How to Monitor Poverty for the Millennium Development Goals

ANGUS DEATON
Angus Deaton is Dwight D. Eisenhower Professor of Economics and International Affairs at Princeton's Woodrow Wilson School and Department of Economics

Abstract   I consider two issues concerning how to monitor global poverty for the Millennium Development Goals, the selection of poverty lines, and the data sources for monitoring poverty over time. I discuss the choice of a single international line, converted using purchasing power parity (PPP) exchange rates, versus the use of country-specific poverty lines. I note the difficulties in constructing PPP exchange rates but argue in favor of a single international line, converted at PPP rates, which would be regularly updated using domestic price indexes. Re-basing, using updated PPP rates, would be done infrequently. For example, if the global poverty numbers were estimated annually, the PPP rates might be updated once a decade. In any case, it is important that the poverty estimates be calculated much more frequently than the PPP rates are revised. I discuss whether monitoring should be performed using national accounts data on income or consumption, supplemented by distributional data so as to make inferences about poverty, or using data from household survey data. I argue that data from the national accounts are not suitable for measuring poverty and that their use requires assumptions that are unlikely to hold. In particular, monitoring poverty through the national accounts runs the risk of pre-judging important issues that are properly the subject of measurement, not assumption, such as the extent to which aggregate growth benefits the poor. I argue that poverty should be directly measured using household survey data, and I discuss what needs to be done to enable such monitoring to be placed on a sounder basis.

Key words: Poverty, Monitoring, Measurement, Purchasing Power Parity, National Accounts, Household Surveys

Introduction

I consider two important issues in monitoring world poverty for the Millennium Development Goals (MDGs). The first section is concerned with the selection of poverty lines, whether they should be national or international;

and if the latter, how the line should be converted to local currencies. It concludes with a proposal for modifying current practice in which purchasing power parity (PPP) exchange rates are used to set baselines, but not for frequent updating. I also propose that the local equivalents of the international line be validated within countries. The next section concerns data sources, and whether monitoring should be carried out using national accounts and distributional data to infer poverty, or from direct measurement using household surveys. I argue in favor of household surveys, and again end with a series of proposals for survey work. The final section summarizes some current priorities.

How to draw the line: poverty lines for monitoring the MDGs

National or international lines?

The World Bank's world poverty counts use international poverty lines of approximately $1-a-day and $2-a-day at 1993 international purchasing power parity (PPP) prices. The notion behind such lines is a simple one; that for the purpose of the global counts, or indeed for monitoring poverty reduction, we need a common international standard. While a common money standard is certainly not the only possibility, the rhetoric of a $1-a-day line has been enormously successful, particularly for the first-world and international agency audience to which it is primarily directed.

Yet, as we shall see, the construction of such lines is a complex and difficult task, and the attraction of even the simple concept is apt to diminish once the details of its implementation are exposed to close inspection.

For most of its country work, the World Bank (like policy-makers in each country) uses the *national* poverty lines that exist for many countries, although not all. The MDG target, which involves cutting poverty rates by one-half, could conceivably be applied to poverty measures based on either national or international poverty lines. So there is a choice.

There are strong arguments for an international line, if only because of the proven appeal of something like the $1-a-day concept, even if the reality falls some way short of the rhetoric. To start with an area where there is a wide degree of agreement, no one would argue that the US poverty counts, as computed by the US Census Bureau, should be included in world counts or in the MDGs. Of course, the counts as computed by the World Bank do not include the rich countries, but the obvious interpretation of the exclusion is that the US poverty line is too high to be used in international comparisons, and that if a suitable international line were applied to the US, then there would be little or no poverty. Martin Ravallion's (1994) survey of international poverty lines shows that the US phenomenon is quite general. While national poverty lines do not vary with the level of Gross Domestic Product (GDP) among the poorest countries, after a little less than $3 per person per day, national poverty lines are higher when the country is richer. Evidence from surveys within rich countries also shows that people's own notions of the poverty line rise as they get better-off. This is one reason why some countries,

354

such as Canada, and the European Council of Ministers have formalized the process by defining a poverty line in relation to the mean or median of the income distribution. These relative poverty lines clearly make sense for each country considered in isolation. But they hardly correspond to a universal notion of a state of extreme poverty and deprivation that is recognizable irrespective of average living standards in each country.

Atkinson and Bourguignon (2001) have suggested a way of using information from both national and international lines within a common global framework. They propose a lexicographic treatment of absolute and relative poverty, focusing on the former — with a fixed common poverty line — among the poorest countries, and then switching to the latter — using national poverty lines — among better-off countries, including the richest. A relative view of poverty, focusing on participating in the life of the country in which one lives, is something that only becomes relevant once absolute deprivation has been dealt with. Atkinson and Bourguignon's synthesis provides an elegant way of thinking about global poverty in a unified way. In the context of the MDGs, however, it is absolute poverty that is relevant and the emphasis is on the absolute level of living standards rather than the relative income level that permits participation.

It might be thought that, for monitoring change, the choice of line is not very important in practice. The poverty MDG is defined in terms of poverty *reduction*, not in terms of the poverty *level*, so that even if national poverty lines are not comparable, they do not prevent us from looking at poverty reduction in each country using its own individual poverty line. But this argument will not take us very far. Suppose, to illustrate, that all countries have the same log-normal distribution of income, with the same degree of inequality, differing only in the mean of log income. Then the amount of growth needed to halve the poverty rate is larger in a country with a higher poverty rate. For example, in a country where the poverty line is equal to median income, the poverty rate is 50%. To cut that to 25%, mean income has to increase by two-thirds of the standard deviation of log income; if the standard deviation is 0.5 (the figure for consumption per head in India), we need income to increase by about one-third. For a country that starts at 25%, cutting poverty by one-half requires that mean income increases by only one-quarter; and if the country is fortunate enough to have a poverty rate of only 1%, cutting it by one-half needs only a 12.5% increase in mean income. Countries that are richer at the start, and have adjusted their poverty lines upward, have a stiffer task to meet the goal, but it is not a task that necessarily has much to do with eliminating the extreme poverty and deprivation with which we are concerned.

Which leads me to the position that *some* degree of international comparability is desirable, and that *some* kind of international poverty line is required. So we must either defend the $1-a-day line, or come up with something better.

The $1-a-day line has been criticized because it is unrelated to any clear conception of international poverty (Reddy and Pogge, 2002). Their alternative is to specify some minimum level of living in terms of "the resources necessary to achieve a salient set of elementary capabilities". Many national

poverty lines, in both rich and poor countries, are constructed in a way that is consistent with such an idea. For example, both the Indian and US poverty lines were originally set with reference to the ability to buy a minimal food bundle or number of calories; for comparative discussions of those histories, see Deaton (1997, pp. 141–144). However, a closer examination of these and other cases suggests that the link with food is more rhetorical than real. In both India and the US, the poverty line has been held constant in real terms, updated in nominal terms by a price index, and no attempt has been made to preserve the original link with food. And in both countries, updating according to the original methodology would generate poverty lines that are quite different from those currently in use. A more accurate interpretation of the history would be that the Indian and American poverty lines were originally accepted because they fell within the range of what was generally acceptable as a poverty line, a range that included, but was far from determined by, the incomes of those currently buying some minimally acceptable food bundle. These lines survived, not because of their link to any prescribed level of living, but because they have continued to fulfil a useful purpose for policy-makers and in the policy dialog more broadly. Many other lines could presumably have served the same function, at least for a time, and all lines are continually up for challenge by politicians, commentators, and expert panels. Proposals to re-set the lines to a meaningful poverty standard are frequently advanced, but rarely carry the day, in part because the original justification of the lines tends not to be convincing on close inspection, but also because it is so difficult to come to agreement on what it costs to maintain a minimal standard of living. A committee of experts sitting in judgment on what poor people need to consume, or worse still what they *ought* to consume, is a particularly objectionable mechanism for setting poverty lines.

