Estimating standard errors for poverty measures

Kenneth R. Simler*
Research Fellow
Food Consumption and Nutrition Division
International Food Policy Research Institute
2033 K Street, NW
Washington, DC 20006-1002 USA
k.simler@cgiar.org
Tel: +1.202.862.5636
Fax: +1.202.467.4439

Channing Arndt
National Directorate of Studies and Policy Analysis
Ministry of Planning and Development
carndt@mpd.gov.mz

November , 2005

Poverty comparisons with absolute poverty lines estimated from survey data

Abstract

The objective of measuring poverty is usually to make comparisons over time or between two or more groups. Common statistical inference methods are used to determine whether an apparent difference in measured poverty is statistically significant. Studies of relative poverty have long recognized that when the poverty line is calculated from sample survey data, both the variance of the poverty line and the variance of the welfare metric contribute to the variance of the poverty estimate. In contrast, studies using absolute poverty lines have ignored the poverty line variance, even when the poverty lines are estimated from sample survey data. Including the poverty line variance could either reduce or increase the precision of poverty estimates, depending on the specific characteristics of the data. This paper presents a general procedure for estimating the standard error of poverty measures when the poverty line is estimated from survey data. Based on bootstrap methods, the approach can be used for a wide range of poverty measures and methods for estimating poverty lines. The method is applied to recent household survey data from Mozambique. When the sampling variance of the poverty line is taken into account, the estimated standard errors of Foster-Greer-Thorbecke poverty measures increase by 15 to 30 percent.

JEL codes: I32, C13, 012

Keywords: poverty measurement, bootstrap, Mozambique

1. Introduction

A principal objective of poverty measurement is to make comparisons between groups. Analysts and policymakers are generally interested less in the absolute level of poverty at a given place and time than they are in knowing how measured poverty levels compare to levels observed in other settings or at other points in time. Is poverty higher in the hills or on the coast? Did poverty decline following implementation of a poverty reduction program? These questions have become more and more important in recent years. Besides the high profile Millennium Development Goal of halving world poverty by 2015, country development programs and donor support are increasingly driven by the Poverty Reduction Strategy Paper (PRSP) process, which requires close monitoring of poverty levels and detectable progress in reducing poverty. For example, indications of a recent increase in poverty in Uganda sparked a debate about whether the national development strategy, which had steadily reduced poverty in the 1990s, needed an overhaul (Kappel et al. 2005).

There are many ways to define and measure poverty, but with few exceptions the empirical basis for poverty comparisons is statistical, employing point estimates of relevant poverty measures and their associated standard errors. These are generally estimated from household survey data. Statistical tests are applied to assess whether differences or changes in poverty levels are significant. Research over the past 15 years has increasingly refined statistical inference methods for poverty measures. Kakwani (1993) develops distribution-free asymptotic standard errors for several additively decomposable poverty measures. Bishop et al. (1995) provide asymptotic theory for testing poverty measures decomposed by sub-group. Ravallion (1994a) examines the effect of errors in consumption data on poverty comparisons, finding that noisier data for some sub-groups can lead to re-rankings of poverty measures, with the exact nature of the re-ranking dependent upon the poverty measure used. These approaches assume that the data are generated by simple random sampling, but Howes and Lanjouw (1998) note that most poverty data come from stratified cluster sample surveys. Using data from Pakistan and Ghana, they find that when complex sample design is taken into account, estimated standard errors of FGT poverty measures increase by 26 to 33 percent in Pakistan and 45 to 64 percent in Ghana.

All of these approaches tacitly assume that the source of statistical error is in the welfare metric, e.g., income, expenditure, or consumption. That is, the welfare metric is treated as a random variable, but the poverty line is treated as a fixed constant. In fact, poverty lines are often estimated from the same survey data as the welfare measure, and thus are also random variables. This is clearest in the case of relative poverty lines that are derived from empirical income distributions, such as one-half of median income. Preston (1995) presents standard error formulae for poverty measures that incorporate simple random sampling error in the relative poverty line as well as the welfare measure. He observes that the two sources of error could reinforce or offset one another, so that one cannot say a priori whether accounting for sampling error in the poverty line will increase or reduce the precision of poverty estimates. Zheng (2001) builds on this work to develop asymptotic distribution-free inference methods for several additively decomposable poverty measures when relative poverty lines are set as percentages of mean income or percentages of quantiles, and allows for cluster sampling. In his empirical applications, Zheng (1997, 2001) finds that the sampling error of the poverty line always increases the standard errors of poverty estimates.

