# CHAPTER V. STATISTICAL TOOLS AND ESTIMATION METHODS FOR POVERTY MEASURES BASED ON CROSS-SECTIONAL HOUSEHOLD SURVEYS

*John Gibson*

## Introduction

Most of what is known about poverty and living standards in developing countries comes from household surveys. A household survey can provide data on many topics related to poverty, especially on some monetary indicator of welfare (expenditure on household consumption is the preferred indicator, for reasons discussed below). Advantages of a quantitative indicator are that it can be generalised from a sample to national totals; it can enable consistent comparisons of poverty through time, across a country's regions, and potentially across countries; and it is amenable to simulation and prediction, which are needed when studying the potential impact of proposed policies on poverty. Priority is placed on a monetary indicator because ultimately poverty alleviation programs have to be budgeted for, which is easier for monetary indicators than non-monetary ones.

Nevertheless, it is usual for a poverty-focused household survey to include non-monetary indicators, both of a quantitative nature (e.g., the height of young children, as an indicator of nutritional problems) and of a qualitative nature (e.g., perceptions about the adequacy of health care). Use of selected qualitative indicators raises issues of balance between survey and non-survey approaches that go beyond this chapter (see Chapter 6). But one point should be made here about these non-survey methods: while case study and participatory approaches may provide insights about poverty in a form more readily understood by policymakers it is important that they are backed up by survey evidence (see Box 1) in case they are given too much weight. Of course, these methods can also reveal the limitations of surveys by illustrating aspects of poverty that go beyond insufficient consumption and poor access to health and education – issues such as lack of safety and lack of power within families or communities. Hence, even though this chapter is only about household surveys, it should be considered in tandem with other methods for studying poverty.

**Box1:The Importance of Water: Survey and Case Study Evidence from Papua New Guinea**

A poverty assessment in Papua New Guinea relied on a multi-topic household survey that was backed up with various case studies (World Bank, 1999). The participatory study of health and nutrition showed that difficulties in accessing clean drinking water were a major problem for the poor. This was backed up by the education case study, which found lack of water as one of the most common reasons for the frequent closure of rural schools. These observations were supported by qualitative questions in the household survey, where improved water supply was listed as the most important priority by men and women when asked "*what in your opinion could government do to most help this household improve its living conditions?*". Finally, the quantitative component of the household survey confirmed the significant impact that poor access to water has on households: the poorest one-quarter of the population live in households where one hour per day was spent fetching drinking water. The survey also showed that this burden was borne overwhelmingly by women and girls.

This chapter is divided into four sections. The first studies several cross-cutting issues that may have to be considered--irrespective of the particular type of cross-sectional survey used--for poverty measurement. These issues are the choice between consumption and income as welfare indicators for measuring poverty, the importance of consistency of household survey methods when making poverty comparisons, methods of restoring comparability to inconsistent surveys, the effects of measurement errors, and the variance estimators that are appropriate for the complex sample designs that are used for household surveys. The second section discusses the particular types of surveys that statistical agencies and poverty analysts may have available to them. This includes discussion of different requirements of poverty-focused surveys compared to more traditional surveys that are used for gathering means and totals (e.g., expenditure weights for a Consumer Price Index). The third section discusses price data and how they can be collected and used to place a monetary value on either poverty lines or the change over time in the cost of reaching a poverty line standard of living. The final section discusses the difficult issues associated with assessing individual welfare and poverty from data that are collected on households.

## 5.1 Cross-cutting issues in poverty measurement

This section considers issues in poverty measurement that are largely independent of the particular type of household survey used.

### 5.1.1 Reasons for favoring consumption expenditure as a welfare indicator

The most common welfare indicators for poverty measurement are expenditure on household consumption and household income. The trend is for increased reliance to be placed on consumption-based measures for poverty analysis. For example, in a compilation of household surveys from 88 developing countries, which was originally constructed for establishing world poverty counts, 36 of the surveys use income as their welfare measure and 52 use expenditures (Ravallion, 2001). Similarly, the statistics offices in a majority of the developing countries providing metadata in the Statistical Addendum use either consumption expenditures solely or in combination with income as their welfare measure. The only region with a high reliance on income surveys is Latin America, although even in that region there is an increased use of expenditure surveys for poverty measurement (Deaton, 2001). Growing use of household consumption expenditure as the welfare indicator for poverty measurement reflects both conceptual and practical reasons. Conceptually, consumption expenditure is a better measure of both current and long-term welfare. Practically, income is considerably more difficult to measure.

In principle, the best measures of a household's long-term economic resources are either wealth or permanent income, which is the yield on wealth. Important components of wealth, such as the present value of expected labour earnings, are unobservable. While current income is observable, it has a transitory component, which obscures any ranking of households based on permanent income. However, consumers have some idea about their permanent income, and so are unlikely to make lasting adjustments to their spending if they believe that the changes in their income are transitory. Consequently, consumption is a function of permanent but not of current income. This reliance of consumption on permanent income also means that consumption levels are less variable over time than are income levels. In other words, because the transitory component of consumption is small, current consumption is a good measure of permanent consumption, which in turn is proportional to permanent income.

The choice of consumption rather than income indicators can affect the temporal trends in poverty rates. Because of transitory income fluctuations, income-poor households include those who have suffered temporary reductions in their incomes, while their consumption level may stay close to its long-run average (depending on the options for consumption smoothing). Such households have high ratios of consumption expenditures to income. For example, in Thailand, the expenditure to income ratio ranges from 2.0 in the poorest income decile to 0.8 in the richest decile (Deaton, 1997). Thus, if the poverty line remains fixed in real terms while the society enjoys an increase in average income, the ratio of consumption to income at the poverty line will grow over time because the poverty line is cutting at a lower and lower point in the cross-sectional income distribution. Therefore, the poor will increasingly be those with high permanent incomes who happened to suffer transitory shocks to their income during the reporting period. Because the measured consumption expenditure of this group is high relative to their income, a wedge is driven between the time-path of income-based and consumption-based poverty measures (Jorgenson, 1998). For example, the U.S. poverty rate fell by 2.5 percent per year from 1961 to 1989 when real total expenditure is used as the welfare measure. However, it declined by only 1.1 percent per year when income is used (Slesnick, 1993).

In addition to affecting the trend in poverty, transitory income fluctuations also affect the precision of the cross-sectional poverty profile. The high transitory component in measured income means that a poverty profile based on income is less likely to identify the characteristics of the long-term poor. Instead, it will mix together households with low permanent incomes and those with temporary reductions in income. For example, Slesnick found that the U.S. poverty profile shows surprisingly high homeownership rates and low food budget shares when income is used to define the poor. This goes against the expectation that the poor have few assets and devote most of their budgets to necessities like food (Slesnick, 1993).

In terms of practicalities, at least three factors make household income more difficult to measure than household consumption expenditures. These difficulties are likely to impair the accuracy of the income data gathered and are especially apparent in developing and transition countries. First, survey questions on income typically require a longer reference period than is needed for questions on expenditures because income estimates for periods less than a year will be affected by seasonal variation, especially for agricultural households. While there may be seasonal and other short-term temporal patterns in consumption expenditures, they will normally be less marked if households have access to consumption-smoothing devices such as savings, credit, storage, and exchange networks. The longer reference period needed for measuring income introduces greater problems of recall error.

Second, household income is hard to construct for self-employed households and those working in the informal sector because of the difficulty in separating out business costs and revenue. Frequently, arbitrary assumptions are needed to measure the income streams from assets such as agricultural livestock, and there can be difficulties in valuing the receipt of in-kind payments and self-produced items. These problems are less severe, although not absent, when household consumption is measured. Moreover, in developing and transition economies, the sources of household income are more diverse than the categories of household consumption so it is harder to design and implement questions for all of these sources.[34]

Third, questions about consumption are usually viewed as less sensitive than questions about income (although alcohol, tobacco and narcotics, and sexual services are usually viewed as sensitive and so expenditure on these is unlikely to be reliably measured), especially if respondents are concerned that the information will be used for tax collecting purposes or where illegal or barely legal activities provide a substantial portion of household income.

Given this preference for using consumption expenditures as the welfare indicator for poverty measurement there are a number of practical issues about how to calculate this expenditure. These include the calculation of the user cost for durable goods and what to do about expenditures on taxes and other government charges, and on financial instruments and insurance that allow a reallocation of consumption over time. A comprehensive set of recommendations on these issues is provided by Deaton and Zaidi (2002).

[34] While consumption surveys may be longer, they essentially repeat the same question on potentially hundreds of detailed consumption items. This is tedious but not conceptually difficult.

### 5.1.2 Consistency of household survey methods and poverty comparisons

Has poverty increased? This is one of the most important questions that household survey data should answer. It is a question that will be more commonly asked as progress toward the Millennium Development Goals is monitored and as the number of countries with nationally representative surveys in at least two different years increases. Because it is rare for household surveys to use identical methods, answers to questions about poverty changes may not be robust. Ideally, detailed experiments should assess the effect on measured poverty rates of changes in survey methods so that adjustment factors can be calculated and robust poverty trends retrieved.

Such experiments are rarely carried out as a part of poverty monitoring. However, recent methodological experiments demonstrate the tremendous sensitivity of estimates from household surveys to changes in key design features. Amongst these key features are different fieldwork methods (diaries versus recall), longer (more detailed) versus shorter (less detailed) recall questionnaires, and different reference periods over which expenditures are meant to be recalled. For example, in an experiment in Latvia, one-half of the households were given a diary for recording expenditures and in a subsequent period they were given a recall survey, while the other half had the recall first and then the diary. Reported food expenditures were 46 percent higher with the diary, regardless of whether the diary was used first or second (Scott and Okrasa, 1998).

An experiment with a recall survey in El Salvador gave a long questionnaire (75 food items and 25 non-food items) to one-quarter of a sample, with the rest given a short questionnaire (18 food items and 6 non-food items) that covered the same items but more broadly. Average per capita consumption was 31 percent higher with the long questionnaire (Jolliffe, 2001). An experiment in Ghana varied recall periods, with reported spending on a group of frequently purchased items falling by 2.9 percent for every day added to the recall period, with the recall error levelling off at about 20 percent after two weeks (Scott and Amenuvegbe, 1991).

Perhaps the most well known evidence on the sensitivity of poverty estimates to changes in survey design comes from India. Between 1989 and 1998, the National Sample Survey (NSS) in India experimented with different recall periods for measuring expenditure, replacing the previously used 30-day recall period with a 7-day recall for food and a one year recall for infrequent purchases. The shorter recall period raised reported expenditure on food by around 30 percent and on total consumption by about 17 percent. As Deaton (2005, p. 16) points out, "because there are so many Indians close to the poverty line, the 17 percent increase was enough to reduce the measured headcount ratio by a half, removing almost 200 million people from poverty."

Because of the policy significance of this statistical artifact, both Indian and foreign economists and statisticians developed adjustment methods that attempt to restore comparability to Indian poverty estimates (see Section 5.1.3 for details on some of these methods). However, it is likely that in many poorer, smaller, and less significant countries there is neither the expertise nor the foreign interest to correct such noncomparabilities (Box 2) .This gives all the more reason for such countries to be careful when changing their survey design, ideally using controlled comparisons where random sub-samples are given either the old design or the new design, so that adjustment factors can be calculated to restore temporal comparability.

**Box2: Incomparable Survey Designs and Poverty Monitoring in Cambodia in the 1990s**

Three socio-economic surveys were carried out in Cambodia during the 1990s to measure living standards and monitor poverty. Despite this active investment in data gathering, all supported by international donors, each survey was inconsistent with previous and subsequent surveys so no firm evidence exists on whether poverty rose or fell. The initial 1993-94 survey had a very detailed consumption recall list (ca. 450 items) to provide weights for a national Consumer Price Index (CPI). This detail was not needed for most of the population because the CPI was only ever compiled for the capital city, and it lead to an excessively detailed basket of foods ($n$=155) for the poverty line. Subsequent surveys gathered data on prices for less than one-third of the items in the basket, so updating of the poverty line relied heavily on assumptions.

The second survey in 1997 used only 33 broadly defined items in the consumption recall, and was fielded at a different time of the year. Consumption estimates from this survey were adjusted upwards (and poverty rates downwards) by up to 14 percent for rural households to correct for a perceived under reporting of medical expenses. This under reporting was estimated by comparing health spending in the short questionnaire with estimates from a more detailed health expenditure module fielded with the survey. The apparent fall in the headcount poverty rate from 39 to 36 percent between 1993 and 1997 is reversed if this adjustment is not applied.

The third survey in 1999 used 36 items in the consumption recall and was in conjunction with a detailed income and employment module. It was again conducted in different months than the earlier surveys. But this time, it was randomly split into two rounds, with half the sample in each. Greater efforts to reconcile consumption and income estimates at a household level in the second round led to dramatic changes in poverty estimates. In the first round, the headcount poverty rate was 64 percent, and in the second round it was only 36 percent. The dramatic fall in the poverty rate came from higher recorded expenditures and lower inequality in the second round. No robust poverty trend for the 1990s can be calculated from these irreconcilable data (Gibson, 2000)

### 5.1.3 Correction methods for restoring comparability to incomparable surveys

When controlled comparisons are not available, other methods have to be considered for restoring temporal comparability to incomparable surveys. Correction methods have been developed for at least two sources of incomparability: changes in the commodity detail of an expenditure recall questionnaire, and changes in the reference period over which expenditures are meant to be recalled. While these methods have been developed because of problems in specific surveys, they could be applied more widely and so are briefly discussed here.