By these criteria, the $1-a-day line has a good deal to commend it. On the one hand, it corresponds, at least approximately, to the national poverty lines of a number of the poorest countries. And this is surely the right place to look. Indeed, it is hard to think of a more appropriate definition for international poverty than being poor in the poorest nations, and we can rely on the political processes in these countries to throw up appropriate poverty lines. But just as important is that such lines are useful for the purposes to which they are put, in this case for monitoring by the inter-national community the number of very poor people in the world, and the $1-a-day concept has proven itself by the extraordinarily extensive use to which it has been put in the decade since it was first constructed. Indeed, its use in the MDGs is itself a testimony to its value.

If we accept the $1-a-day or $2-a-day lines as a reasonable starting point, then we need a method for converting them into the local currency equivalents that are required for measuring poverty in each country.

PPP price indexes

One thing that we cannot do is to convert a common international poverty line using market exchange rates. In poor countries, because labor is

356

relatively cheap (because poor people are poor), the prices of non-tradable goods, especially those with a high labor content, are low relative to the prices of tradeable goods. In consequence, the market exchange rate between the poor country and the US (for example), which is determined by trade, is unlikely to be an accurate measure of the cost of living difference between the countries. A dollar buys 50 rupees in the foreign exchange market, but food that would cost $10 in the US can be purchased for only 100 rupees, so that the implicit exchange rate is not 50, but 10. An Indian consumption level of 500 rupees a month would convert to only $10 a month at the official exchange rate, but would be worth $50 at the 'food' exchange rate. In consequence, using the official exchange rate to convert consumption or income would vastly exaggerate the difference in levels of living between the two countries, and more generally between poor and rich countries, systematically overstating poverty in the former.

This account also suggests the remedy, which is to take a bundle of goods consumed by poor people, and 'price it out' in each country. If the cost of the bundle of goods in India (say) is the poverty line, the cost of the bundle in the other countries gives the appropriate value of the common poverty line in those countries. Alternatively, we can select a base country and calculate Laspeyres price indexes for all other countries relative to that base, so that if we have something that is approximately a $1 per person per day poverty line for the base country, we can convert it into other currencies by multiplying by the price indexes. Such price indexes are examples of the PPP exchange rates that we need if we are to have a common international poverty standard.

The construction of Laspeyres indexes by 'pricing out' a poverty bundle can be formally linked to methods that rely on calculating the cost of a minimal standard of living, or indeed of achieving a minimal set of capabilities. Such cost-of-living price indexes can be approximated by pricing out the goods purchased by people at the relevant level of living, and such approximations are routinely relied upon by statistical offices who see their consumer price indexes in cost-of-living terms. The accuracy of approximation can be improved by calculating various 'superlative' index numbers (Diewert, 1976), of which Fisher's ideal index is one, and these are often used in one form or another by the statisticians who construct PPP price indexes.

General problems with PPP exchange rates

Because PPP exchange rates are price index numbers, we can use the extensive body of theory about price indexes to explore their strengths and weaknesses. In particular, it is important to separate the *general* problems that are common to all price indexes, and which are therefore inherent in any attempt to convert a common international poverty standard into local currencies, from problems that are *specific* to current PPP exchange rates that are used for poverty measurement, and which might be ameliorated by better practice.

357

If poor people the world over consumed a single commodity (rice, say), the construction of PPP exchange rates for poor people's consumption would be a simple matter. All that we would need would be the price in local currency. This is the case where the price index works the best, which is when it is unnecessary, because we have a *price*, not a *price index*. In reality, poor people consume a range of goods, whose relative prices are different in different places and times.

Patterns of consumption are different in different parts of the world, and often in different parts of the same country. There are several different conceptual bases for constructing price indexes. Perhaps the simplest is to price out a common bundle that contains the kinds of goods consumed by poor people around the world. Alternatively, we can price not a bundle of goods, but the cost of attaining a common standard of living. For price indexes appropriate for international counts of extreme poverty, this would be the living standard of the poorest, and there is nothing that stops us from thinking of that living standard in terms of capabilities. But the exercise is clearly a difficult one. We are asked to consider the conceptual experiment of how many pesos it would cost a landless laborer from Bihar to live as well (or as poorly) in Mexico as he/she lives in India. Once stated thus, it is not at all clear that we should expect any satisfactory answer.

The standard economic theory of cost-of-living indexes is cast in terms of utility theory, with price indexes defined by comparing the cost of attaining the same standard of living (utility) at a different set of prices. When people consume only one good, such as rice, the standard of living maps directly into rice consumption, and cost-of-living indexes are just the relative costs of a unit of rice. The same is true if everyone consumes the same bundle of goods (or composite commodity), always in the same proportions, but in different amounts. When there are many goods but the relative prices are the same in all countries, the cost of a bundle of goods in country A relative to its cost in country B will be the same no matter what bundle we choose, and in this case too there is a straightforward way to measure price indexes. But the reality is that the relative prices of goods are quite different in different countries. Consumption patterns are also quite different, presumably in part because of international differences in incomes and relative prices, but also because of differences in tastes. When tastes differ, the theory offers no consistent yardstick with which to measure the cost of living, and there is no reason why the relative costs of living in India and Mexico should appear the same to Indians as they do to Mexicans. For an international subsistence poverty line, we can perhaps (although only perhaps) ignore taste differences, specifying a basic level of human functioning, and calculate the minimum cost of achieving it in different countries. The extent to which such an index would be feasible is substantively different from currently existing indexes, and the properties it would possess in practice are topics that would be worth more research.

There is a wide variety of price indexes. The simple Laspeyres index outlined in the previous subsection is not always feasible — goods consumed by the poor in one country may not be available in another country — and

358

when it is feasible, it is unclear why one country rather than another should serve as the base. We could consider using each country in turn as base, but we will generally get as many different sets of PPPs as there are countries. These multiple indexes can then be 'reconciled' by various kinds of averaging, sometimes leading to 'better' index numbers according to some criteria. The important point here is the one noted long ago by Irving Fisher: that there does not exist any index number that satisfies all the 'reasonable' requirements that might be placed on it. For example, there is no guarantee that the price level in country A relative to country B is the reciprocal of the price level in country B relative to country A, or that if we go from country A to country B, from country B to country C, and from country C back to country A that we will finish up where we started. One desirable property can be guaranteed by choosing an appropriate formula, but only at the price of losing another. Price indexes are not prices. The constructors of PPP exchange rates, like those who are responsible for national price indexes, know this, and make compromises. But they are always open to criticism that *some* 'obvious' or 'elementary' requirement has been violated. There is no way of constructing a PPP index number that is immune to all such criticism.

These problems exist even in the (possibly) conceptually easier case of within-country comparisons. Until the early 1990s, India had two poverty lines: one for rural and one for urban households. Over the past decade, however, there have been different poverty lines for each state, and for each sector within each state. The methodology under which these lines were constructed and are updated is explained in Government of India (1993) and appears to be sound in principle. Yet the results have brought official Indian poverty measurement into something close to disrepute. Urban poverty lines are so much higher than rural lines that, in many states, the urban poverty rate is higher than the rural rate, something that few independent observers would accept. In addition to these unsatisfactory intrastate price indexes, the pattern of prices across states also appears to make very little sense. There is an internal Indian debate on price indexes that parallels the international debate about unsatisfactory PPP exchange rates. And at least part of the reason lies in the inherent difficulty of constructing such indexes.