The objective of this paper is to bring these two strands of the literature together to provide a method for more accurately assessing the precision of estimates of absolute poverty, leading to more reliable poverty comparisons. It argues that like relative poverty lines, absolute poverty lines that are estimated from sample data (which we believe is the norm) have a sampling error that needs to be included in the calculation of standard errors for poverty measures. We present a bootstrapping procedure for estimating the sampling error of absolute poverty lines, and assess its effect on the standard error of Foster-Greer-Thorbecke (1984) poverty measures. The bootstrap is, of course, a familiar approach with well-known properties. We use recent household survey data from Mozambique to explore the impact on poverty comparisons, and conclude with a discussion of the circumstances under which this approach is appropriate for measuring the precision of estimated poverty measures.

The remainder of the paper is structured as follows. Section 2 considers challenges in estimating poverty and assessing the precision of estimated poverty measures. This is followed by a description of the methods and data in section 3. Section 4 presents empirical results. Section 5 summarizes and concludes, including remarks about the scope for wider application of this procedure.

2. Estimating poverty

The measurement of poverty poses two fundamental questions (Sen 1976). First, how does one identify the poor among the total population? Second, how does one aggregate information on individuals and households into a scalar measure of poverty? The first question has two components, namely, how do we measure individual welfare and, using this same metric, how do we determine the threshold that separates the poor from the nonpoor? There exist various approaches, but most often the welfare metric is based on a comprehensive measure of household consumption, normalized to a per capita or per adult equivalent basis. This measure includes not only cash expenditure, but also auto-consumption and imputed use-values for consumer durables and owner-occupied housing. Deaton and Zaidi (2002) provide an extensive review of the conceptual framework, underlying theory, and practical issues related to constructing a measure of household welfare based on consumption.

Setting the threshold that defines poverty is a contentious issue with a large literature, and we will not attempt a thorough review of it here. Low-income countries typically employ absolute poverty lines, which attempt to reflect a fixed standard of living across the domain of the poverty comparisons (Ravallion 1994b).[1] Although methods vary, all base the level of absolute poverty lines on the cost of acquiring basic food and nonfood needs. For example, the Cost of Basic Needs (CBN) approach calculates a food component of the poverty line from the cost, in local prices, of acquiring a diet sufficient in food energy, based on consumption patterns of poor households in that area (Ravallion 1998). The food component is complemented by a nonfood component, which is typically estimated from the nonfood budget share of

---

[1] Industrialized countries often use relative poverty lines. For example, the EU uses 60 percent of national median income as an indicator of "poverty risk" (Guio 2000).

households that are at or near the poverty line.[2] An alternative approach is the Food Energy Intake (FEI) method, which uses regression analysis to set the poverty line at the monetary value of the total expenditure or income at which an average (or representative) household just manages to meet stipulated caloric requirements (Greer and Thorbecke 1986). Other methods of setting absolute poverty lines include subjective poverty lines and the dollar-a-day poverty line that is sometimes used for international comparisons.[3]

The methods presented in this paper can be applied to any poverty line that is estimated statistically. Similarly, they can also be used for an array of additively decomposable poverty measures, such as those proposed by Watts (1968), Clark et al. (1981), and Foster et al. (1984).[4] In the empirical application in this paper, we focus on one method for estimating poverty lines (CBN), and one set of poverty measures, the Foster-Greer-Thorbecke (FGT) or P_α class of additively decomposable poverty measures.

At the household level, the general form of the FGT measure can be written

---

[2] One approach is to use a restrictive nonfood poverty line, based on the nonfood budget share of those households whose total food and nonfood consumption is in the neighborhood of the food poverty line. A higher nonfood poverty line may be obtained by using the nonfood budget share of households whose food consumption is in the neighborhood of the food poverty line. Both approaches are seen in empirical work.

[3] Stochastic dominance approaches (Atkinson 1987, Foster and Shorrocks 1988, Davidson and Duclos 2000) make poverty comparisons across a range of possible poverty lines, rather than specific discrete poverty lines. These approaches are discussed further in the concluding section of this paper.

[4] To the best of our knowledge, this approach cannot be applied to poverty measures that are not additively decomposable, such as the Sen (1976) or Kakwani (1980) indices.