A frequent feature of household surveys is that the consumption aggregates differ in their composition and coverage. For example, one survey may have “rice” as an item, but this is broken down in a subsequent survey into basmati rice and plain rice. This greater detail would be expected to raise measured consumption because it prompts respondents to remember some expenditure that they would otherwise forget. Similarly, one survey may cover a wider range of foods eaten out of the home than an earlier survey, also inflating estimates of consumption growth. In cases such as this, the bundle of foods in the poverty line should be recalculated, restricting attention just to items that are common to both surveys (Lanjouw and Lanjouw, 2001).

This abbreviated food poverty line (abbreviated because it excludes items whose definition changed between surveys) is then scaled up to provide a total poverty line. The particular method of scaling which is appropriate is associated with what is sometimes called the “upper poverty line”. This is an example of the Engel method, talked about more generally in Chapter 4.

The "upper poverty line" uses a non-food allowance that is calculated from the food budget share of those households whose food spending exactly meets the (abbreviated) food poverty line, $w^{\mathrm{U}}$. Specifically, the food poverty line, $z^{\mathrm{F}}$, is inflated upwards by this budget share: $z^{\mathrm{U}}=z^{\mathrm{F}}/w^{\mathrm{U}}$. In contrast, the "lower poverty line" adds to the food poverty line the typical value of non-food spending by households whose total expenditure just equals $z^{\mathrm{F}}$. This is more austere because these households would displace some required food consumption, given that they don't actually spend their total budget on food (Ravallion, 1994). If the food budget share of households whose total expenditure just equals $z^{\mathrm{F}}$ is $w^{\mathrm{L}}$, the "lower poverty line" is calculated as: $z^{\mathrm{L}} = z^{\mathrm{F}} + z^{\mathrm{F}}\,(1\text{-}w^{\mathrm{L}})$.

The different food shares that are needed for these two different poverty lines can be found from the following Engel curve:

$$w = \alpha + \beta \ln\left(\frac{x}{n \cdot z^{F}}\right) + \sum_{k=1}^{K} \gamma_k n_k + \varepsilon \tag{1}$$

where $w$ is the food budget share, $x$ is total expenditure, $n$ is the number of persons, $z^{\mathrm{F}}$ is the food poverty line, and $n_k$ is the number of people in the $k^{\mathrm{th}}$ demographic category. If total expenditure equals the cost of the food poverty line, $\ln\left(x/\left(n \cdot z_j^{F}\right)\right) = 0$, so $w^{\mathrm{L}} = \hat{\alpha} + \sum_{k=1}^{K} \hat{\gamma}_k \bar{n}_k$ where $\bar{n}_k$ is the mean of the demographic variables for the reference household used to form the poverty line basket of foods. Finding $w^{\mathrm{U}}$ requires a numerical solution, characterised by $n{\cdot}z^{\mathrm{F}}{=}x{\cdot}w^{\mathrm{U}}$. This can be substituted into equation (1) to give:

$$w^{\mathrm{U}} = \alpha + \beta \ln\left(w^{\mathrm{U}}\right)^{-1} + \sum_{k=1}^{K} \gamma_k n_k \tag{2}$$

Using $w$-1 to approximate $\ln w$, an initial solution of $w_0$=($\alpha_k$+β)/(1+β) can be found, where $\alpha_k = \hat{\alpha} + \sum_{k=1}^{K} \hat{\gamma}_k \overline{n}_k$ gives the combined effect of the intercept and the demographic variables for the reference household. This estimate can be improved upon by iteratively solving the following equation, $t$ times (Ravallion, 1994):

$$w_t^{\mathrm{U}} = w_{t-1}^{\mathrm{U}} - \frac{(w_{t-1}^{\mathrm{U}} + \beta \ln w_{t-1}^{\mathrm{U}} - \alpha_k)}{1 + \beta / w_{t-1}^{\mathrm{U}}}. \tag{3}$$

This upper poverty line can yield robust comparisons between the two surveys, under the assumption that the relationship between food spending and total spending stays the same over time. The other requirement for the comparisons to be robust is that only the head count measure of poverty is used. The problem with higher order poverty measures is that the relative distance between the consumption level of the poor and the poverty line may increase as the components in the consumption aggregate become more comprehensive. Thus, moving to an increasingly broad definition of consumption could show higher poverty, even if the same households are considered poor under each definition (Lanjouw and Lanjouw, 2001).

Another way in which one survey can be incomparable with an earlier one is if there are changes in the length of the reference period over which expenditures are meant to be recalled. But if at least a subset of expenditures maintain the same reference period it may be possible to restore comparability. For example, while the National Sample Survey in India adjusted the reference period for most survey items during the 1990s, fuel and light, miscellaneous goods, and a few other items maintained a consistent 30-day reference period in all of the surveys. In total, these items with the consistent reference period, which can be called the "30-day goods," account for about 20 percent of expenditures. Deaton (2003) uses expenditures on these items in the 50th Round of the NSS (in 1993-94) to predict the probability of being poor in that round of the survey. The estimated relationship from that year is then applied to the distribution of 30-day expenditures in the 55th Round of the NSS (in 1999-2000) to predict the probability of being poor in the 55th Round. This estimated poverty rate in the 55th Round should then be comparable to that from the 50th Round, as long as there is a stable relationship between spending on the 30-day goods and total spending, and as long as the density of spending on the 30-day goods is not affected by the changes in other parts of the questionnaire.

The specifics of the approach are described by Deaton (2003, pp. 323-4) and are summarized here. Let $F(\,)$ be the cumulative distribution function of per capita expenditures. The poverty rate, *P,* is given by $F(z)$, the fraction of people living in households where per capita expenditure is below the poverty line, *z*. The probability of being poor, conditional on spending amount *m* on the 30-day goods, is $F(z\,|\,m)$ so that the poverty rate is: $P = \int_0^\infty F(z\,|\,m)g(m)dm$ where *g*(*m*) is the density function of expenditure on the 30-day goods.

Although this equation cannot be evaluated using data from the survey with the changed recall period, it is possible to use the conditional headcount function, $F(z\,|\,m)$

from the earlier survey in conjunction with the actual distribution of 30-day expenditures from the later survey. In particular, Deaton (2003, p. 324) uses data from the 50th Round survey to compute the headcount conditional on *m* and then estimates the poverty rate in the 55th Round according to $\hat{P}_{55} = \int_0^\infty \hat{F}_{50}(z \mid m)\hat{g}_{55}(m)dm,$ where the "hats" denote estimates and the subscripts denote either Round 55 or Round 50 on the NSS.

When this correction method is applied to the Indian data, it shows that most of the observed decline in poverty between the two incomparable surveys in the 50th and 55th Rounds appears to be a real change and not a statistical artefact of the variation in the recall period. A similar conclusion is reached by Tarozzi (2004) who uses a more flexible procedure that can be conditional on more than one auxiliary variable. This more flexible procedure may be able to do more than just re-establishing comparability over time for statistics estimated using surveys of different design. It is possible that it could be applied to the problem of combining data from a survey and census to provide precise measures of poverty for small areas (see Chapter 7 for a discussion of poverty mapping).

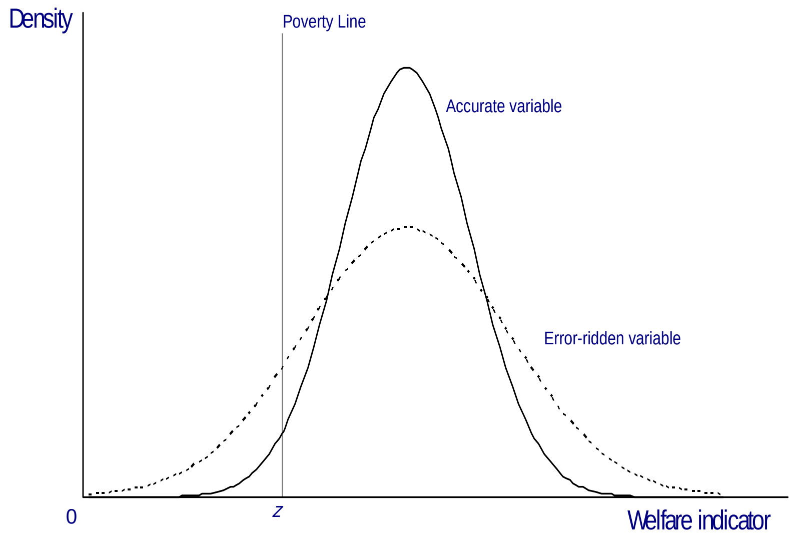
### 5.1.4 Measurement error in cross-sectional survey data

The sensitivity of poverty estimates to changes in household survey design discussed in Section 5.1.2 points to the problem of measurement error in cross-sectional survey data. (This issue is also addressed in the context of panel surveys in Chapter 8.) The widely different estimates of consumption and poverty resulting when two survey designs are used suggest that both estimates cannot be right and possibly neither are.

Measurement error in surveys poses a special challenge to statistical agencies when the focus is on poverty and other distributional statistics, rather than on means and totals which are the traditional statistics of interest. While random measurement error should not affect estimates of the mean or the population total if the sample is large enough, such errors will systematically bias poverty estimates.

In particular, the headcount index of poverty will be higher with a more variable welfare indicator, if the poverty line is below the mode of the welfare indicator. It will be lower if the poverty line is above the mode (Ravallion, 1988). This is illustrated in Figure 1, where an accurate welfare indicator is compared with an error-ridden indicator. The density functions of the two indicators have the same shape and same mode if the measurement error is random (that is, has a mean of zero) but there are wider tails for the error-ridden indicator. Thus, if the poverty line is located below the mode of these two distributions, there is a greater area under the density function of the error-ridden indicator (between 0 and $z$) than under the density function of the accurate indicator. Consequently, the value of the headcount index calculated with the error-ridden indicator will exceed that calculated with the accurate indicator. Higher order poverty statistics, such as the poverty gap index ($P_1$) and the poverty severity index ($P_2$), will also be overstated.

**Figure 1:** The effect of random measurement error on poverty estimates




To illustrate the possible effects of measurement error, household survey data from Papua New Guinea are used to calculate poverty statistics. In the original data, the mean consumption level is K911 per person per year, and the headcount index of poverty is 37.4 percent. A proportionate error was added to the survey data on consumption, $x$, so that the error-ridden indicator, $x_e$ was $x_e = x \cdot (0.5 + v)$ where $v$ was a uniformly distributed random number distributed between zero and one. The error-ridden indicator has the same mean level of consumption, but all poverty statistics are biased upwards, ranging from a 6.8 percent error for the headcount index to a 34.6 percent error for the poverty severity index (Table 1).

| Table 1: Example of the Effect of Measurement Error on Poverty Estimates | | | | |
|---|---|---|---|---|
| | Consumption (Kina/capita/year) | Headcount ($P_0$) | Poverty gap ($P_1$) | Poverty severity ($P_2$) |
| Original data | 911.0 | 37.4 | 12.4 | 5.6 |
| Adding measurement error | 911.6 | 40.0 | 14.9 | 7.5 |
| Percentage error | 0.0 | 6.8 | 20.4 | 34.6 |

Note: Poverty rates are calculated from poverty lines set for five regions of Papua New Guinea and are based on baskets of locally consumed foods providing 2,200 calories per day, with an allowance for non-food spending. The (population-weighted) average value of the poverty lines is K461 per person per year.
Source: Authors calculation from Papua New Guinea Household Survey data.

### 5.1.5 Variance estimators for complex sample designs

Household surveys are based on samples, but interest is in the underlying population. Hence, sampling errors are needed, especially when comparing poverty estimates between two groups or two time periods because these errors affect the confidence with which we can claim that poverty is higher in region *A* rather than region *B*, or in year 1 compared with year 2.

There are three essential features of complex sample designs:

- Weights, where some sampled observations represent more members of the population than do others,
- Two-stage sampling, where Primary Sampling Units (PSU) are first selected and then certain households within those PSUs are surveyed, and
- Stratification of the sample.

Weights may be needed either by design, to get larger samples for sub-groups of particular interest (e.g. a capital city), or to restore the representative nature of the sample if there is non-response (e.g., up-weighting the remaining observations from the group with high non-response rates). Two-stage sampling occurs because it is a cost effective way of carrying out fieldwork; it is cheaper to get a sample of 100 by visiting just 10 villages and selecting 10 households from each rather than visiting 100 villages and selecting just one household in each village. Stratification occurs because survey designers find that if they use prior information on factors that are likely to be associated with poverty (e.g., geographical remoteness) they can draw a sample in closer accordance with the proportions in the population rather than leaving this to chance.

Two-stage sampling is less efficient than simple random sampling in statistical terms (which causes larger standard errors). This is because the households within a PSU tend to have similar characteristics, so a sample drawn from them reflects less of the population's diversity than would a simple random sample with the same number of households. At the same time, stratification reduces sampling errors because it reduces the chance that a relevant part of the sampling frame will go unrepresented. Ignoring these complex design features can considerably bias estimates of sampling error. Howes and Lanjouw (1998) find the standard error of the headcount poverty rate in Ghana is 45 percent higher when clustering and stratification are accounted for compared with wrongly assuming simple random sampling.