Deaton and Tarozzi (2000) report the results of a project to re-calculate price indexes for Indian states and sectors for 1987–1988 and 1993–1994. The conceptual problems are immediately relevant in practice. If, for example, we wish to construct a price index for Kerala relative to Uttar Pradesh, we can start by calculating the bundle of consumption at or around the poverty line in Kerala, and then price it out in Uttar Pradesh. But this is immediately problematic, because some of the goods that are heavily consumed in Kerala (e.g. fresh fish, coconuts, and coconut oil) are rarely purchased in Uttar Pradesh. The few recorded purchases are at very high prices, because these goods are exotica in Uttar Pradesh, and cater only to a small minority. Conversely, *atta* is a basic staple in Uttar Pradesh, but rarely shows up in Kerala. The cost of the Keralan poverty bundle in Uttar Pradesh

is therefore not a very good measure of the relative costs of being poor in Uttar Pradesh relative to Kerala. If we start from Uttar Pradesh, we have the same problems in reverse, and we get an answer for the cost of living in Kerala relative to Uttar Pradesh that is not the reciprocal of the cost of living in Uttar Pradesh relative to Kerala.

One answer to these difficulties is to average the two; indeed, the geometric average is the Fisher ideal price index. There is a lot to be said for the Fisher index in this context; it moderates the extremes that we get from either of the two Laspeyres indexes and, because it is a superlative index in the sense of Diewert (1976), it captures at least some of the change in consumption patterns that would take place if the Keralan actually moved to Uttar Pradesh, or *vice versa*. But the Fisher ideal index is subject to criticisms of its own; see, for example, Reddy and Pogge (2002) who show that changes in quantities (driven by factors other than prices) can lead to counter-intuitive effects on the price index. Another partial solution that is sometimes proposed is geographical chaining. We can compare Kerala with Uttar Pradesh by passing through a sequence of neighboring states — from Kerala to Andhra, to Maharashtra, to Rajasthan, and eventually to Uttar Pradesh — calculating an index only for pairs of nearby states with consumption patterns that are more similar to those at the beginning and end of our journey, and then chaining the results. Such chained indexes also have many desirable properties. But they can also be criticized on the grounds that such an index has the odd feature that the price index of Uttar Pradesh relative to Kerala depends on what happens to prices and consumption patterns in Maharashtra, a state that is adjacent to neither.

There are no general solutions to the conceptual problems of constructing PPP price indexes. Those who construct them know this, and make compromises, at least some of which are undesirable. There is no alternative.

Specific problems of the PPP indexes used in the world poverty counts

The PPPs used in the world poverty counts originally came from the Penn World Tables, but more recent poverty counts have used PPPs prepared by the World Bank itself. Constructing PPP index numbers is a major undertaking, requiring each participating country to supply information that is not part of its usual activities, and not all countries have done so. The number of these 'benchmark' countries has increased over time, from 60 in the 1985 index to 110 in the 1993 version, but excluding China in all rounds, and India in the most recent. Apart from China, which is treated specially, non-benchmark countries have PPP exchange rates imputed to them through a regression procedure that effectively adjusts their official exchange rate by an amount that is similar to the adjustment for countries at a similar level of development for which benchmark data exist. This procedure, while clearly sensible, can also be subject to substantial error, and in some extreme cases in the past the Bank has published PPPs and associated poverty rates that

360

were clearly incorrect. There are also concerns that country statistical offices do not always take these tasks very seriously; the data for international price comparisons do not feed into domestic policy-making and they have no domestic constituency, so they tend to be assigned low priority and low funding.

The latest World Bank poverty counts, detailed in Chen and Ravallion (2001), show some very large changes compared with earlier counts, even when computed for the same country in the same year, and even on a broad regional basis, and the changes are mostly attributable to the switch to new PPP numbers. Some of these changes almost certainly reflect a move from worse to better data — it is surely better to have 115 countries than 65, and there have certainly been improvements within countries — but some are probably due to 'noise', essentially measurement error. Certainly the new figures paint a different picture of global poverty than do the old ones (see Deaton, 2001). And although it is hard to predict what might happen with a new round of PPP numbers — and it is possible that they will be much more stable from now on — it is surely undesirable to measure world poverty with methods that are so unstable and unreliable.

There are also concerns about the way the PPPs are constructed. Price indexes are essentially the product of *weights* and of *prices*, and there are questions about the appropriateness of both.

The PPPs in the current poverty counts are *consumption* PPPs, rather than the earlier *national product* PPPs, and this change is clearly an improvement. However, there remain questions about whether these consumption PPPs are really relevant for the poor. The consumption bundles of the poor are not the same as the average consumption bundle, and price movements in the latter can be different from price movements in the former; for example, if the relative price of food increases. Furthermore, national price indexes are also computed on a *plutocratic* basis, with weights that are proportional to *aggregate* expenditures, and not on a *democratic* basis, in which each household's own price index is averaged to get a national index. As a result, the national price index tends to be representative of relatively well-off households —Deaton (1998) calculates that the US Consumer Price Index is representative of a household at around the 75th percentile — further reinforcing the divergence between what we get (the national consumer price index) and what we want (a price index for the poor). This plutocratic bias will be moderated, at least to some extent, if the weights use household surveys in which rich households are under-represented.

There is also a concern about world commodity prices. Some foodstuffs that are important to the poor, such as rice, wheat, or sugar, have active world markets, in which prices are (in)famously volatile. If the world price of rice rises relative to the world price of wheat, for example, the consumption of countries where people eat rice will become more valuable at international prices relative to the consumption of countries where people eat wheat. With a constant real PPP poverty line, the rice-eating consumers will become less poor relative to the wheat-eating consumers. If markets

work well, so that the world prices are actually the prices that poor people face, this is what we want to happen; after all, the rice consumers always have the opportunity to switch to other commodities. But it is not clear that these world price fluctuations are indeed relevant for the majority of poor people in the world. Most rice in the world is eaten within a short distance of where it is grown and, for many rice consumers, the world price might as well be the price on Mars. The severity of this problem depends on just how the compilers of the PPPs go about collecting local prices, and the extent to which the prices used to construct the benchmarks are affected by fluctuations in world prices. The instability of actual PPPs from one revision to the next suggests that commodity price instability may be playing a role, but instability may also come from other sources. In any case, the ideal PPP would not only use weights that are appropriate for poor people, but would collect prices that are appropriate for them. Statistical offices rarely collect different prices for different people, but work on the premise that all consumers face the same prices.

What to do? A proposal

The PPPs currently used to covert the $1-a-day lines into local currencies suffer from all the generic problems of price indexes. More specifically, there are questions about their relevance to poor people, and the quality of their construction. A constantly shifting, erratic, and doubtfully relevant yardstick seems like a poor instrument for monitoring the MDGs.

Yet it is clear that we cannot do without PPPs altogether, at least not without abandoning altogether the idea of a genuine international poverty line. Deaton (2001) proposes a procedure that combines the PPP method with the idea of national poverty lines. It works something like this:

1. start from the $1993 PPP poverty lines in Chen and Ravallion (2001);
2. ask United Nations Development Programme and World Bank offices in each country to check these lines;
3. modify the lines to correct serious errors revealed at the country level;
4. update the lines over time using domestic price indexes, without further reference to PPP exchange rates; and
5. if step 4 is carried out on an annual basis, as is warranted by the importance of the counts, then major improvements to PPP exchange rates could be incorporated infrequently, no more than once a decade.

Step 1 recognizes the rhetorical value of the $1-a-day concept, retaining the trademark. It is worth bearing in mind that poverty lines always have a degree of arbitrariness, and we should use this to our advantage. The hope is that within that range of arbitrariness, we can find a poverty line for each country that is: (a) close to $1-a-day at 1993 PPP, and (b) meaningful and that makes sense to people in the country for the purposes of the international (not domestic) counts.

Step 2 recognizes that it is inappropriate for these lines to be exclusively set by 'experts' in New York, Washington or Geneva. Poverty lines work

362

because they seem reasonable to the people who use them. The $1-a-day poverty line should not be very far from national poverty lines in the poorest countries (as is, in fact, the case in India), although they will (and should) appear ungenerous in better-off countries. At the minimum, the local consultations will check that the PPP starting point poverty line is not absurd in terms of local purchasing power and local expenditure patterns, thus exercising a check on the PPP calculations. Better still would be more extensive local consultations, always informed by the fact that the line is supposed to be an *international* line, identifying people who are poor by (say) Indian standards, not necessarily local standards. I do not think that it is impossibly difficult to open a debate in Thailand, say, on what it would cost there to maintain someone who is regarded as poor in India.

