed respondent without an aid (as was done in the Indonesia Family Life Survey), using pictures, and using three-dimensional models to elicit price opinions. It would also be interesting to learn whether certain types of respondents have more informed opinions than other household members. Such comparisons are precluded by the design of this study, 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 provide a way to ask questions about hypothetical prices. Households could be asked how some hypothetical price changes would affect their demand for a pictured item. It would be an interesting experiment to find out how well the direct approach approximated the econometric estimates of price elasticities of demand. Such willingness-to-pay questions could be applied to more than food, with medicines and other health interventions being plausible candidates.³¹

REFERENCES

- Ahmad, E., and N. Stern. 1984. "The Theory of Reform and Indian Indirect Taxes." *Journal of Public Economics* 25(3):259–98.
- . 1991. *The Theory and Practice of Tax Reform in Developing Countries*. Cambridge: 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., E. Frankenberg, and D. Thomas. 1999. "Measuring Change in Indonesia." Labour and Population Program Working Paper 99-07. RAND, Santa Monica, Calif.
- Capéau, B., and S. Dercon. 1998. "Prices, Local Measurement Units and Subsistence Consumption in Rural Surveys: An Econometric Approach with an Application to Ethiopia." Working Paper 98-10. Oxford University, Centre for the Study of African Economies, Oxford.
- Cox, T., and M. Wohlgenant. 1986. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68(4):908–19.

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

- Deaton, A. 1988. "Quality, Quantity, and Spatial Variation of Price." *American Economic Review* 78(3):418–30.
- . 1989. "Household Survey Data and Pricing Policies in Developing Countries." *World Bank Economic Review* 3(2):183–210.
- . 1990. "Price Elasticities from Survey Data: Extensions and Indonesian Results." *Journal of Econometrics* 44(3):281–309.
- . 1997. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore, Md.: Johns Hopkins University Press.
- . 2003. "Prices and Poverty in India, 1987–2000." *Economic and Political Weekly* 25(January): 362–68.
- Deaton, A., and F. Grimard. 1992. "Demand Analysis for Tax Reform in Pakistan." LSMS Working Paper 85. World Bank, Washington D.C.
- Deaton, A., and M. Grosh. 2000. "Consumption." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries*. Washington, D.C.: World Bank.
- Deaton, A., J. Friedman, and V. Alatas. 2004. "Purchasing Power Parity Exchange Rates from Household Survey Data: India and Indonesia." Princeton University, Princeton, N.J.
- Frankenberg, E. 2000. "Community and Price Data." In M. Grosh and P. Glewwe, eds., *Designing Household Survey Questionnaires for Developing Countries*. Washington, D.C.: World Bank.
- Friedman, J., and J. Levinsohn. 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 Cambodian Ministry of Planning, Phnom Penh.
- Gibson, J., and S. Rozelle. 2002. "Demand Systems with Unit Values: Comparisons with Elasticities from Market Prices." University of Waikato, Department of Economics, Hamilton, New Zealand.
- Gibson, J., S. Rozelle, and T. Le. 2002. "Evaluating the Use of Unit Values and Community Prices in Demand Analysis." University of Waikato, Department of Economics, Hamilton, New Zealand.
- Glewwe, P. 1991. "Investigating the Determinants of Household Welfare in Côte d'Ivoire." *Journal of Development Studies* 35(2):307–37.
- Gracia, A., and L. Albus. 1998. "The Demand for Meat and Fish in Spain: Urban and Rural Areas." *Agricultural Economics* 19(3):359–66.
- Jolliffe, D., and A. Semykina. 1999. "Robust Standard Errors for the Foster-Greer-Thorbecke Class of Poverty Indices: SEPOV." Stata Technical Bulletin 51. Stata, College Station, Tex.
- 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, Centre for Research in Economic Development and International Trade, Nottingham, U.K.
- 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 F. Goletti. 2000. *Rice Market Liberalization and Poverty in Viet Nam*. Research Report 114. Washington, D.C.: International Food Policy Research Institute.
- Nicita, A. 2004a. "Efficiency and Equity of a Marginal Tax Reform: Income, Quality and Price Elasticities for Mexico." Policy Working Paper 3266. World Bank, Washington, D.C.
- . 2004b. "Who Benefited from Trade Liberalization in Mexico? Measuring the Effects on Household Welfare." Policy Working Paper 3265. World Bank, Washington, D.C.
- Ravallion, M., and L. Dearden. 1988. "Social Security in a 'Moral Economy': An Analysis for Java." *Review of Economics and Statistics* 70(1):36–44.

- Saunders, C., and C. Grootaert. 1980. "Reflections on the LSMS Group Meeting." Living Standards Measurement Study Working Paper 10. World Bank, Washington, D.C.
- Wood, D., and J. Knight. 1985. "The Collection of Price Data for the Measurement of Living Standards." Living Standards Measurement Study Working Paper 21. World Bank, Washington, D.C.
- World Bank. 1999. *Papua New Guinea: Poverty and Access to Public Services*. Washington, D.C.