Techniques for calculating sampling variance and standard errors from complex sample designs fall into two general categories: Taylor series linearization and replication techniques. A Taylor series expansion is a linear approximation to a nonlinear function, and this is relevant because many estimates of interest in sample surveys are nonlinear.

Formally, $f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)(x - x_0)^2/2! + \ldots$ which says that the function $f(x)$ can be approximated at one point, *x*, by taking its value $(f(x_0))$ at a nearby point, $x_0$, and using the slope at that point, $f'(x_0)$, to extrapolate to the point where we want to evaluate the function.

An improvement in the approximation comes from the second order term $f''(x_0)(x - x_0)^2/2!$ ($f''$ is the second derivative and ! is the factorial, so 2! is $1 \times 2 = 2$ and 3! is $1 \times 2 \times 3 = 6$) and the higher order terms. Variance estimators used with survey data assume that the second and higher order terms are of negligible size, leaving only the first-order, linear, portion of the expansion, $\text{var}(f(x)) \approx \text{var}[f(x_0) + f'(x_0)(x - x_0)]$. In other words, the variance estimate for a linear approximation to the estimator is used to estimate the variance of the estimate itself.

A wide range of software is available to calculate the variance of survey estimates using this linearization technique. For example, CENVAR within the IMPS package provided by the US Census Bureau and CSAMPLE within the EPI-INFO package provided by the US Center for Disease Control use linearization. This is also the main method used in the survey analysis procedures for general purpose econometric software like SAS and STATA. Two features of this estimation approach are relevant. First, a separate formula for the linearized estimate must be developed for each type of statistical estimator (such as a mean or a ratio). This is not a binding constraint because all of the widely used poverty measures can be expressed as the mean of a suitably transformed variable. For example, the poverty severity index ($P_2$) is just the mean of the squared proportionate poverty gaps, $\left[(z-y)/z\right]^2$ where $z$ is the poverty line, $y$ is the welfare indicator, and the squared proportionate gap is zero if $y \geq z$.[35] The second feature is that these estimators require at least two PSUs per stratum, which will usually be achieved by the sample design although it can be violated when examining narrow sub-populations.

Replication techniques take repeated sub-samples, or replicates, from the data. These replicates are then used to recompute the weighted survey estimates. For example, 50 replicate samples might be drawn from the original sample, and the poverty rate is calculated from each of these 50 replicates. The variance is then computed in terms of the deviations of these replicate estimates from the whole-sample estimate. The two main replication methods are Balanced Repeated Replication and Jackknife Repeated Replication. The basic idea of jackknife replication can be illustrated for the sample variance of the mean in a simple random sample. Suppose $n$=5 and sample values of $y$ are 6, 10, 4, 2, and 8. The sample mean $\bar{y} = 6,$ and its sampling variance is $\text{var}(\bar{y}) = \left(1/n\right)\sum(y_i - \bar{y})^2 \big/ (n-1) = 2.$ As an alternative to this analytical formula for the variance, the jackknife variance of the mean is obtained as follows:

1. Compute a pseudo sample mean by deleting the first sample value, which results in $\bar{y}_{(1)} = (10+4+2+8)/4 = 6.$ By deleting the second sample value instead, the second pseudo mean is $\bar{y}_{(2)} = (6+4+2+8)/4 = 5;$ and similarly, $\bar{y}_{(3)} = 6.5,$ $\bar{y}_{(4)} = 7,$ and $\bar{y}_{(5)} = 5.5.$

[35] Variations in household size and in household sampling weights may require a weighted mean to be used.

2. Compute the mean of the five pseudo-values $\bar{\bar{y}} = \sum \bar{y}_{(i)} / n = 30/5 = 6,$ which is the same as the sample mean, and
3. Estimate the variance from the variability among the five pseudo-values, $\text{var}(\bar{\bar{y}}) = [(n-1)/n] \sum (\bar{y}_{(i)} - \bar{\bar{y}})^2 = 2,$ which gives the same result as the analytical formula above.

Obviously there is no need to use jackknife replication for the variance of the mean of a simple random sample because an analytical formula is available. But the same idea can be extended to clustered samples. Specifically, a replicate can be formed by removing one PSU from a stratum and weighting the remaining PSUs in that stratum to retain the stratum's share of the total sample, and a pseudo-value can be estimated from each replicate. With the Balanced Repeated Replication, the replicates are formed by dividing each stratum into two PSUs and randomly selecting one of the two PSUs in each stratum to represent the entire stratum. Clearly, both replication techniques require at least two PSUs in each stratum.

Fewer software packages appear to use replication techniques compared with those using the linearization approach. Among those that do are VPLX which is supplied free by the US Census Bureau and WesVar, while a replication add-on has recently been made available for STATA.[36] The difference in availability of software for the two methods is unlikely to reflect any belief that one method for dealing with complex sample

[36] The linearization method has been available in Stata since version 5 (ca. 1996) under the command prefix svy, while a freely available add-on for the replication methods under the command prefix svr is available at http://econpapers.repec.org/software/bocbocode/s427502.htm

date is superior to the other. According to Korn and Graubard (1999), estimators based on smooth functions of the sample data (e.g., totals, means, proportions, and differences between proportions) have comparable variance estimates under both replication and linearization methods.

Regardless of the method used to calculate the sampling variability for complex samples, obtaining correct variances is especially important in the context of poverty monitoring. In monitoring, the main interest is the change in poverty levels--if any--between measurement periods, say $t_1$ and $t_2$. If $Y_{t1}$ and $Y_{t2}$ are the poverty statistics, we would like to know whether the observed difference, $Y_{t2} - Y_{t1}$, is indicative of a real change in the population rather than just reflecting sampling variability. Thus what is required is an estimate of the variance of the difference: $V(Y_{t2} - Y_{t1}) = V(Y_{t2}) + V(Y_{t1}) - 2\,Cov(Y_{t2},Y_{t1})$. The terms on the right-hand side can be estimated as design-based variance estimates of means or of ratio estimates. Let the square root of the resulting estimate be $se(Y_{t2}-Y_{t1})$, i.e., the standard error of the difference. The interval, $Y_{t2} - Y_{t1} \pm 1.96\, se(Y_{t2} - Y_{t1})$ defines a 95 percent confidence interval about the true difference (it would be 90 percent if 1.64 were used instead of 1.96). A confidence interval that is to the left of zero is indicative of an increased poverty rate. One that captures zero supports a "no change" hypothesis. An interval to the right of zero provides empirical evidence for a reduced poverty rate.

Under normal conditions wherein the poverty situation changes slowly, the real difference in poverty incidence narrows as the interval between $t_2$ and $t_1$ is shortened. This means a commensurately very small standard error is required to detect a small change in the poverty incidence for the population. Thus, more frequent monitoring does not mean a smaller sample size for each survey round. On the contrary, a more efficient sampling design and bigger sample are needed to reduce the noise (sampling error) to a level that would provide a good chance of detecting a weak signal (change in poverty incidence). Otherwise, there would be no point in the monitoring exercise if it were known *a priori* that the computed confidence interval will most likely straddle zero. It is to be noted also that all these considerations, including sample size, pertain equally if not more to sub-national domains of interest, e.g., urban-rural and regions, rather than to national level estimates.

## 5.2 Types of surveys

Several different types of household survey can be used to measure and analyze poverty. Very few of these surveys have poverty measurement as their primary objective. Thus statistical agencies have to carefully evaluate whether surveys that have other (or multiple) objectives can provide reliable data for measuring poverty.

### 5.2.1 Income and expenditure (or budget) surveys

Almost all countries have either a Household Income and Expenditure Survey (HIES) or a Household Budget Survey (HBS). Methods used to measure consumption expenditures in these surveys vary widely, in terms of data collection (recall, family diaries, and individual diaries), reference periods over which consumption is observed, and whether households are observed only once or revisited during a year. But one common feature is that in almost all cases the HIES and HBS are designed mainly to provide expenditure weights for a Consumer Price Index (CPI) and to assist in the calculation of National Accounts. For these tasks a survey only needs to provide estimates of means and totals. But there are important differences between the needs of CPI-focused and poverty-focused surveys, involving topical coverage, reference periods, and the need for revisits. Consequently, if statistical agencies are to place more weight on the objective of improving poverty measurement, certain changes to the design of these surveys may be warranted. An immediate problem in using HIES and HBS for poverty analysis is that because of the burden of remembering expenditures on so many items, respondents are typically asked about few other topics. Thus, there are often few variables available from the survey that can either help explain the poverty status of the household or assist in the more general objective of modelling household behaviour.

In contrast, poverty-focused surveys typically obtain measures of total consumption that do not have the level of commodity detail sought in an HIES or HBS. The reduced effort spent gathering the consumption data allows more attention to be paid to a broader array of topics that can assist in modelling the effect of various anti-poverty interventions. One key topic needed for poverty-focussed surveys is local prices which are rarely collected by HIES and HBS. Section 5.3 discusses this fully.

Although poverty-focused surveys do not need a lot of commodity detail, they do have to provide an accurate estimate of long-run welfare for each household in the sample. Such accurate estimation at the household level is not required for surveys that focus only on population means and totals because the effects of random errors can be expected to cancel each other out in the estimation of the mean. But for poverty rates and other variance-based statistics, the effect of random errors accumulates so errors in measuring household level welfare will be reflected in inaccurate estimates of aggregate poverty rates.

While the limited topical coverage of HIES and HBS restricts poverty analysis, the major problem with these surveys is the short period over which consumption is observed. Because respondents find it hard to remember spending on frequent purchases, HIES and HBS typically use a very short reference period (e.g., a one-week recall or a two-week diary), which may be atypical of the household's usual standard of living. This short observation period is sufficient if the goal is just to measure the average shares of household expenditure devoted to each good and service, which is all that CPI expenditure weights are. Specifically, if the sample is spread evenly over the months in the year, it is possible to get an annual average for a synthetic "representative household" without accurately estimating the annual expenditures of each household. In contrast, poverty measurement requires accurate estimates of long-run welfare for each household.

Such long-run measures appear to be provided by some surveys that report expenditures and poverty on an annual basis. But many of these surveys simply observe households for a week, fortnight, or month, with consumption from these periods annualised by multiplying by 52, 26, or 12. The length of the reference period may vary with the category of consumption, being longer for costly and/or infrequently consumed items and shorter for frequently consumed and minor items that would be easily forgotten. While the scaling factors that convert these short duration observations into annual figures vary, the principle in all cases is the same: an estimate of annual expenditures can be made by simple extrapolation from shorter observation periods.

What is the problem with these annualised estimates and also with estimates that are collected and reported for shorter periods like a fortnight or a month? Random shocks, which occur during the observation period and are subsequently evened out over the rest of the year, get included along with the genuine between-household inequality in annual expenditures. Consequently, estimates of annual inequality are overstated. In any setting where the poverty line is below the modal value of per capita expenditure, the overstated dispersion will also lead to an overstatement of the poverty head-count and other measures of poverty.

The degree to which measured annual inequality and poverty are overstated when short reference periods are used can be seen in urban China (Table 2). China is of interest in this regard because respondents in the HIES in China keep a daily expenditure diary for a full 12-month period, which provides a benchmark to evaluate estimates that are based on extrapolations from shorter periods. For example, if expenditures for each household were only observed for one month (but the sample is spread over the year) and multiplied by 12 to give an annualised estimate, inequality in annual expenditures would be overstated by over 60 percent, annual headcount poverty by over 50 percent, and the poverty gap index by 150 percent.

The upward bias is roughly halved if expenditures are annualised from two months of data (collected six months apart) and declines further if the survey collects either four or six months of expenditure data. It is notable that there is no overstatement in estimates of mean annual expenditure when any of the short-period data are extrapolated to annual totals. This emphasises the fact that a survey design that does a good job of estimating the mean will not necessarily be accurate for variance-based measures like poverty and inequality.

Table 2: Percentage Overstatement in Inequality and Poverty Measures for Urban China when Annual Expenditures are Obtained by Extrapolating from Monthly Data

| | Extrapolation based on observations in: | | | | | Corrected extrapolation |
|---|---|---|---|---|---|---|
| | 1 month | 2 months | 4 months | 6 months | | |
| Mean annual expenditure | 0.1 | 0.1 | 0.1 | 0.1 | | 0.1 |
| Gini index of inequality | 64.6 | 36.4 | 17.7 | 11.6 | | 6.4 |
| Head-count poverty rate | 53.1 | 32.2 | 14.0 | 15.0 | | 0.1 |
| Poverty gap index | 149.8 | 77.8 | 34.2 | 19.4 | | 5.0 |

Note: Corrected extrapolation uses correlation from a single revisit (i.e., two months of data).
Source: Gibson, Huang and Rozelle (2003).

One response to exaggerated poverty estimates that come from extrapolated annual expenditures is to only report poverty for shorter periods, corresponding to the reference period used by the HIES. For example, if a survey observes most household consumption for only a week, the poverty estimates would also be reported on a weekly basis. However, such short-period estimates may be dominated by transitory fluctuations.