Step 3 should be performed centrally to ensure conformity across countries. The concern is that people in better-off countries may wish to tie the poverty line to their local standard of living, with the result that we overestimate their poverty by the standards that are required for this work. If step 2 works well, few alterations should be made at stage 3, apart from those that reflect errors or approximation failures in the process of calculating consumption PPP exchange rates.

Step 4 means that the international lines, although different from the local lines, are updated in exactly the same way. The desirable feature of this scheme is that we avoid the possible effects of fluctuations in international commodity prices that are not relevant for domestic consumers, as well as possible errors in PPP calculation, errors that are likely to be more severe in changes than in levels. The undesirable feature is that we may lose touch with whatever components of the PPPs capture real changes in the living standards of poor people in one country relative to another. Yet, such genuine changes should also show up in domestic prices, and will be captured by local updating. In many of the cases where a properly measured PPP inflation rate differs from a properly measured domestic inflation rate, it is the latter that is appropriate for our purpose. We must also recognize that, in a world of less than perfect measurement, we may wish to use a different procedure for measuring levels than for measuring changes. The PPPs, imperfect although they are, are better than market exchange rates for setting international poverty lines. But the changes in PPPs over time are likely to be less reliable for measuring poverty than the changes in domestic consumer price indexes. It should also be emphasized that domestic price indexes are important statistics that are carefully scrutinized and debated by politicians and interest groups in each country. Not only is this likely to be some guarantee of quality, but the use of these domestically produced numbers in calculating poverty generates a consistency between the national and international poverty counts — because the two sets of poverty lines move in parallel — which will enhance the transparency and credibility of the latter both domestically and internationally.

Step 5 deals with the conflict between, on the one hand, not wanting to allow PPP inaccuracies or inappropriate price variations unduly to affect the estimates, and on the other, taking advantage of whatever improvements

take place in PPPs over time. In the past, the global poverty counts have been done only irregularly, as have PPP updates and extensions, so that the latter have greatly interfered with the stability and interpretation of the former, as well as generating a good deal of confusion among unsophisticated users. However, given the greatly increased attention that has come with the MDGs, the poverty counts must surely now be performed on an annual basis. If so, then it would make sense to revise the PPP basis; for example, once a decade. A useful analogy is with national accounts, which are infrequently but regularly re-based. While such re-basing is desirable, it cannot be done too often. If the re-basing were done every time the national accounts were published, they would become almost unusable.

How to monitor poverty for the MDGs

Two methods

There are two possible methods for calculating poverty and monitoring its trend over time, and both have been widely employed. They give quite different answers.

- Use either consumption or GDP data from national income and product accounts (NIPA), together with some measure(s) of distribution, or with survey data corrected to match the NIPA, to infer poverty rates and their trends over time. The measures of distribution typically come from household survey data.
- Ignore the NIPA information, and use household survey data to estimate the fraction of people living in households whose income or consumption is below the income or poverty line for the corresponding household type.

The survey method is used by the World Bank for the $1-a-day counts, although it makes some use of the NIPA method for projections; for example, those in its *Global Economic Prospects*.

The survey method is also used by many governments, as diverse as the US, based on data from the March round of the monthly Current Population Surveys, and India, based on data from the quinquennial large National Sample Surveys. The NIPA-based method was once used in India, until the Planning Commission was heavily criticized for doing so — on grounds similar to those reviewed in the following — by an Expert Group Report (Government of India, 1993). According to Szekely *et al.* (2000), the Economic Commission for Latin America and the Caribbean (ECLAC) also uses the national accounts to adjust survey data. Recently, the method has been used by Bhalla (2002), Sala-i-Martin (2002), and UNCTAD (2002) based on work by Karshenas (2001, 2002).

The two methods give different answers. In the 1990s, according to the all-survey method, there was a good deal less poverty reduction than might have been expected if the whole population had benefitted equally from the average rate of economic growth in the world as a whole, and in some large countries, most notably India. Bhalla and Sala-i-Martin 'resolve' the puzzle in

364

favor of growth using the NIPA method, which ensures that the growth in the NIPA magnitudes is fully reflected in the poverty measures. According to these procedures, the poverty MDG has already been met. UNCTAD (2002) uses the NIPA-based measure to argue not so much that the trend of poverty reduction has been understated, but to note that the geographical pattern of poverty reduction is different on a NIPA basis than on a survey basis, with Africa notably poorer. The poorest poor countries are relatively poorer using NIPA-based poverty measures. The two outcomes, over time and over countries, are of course related. In both, poverty declines more rapidly with economic growth using the NIPA methodology.

Each method has variants. For example, Bhalla and UNCTAD both use consumption data from the NIPA as their base, while Sala-i-Martin uses GDP data. The distributional measures that are required to calculate poverty from the NIPA averages sometimes use many points on the distribution (e.g. quintiles or deciles), or even a distributional assumption coupled with a single summary measure such as the gini (see later). When the full survey data are available, each household's reported income (or consumption) can be multiplied by the ratio of mean income (consumption) according to NIPA to mean income (consumption) according to the survey, and poverty measures calculated from the scaled household-level data. In the ECLAC case, scaling is carried out by components of income so that, for example, property income and wage income are subject to different scaling factors.

Within the survey method, there are different survey designs and different measures. The resource measure can be income or consumption. Questionnaire design differs from country to country, and sometimes over time within a country. For example, Indian poverty estimates from the latest large survey in 1999–2000 are not compatible with those for earlier years.

Official US inequality measures are not comparable before and after 1993. Sampling designs differ from one survey to another, as do refusal rates and the zeal with which enumerators pursue non-respondents. Some surveys use 'panels' of households, who are retained for more than one survey, whereas others draw a fresh sample for each survey.

For reference, and to illustrate, it is useful to record some of the formulae that might be used. In the NIPA method, suppose that mean consumption is estimated from the national accounts, and we have a single inequality measure (e.g. the gini) from a household survey. Suppose that consumption is approximately log-normally distributed in the population. The poverty rate is given by:

$$P_0 = \Pr(x \leqslant z) = \Pr(\ln x \leqslant \ln z) = \Phi[(\ln z - \mu)/\sigma] \qquad (1)$$

where x is the measure of resources, typically income or consumption per head, z is the poverty line, μ is the mean of the logarithm of x, σ is the standard deviation of the logarithm of x, and Φ is the cumulative distribution function (cdf) of the normal distribution. Under the log-normality assumption, the mean of logarithms can be calculated from the logarithm of the means $\ln \bar{x}$ using the formula:

$$\mu = \ln \bar{x} - (\sigma^2/2) \qquad (2)$$

365

and σ can be calculated from:

$$\sigma = \sqrt{2}\Phi^{-1}[(g+1)/2] \tag{3}$$

where g is the gini coefficient (see Aitchison and Brown, 1969, pp. 13, 113). These formulas are only one way of turning means into poverty counts. Another way, which obviates the need for the log-normal assumption, is to use points on the empirical distribution; for example, the quintiles provided by Deininger and Squire (1996). These are used to construct approximations to the cumulative distribution function, sometimes with supplementary functional form assumptions, from which the headcount ratio can be read off.

In a household survey, the quantity x is measured directly, typically at the household level. The headcount ratio is then estimated by the number of persons in the population who live in households whose x is below the poverty line. This version of the headcount ratio would be estimated by:

$$P_0 = [\Sigma_1^n w_h n_h 1(x_h \leqslant z)]/(\Sigma_1^n w_h n_h) \tag{4}$$

where w_h is the survey weight or inflation factor for household h, and n_h is the number of household members — an indicator function that takes the value 1 if its argument is true and 0 if it is false. In the NIPA 'corrected' version of this, equation (4) is applied not to x, but to αx where α is chosen so that the mean of αx is equal to the mean from the NIPA accounts.