Cross-country comparisons will also be difficult unless a standard reference period is agreed to, although this problem already exists because extrapolated annual estimates are not comparable to proper annual data like those available from China. Annual reporting periods are likely to continue to be used while agriculture remains an important source of household income because of the resulting seasonality in consumption and poverty.

### 5.2.2 Correcting overstated annual poverty from short reference period HIES and HBS data

One method that may combine the practicality of short observation periods with the need for annual estimates of expenditures and poverty is to revisit some surveyed households at least once during a year. Rather than simply adding the two estimates of the household's expenditure and naively extrapolating to an annual total (as was done in Table 2), Scott (1992) suggests a "corrected extrapolation" based on correlations between the same household's expenditures in different periods of the year – correlations implicitly assumed to be 1.0 by simple extrapolation.

For example, consider a survey that gathers all expenditure data using a one-month reference period (as the National Sample Survey in India did until recently). Let $\bar{x}_m$ refer to the average, and $V(x_m)$ the variance, of monthly expenditures across all $i$ households and $t$ months in the year. Extrapolating to annual expenditure totals by multiplying monthly expenditures by 12 gives an estimated variance of annual expenditures of $144 \cdot V(x_m)$. As indicated in Table 2, this extrapolation overstates the variance in the annual expenditures that would be recorded if each household was observed for a full 12-month period:

$$V(x_a) = \frac{1}{N}\sum_{i=1}^{N}\left(x_{i,a} - \overline{x}_a\right)^2 \tag{4}$$

where $x_{i,a}$ is annual expenditure by the $i^{\text{th}}$ household and $\overline{x}_a$ is average annual expenditures. Equation (4) can be expressed as:

$$V(x_a) = \sum_{t,t'=1}^{12} r_{t,t'}\sigma_t\sigma_{t'} \tag{5}$$

where $r_{t,t'}$ is the correlation between expenditures in month $t$ and month $t'$ and $\sigma_t$ is the standard deviation across households in month $t$. This follows because $x_{i,a} - \overline{x}_a$ in equation (4) can be expressed as the sum of the deviations of each household's monthly expenditure from the mean for that month, $d_{it} = x_{it} - \overline{x}_t$ and the $d_{it}$ terms are components of the correlation coefficient:

$$r_{t,t'} = \frac{1}{N}\sum_{i=1}^{N} d_{it}d_{it'} \Big/ \sigma_t\sigma_{t'}\,. \tag{6}$$

Assuming that the dispersion across households does not vary from month to month, i.e., $\sigma_t = \sigma_{t'}$ equation (5) can be expressed as:

$$V(x_a) = \left[12 + 132\cdot\overline{r}\right]V(x_m)\,. \tag{7}$$

where $\overline{r}$ is the average correlation between the same household's expenditures in all pairs of months in the year. Equation (7) shows that the variance from simple extrapolation to annual totals, $144\cdot V(x_m)$, equals $V(x_a)$ only in the special case of $\overline{r} = 1$.

The corrected extrapolation uses estimates of $\bar{r}$ to scale the $i^{\text{th}}$ household's deviation from the overall monthly average $(x_{it} - \bar{x}_m)$, up to an annual value. Adding this to the annual average across all households, $\bar{x}_a = 12 \cdot \bar{x}_m$, gives:

$$x_{i,A} = (x_{it} - \bar{x}_m)\sqrt{12 + 132 \cdot \bar{r}} + 12 \cdot \bar{x}_m . \tag{8}$$

For example, if $\bar{r} = 0.5$, the scaling factor is only 8.8 (=√78), rather than the scaling factor of 12 implied by simple extrapolation. Thus, the deviation of a household's one-month expenditures from $\bar{x}_m$ has a smaller effect than under simple extrapolation, leading to a less dispersed distribution of annual expenditures and a lower poverty estimate (if the poverty line is below the mode of the expenditure distribution).

While the most reliable estimate of $\bar{r}$ would use the 66 correlation coefficients $r_{t,t'}$ between all $i \neq j$ pairs of months, this provides no practical advantage because it requires observing each household in every month in the year, as is done, for example, by the HIES in China. However, even getting an estimate of $\bar{r}$ from just two, non-adjacent months may be sufficient.

The final column of Table 2 shows that this method gives estimates that are quite close to those obtained from observing each household's expenditure for all 12 months of the year. In urban China, the errors from this corrected extrapolation method never exceed 6 percent and are much smaller than the errors generated by multiplying monthly data by 12, as was done in the first column of Table 2. Using revisits in more months to form a more reliable estimate of $\bar{r}$ does not significantly improve estimates (Gibson et al., 2003).

Thus, a single revisit about six months after the first survey of the household's expenditure may give a good estimate of $\bar{r}$ so that equation (8) can be used to improve estimates of annual poverty, even when a HIES or HBS uses short observation periods. This economical approach to estimating $\bar{r}$ will be valid if the correlations among non-adjacent periods vary little as the gap between observations increases, as was found by the 1993-94 Household Budget Survey in Zambia where $r_{t,t'}$ fell by just 0.0078 for each month that the gap between $t$ and $t'$ increased (CSO, 1995).

Further savings may be made by restricting the repeated observations to a random subset of the sampled households to lessen the cost of getting the parameter $\bar{r}$. This random sub-sample should be large enough to allow $\bar{r}$ to be calculated separately for major groups of the population (e.g., rural and urban, and rich and poor) because the extent to which expenditures fluctuate within the year may differ between these groups. For example, in a survey in Papua New Guinea, households in 20 percent of the primary sampling units in the sample were revisited about six months after the initial survey to estimate $\bar{r}$, and this only added about 10 percent to the cost of the survey (compared with just using a cross-section) while substantially improving poverty estimates (Gibson, 2001).

### 5.2.3 Living Standards Measurement Study surveys

In contrast to the HIES and HBS, both of whose main objective is to measure means and totals, the Living Standards Measurement Study (LSMS) surveys of the World Bank have a primary focus on measuring the distribution of living standards.

Consequently, the design of the LSMS has been dictated by the need to have accurate measures of monetary living standards for each household in the sample, not just for a representative household. Even though the LSMS surveys collect information on both income and consumption, poverty measurements from these surveys have always used consumption data. In contrast, some analysts choose to measure poverty using income data from HIES, even when consumption expenditure data may be available.

A further difference is that the LSMS surveys are explicitly multi-topic surveys. In addition to income and consumption, they collect detailed data on education, health and anthropometry, employment, migration, agriculture, non-farm enterprises, savings and credit, and community-level data on public services and local prices. This more extensive coverage is achieved by reducing the commodity detail required in the consumption module.

Besides providing alternative indicators of poverty (such as lack of education, poor access to water, and malnutrition of children), the broader topic coverage of LSMS surveys enables household behaviour to be modelled. This can help in the formation of policies to break the intergenerational cycle of poverty (Box 3). For example, households where adults have low levels of education tend to be poor. Hence, LSMS surveys include considerable detail on educational expenses, distance to schools, and quality of school materials for current students. These data can help explain factors that limit enrollment of certain groups of students (e.g., girls, and students from particular regions or income groups). Once those factors are identified, interventions can be designed to improve current enrollments and reduce the likelihood of future poverty.

**Box3: Mother's Education, Child Stunting, and Intergenerational Poverty in Papua New Guinea.**

Analyses of LSMS survey data from Papua New Guinea have identified one mechanism through which poverty and ill-health are transmitted across generations and suggests an intervention that could break this cycle (Gibson, 1999). The low levels of education of mothers compared with fathers (a gap of two school years, on average) contributes to the stunted growth of children (i.e., children are shorter for their age). Parental education affects stunting by improving knowledge of health and nutrition, as well as by increasing incomes. In fact, an additional year of schooling for mothers is three times more effective at reducing stunting than is a year of schooling for the father (with or without controls for income). Stunting matters to poverty because stunted children have higher risk of sickness and death and poorer mental development. In addition, stunted girls grow up to be stunted mothers, who are more likely to give birth to underweight babies that have a greater risk of being stunted (UNICEF, 1998). Hence, the vicious circle, caused partially by gender bias in schooling, continues across generations.

A very detailed description of all modules in the LSMS surveys is available in Grosh and Glewwe (2000). The most important module from the point of view of poverty measurement is the consumption module, fully described by Deaton and Grosh (2000). Only two aspects of LSMS surveys are considered here: use of bounded recall and use of recall questions designed to provide information for an annual reference period.

To prevent telescoping errors, which are a mis-dating of expenditures, some LSMS surveys used a bounded recall where interviewers first visited respondents to administer modules of the survey other than the consumption recall. A subsequent visit was then made one or two weeks later and respondents were asked about consumption since the previous visit. The expectation was that the initial visit would clearly mark the beginning of the recall period and reduce the mis-dating of consumption. There does not appear to have been an evaluation of this design, although it was consistent with findings in the literature on telescoping (Neter and Waksberg, 1964), and it was not used in all LSMS surveys, creating some non-comparability.

In addition to either a bounded or unbounded recall of consumption over an immediately previous period like a month, some LSMS surveys attempted a longer term recall. Following a screening question on whether the household consumed the particular item during the past year, respondents who had were asked about the number of months they purchased the item, the number of times per month they purchased the item, and the usual quantity and value of this usual purchase. A similar set of questions was asked about own-production and other non-purchases (such as gifts received). The product of usual purchase value, times per month usually purchased, and months per year purchased may give an estimate of annual expenditure on the item.

If these questions are answered accurately they solve the problem of overstated inequality and poverty when annualizing consumption estimates from short reference periods. Deaton and Grosh (2000) present evidence that suggests this form of annual recall provides similar data to recall over the previous month. However, this is not a firm verification because the two types of data are gathered in the same interview and are likely influenced by each other. This is an area where statistical agencies could usefully carry out further experiments.

#### 5.2.4 Core and module designs

While multi-topic surveys are useful for poverty measurement and distributional analysis, they are hard to conduct. Therefore, data are normally available only at low frequency and for small samples, making them less useful for poverty monitoring. Some statistical agencies deal with this problem by using a core-module design. A simple core survey is fielded frequently and a variety of rotating modules are appended to the core survey. For example, the Indonesian SUSENAS has an annual core with questions on demography, education, labour market activity, and an abbreviated consumption recall that covers 23 broad categories. This is supplemented with a detailed consumption module, using 320 detailed categories, that is given to a subset of respondents every third year. In the intervening years, modules on other topics are used.

Although the core-module design is popular, it has at least two drawbacks that can cause inconsistent poverty comparisons. First, estimates from detailed consumption modules are often inconsistent with the results from abbreviated consumption questions in a core. For example, in SUSENAS the consumption estimates in the core appear to be understated, particularly for households with higher true consumption (mean reverting error) and for larger households (Pradhan, 2001). It is therefore not possible to create a consistent annual series of consumption and poverty estimates by using results from the core survey in two years and from the module survey in the third year. Second, contents of rotating modules can affect the core so even core-to-core temporal comparisons may be inconsistent. For example, in the Cambodia Socio-Economic Survey (CSES) of 1999, the addition of a detailed income module affected the consumption data in the core because of a desire by either respondents or interviewers to reconcile consumption and income at the household level (see Box 2).

The behaviour of poverty analysts can also be affected by the contents of a module. A detailed social sector module in the 1997 CSES had estimates of health expenditures that were much higher than the health spending recorded in the core, so the estimate of total expenditure for the core survey was adjusted higher (by up to 14 percent) because of the presumed undercount. This destroyed the comparability with consumption and poverty estimates from previous and subsequent core surveys where this adjustment had not been made (Gibson, 2000). These examples suggest that care is needed in the use of core-module surveys.

#### 5.2.5 Demographic and Health Surveys

Demographic and Health Surveys (DHS) now cover more than 170 surveys in 70 countries throughout the developing world. Country-specific details of these surveys can be found at www.measuredhs.com. A somewhat similar, though less well known set of surveys, are the Multiple Indicator Cluster Surveys (MICS) that are carried out by UNICEF. These surveys have three potential advantages over more traditional sources of household data for poverty analysis.

- They are available for a wider range of countries, especially in Africa;
- In many countries they are available at two or more points in time, allowing temporal comparisons; and

- Key survey instruments are standardized for all countries so cross-country comparability is much greater than in any other type of household survey.

Offsetting these potential advantages, a very major drawback of DHS and MICS is that, except for a few experimental modules, they do not collect information on either incomes or consumption. Consequently it is not possible to use this rich source of data for conventional poverty measurements. However, recent research suggests that the information collected by these surveys on dwelling facilities (e.g., presence of piped drinking water) and asset ownership (e.g., radios and bicycles) may provide a measure of household economic status that may be useful for distributional and poverty analysis.

There are two lines of this research, only one of which has proceeded directly to poverty measurement. The most well known statistical method for using these surveys in place of consumption data is based on research by Filmer and Pritchett (2001). These authors use both household consumption expenditure and an “asset index” to see which is better at explaining patterns of children’s school enrollments in Nepal, Indonesia, Pakistan, and states of India (using the National Family Health Survey for India, which is similar to the DHS). They find that the asset index is a proxy for economic status that is at least as reliable as conventionally measured consumption expenditures. This asset index uses the method of principal components, which is a mathematical technique for transforming several correlated variables (on household asset ownership and dwelling facilities in this case) into a smaller number of uncorrelated variables. Only the first principal component, which accounts for as much of the variability in the data as possible, is used by Filmer and Pritchett (2001) and others who follow their approach. Typically this component accounts for about 25 percent of the variation in asset ownership and facilities in a DHS. There are no units for interpreting this asset index, so it is used only for ordinal comparisons. One common use has been to compare educational attainment of the richest 20 percent of households and the poorest 40 percent (see http://www.worldbank.org/research/projects/edattain/edattain.htm).

While the asset index approach of Filmer and Pritchett (2001) has not been used to directly study poverty, a related method has been developed by Sahn and Stifel (2000) to make poverty comparisons across time and space for 11 African countries. In this method, DHS data from all 11 countries are pooled and an asset index is formed using the method of factor analysis. Unlike the method of principal components, which uses all the variability in an item, factor analysis allows some variability to be unique, with only the variability that is common with the other items used to form the asset index.

Relative "poverty lines" are created from the asset index, based on the values of the index at the 25th and 40th percentile of the pooled sample. Poverty comparisons are made across countries, and especially over time for each of these countries by seeing what proportion of the population in a subsequent DHS have an asset index that is below the values that were at the 25th and 40th percentiles in the first survey. The change in poverty over time is also calculated with the poverty gap and squared poverty gap measures, and this change is decomposed by sector.

There would need to be a validation of this method to see whether the results closely mimic those calculated with more typical consumption data before any recommendations could be made about its wider use. Even in the absence of such a validation, there are at least three concerns with the approach:

- An index based on the principal components approach (and presumably also the factor analysis approach) appears to put higher weights on durable goods that are easier to own which is not the pattern that occurs for an index based on a more explicit model for the ownership of durables (Mukherjee, 2005);
- The link between assets and expenditures is likely to be non-linear, so the ability of an asset index to serve as a proxy for unmeasured consumption is likely to vary over the income distribution and through time; and
- The very simplicity of the questions that underlie the asset index could also prove to be a weakness because yes/no questions on ownership of an asset do not distinguish between the wide variations in quality of these assets.

### 5.3 Pricing and updating the value of poverty lines

Information on the prices that households pay for items they consume is crucial for poverty measurement. Most obviously these prices are needed to place a monetary value on the food basket for a Cost of Basic Needs (CBN) poverty line. Prices are also needed to calculate the change over time in the cost of reaching a poverty line standard of living. Even methods for constructing a poverty line that seem to rule out the need for prices, such as the Food Energy Intake (FEI) method, prove on further examination to require information on prices.[37] In fact, measurement of local prices is needed for some or all of the following three tasks:

1. pricing the food basket for the Cost of Basic Needs (CBN) poverty line,
2. forming spatial deflators, so that any ranking of household consumption expenditures is in real rather than nominal terms, and
3. imputing values either when the survey only collects quantities or when checking the sensitivity of the consumption estimates to the use of respondent-reported values.

The methods used to calculate a CBN poverty line are discussed in Chapter 4 so attention here is restricted to the calculation of spatial price deflators and the use of price data for imputing values when only quantities are collected.

### 5.3.1 Spatial price deflators

Spatial price deflators are needed because price differences between regions may make between-household comparisons of nominal consumption expenditures misleading.[38] For example, in the CBN method of setting poverty lines it is typical to base the poverty line basket of foods on the actual consumption pattern of a group of poor

[37] The FEI method relies on a regression of calorie intakes on a welfare indicator like per capita expenditures. Once a calorie target is set (say, 2000 calories per person per day) the regression is inverted to solve for the required expenditure to meet the calorie target. However there will be a measurement error in this regression if it is carried out in terms of nominal expenditures when there are large price differences between regions. This error will tend to reduce the magnitude of the regression coefficient, causing an overstatement in the level of expenditures required to reach the calorie threshold and hence an overstatement in the value of the poverty line. This error could be reduced if price data were available to calculate real expenditures that reflect regional differences in the cost of living.

[38] Temporal price deflators may also be needed. It is typically assumed that prices do not vary over time within a cross-section but in inflationary environments even a few months between the time of the first and last household being surveyed could cause a difference between nominal and real expenditures.

households.[39] But in order to identify this group of poor households, some ranking must be used and this needs to control for spatial price variation. Otherwise poor households from regions where prices are high are less likely to be included in the reference group than are poor households in regions where prices are low because those from the higher priced region will have higher nominal expenditures.

The ideal way to control for spatial differences in the prices facing households is to calculate a "true cost-of-living index". This true cost-of-living index is based on the *expenditure function*, $c = c(\bar{u}, \mathbf{p})$, which gives the minimum cost, $c$ for a household to reach utility level $\bar{u}$ when facing the set of prices represented by the vector $\mathbf{p}$. For two, otherwise identical households, one living in the base region and facing prices $\mathbf{p^0}$, and the other living in another region facing prices $\mathbf{p^1}$, the true cost-of-living index is:

$$\text{True cost-of-living index} = \frac{c(\bar{u}, \mathbf{p^1})}{c(\bar{u}, \mathbf{p^0})}$$

which can be interpreted as the relative price in each region of a fixed level of utility. Although this is the ideal spatial price index, it is not commonly calculated, even in developed countries.

Instead the usual approach to controlling for spatial price differences is to use a price index formula that approximates the true cost-of-living index. A common choice is

[39] Exactly how many households should be in this group depends on prior notions of the poverty rate. For example, if it was believed that the poverty rate was 0.25 it would be likely that an analyst would use the food consumption patterns of the poorest quarter of households for obtaining the poverty line basket of foods. If this prior estimate of the poverty rate turns out to be quite different than the subsequently calculated one, it may be necessary to revise the calculations, using a different definition of the starting group (Pradhan, Suryahadi, Sumarto and Pritchett, 2001).

the Laspeyre's index, which calculates the relative cost in each region of buying the base region's basket of goods:

$$L = \frac{\sum_{j=1}^{J} Q_{kj} P_{ij}}{\sum_{j=1}^{J} Q_{kj} P_{kj}},$$

where *k* is the base region, *i* indexes every other region, *j* indexes each item in the consumption basket, and *Q* and *P* are quantities and prices.

The Laspeyre's index overstates the cost-of-living in high price regions. It does not let households make economising substitutions away from items that are more expensive in their home region than they are in the base region. For example, ocean fish are usually more expensive in the interior of a country than on the coast, so the quantity of fish consumed would typically be lower in the interior than on the coast. But if a coastal region is the base region, the Laspeyre's index calculates the cost of purchasing the coastal level of fish consumption at the high prices prevailing in the interior. Instead, a true cost-of-living index would calculate the cost of obtaining the coastal level of utility when facing the high prices for fish that prevail in the interior, letting the household rearrange its consumption bundle to minimise cost.

Another commonly used price index, the Paasche index understates the cost of living in high price regions because it evaluates relative prices using a basket of goods that varies for each of the *i* regions:

$$P = \frac{\sum_{j=1}^{J} Q_{ij} P_{ij}}{\sum_{j=1}^{J} Q_{ij} P_{kj}}.$$

In other words, the Paasche index takes a weighted average of relative prices, where the weights reflect prior economising substitutions by households. Continuing the above example, the Paasche index weights the high price of fish in the interior with the (low) quantity of fish consumed by interior households. This understates the cost of living disadvantage in the interior compared with the coast because it puts a smaller weight on the items with the highest prices relative to other regions.

A geometric average of the Laspeyre's and Paasche indexes gives a Fisher index: $F = (L \times P)^{1/2}$. This is a superlative price index which will closely approximate a true cost-of-living index. Another superlative price index that is sometimes used is the Törnqvist index:

$$T = \exp\left[\sum_{j=1}^{J}\left(\frac{w_{kj} + w_{ij}}{2}\right)\ln\left(\frac{P_{ij}}{P_{kj}}\right)\right]$$

where $w_{ij}$ is the average share that item *j* has in the consumption basket in region *i*, and region *k* is the base region.

One practical difficulty with all of these price index formulae is that they require a full set of prices for all items in the consumption basket. Household surveys are typically not able to collect prices for all consumption items (for example, prices for services are hard to measure) so assumptions are needed about the regional pattern of prices for the items that are not observed. One solution to this problem is to derive the spatial price index from the regional poverty lines because poverty lines can be calculated when there are missing non-food prices (see equation (1)). But even for consumption items where it is conceptually possible to gather price data, there are often practical difficulties that result in very many missing prices. 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).

### 5.3.2 Whose cost of living?

The possibility of deriving a spatial price index from regional poverty lines raises the important issue of whose cost of living is being measured by the price index. A price index derived from poverty lines would typically measure regional differences in the cost of living amongst the poorest *x* percent of the population, where *x* is either the fraction of the population below the poverty line or the fraction whose food budgets were used to create the poverty line food basket. The regional pattern of cost of living differences for this group could be quite different to the pattern shown by a price index that places greatest weight on households who are either in the middle or the upper parts of the income distribution.

There are three sources of possible difference between a price index for the poor (such as one derived from a set of regional poverty lines) and a more general purpose price index that reflects the cost of living for the middle or upper parts of the income distribution:

1. the composition of household budgets changes when moving up the income distribution, so a price index for the poor would put more weight on basic necessities,
2. for a given category of consumption (say, rice) the particular brands, grades, varieties and outlets where rich and poor purchase may differ and also may have different prices, and
3. various formulae that combine price data with information on the importance of each commodity in household budgets can place more weight on either rich households or poorer households.

This question of weighting matters because, as shown above, a price index is essentially a weighted average of relative prices where the weights reflect the average importance of the commodity in household budgets. One way to calculate this average importance for a commodity would be to add up expenditure on that item across all households, and then calculate the ratio of the total expenditure on that item to the total expenditure on all items. This is the approach used in the calculation of Consumer Price Indexes around the world. One feature of this method is that it gives more weight to the rich, because they have more total spending. Consequently the resulting price index is sometimes called a "plutocratic price index" (Prais, 1958).

Rather than taking ratios of total spending, another method of calculating the average importance of a commodity would be to first calculate budget shares for each household. In the second step these budget shares would be averaged across all households. This average of shares approach gives every household the same weight (except for any variation due to household size and sampling weights). Thus it can be considered a democratic price index because a rich household has no more impact on the finally calculated average than does a poor household.

A hypothetical example showing the difference between these two types of averages is presented in Table 3. There are two households, with one having three times the total spending of the other. Only two commodities are available to consume: cassava, which is a necessity and ice cream, which is a luxury. If the average importance of each commodity is calculated in terms of the shares of total expenditure, the resulting price index would put 25 percent of the weight on the price of cassava and 75 percent on the price of ice cream. This is much closer to the consumption pattern of the rich household than the poor household. But if the average of shares approach was used, the weights would be 30 percent on cassava and 70 percent on ice cream which is halfway between the consumption patterns of the two households.

Table 3: Example of Two Different Weighting Methods for a Price Index

| | Cassava | Ice Cream | Total Spending | Cassava Share | Ice Cream Share |
|---|---|---|---|---|---|
| Poor household | $40 | $60 | $100 | 0.40 | 0.60 |
| Rich household | $60 | $240 | $300 | 0.20 | 0.80 |
| Total | $100 | $300 | | | |
| Share of total | 0.25 | 0.75 | | | |
| Average of shares | | | | 0.30 | 0.70 |

Source: Author's example.

There is one other implication of the result that a typical Consumer Price Index uses weights that are closer to the consumption patterns of rich households. To the extent that the price trends for items consumed by the rich differ from the trends of those consumed by the poor, a CPI may be a poor choice for updating poverty lines to account for price changes over time.

### 5.3.3 Using prices to impute the value of consumption

Self-produced items, and especially food, are a major component of consumption in rural areas of many developing countries. The monetary values placed on these self-produced items in surveys are often the values that respondents themselves suggest. There are grounds for questioning the reliability of these respondent-reported values. Many households who produce a food do not buy that same food, so they may not be well informed about prices when they assign a value to their own food production. Moreover, the items available for sale in markets may be of a different quality than their own production so even if they are aware of prices in the market they may not be able to accurately impute a value for their own production.

There are two concerns about relying on respondent-reported values for self-production. First, they introduce an additional, and extraneous, source of inequality into measured consumption. If the poverty line is below the mode of the welfare indicator, this increase in measured inequality will raise the measured poverty rate (see Figure 1). For example, it may seem unreasonable that two households, who produce the same quantity of a food in the same location, can value that production differently. A household might fall below the poverty line just by being too pessimistic when valuing their own food production because they think prices are lower than they truly are.

Second, the values applied to self-produced food items could differ, systematically, from market prices. Such discrepancies could drive a wedge between the market prices used to form a Cost of Basic Needs food poverty line and the values used to form estimates of consumption. If respondents tend to report values for their self-produced foods that are lower than market prices, estimates of the incidence of poverty could be inflated, especially in rural areas where subsistence food production is important.

There are two alternatives to respondent-reported values, as measures of the value of self-produced food items. The first is to value self-produced foods with the average of the implicit unit values used by other households in the same cluster (that is, Primary Sampling Unit) as the respondent. These implicit unit values are the ratio of value to quantity reported by each respondent, and are similar to a price except that they may reflect quality variation and also measurement error. Replacing respondent-reported values with a cluster average removes the within-cluster variability in valuations. However, it does not address any discrepancy between these average unit values and market prices which may drive a wedge between the prices used for the poverty line and the implicit prices used when valuing consumption.