Are national accounts or survey data more reliable?

There is a longstanding prejudice, at least in the West, that national accounts data are more reliable than data from household surveys. UNCTAD (2002) uses two arguments; that NIPA construction is more likely to be standardized across countries, and that NIPA data are available for virtually all countries. Average consumption from household surveys is often lower than average consumption from the NIPA, something that is as true in India as it is in the US, and this is interpreted as evidence that household surveys systematically understate consumption. And indeed, there are well-known difficulties with surveys. Rich households are often hard to reach; rich people who live in gated communities cannot be reached by surveyors, and when rich households are approached they often refuse to cooperate, or send servants to speak to the surveyors. Szekely and Hilgert (1999) show that, in a number of Latin American surveys, the highest recorded incomes are around (and often less than) the typical salaries of a typical manager as calculated by an international consulting firm. Almost anyone who has tried to look at saving in a household survey has discovered that income is implausibly low relative to consumption, with a large fraction of households at the bottom of the income distribution reporting that they spend more than they earn. While such an outcome is consistent with a permanent income story of consumption, with higher fractions of people at the bottom temporarily poor and dissaving out of previously accumulated assets, it might also come from gross understatement of income.

There are also problems at the bottom of the distribution. People who

366

do not live in households (street people, or beggars) cannot be reached by household surveys. It is often expensive, difficult, or even dangerous, to travel to remote areas or to areas where there may be a concentration of poor people. Statistical offices typically undersample poor areas, if only on cost grounds, and although this can be corrected for, accuracy is likely to suffer. Indeed, in some countries (such as Korea for many years) surveys are confined to urban areas. Because income, and to a lesser extent consumption, is unevenly distributed throughout the year, particularly in agricultural economies, the measured distribution of resources, and sometimes its mean, can be sensitive to the timing of visits throughout the year. Other issues concern the sensitivity of results to apparently minor changes in design; for example, when the Indian National Sample Survey (NSS) experimentally changed the recall period for food from 30 to 7 days, the estimated poverty rate was cut by one-half.

Some household surveys were originally designed not to collect data on poverty or inequality, but for other purposes; for example, for calculating average consumption patterns to be used in the construction of consumer price indexes. This is important because the design of a survey is often tailored to its purpose. In particular, designs for measuring *means* are likely to differ from designs for measuring *dispersion.* For example, it is often thought that people progressively forget purchases as the date of the purchase recedes. If so, the most accurate estimates of mean consumption will come from asking people how much they spent in the last day. But because many people do not spend every day, and because no one buys everything every day, inequality in reported purchases will be larger the shorter is the reporting period. In most circumstances, shorter reporting periods will also mean higher measured poverty rates. Given that different surveys use different reporting periods, the comparability of *means* from one survey to another is likely to be better than the comparability of measures of *dispersion*. The same is true for the treatment of seasonality in collecting income data where it is much easier to obtain an accurate estimate of the mean than the dispersion over households.

While it is true that the conceptual framework of the NIPA is more *standard* across countries and over time than is the design of household surveys, that does not imply that NIPA data are necessarily more *accurate*. NIPA consumption estimates are typically constructed following a 'commodity flow' approach. Consider, for example, consumption of a cereal, such as rice. In well-run offices, the statisticians have data on production — for example, from crop cutting, aerial, or farm surveys — as well as data on imports and exports. In less fortunate cases, production is often projected from the rural population, on the assumption that output per head is roughly constant. From the estimates of production, imports, and exports, the statistical office computes net domestic consumption. From this, it must deduct business (including cereals fed to animals) and government consumption; data on the latter may be quite good, but data on the former are typically very poor. The best that can be done is often an allowance using some 'multiplier' or 'fudge factor', which ideally would be based on a

367

business survey, but is often seriously outdated. An allowance also has to be made for investment uses, in this case inventories, which are also difficult to measure. The final total is a *residual*; there is no direct measurement of consumption. As a result, errors in measuring exports, production, inventories, or animal feed are all cumulated into the estimate of consumption.

Other problems with the NIPA are documented by Seers (1983), who notes that, in many countries, less than one-half of the national income estimate is derived from primary sources. More important for my current purposes, he points to the difficulty of capturing 'informal' (including, but not confined to, illegal) income-generating activities, and argues that the share of such unmeasured activities in GDP falls as the economy develops. In consequence, the growth rate of GDP as measured will be overstated, as more and more 'informal' activity becomes formal and is captured in the statisticians' net. To the extent that household surveys measure consumption, as opposed to income, at least some of the informal and illegal activity will be captured. This could be thought of as the 'Al Capone' effect; although the authorities could not prove that Capone was engaged in illegal activities, they could show that his expenditure was greatly in access of his reported income, and could convict him on charges of income tax evasion. As a result, not only will the measured growth rate of GDP be too high, but the ratio of survey-measured consumption to NIPA-measured consumption will decline with economic development, and may well be greater than one in the poorest countries. That this is in fact the case is documented by Karshenas (2001), and is at least in part responsible for the finding in UNCTAD (2002) that the poorest countries are relatively poorer when poverty is measured using the NIPA method in place of the World Bank's survey methods.

In India there has been a vigorous debate about poverty in the 1990s, with much attention given to the discrepancy between the NIPA and the surveys. According to some accounts, the ratio of NIPA to NSS consumption fell from around 90% in the 1970s to less than 60% in the late 1990s, and this discrepancy was attributed by some to the failings of the NSS data, more or less in the terms already outlined. (Some of it is also attributable to comparison with incompatible NIPA data at the beginning and the end of the period.) But recent work within the government has revealed serious problems with the accuracy of the NIPA measures of consumption. Recent revision and re-basing of the accounts generated large differences in some categories of consumption expenditures, even in nominal terms, and even though there was no conceptual change. Kulshreshtra and Kar (2002) and Sundaram and Tendulkar (2003), reviewing the work, show that the NIPA estimate for consumption of fruits and vegetables in 1993–1994 in nominal rupees more than doubled between the 1998 and 1999 versions of the national accounts. The estimate for clothing fell by about one-half, and that for rent, fuel, and power rose by more than 40%. Even with some canceling out of pluses and minuses, total consumption was revised upwards by 14%. (This revision, if it were carried through to poverty construction using the NIPA methodology, would cut the Indian poverty rate by a little less than one-half.) These revisions reflect changes in data collection practices, in many cases from firm or farm surveys, that are evi-

368

dently no more reliable than the household surveys. Indeed, it is hard to see why we should expect them to be.

Household surveys vary a great deal in design and reliability from one country to another. In some countries, India and Indonesia being two notable examples, there is a long tradition of high-quality survey practice. I have spent a great deal of time working with Indian NSS data and with Indian NSS officials, and there is no doubt that tremendous care and expertise goes into the work. Interviewers are well trained — for example, all have masters degrees in statistics — and devote great efforts to ensuring cooperation, returning many times to households who at the first visit are unavailable. The NSS claims that it has essentially zero refusal rates, and while this is probably not literally correct, it is clear that refusals are low. In India, the perception is usually the opposite of that in the West, with the survey data typically thought to be of higher quality than the NIPA accounts. I see no reason to disagree with that supposition. Of course, India is not everywhere, and there are undoubtedly countries where the surveys are genuinely less reliable than the NIPA. Indeed, there are several examples of 'failed' household surveys, where the results are simply not credible; see the examples in Karshenas (2002), and Deaton and Grosh (2000). But I do not believe there should be any general presumption that NIPA accounts are more reliable nor more accurate than are surveys. Indeed, as I shall argue in the following, there are some general reasons to suppose that surveys are likely to be *more* accurate when it comes to the measurement of poverty.