The second alternative is feasible only if a survey has collected prices from local markets. In this case it is possible to value self-produced foods with the average price that was observed during the survey in the market closest to the respondent. It is notable that both of these alternative ways of valuing self-produced foods switch the cornerstone of consumption measurements from the respondent reports of values to the survey estimates of food production quantities. This does place a lot of faith in quantity measurements, and these measurements are not necessarily the ones where statistical offices have the greatest expertise, compared to, say, agriculture ministries and others who do crop surveys. But unless statistical offices collect prices in local markets it is impossible to know how sensitive the estimates of consumption and poverty are to the various assumptions made when valuing self-produced items.

### 5.3.4 Practical issues in collecting price data

Once a decision has been made to obtain price data, either for setting the food poverty line, calculating a spatial price deflator or placing a value on non-purchased consumption, there are three practical questions that a statistical agency must consider:

1. How many prices to collect, in terms of the number of items and the number of individual price observations per item,
2. Where to collect prices, and at what geographical scale to calculate and report any resulting price indexes, and
3. How to collect the price information, in terms of the following four choices:

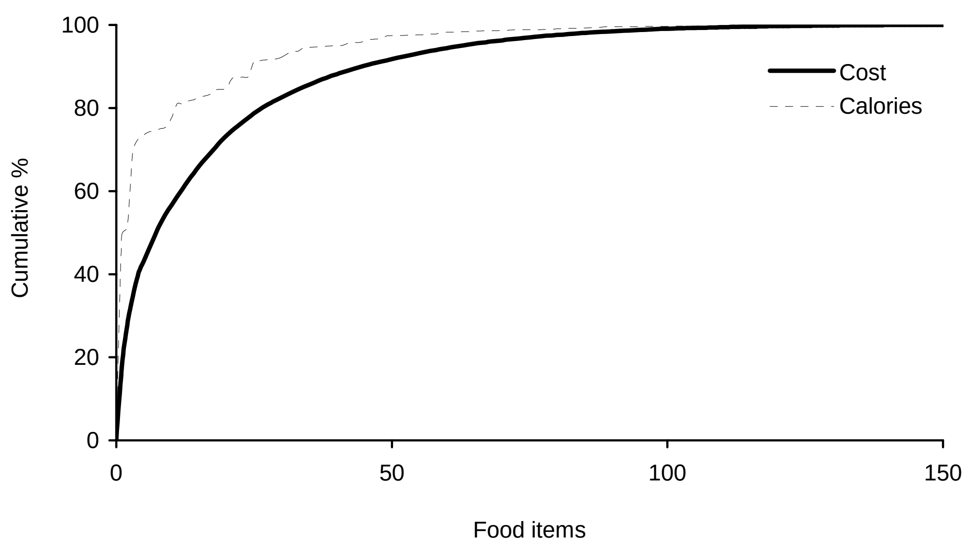
a. Unit values (that is, the ratio of expenditure to quantity) coming typically from a consumption recall but potentially also from individual transaction records in expenditure diaries,
b. Price surveys in community markets, such as those typically done by LSMS surveys,
c. Surveys of opinions about prices from either sampled households or community leaders, and
d. Existing price collection efforts, as might already be occurring for a Consumer Price Index.

In terms of the number of items to collect prices for, ideally there should be full coverage of all of items in the poverty line (if it is a CBN line with a specified basket of food) and all of the items specified in the consumption recall (if diaries are not used). The prices of key non-foods should also be collected even if an Engel method is used to scale the food poverty line up to the total poverty line (see equation (1)) without using any non-food prices.

This recommendation for the number of items to collect prices for switches attention to the issue of how many items to specify in a food poverty line (and the related issue of how disaggregated are the commodities in the expenditure questionnaire). One useful tool in this regard is the concentration curve for the foods in the poverty line basket. This curve starts with the most important food and plots the cumulative contribution to either the total cost or the total calorie content of the poverty line basket.

Figure 2 presents an example from Cambodia, where the poverty line calculated from a 1993/94 survey had 155 separate food items. This detailed food basket was never fully priced in subsequent surveys, which only gathered data on the prices of about 30 foods. In fact this more abbreviated level of price collection would have been an appropriate level of detail for the poverty line food basket. According to Figure 2, a basket with just the 20 most important items would give 73 percent of the total cost and 85 percent of the total calories in the food poverty line. A basket with 45 items would give 90 percent of the total cost and 96 percent of the calories. In other countries there may need to be more foods in a poverty line basket, depending on the importance of the basic staples, but constructing curves like Figure 2 would be a sensible first step for designing both the poverty line basket and the level of detail required in food price surveys.

**Figure 2:** Concentration curves for poverty line food basket



In terms of how many observations to make on the price of each item, the standard in most LSMS surveys is three observations per village (that is, per cluster). It is not clear if a fixed number of observations per item is the best approach, although it does have the advantage of simplicity. A CBN food poverty line is a statistic (essentially a weighted average of a set of average prices) although it is rare to see standard errors reported for poverty lines. This statistic would be more precisely estimated if the prices for the items contributing the most weight were based on larger samples than the samples used to measure the price for minor items.

The variability across time and space should also be considered when deciding how many observations to take on the price of each item. Some items may be subject to price controls (for example, fuels) so the same price might be observed over all outlets and across short time spans. Other items, and particularly informally marketed foods, may have prices that vary from day to day and from seller to seller, so more observations are required to precisely measure the prices for such items.

In terms of where to collect prices, the aim should be to observe prices in the markets actually used by the households in the sample. Thus it is a valuable addition to a household questionnaire to enquire of respondents where they actually buy the items they consume. Otherwise an approach of just visiting markets and asking vendors the price of particular goods (as was done by the LSMS surveys) can be subject to certain criticisms. In particular, the prices that are gathered may be from the wrong market, or for the wrong specification of goods, or the prices quoted may not be the prices actually paid by local residents because of bargaining (Deaton and Grosh, 2000).

In terms of the geographical scale at which to calculate average prices (as an input to the poverty lines), most surveys report these for only a few major regions despite prices being collected from a far larger number of communities. Consequently these regional average prices will overstate the cost of buying the poverty line basket of foods in some communities within each region, while understating it for others. Measured poverty will be too high in the communities where regional average prices overstate the cost of the basket of foods because these same (high) prices are not used for valuing food consumption. Hence, the value of some households' consumption will be above the poverty line if that line is priced using local (i.e., cluster-level) prices, but below the poverty line if regional average prices are used. Bias in the opposite direction (measured poverty too low) will occur in clusters where regional average prices understate the local cost of the poverty line basket of foods.

It would seem that there is no net effect of using regional average prices because the overstatement of poverty in some communities within the region is cancelled out by the understatement in others. This would only be true if the distribution of food prices within each region is symmetric, with the mean equalling the median (e.g., a Normal distribution). However, if the within-region distribution of prices is positively skewed, with the mean exceeding the median, there will be fewer communities with prices above the regional average than below the regional average. Consequently there will be more communities where poverty is overstated than understated. Hence it is important to examine the within-region distribution of prices. It would also be a useful sensitivity analysis to calculate poverty lines and poverty rates using cluster level prices to see how they differ from the estimates based on regional average prices.

Surprisingly little is known about the last practical question, of what is the best way of collecting price information. The available choices are community price surveys, unit values, price opinions and relying on existing price collection efforts. Unit values are often used in poverty studies because very few HIES and HBS collect local prices when gathering household expenditure data. Many of these surveys do collect food quantities in addition to expenditures so that unit values can be calculated. But unlike prices, unit values are available only for purchasers. Furthermore, they are subject to quality effects if some households buy better varieties within a commodity category. The final problem with unit values is that they reflect measurement errors in quantities, expenditures, or both. There is no consensus about the use of unit values in poverty studies. Deaton (1997) reports evidence from India that indicates that unit values are a reasonable proxy for prices whereas Capéau and Dercon (1998) find that the poverty rate in Ethiopia is overstated by 20 percent when unit values are used instead of prices.

One detailed experiment compared both unit values and price opinions against a standard of prices gathered by surveying local stores and markets (Gibson and Rozelle, 2005). These market surveys were argued to provide a good standard in their setting (Papua New Guinea) because there is no haggling, local markets are well defined and geographically separated, and there is not much quality variation amongst goods across the various markets. The price opinions were obtained by showing respondents in the sampled households photographs of a variety of different items and asking them their opinion about the price of the same items in local markets. Other surveys, such as the IFLS in Indonesia, obtain data in a somewhat similar way but only ask key informants (such as the head of the local women's group) rather than all sampled households and don't necessarily use photographs to aid the recall.

The results of this experiment show that estimated poverty rates are considerably overstated when unit values are used to construct the poverty lines. For example, when unit values are used the head-count index is estimated to be 28 percent rather than the actual figure (based on market prices) of 22 percent (Table 4). This difference is statistically significant. In contrast, when the price opinions are used, there is only a slight overstatement of the poverty rates. The price opinions in this experiment took about two hours per cluster to collect, which was somewhat shorter than the time taken to gather the prices from local stores and markets. Thus, relying on informed opinions about prices may be an economical and reasonably accurate way of obtaining local prices, although more experiments would be needed to establish this.

Table 4: Poverty measures with different method of collecting prices, 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) |

*Source:* Gibson and Rozelle (2005).
*Note:* 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.

The final choice, of relying on existing price collection efforts, is unlikely to work in many settings. The Consumer Price Index in many countries relies almost solely on urban prices, so these would not be applicable for calculating either poverty lines or spatial deflators and for imputing the value of consumption for rural households. Moreover, as explained above, the commodity weighting in a CPI is much more towards the consumption pattern of richer households, so the index values are unlikely to be relevant to poverty-related analysis.

Given the need for price data and the concerns about both unit values and relying on existing price collection efforts, it would be worthwhile for statistical agencies to invest more effort in gathering prices from local stores and markets and opinions about prices when their household surveys are fielded.

### 5.4 Assessing individual welfare and poverty from household data

Poverty is experienced by individuals, but information on total consumption can only be collected from households. While individual income data are regularly collected, they are not useful for poverty measurement until further assumptions are made about sharing within households. Thus the usual method of measuring poverty is to count the number of (or sum the poverty gaps for) people whose collective household consumption expenditure (or income) is below the poverty line. Results may be presented on an individual basis by weighting by household size. But the calculations are still fundamentally household-based. The disconnect between the level at which data are collected compared with the level at which analysis is desired raises two questions:

- Are there reliable methods of observing whether some types of individuals within households, such as women or the elderly, are poorer than others in the same household?
- How should adjustments be made for differences in household size and composition when determing individual welfare and poverty status based on household data?

The literature on intra-household inequality addresses the first issue. This literature has yet to make much impact on the activities of statistical offices, partly because of the practical difficulties involved. The second issue, which is addressed by the literature on equivalence scales, is more widely recognised by statistical offices. Indeed, approximately 30 of the countries providing metadata in the Statistical Addendum make some allowance for equivalence scales when setting poverty lines and measuring poverty. Because of this wider use by survey agencies, and also because equivalence scales have a longer history, they are discussed first.

5.4.1 Equivalence scales

A common method of taking account of households of differing size and composition is to convert each household into a number of equivalent adults, $N_e$, using a formula like:

$$N_e = (A + \varphi C)^{\theta} \qquad \varphi \leq 1,\ \theta \leq 1 \tag{9}$$

where the household is comprised of A adults and C children. The parameter φ is the adult-equivalence of a child, and the parameter θ reflects possible economies of scale favoring larger households due to the allocation of fixed costs (such as heat and light) over a greater number of people. For example, the Luxemburg Income Study calculates adult equivalents by taking the square root of household size, so $\varphi=1$ and $\theta=0.5$. In developing countries, per capita consumption (or income) is widely used as the welfare measure, so $\varphi=1$ and $\theta=1$. This implicitly assumes that it is as costly to provide for a child as an adult, and that the cost of living for, say, ten people is ten times the cost for one person. Both assumptions are likely to be contentious.

It would be desirable to have simple and reliable methods for estimating φ and θ. However, empirical data alone cannot reveal equivalence scales. For example, knowing the consumption patterns for households with different numbers of children is not enough information for estimating child costs, φ. As Pollak and Wales (1979, p.216) note:

> "The expenditure level to make a three-child family as well off as it would be with two children and $12,000 depends on how the family feels about children. Observed differences in the consumption patterns of two- and

> three-child families cannot even tell us whether the third child is regarded as a blessing or a curse."

More formally, the problem is one of under-identification. It is possible to construct two different cost functions that a household faces to reach a given utility level and derive the same demand function from each one (Deaton, 1997). These different cost functions can embody different attitudes of parents toward their children and different elasticities of cost with respect to household size. Accordingly, observed demands do not provide sufficient information to identify either the costs of children or their related economies of scale.

Additional assumptions are needed to identify equivalence scales from observed data on household consumption patterns. One approach is based on what is sometimes called Engel's second law, the assertion that the food share is an inverse indicator of welfare across households of different sizes and compositions. There is no theoretical justification for this Engel approach. Moreover, its empirical results are highly sensitive to the measurement errors associated with certain data collection methods (Gibson, 2002). Thus, even though the approach is sketched below, it is not recommended.

Another approach is known as the Rothbarth method, where identification is obtained from the assumption that adults' standard of living is indicated by the value of expenditure on "adult goods" (goods not consumed by children). This approach can measure the costs of children but not economies of scale. The equivalence scales from this method are typically smaller than those calculated from the Engel method, which is known to overstate the costs of children (Nicholson, 1976).

A third method compares the recommended daily allowances of nutrients, and especially calories, for different age and gender groups to determine the adult equivalence of a child. This assumes that relative food needs for children are the same as relative non-food needs, which seems unlikely. Moreover, some controversy surrounds the definition and use of nutrient requirements because it is not clear whether the lower requirements for, say, women reflect lower needs or just the adaptation to receiving less by a historically discriminated-against group (Sen, 1984).

Given these limitations, an appropriate goal for many statistical agencies may simply be to use equation (9) to carry out sensitivity analyses, trying different values of $\varphi$ and $\theta$ to see whether any conclusions reached previously using per capita consumption are overturned. This approach has highlighted, for example, that people in widow-headed households in India are more likely than people in other households to be poor only if economies of scale are important (Dreze and Srinivasan, 1997). Thus, conclusions about gender and poverty may have to be conditioned on assumptions about economies of scale.

An important detail when using equivalence scales, either for sensitivity analysis or for the main poverty calculations, is that the scales should be applied symmetrically to both the poverty line and the welfare indicator. This follows from the fact that the poverty line is just a point (or a threshold) on the distribution of the welfare indicator, and thus should be subject to the same measurement definitions. An example of this point is provided in Table 5, based on results for Papua New Guinea where the main analysis was based on equivalence scales with φ=0.5 and θ=1, and with children being defined as those aged from 0-6 years.[40]

Table 5: Effect of Assumptions about Adult Equivalence and Scale Economies on Calculated Poverty Rates in Papua New Guinea

| Adult equivalence | Scale economies | Mean expenditure[a] | Poverty Line[a] | Headcount Poverty Rate | | |
|---|---|---|---|---|---|---|
| | | | | National | Urban[b] | Rural[b] |
| *With no adjustment to poverty line* | | | | | | |
| φ=0.5 | θ=1.0 | 911 | 399 | 30.2 | 11.4 (5.7) | 33.5 (94.3) |
| φ=0.5 | θ=0.5 | 2173 | 399 | 3.7 | 0.4 (1.7) | 4.3 (98.3) |
| *Adjusting the poverty line* | | | | | | |
| φ=0.5 | θ=0.5 | 2173 | 1016 | 30.2 | 6.5 (1.3) | 34.4 (6.7) |

Note:
[a] In Kina per year per adult equivalent (or *effective* adult equivalent) when scale economies are assumed.
[b] Shares of national poverty in ( ).
Source: Authors calculation from Papua New Guinea Household Survey data.

If scale economies are introduced by just dividing household expenditure by $n^{0.5}$ -- the effective number of adult equivalents – the estimated poverty rate falls dramatically from 30.2 to 3.7 percent. The reason is that the effective size of all households with more than one member (99.5 percent of the population) falls, giving apparently higher living standards to almost everyone. This approach is flawed. The poverty lines are based on the consumption patterns of households with an average of almost six members. The poverty

[40] The estimate of φ=0.5 was the (rounded) average of the results from using the Engel and Rothbarth methods and the Recommended Daily Allowance of calories for children and adults (World Bank, 1999).

lines would be higher if they were just based on single-person households because of the diseconomies of living alone.

Ideally, all calculations used to derive the poverty line should rely on the same equivalence scale applied to the welfare indicator. But in the absence of this comprehensive approach, the poverty line may be adjusted in the following manner:

(i) find a household of average size and composition whose per capita expenditure is equal to the poverty line, and

(ii) set the adjusted poverty line equal to whatever value their per-effective-adult-equivalent expenditure becomes after the introduction of the equivalence scale.

This rule ensures that a household of average size and composition remains above or below the poverty line irrespective of the choice of equivalence scales (Dreze and and Srinivasan, 1997). In the example in Table 5, this adjustment raises the poverty line from K399 to K1016, and the national poverty rate returns to the previously calculated level of 30.2 percent.

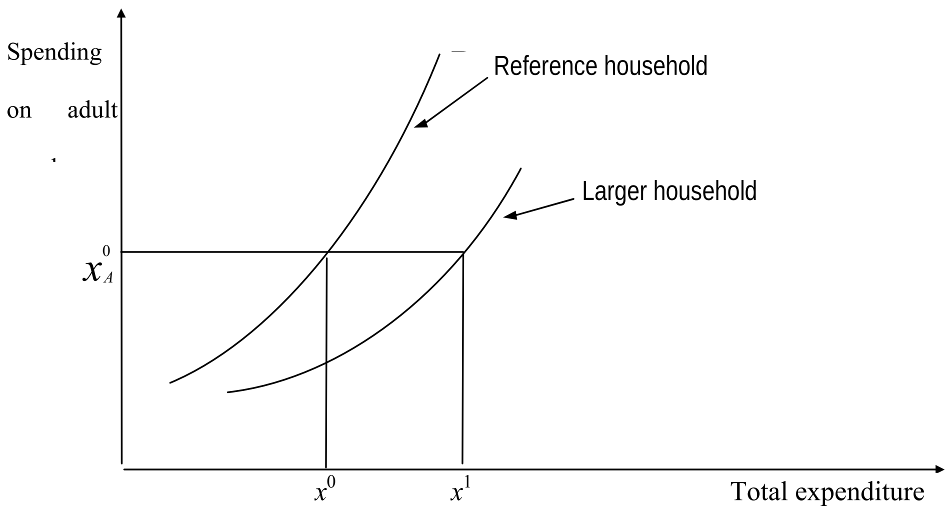
Once a similar equivalence scale is applied to both the welfare indicator and the poverty line, the main effect of assumptions about child costs and scale economies should be to alter the poverty profile, rather than the aggregate poverty measurements. The poverty profile for any characteristics associated with differences in household composition and size, such as sector of residence, and the age and marital status of the household head are likely to be sensitive to assumptions about equivalence scales. Thus, in the example in Table 5, the introduction of scale economies reduces the share of poverty in urban areas from 5.7 percent to 3.3 percent because urban households are larger than rural households (7.0 versus 5.7 residents).

5.4.2 The Rothbarth method of measuring child costs

The Rothbarth method of measuring child costs starts, somewhat paradoxically, with expenditures on goods that are not consumed by children--for example alcohol, gambling, and tobacco. Expenditures on these goods should fall when children are added to the household. Children bring additional consumption needs without any offsetting increase in income so there is effectively less income available for the adults to spend on these "adult goods". Moreover, unlike other goods such as food, it is possible to rule out a direct demand effect causing increased spending on these goods since child don't gamble, smoke, and drink alcohol. Therefore, the cost of a child can be measured by calculating the amount of compensation that would have to be paid to parents to restore expenditure on adult goods to the former level before the child was added to the family.

The Rothbarth method is illustrated in Figure 3, showing the relationship between total household expenditure and expenditure on adult goods. Spending on adult goods rises as total household expenditure increases, according to the schedule AB. For a reference household composed of two adults, total expenditure is $x^0$ and adult goods expenditure is $x_A^0$. In comparison, a two-adult and one-child household spends less on adult goods at the same level of total outlay because of the competing needs of the child. The household would require total outlay of $x^1$ to restore adult goods spending to its previous level. Thus, $x^1$-$x^0$ is the cost of the child and its adult-equivalence is ($x^1$-$x^0$)/($x^0$/2).

**Figure 3:** Rothbarth method for measuring child costs




A major difficulty in implementing this method is finding a set of valid adult goods. It is essential to first specify the appropriate consumption categories when designing surveys. This means, for example, separating adult clothing from children's rather than men's from women's. But even with a good number of candidate adult goods, it is necessary to test that they meet the appropriate statistical requirements.

One test uses the insight that because the child acts like a reduction in income, the reduced expenditure on each individual adult good ought to be in proportion to the marginal propensities to spend on each good (Deaton, 1997). This test can be implemented using the concept of an "outlay equivalent ratio" (also used below in the discussion of intra-household inequality), which can be obtained from an estimated regression of the budget share equation for a good:

$$w_i = \frac{p_i q_i}{x} = \alpha_i + \beta_i \ln\left(x/n\right) + \eta_i \ln n + \sum_{j=1}^{J-1} \gamma_{ij}\left(n_j/n\right) + u_i \tag{10}$$

where the product (of price and quantity) $p_i q_i$ is the expenditure on good $i$, $w_i$ is its budget share, $x$ is the value of total household consumption expenditure, $n$ is total household size, $n_j$ is the number of people in the $j^{th}$ demographic group, and $u_i$ is a residual. Coefficients from equation (10) can be used to calculate outlay equivalent ratios for each good:

$$\pi_{ir} = \frac{(\eta_i - \beta_i) + \gamma_{ir} - \sum_{j=1}^{J} \gamma_{ij}\left(n_j/n\right)}{\beta_i + w_i} \tag{11}$$

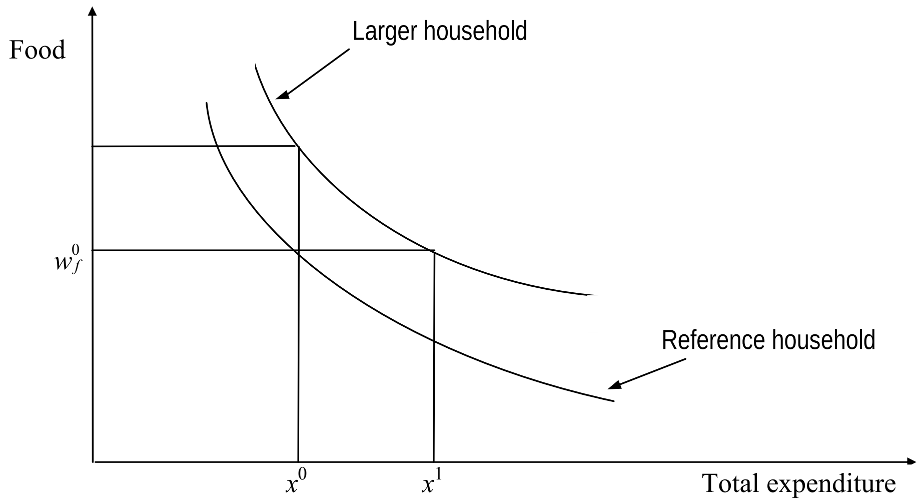
(sample means can be used for the $w_i$ and the $n_j/n$ ratios). The ratio, $\pi_{ir}$ measures the effect of an additional person of type $r$ on the demand for good $i$ in terms of the percentage change in outlay (expenditure) per person that would have been necessary to produce the same effect on demand. For any particular type of child group (say, 0-6 year-old boys) the outlay equivalent ratios should be the same across a set of valid adult goods (subject to sampling variability).

Once a set of adult goods have been identified, equation (10) can be used to find the budget share and expenditure on an adult good for a reference household. In principle, this can be calculated with a single adult good. But improved statistical precision may occur if all of the valid adult goods are aggregated into a combined category. The equation is then used to recalculate the adult goods expenditure after a child is added to the reference household. The final step is to calculate how much household total expenditure would have to increase to restore adult goods expenditure to its initial level. For example, a poverty assessment in Papua New Guinea used this approach and found that adding an older child (7-14 years of age) to a 2-adult household would require a 31 percent increase in total expenditures to restore adult goods expenditure to their previous level (World Bank, 1999). Thus, the adult-equivalence of an older child was approximately 0.6 of an adult.

### 5.4.3 The Engel method of measuring child costs

Figure 4 shows how the Engel method works. The food share is plotted against total household expenditure for a reference household with two adults and for a household that also has a child. At any given level of total expenditures, for example $x^0$, the household with children has a larger food share than does the reference household. Assuming that the food share is an inverse welfare measure across household types, individuals in the household with children appear worse off. The household with children would need total expenditures of $x^1$ to have the same food budget share, and thus the same welfare level as the reference household. Therefore, $x^1$-$x^0$ is a measure of child costs and the adult-equivalence of a child is $(x^1\text{-}x^0)/(x^0/2)$. This can be worked out from the parameters of a food Engel curve like equation (10).

**Figure 4:** Engel's method for measuring child costs




The Engel method overstates the cost of children. The family's food budget share will rise even if the parents are given the exact amount of money needed to provide for the child while maintaining their own consumption. The rise in the food share occurs because the child's consumption is concentrated more on food than is the consumption of the parents. But under the logic of the Engel method, this rise in food share indicates a need for further compensation, which amounts to an over-compensation (Nicholson, 1976).

5.4.4 The Engel method of measuring scale economies

Larger households devote more of their budgets to food than do smaller ones, holding total outlay constant. In this respect, they are like households that have children (whose consumption is concentrated more on food than is the consumption by adults), so the Engel method of measuring child costs is readily adapted to the measurement of scale economies. For example, the approach illustrated in Figure 4 could be extended by plotting a family of Engel curves and calculating the extra expenditure ($x^k$-$x^0$) needed for households of size $n^0$+$k$ (where $k$=1,2,3,…).