Although there has been little formal work in poor countries, there is a literature for the US and the UK, for example, documenting the characteristics of those who refuse to cooperate with surveys; see, for example, Groves and Couper (1998). The findings are consistent with the view that is the rich, not the poor, who are more likely to be missed. I suspect that this result carries through to poor countries; it is certainly consistent with the results for Latin America in Szekely and Hilgert (1999). Another example comes from Jamaica where the statistical office faces the problem of collecting survey data in cities where there are no-go areas controlled by political warlords. Yet the statistical office goes to such places, with whatever protection it needs. But armored surveyors do not help penetrate the walled-off estates and gated communities of the rich. South African surveyors tell similar stories. So if there is a systematic bias in household surveys, it is most likely to come from understating the consumption and income of the rich.

NIPA consumption/income is not the same as survey consumption/income

There are a number conceptual differences between NIPA and survey measures. Items that are included in the NIPA, but not the surveys are:

- consumption in the form of imputed rents of owner-occupied housing;
- consumption in the form of imputed charges for financial intermediation;
- consumption by non-profits;

369

- residual business consumption; and
- incomes from employer's contributions to pension funds.

The item included in the surveys, but not in the NIPA is:

- the component of annuity incomes that represents rundown of assets, as opposed to income from assets

Neither NIPA nor survey consumption measures include those components of government expenditures that contribute to living standards, such as education or health, nor more arguable items such as defense. If the NIPA methods uses GDP, not consumption, all of these items are implicitly included, as well as the net savings of firms, government, and foreigners; as far as I am aware, only Sala-i-Martin uses this methodology explicitly.

The first two items in the list of items in NIPA but not the surveys are conceptually appropriate, so that NIPA is correct and the surveys wrong. Consumption of financial intermediation is the difference between the interest charged by financial institutions to consumers as borrowers and the interest paid to them as depositors, a difference that is accounted as the implicit value of the financial intermediation (United Nations, 2000, p. 47). That same source says that "the allocation of this imputed service depends on the country's policy", suggesting that the treatment of this item is not consistent across countries. In India, this item alone has grown rapidly over time, accounting for 0.5% of consumers' expenditure in 1983–1984, but more than 2% in 1993–1994 (see Kulshreshtra and Kar, 2002).

If expenditures by non-profits are on behalf of households, the third item could also be arguably correct. (The argument is better for a non-government organization running a soup-kitchen than for a political party.) If we are using the NIPA to compute household disposable incomes, the last item in the list will be included (again correctly), and it is missed by the surveys. Because household respondents cannot typically parse out the asset and income components of annuities, households will almost certainly misreport the total amount as income, which overstates income and saving, and understates dissaving.

There are two important points to note.

- The items that are missed by the surveys, and captured by the NIPA, are likely to be more important for richer than for poorer households. Poor households in India rarely have bank accounts (even in the US, around 30% of households do not have a checking or saving account), yet the implicit financial intermediation turns out to be the largest single item in the discrepancy between NIPA and survey consumption.
- If the items missed by the surveys are income elastic, as is plausibly the case, economic development, over time or over countries, will tend to widen the gap between NIPA and survey consumption. This is in addition to the similar effect caused by the gradual incorporation of more and more informal activity into the national accounts.

By the first point, and even if the survey mean is an unreliable underestimate

370

of total (or average) consumption, it may nevertheless be accurate for the poor. Even if we believe the NIPA, there is no implication that underestimation of the survey mean implies overestimation of poverty. This has been repeatedly noted in the Indian debate. If we also accept the second point, then the growth of average NIPA consumption will overstate the growth of consumption of the poor, and use of the NIPA method will risk overestimating the rate of poverty decline. Ideally, this would not happen if consumption inequality were being correctly measured, because the effects of the increase in NIPA consumption on poverty reduction would be moderated by widening consumption inequality. *But this cannot happen in practice.* The only source of distributional information is the surveys, which exclude the items in question, so that their faster growth than average cannot be recorded, even in principle. The situation is almost certainly worse still if we were to apply the NIPA method using broader measures, such as GDP. Furthermore, the most likely reason for surveys to underestimate mean consumption, the under-representation of the rich, means that the surveys will also understate inequality.

The evidence in UNCTAD (2002, p. 48, Chart 8) and Karshenas (2001) shows that, compared with survey measures, NIPA-based poverty measures are relatively higher for the poorest of poor countries, and relatively lower for the less poor countries. This presumably comes from the fact that the ratio of NIPA to survey consumption increases with the level of economic development, which is consistent with the earlier argument and, by itself, militates in favor of using the survey estimates for poverty measurement. Because the items that are both included in the NIPA and excluded from the surveys are not consumed by the poor, this evidence is an argument not for the use of the NIPA, but for the surveys. As we have seen, the changing ratio of NIPA to survey consumption can also come from the fact that the surveys capture more informal activities; if not their proceeds, at least the consumption based on them. Once again, the poverty measures from the surveys are right and those from the NIPA are wrong.

From NIPA means to poverty: using distributional information

In the NIPA method, mean consumption (or income) is used to yield a poverty measure using data on the distribution of consumption (or income); for example, using formulas such as equations (1), (2) and (3).The measure of distribution, σ, the standard deviation of logs in equation (1), the gini coefficient g, or more comprehensive information on distribution all come from household surveys that are essentially the *only* source of distributional information. In principle, if NIPA consumption (say) could be 'corrected' to place it on a survey basis, if the distributional information were up to date, and if the NIPA information was consistent with that from the surveys, it would not matter whether the NIPA or survey method were followed. It should be noted that those who endorse the NIPA method and reject the accuracy of the surveys are nevertheless prepared to accept and use the distributional information that they contain. It is quite unclear whether there are any reasonable assumptions under which such a schizophrenic treatment

will be valid (Ravallion, 2002). The most frequently cited reason for survey failure is that the consumption of the rich is missed, and it is far from clear that such selective non-reporting will not bias estimates of inequality (see Mistiaen and Ravallion, 2003; Deaton, 2004). If the argument with the surveys concerns issues such as variable reporting periods or the treatment of seasonality and their effects on measurement, then the mean is likely to be more accurate than the dispersion, so that there is little argument for retaining the latter instead of the former.

In many cases, survey data are only irregularly available — which is one of the reasons why the NIPA method is used — so that the information on distribution is out of date. It is often argued that this matters hardly at all, because income distributions change only slowly, and because studies, usually based on the Deininger-Squire data, have shown that there is little or no systematic relationship between growth and changes in distribution. In terms of equations (1)–(3), we can take a fixed value of σ or g, and simply use the trend in NIPA consumption (or in GDP if consumption is a fixed proportion of GDP) to update the poverty numbers.

These arguments are dangerous. At least some of the debate on globalization, growth, and poverty hinges on whether or not growth is equally shared, and whether it is indeed true that incomes of the poor grow at the same rate as the economy as a whole. We do no justice to that debate if we wish it away by *assuming* that the poor proportionately share the benefits. And we can hardly expect our poverty measurements to be taken seriously if the way in which the measurement is done *assumes away* the possibility that the benefits of growth are unequally shared. Whether or not it is true that growth increases inequalities, we need a method of keeping score that does not assume the answer. We cannot argue that growth reduces poverty without having demonstrated the fact.

That distribution should be impervious to growth is somewhat implausible. It is also implausible that the errors that cause the discrepancies between the NIPA and surveys contaminate only the survey mean, while leaving unimpaired measures of distribution. By distributing NIPA consumption to people in a fixed way, while ignoring the survey mean, we are assuming that the discrepancy between the survey and NIPA estimates can plausibly be assigned proportionately to everyone. That is inconsistent with evidence that the rich are less likely to be covered in the survey. It would also require that the discrepancies are proportionately the same for the different goods and services that comprise consumption, as well as between rural and urban sectors, and the different regions of the country. In many cases, we have good evidence that these assumptions are not true. Drawing again on India, Sundaram and Tendulkar (2001) have shown that the differences between NIPA and survey estimates differ systematically across commodity groups so that, because the poor spend their budgets in different ways from the rich, it is certainly incorrect to assume that the discrepancies are distributionally neutral. By itself, and if it were the only problem, this issue could be dealt with by calculating scaling factors for each commodity group, and applying them to each household's expenditure group by group.