The regression approach in equation (10) can also be adapted, using $n^\theta$ instead of $n$ as the measure of household size. Thus, if $x^0$ is the outlay of a one-person household, an $n$-person household of the same composition needs a total outlay of $x^0n^\theta$ to have the same food share (and the same welfare level, by assumption). For example, Lanjouw and Ravallion (1995) estimate $\theta$ to be 0.6 in Pakistan, so if 10 individuals formed a 10-person household, their per-capita food spending could decline by 60 percent and according to the Engel method they would still have the same level of welfare ($10^{0.6}$=3.98). These large estimates of scale economies have attracted some criticism because they imply improbably large reductions in food spending by consumers in a poor country with considerable under-nutrition (Deaton, 1997).

Unfortunately, the Engel method makes no more sense for measuring scale economies than it does for measuring child costs. Consider a larger household with the same per capita expenditures as a smaller household. If there are scale economies, the larger household is better off. Thus, according to Engel's second law, the larger household should have a lower food share. But a decline in the food share with constant per capita expenditures can occur only if there is a decline in food spending per person. It is very unlikely that people who are better off would spend less on food, especially in poor countries where nutritional needs are not being met.

In addition to this conceptual problem, the Engel method does not give robust empirical estimates of scale economies. In an experiment in Port Moresby, the capital city of Papua New Guinea, the Engel estimate of θ was 0.76 (and not statistically significantly different from 1.0, implying no scale economies) for a half sample whose expenditures were surveyed with diaries. However, in the other half-sample, where a recall survey was used, the Engel estimate of θ was 0.41, implying large economies of scale (Gibson, 2002). This evidence is problematic because estimates of scale economies should not depend on the method used to gather expenditure data. The conceptual and empirical problems with the Engel method suggest that it is a statistical tool that should not be used for poverty measurement.

5.4.5 Adjusting poverty statistics when adult equivalents are units

Poverty gap measures may need modifying when the welfare indicator and poverty line are measured in adult equivalent rather than per capita terms. The standard FGT formula uses the number of people, N, and the number of poor people, Q: $P_1 = \frac{1}{N}\sum_{i=1}^{Q}\left((z-y_i)/z\right)$. The monetary poverty gap can be calculated as: $P_1 \times N \times z$, but this will exaggerate the cost of closing the gap when adult equivalents are used.

For example, consider a two-adult and two-child household with total annual expenditure of $1,200. The poverty line is set at $500 per adult equivalent, and children count as 0.5 adults. Comparing expenditure per adult equivalent ($400) with the poverty line indicates an average gap of $100 and a $P_1$ measure of 0.2. If $P_1$ is multiplied by $N \times z$, the aggregate value of the poverty gap will appear to be $400. But in fact it is only $300. One way to prevent this overstatement is to estimate all poverty measures using adult equivalent numbers rather than person numbers, even though these may not be the most familiar units for communicating the results. An alternative is to use correction formulae suggested by Milanovic (2002).

### 5.4.6 Methods for estimating the intra-household allocation of consumption

Several procedures have been suggested for using household data to see if some types of individuals are poorer than others within the same household. There has been limited success with these procedures, and it is likely to be several years before statistical offices would consider routinely applying them. Nevertheless, greater awareness of these procedures may be helpful, especially if it leads to the collection of data that are better suited to the needs of these methods.

In the Rothbarth method of measuring child costs discussed above, children exert negative income effects on the demand for adult goods. If some types of children have larger income effects than others, it may provide evidence of a gender bias within the household. For example, if outlay equivalent ratios (see equation (11)) are more negative for boys than for girls, this suggests that parents cut back their own consumption more when a boy is added to the household than when a girl is added (Deaton, 1989).

Unfortunately, most applications of the adult goods method have produced puzzling results, sometimes finding bias against girls in locations where it is not expected and no bias in places where other evidence strongly suggests that boys are favoured (Deaton, 1997). It is possible that part of this failure reflects the coarseness of the data collected in many household surveys, which have rather few adult goods disaggregated. In one of the few applications where the method worked as expected, the questionnaire had contained a set of well-defined categories for adult goods because the test for gender bias had been planned when the data were collected (Gibson and Rozelle, 2004).

It is harder to study unequal allocations between adults because differences in demand – even if observed at an individual level – may just reflect differences in preferences. These differences in preferences can be ignored when the adult-goods method is applied to children who exert only income effects (because they don't consume the goods themselves). Some headway in identifying "sharing rules" for the allocation of consumption between adults in the same household has been made by Bourguignon and Chiappori (1992). They identify the sharing rule using either "assignable goods" or "exclusive goods". An assignable good is a private good (that is, a good where the consumption by one person subtracts from the consumption of another) whose consumption by each member of the household can be observed. An exclusive good is a private good used by only one member of the household. Progress in applying these methods may be aided by household surveys that use diaries for each adult, rather than household level reporting, and that also collect information on whether purchases are destined to be consumed by the purchaser (the diary-keeper) or someone else (Browning et al, 2003).

### 5.4.7 Collecting non-monetary data on individuals to estimate gender-specific measures of poverty

Most household surveys collect information on non-monetary welfare indicators such as education and, less frequently, health. These data are usually collected for each individual in the household and offer the possibility of assessing individual poverty, at least in a non-monetary sense. Comparisons of educational attainment and participation for women relative to men are regularly made with such data. Comparisons of health status can also be made, especially using anthropometric data. It has also been claimed by Case and Deaton (2002) that self-reported data on health can prove useful.

In these surveys of self-reported health, respondents are asked to rate their overall health status on a 5-point scale, ranging from "excellent" to "poor." There can be a considerable amount of adaptation to poor living conditions, which hampers comparisons of self-rated health across communities and countries. But within individual communities, comparisons are not affected by this adaptation, and these comparisons suggest that women's self-rated health is worse than men's (Case and Deaton, 2002). At least in the short-run, there may be more success at understanding the gender dimensions of poverty using broader health and education measures than there is from attempting to untangle consumption of individuals within the household.

## 5.5 Conclusion

This chapter described some of the methods and data available for measuring poverty with cross-sectional household surveys. These surveys are the workhorse of poverty analysis, in part because all countries have at least one cross-sectional household survey of one type or other. Many of these surveys were not originally designed with poverty measurement as a key aim and have certain features, such as short reference periods and limited topical coverage, that limit their usefulness as a source of data for understanding poverty. However, some modest modifications of survey practice that are suggested in this chapter could improve the quality of poverty measurements coming from these surveys.

Another theme of the chapter is the need for consistent survey methods so that poverty comparisons uncover real changes in the population rather than artifacts that are due to variation in survey design. Examples from India and Cambodia show how large the effects seemingly minor variations in survey design can have on poverty estimates. It would be a welcome addition to current practice if all statistical agencies carried out detailed experiments to assess the effect on measured poverty rates when they change survey methods, so that adjustment factors can be calculated and robust poverty trends retrieved.

Sensitivity of poverty estimates to variations in survey design also highlights the importance of measurement error. A previous emphasis on means and totals as the statistics of interest may have lulled some survey agencies into a belief that random measurement errors do not matter so long as they cancel out. However this is not true in the context of poverty and inequality measurement. Accordingly, statistical agencies should increase the efforts made to improve the accuracy of their household survey data.

## References


Bourguignon, F. and Chiappori, P.-A. (1992). Collective models of household behaviour: an introduction. European Economic Review, 36, 355-364.

Browning, M., Bonke, J., and Uldall-Poulsen, H. (2003). The intra-household allocation of expenditures: new survey evidence from Denmark. The 2003 IAREP Workshop on Household Economic Decisions: Earning, Sharing, Spending and Investing Money, 5-7 December, Agder University College, Kristiansand, Norway.

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). Oxford : Oxford University, Centre for the Study of African Economies.

Case, A. and Deaton, A. (2002). Consumption, Health, Gender and Poverty. (Mimeo). Princeton, NJ : Princeton University, Research Program in Development Studies.

Central Statistical Office (1995). Republic of Zambia Household Budget Survey, 1993-94, Volume 1 – Main Tables and Report Lusaka.

Deaton, A. (1989). Looking for boy-girl discrimination in household expenditure data. World Bank Economic Review, 3(1): 1-15.

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

Deaton, A. (2001). Counting the world's poor: Problems and possible solutions. World Bank Research Observer, 16(2), 125-147.

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

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). Review of Economics and Statistics, 87(1): 1-19.

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

Deaton, A. and Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. (Living Standards Measurement Study, working paper no. 135). Washington, D.C. : The World Bank.

Dreze, J., and Srinivasan, P. V. (1997). Widowhood and poverty in rural India: some inferences from household survey data. Journal of Development Economics, 54(2), 217-234.

Filmer, D., and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data - or tears: An application to educational enrollments in states of India. Demography, 38(1), 115-132.

Gibson, J. (1999). Can women's education aid economic development? The effect on child stunting in Papua New Guinea. Pacific Economic Bulletin, 14(2): 71-81.

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. (2001). Measuring chronic poverty without a panel. Journal of Development Economics, 65(2), 243-266.

Gibson, J. (2002). Why does the Engel method work? Food demand, economies of size and household survey methods. Oxford Bulletin of Economics and Statistics, 64(4), 341-360.

Gibson, J., Huang, J. and Rozelle, S. (2003). Improving estimates of inequality and poverty from urban China's Household Income and Expenditure survey. Review of Income and Wealth, 49(1), 53-68.

Gibson, J. and Rozelle, S. (2004). Is it better to be a boy? A disaggregated outlay-equivalent analysis of gender bias in Papua New Guinea. Journal of Development Studies, 40(4): 115-136.

Gibson, J. and Rozelle, S. (2005). Prices and unit values in poverty measurement and tax reform analysis. World Bank Economic Review, 19(1): 69-97.

Grosh, M. and Glewwe, P. (2000). Designing Household Survey Questionnaires for Developing Countries, Washington, D.C. : The World Bank.

Howes, S. and Lanjouw, J. (1998). Does sample design matter for poverty rate comparisons. Review of Income and Wealth, 44(1), 99-110.

Jolliffe, D. (2001). Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. Journal of Economic and Social Measurement, 27(1), 1-23.

Jorgenson, D. (1998). Did we lose the war on poverty? Journal of Economic Perspectives, 12(1), 79-96.

Korn, E. and Graubard, B. (1999). Analysis of Health Surveys New York: John Wiley and Sons.

Lanjouw, J. and Lanjouw, P. (2001). How to compare apples and oranges: Poverty measurement based on different definitions of consumption. Review of Income and Wealth, 47(1), 25-42.

Lanjouw, P. and Ravallion, M. (1995). Poverty and household size. The Economic Journal, 105 (433), 1415-1434.

Milanovic, B. (1979). Do we tend to overestimate poverty gaps? The impact of equivalency scales on the calculation of the poverty gap. Applied Economics Letters, 9(1), 69-72.

Mukherjee, S. (2005) On asset indices. *Mimeo* Department of Economics, Princeton University.

Neter, J. and Waksburg, J. (1964). A study of response errors in expenditure data from household interviews. Journal of the American Statistical Association, 59(305): 18-55.

Nicholson, J. L. (1976). Appraisal of different methods of estimating equivalence scales and their results. Review of Income and Wealth, 22(1), 1-11.

Pollak, R. A., and Wales, T. J. (1979). Welfare comparisons and equivalence scales. American Economic Review, 69(2), 216-221.

Pradhan, M. (2001). Welfare Analysis with a Proxy Consumption Measure: Evidence from a Repeated Experiment in Indonesia. (Mimeo) Amsterdam : Free University.

Pradhan, M., Suryahadi, A., Sumarto, S., and Pritchett, L. (2001). Eating like which "Joneses?" An iterative solution to the choice of a poverty line reference group. Review of Income and Wealth 47(4): 473-487.

Prais, S. (1958). Whose cost of living? Review of Economic Studies 26(1): 126-134.

Ravallion, M. (1994). Poverty Comparisons. Chur : Harwood Academic Publishers.

Ravallion, M. (2003). Measuring aggregate welfare in developing countries: how well do national accounts and surveys agree? Review of Economics and Statistics, 85(3): 645-652.

Sahn, D. E., and Stifel, D. (2000). Poverty comparisons over time and across countries in Africa. World Development, 28(12), 2123-2155.

Scott, C. (1992). Estimation of annual expenditure from one-month cross-sectional data in a household survey. Inter-Stat, 8, 57-65.

Scott, C. and Amenuvegbe, B. (1991). Recall loss and recall duration: an experimental study in Ghana. Inter-Stat, 4(1), 31-55.

Scott, K., and Okrasa, W. (1998). Analysis of Latvia Diary Experiment. Washington, D.C. : World Bank, Development Research Group.

Sen, A.K. (1984). Family and food: sex bias in poverty. in Resources, Values and Development, Oxford University Press, Delhi.

Slesnick, D. (1993). Gaining ground: poverty in the postwar United States. Journal of Political Economy. 101(1), 1-38.

Tarozzi, A. (2004). Calculating comparable statistics from incomparable surveys, with an application to poverty in India. Paper presented at the North East Universities Development Consortium Conference, Montreal, October 2004, available at http://www.hec.ca/neudc2004/Programme2.html

UNICEF, (1998). State of the World's Children Report, Oxford University Press: New York.

U.S. Census Bureau. (2003). CENVAR [Software]. http://www.census.gov/ipc/www/imps/cv.htm

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