372

But it is also clear that there were real changes in the pattern of consumption inequality in India in the 1990s, changes that are consistent with what might be expected from an economy under gradual reform and where the new opportunities are far from equally available to everyone. Already successful states in the south and west grew the fastest and systematically diverged from the poorer, less successful states. There were also increases in inequality between urban and rural sectors, and within urban sectors (Deaton and Drèze, 2002). The analysis of income tax data also shows that there was a marked increase in the incomes of those at the very top of the distribution in India (see Banerjee and Piketty, 2001). This widening of the income distribution at the very top, which has also happened (for example) in Britain and the US (Atkinson, 2002; Piketty and Saez, 2003), is not picked up in even large household surveys, either because the rich do not participate or because, even though they possess a substantial fraction of total income, there are too few of them to allow adequate representation in the surveys. To assume that distribution is independent of growth, and that inequality remains unchanged as consumption grows, is not to measure poverty, but to assume it.

Finally, it needs to be reiterated that the inequality measures from household surveys do not match the quantity that is being measured in the NIPA and that they are being used to re-distribute. Quite apart from the exclusion of the rich, who account for much more than their population share of GDP, the measure of resources whose distribution is being measured in the surveys is not the same as the measure of resources that goes into the NIPA, and cannot therefore be used to spread it. The goods that are included in the NIPA, but excluded from the surveys, are not taken into account in survey measures of distribution, so that if they grow more rapidly than GDP as a whole, and accrue mostly to the rich, there is no possibility of the fact being picked up in the inequality estimates. The measures listed by Deininger and Squire, for example, relate to survey consumption, or survey income, sometimes on a household and sometimes on an individual basis. Only under very strong conditions do they tell us anything about the distribution of GDP, or even the distribution of NIPA consumption. And as we have seen, the distribution of NIPA consumption over individuals could be widening, even as the distribution of survey consumption over individuals is constant. Therefore, results that show no relation between growth in the national accounts and inequality in the surveys (quite apart from the fact that they are true on average not for individual countries) are not informative about what we need to know, which is what is happening to the distribution over persons of NIPA consumption, or NIPA disposable income.

Inequality and the discrepancies between NIPA and survey estimates

It is useful to combine some of the elements of the foregoing discussion into a simple formulation of NIPA and survey accounting that allows for the fact

373

that measurement errors are likely to work in different ways on the two estimates. To see how this might work, write:

$$Y = \Sigma_1^N y_b \tag{5}$$

where y_b is the consumption (or income) of each of N households in the economy. Using surveys, we estimate y_b from a random sample of households, and add up to get Y. In the NIPA, by contrast, we add up over all transactions. To make the argument simple, suppose that each household makes a single transaction each year, in which it spends its entire income or consumption at once. In practice, equation (5) is measured with error, and one way to think about it is to re-write equation (5) as:

$$Y = \Sigma_1^N p_n(y_b) y_b \tag{6}$$

for the NIPA, and as:

$$Y = \Sigma_1^N p_s(y_b) y_b \tag{7}$$

for the surveys. The functions p_n and p_s are the probabilities that the 'transaction' y_b is captured in the NIPA and in the survey, respectively. These two types of probability are likely to behave differently from one another. In particular, because the NIPA picks up formal sector transactions, and misses many informal transactions, the probability of picking up the transaction is likely to be *constant* or even *increasing* in its size. In India, the consumption of the Maharajah or of the software millionaire is more likely to show up in the national accounts than is that of a farmer or street trader. By contrast, the surveys are much less likely to capture the Maharajah or the software millionaire, so that the survey probabilities are *decreasing* in the size of each income. Under these assumptions, both the NIPA and the surveys will underestimate total consumption. However, making the distribution of income less equal will increase measured NIPA if large transactions are more likely to be recorded than small ones, while simultaneously decreasing aggregate income from the surveys, because the larger the share of income received by the rich, the less will be captured by the surveys.

If there is any truth to this story, increases in inequality will cause the gap between the NIPA and the survey aggregates to rise. Clearly, this is a testable proposition, and it needs to be examined in the data. If borne out, the implications for measuring poverty are important because the survey estimates, which miss the rich but not the poor, are the right ones to use. The NIPA, on the contrary, by following the money, is more focused on those who spend most money. Using the NIPA method in these circumstances will overstate the rate of decline of poverty by attributing to the poor the growth rate that is, in fact, a characteristic of those nearer the top of the distribution.

Conclusions and recommendations

For all the reasons discussed, NIPA-based poverty measures may be misleading, and using them is likely, in practice, to overstate the reduction in

374

poverty over time, as well as to exaggerate poverty differences across countries. Instead, household surveys should be used to estimate the extent of poverty by directly measuring the living standards of poor households, rather than by imputing them through a long chain of inference, many of whose links are weak or absent. Indeed, this is the fundamental reason for using survey not national accounts data to measure poverty; the former provide direct measures of the living standards of the poor, while the latter do not. The national accounts have many important uses, but the measurement of poverty is not one of them. To be sure, there are cases where surveys do not exist, or are seriously outdated, so that there is little option to measure GDP, or aggregate consumption, and to make the best possible guess about what is happening at the bottom of the distribution. But such measures should be recognized for what they are — guesses — and should not be treated on the same footing as direct estimates of the living standards of the poor. And the appropriate remedy in all but the shortest run is to collect survey data, or to improve existing surveys, not to apply a set of implausible corrections to estimates that are inherently incapable of measuring the living standards of poor people.

What needs to be done is to commit to the construction of an adequate database of household surveys for poverty analysis, and to work on improving it. Some useful steps along that process include the following.

- Make an inventory of where we are now. Some of that information is currently available through the World Bank's poverty website, but we need better documentation of existing surveys. For example, for each survey we need to know the coverage (e.g. national or urban), the resource concept that is collected (consumption or income), and relevant aspects of the questionnaire design, such as the degree of detail in reporting (200 consumption items versus 30, for example, or a single income question versus a dozen), the recall periods over which data are collected, and how seasonality is handled. We need to know how households are chosen, and how many refuse to cooperate. Changes in survey practice need to be documented; they are known to a few local experts, or to those who have devoted a large amount of time to their study, but are hard to find for almost everyone else. Surveys that failed or are otherwise clearly unsafe need to be identified and labeled as such.
- The Bank should issue an annual poverty CD-ROM containing the aforementioned information. When household surveys are publically available, the relevant data can be included, or links provided on the website, but even when they are not, percentiles of the income or consumption distribution can be provided. China is a good example of a country where the surveys are not available, even to the World Bank, but where the Chinese statisticians provide (to the World Bank) percentiles of the distribution of income per head that can then be used to calculate poverty or inequality measures.
- The better documentation and availability of either the surveys themselves, or of summary statistics derived from them, will open up the process of

375

poverty calculation. Currently, world poverty figures, and poverty figures for many countries, can only be calculated by the World Bank. The process needs to become more open if the estimates are to carry more credibility. Currently, the (flawed) methodology used by both Bhalla and Sala-i-Martin is the *only* way that outsiders can calculate world poverty estimates

- Given a good inventory, it will be possible to know where surveys are seriously lacking. This is not just a matter of identifying countries that have no, or at least no current survey. We also need to know which surveys correspond to reasonable best-practice standards to the extent that they could be accepted as a part of a poverty monitoring system. The United Nations and the World Bank should consider whether it would be helpful to re-visit the original goals of the Living Standards Measurement Surveys (LSMS), which was to develop a *common* survey instrument that would support internationally comparable estimates of poverty and inequality, doing for surveys what the Systems of National Accounts (SNA) does for the national accounts. The LSMS surveys have been very successful in generating relatively small-scale surveys that cover a range of interlinked household activities. But they are not standardized, especially in recent years. There are a number of successful standardized surveys around the world — such as the Demographic and Health Surveys — and it is time to consider the development of a uniformly applicable household budget survey.
- Until better and more uniform surveys are available, we need more work on methods of adjusting the existing surveys to make poverty measures more credible and more comparable than is currently the case. One particularly important area is the adjustment of surveys for differential non-compliance by income, and an excellent start on this work is reported in Mistiaen and Ravallion (2003). Another current difficulty is that some surveys, for example from China and from all of Latin America, measure *income*, while most other surveys measure *consumption*. The last set of $1-a-day poverty counts scaled the data in the income surveys by the ratio of consumption to income in the NIPA. It should be clear from the foregoing arguments that this is not likely to be satisfactory, although it is a good deal less obvious how to do better. Indeed, I suspect that income tends to be so poorly measured (and typically understated) in surveys in poor countries that the only ultimately satisfactory solution is to switch to consumption surveys. These and other comparability issues will only be sorted out gradually over time, and it is here that academics can make a useful contribution. But we can only do so if provided with more of the raw poverty data than is currently the case.

What to do now?

In the short run, I believe that the World Bank's survey-based monitoring procedures for the $1-a-day and $2-a-day counts should be endorsed, with the emendations discussed in the first section. To reiterate, this would involve accepting the local poverty lines from the Bank as a starting point,

'editing' them using local knowledge, and updating them using local price indexes, all of which would be feasible for the United Nations system in the short run. The key flaw in the World Bank's monitoring is not the work itself, but its lack of replicability and transparency. When other agencies and scholars can do the monitoring for themselves, we will understand much better than we currently do how the numbers vary with different assumptions and different data, and the assessment of progress toward the MDGs will not be the preserve of a single institution.

An alternative, or perhaps a complement, would be for the United Nations and the World Bank to accept joint responsibility for creating and publishing the poverty numbers. Whatever agency is charged with the task needs to be insulated from policy-making, particularly policy-making in the World Bank. National statistical offices are carefully insulated from political interference — for example, by giving their directors substantial terms of office, and by sets of protocols that prevent politicians or policy-makers interfering with the production of the data — but such safeguards on the poverty data have been notably absent within the World Bank. High-profile World Bank publications have proclaimed quite different trends in world poverty (see Deaton, 2002), and such disagreement undermines respect for the estimates.

Acknowledgements

References

Aitchison, J. and Brown, Alan (1969) *The Lognormal Distribution*, Cambridge University Press, Cambridge.

Atkinson, Anthony B. (2002) 'Top Incomes in the United Kingdom Over the Twentieth Century', *Working Paper*, Nuffield College, Oxford.

Atkinson, Anthony B. and Bourguignon, Francois (2001) 'Poverty and inclusion from a world perspective', in J. E. Stiglitz and P.-A. Muet (Eds), *Governance, Equity, and Global Markets: the Annual Bank Conference on Development Economics — Europe*, Oxford University Press, Oxford.

Banerjee, Abhijit and Piketty, Thomas (2001) 'Top Indian incomes, 1956-1998', *Working Paper* [http://www.worldbank.org/indiapovertyworkshop].

Bhalla, Surjit S. (2002) *Imagine There is No Country: poverty, inequality, and growth in the era of globalization*, Institute for International Economics, Washington, DC.

Chen, Shaohua and Ravallion, Martin (2001) 'How well did the world's poorest fare in the 1990s?', *Review of Income and Wealth*, 47(3), pp. 283-300.

Deaton, Angus (1997) *The Analysis of Household Surveys: a microeconometric approach to development policy*, Johns Hopkins University Press for the World Bank, Baltimore, MD.

Deaton, Angus (1998) 'Getting prices right: what should be done?', *Journal of Economic Perspectives*, 12, pp. 37–46.

Deaton, Angus (2001) 'Counting the world's poor: problems and possible solutions', *World Bank Research Observer*, 16, pp. 125-147.

Deaton, Angus (2002) 'Is world poverty falling?', *Finance and Development*, 39(2), pp. 4-7.

Deaton, Angus (2004) 'Measuring poverty in a growing world: or measuring growth in a poor world', *Review of Economics and Statistics* (forthcoming).

Deaton, Angus and Drèze, Jean (2002) 'Poverty and inequality in India, a reexamination', *Economic and Political Weekly*, 7 September, pp. 3729-3748.

Deaton, Angus and Grosh, Margaret (2000) 'Consumption', in M. Grosh and P. Glewwe (Eds), *Designing Household Questionnaires for Developing Countries: lessons from fifteen years of the living standard measurement study*, World Bank, Washington, DC, vol. 1, pp. 91-133.

Deaton, Angus and Tarozzi, Alessandro (2000) 'Prices and Poverty in India', *Working Paper*, Research Program in Development Studies, Princeton, NJ [http://www.wws.princeton.edu/~deaton].

Deininger, Klaus and Squire, Lyn (1996) 'A new data set measuring income inequality', *World Bank Economic Review*, 10, pp. 565-591.

Diewert, W. Erwin (1976) 'Exact and superlative index numbers', *Journal of Econometrics*, 4, pp. 115-145.

Government of India (1993) *Report of the Expert Group on Estimation of the Proportion and Number of Poor*, Planning Commission, Delhi.

Groves, Robert M. and Couper, Mick P. (1998) *Nonresponse in Household Interview Surveys*, Wiley, New York.

Karshenas, Massoud (2001), 'Measurement and Nature of Absolute Poverty in Least Developed Countries', *Working Paper*, School of Oriental and African Studies (SOAS), Department of Economics, London.

Karshenas, Massoud (2002) 'Global poverty: new national accounts consistent and internationally comparable poverty estimates', *Working Paper*, SOAS, Department of Economics, June.

Kulshreshtra, A. C. and Kar, A. (2002) 'Estimates of food consumption expenditure from household surveys and national accounts', *Working Paper* [http://www.worldbank.org/indiapovertyworkshop].

Mistiaen, Johan A. and Ravallion, Martin (2003) 'Survey Compliance and the Distribution of Income', *Working Paper*, World Bank, Washington, DC.

Piketty, Thomas and Saez, Emmanuel (2003) 'Income inequality in the United States, 1913-1998', *Quarterly Journal of Economics*, 118, pp. 1-39.

Ravallion, Martin (1994) *Poverty Comparisons*, Harwood, Philadelphia.

Ravallion, Martin (2002) 'Have we already met the Millenium Development Goals?', *Economic and Political Weekly*, 16 November.

Reddy, Sanjay and Pogge, Thomas (2002) 'How *not* to count the poor', *Working Paper*, Barnard College and Columbia University [http://socialanalysis.org].

Sala-i-Martin, Xavier (2002) 'The disturbing "rise" of global income inequality', *NBER Working Paper No. 8904*, NBER, Cambridge, MA.

Seers, Dudley (1983) *The Political Economy of Nationalism*, Oxford University Press, Oxford.

Sundaram, K and Tendulkar, Suresh (2001) 'NAS-NSS estimates of private consumption for poverty estimation: a disaggregated comparison for 1993-1994', *Economic and Political Weekly*, 13 January, pp. 119-129.

Sundaram, K and Tendulkar, Suresh (2003) 'NAS-NSS estimates of private consumption for poverty estimation: a further comparative estimation', *Economic and Political Weekly*, 25 January, pp. 376-384.

Szekely, Miguel and Hilgert, Marianne (1999) 'What's behind the inequality we measure: an investigation using Latin American data', *Working Paper #409*, IADB, Washington, DC.

Szekely, Miguel, Lustig, Nora, Cumpa, Martin, *et al.* (2000) 'Do we know how much poverty there is?', *Working Paper*, Inter-American Development Bank, Washington, DC.

United Nations (2000) *Household Accounting: experiences in concepts and compilation: Volume 1: household sector accounts*, Department of Economic and Social Affairs, Statistics Division, Studies in Methods, New York, series F, no. 75, vol. 1.

UNCTAD (2002) *The Least Developed Countries Report, 2002*, United Nations, Geneva.