(1) $$P_{\alpha,j} = \max\left[0, \left(\frac{z - y_j}{z}\right)^{\alpha}\right] \quad , \quad \alpha \geq 0$$

where y_j is a money-metric welfare measure for household j and z is the poverty line. The poverty headcount index and poverty gap index are obtained when $\alpha = 0$ and 1, respectively. Poverty in a population of n households is the weighted mean of (1) over all households, with the number of members in each household (h_j) as the weights, or

(2) $$P_{\alpha} = \frac{\sum_{j=1}^{n} h_j P_{\alpha,j}}{\sum_{j=1}^{n} h_j} .$$

In the case of non-self-weighting sample surveys, which is the typical source of poverty data, sample weights (or expansion factors) w_j must be employed to arrive at an unbiased estimator of individual-level poverty measures, written

(3) $$P_{\alpha} = \frac{\sum_{j=1}^{n} w_j\, h_j P_{\alpha,j}}{\sum_{j=1}^{n} w_j\, h_j} .$$

By using h_j as weights, equations (2) and (3) assume that poverty is distributed equally within the household. Although this may be a strong assumption, it is difficult to avoid because individual-specific information on the welfare metric is rarely available. If such data were available then poverty could be measured by equation (3), but with j indexing individuals instead of households and $h_j = 1$.

Howes and Lanjouw (1998) define the estimator of the poverty measure in (3) as $\pi_{\alpha} = t/p$, where p (the denominator in (3)) is the sample estimate of the population size and t (the numerator in (3)) is the sample estimate of "total poverty." They show that under fairly weak assumptions that also conform well to the non-self-weighting stratified multiple stage cluster sampling procedures that are common among household living standards surveys, a Taylor series expansion provides a consistent estimator of the variance of π_α. More specifically, for survey stratum k, cluster c, and n_k cluster samples drawn in the survey sample, a consistent estimator of the variance of π_α is:

$$(4) \qquad \hat{Var}(\pi_\alpha) = \frac{1}{p^2}\left[\hat{Var}(t) + \pi_\alpha^2\, \hat{Var}(p) - 2\pi_\alpha \hat{Cov}(t,p)\right],$$

where

$$(5) \qquad \hat{Var}(t) = \sum_{k=1}^{K} \hat{Var}(t_k) = \sum_{k=1}^{K} \frac{1}{n_k(n_k-1)} \sum_{c=1}^{n_k} (t_{kc} - t_k)^2 ,$$

$$(6) \qquad \hat{Var}(p) = \sum_{k=1}^{K} \hat{Var}(p_k) = \sum_{k=1}^{K} \frac{1}{n_k(n_k-1)} \sum_{c=1}^{n_k} (p_{kc} - p_k)^2 \text{, and}$$

$$(7) \qquad \hat{Cov}(t,p) = \sum_{k=1}^{K} \hat{Cov}(t_k, p_k) = \sum_{k=1}^{K} \frac{1}{n_k(n_k-1)} \sum_{c=1}^{n_k} (t_{kc} - t_k)(p_{kc} - p_k) .$$

The crux of our argument goes back to equation (1). Whereas the welfare metric y_j is treated as a random variable with a sampling error, the absolute poverty line z is routinely treated as a fixed constant, even though it is also estimated from the survey data. Ignoring this variance component leads to incorrect estimates of the precision of poverty estimates, and potentially misleading poverty comparisons over time and space.

The intuition of the argument is seen in Figure 1. In both panels of the figure, the horizontal axis is the welfare measure, the vertical axis is the proportion of the population, the dark curved line is the empirical cumulative density function (CDF) of the welfare measure[5], and the vertical line labeled z is the poverty line. The dotted lines on either side of the CDF are an indicative confidence interval for the cumulative density of the welfare measure. The point estimate of the poverty headcount, $\hat{P}_0$, is read from the vertical axis, at the level where the poverty line intersects the CDF. If the poverty line is assumed to be fixed, then the confidence interval for $\hat{P}_0$ is the interval AB on the vertical axis, corresponding to where the upper and lower bounds of the CDF confidence interval intersect the poverty line.[6] In the lower panel of Figure 1, the assumption of a fixed poverty line is relaxed, and the dashed vertical lines represent the confidence interval around the poverty line. If the estimated welfare metric and the estimated poverty line are independent (i.e., the covariance is zero), then the confidence interval around $\hat{P}_0$ would expand, as shown by the interval CD along the vertical axis of the lower panel. In practice, the poverty line and the welfare metric are unlikely to be independent, and the overall effect on the precision of the estimate of $\hat{P}_0$ will depend on the joint distribution of the two random variables, including the possibility that the estimate of $\hat{P}_0$ will be more precise when the variance of the poverty line is taken into account, as noted by Preston (1995). Figure 1 also illustrates that whether or not the poverty line variance is included in the poverty measure's standard error, the precision of the poverty estimate also

---

[5] The upper end of the CDF is truncated to focus on the region around the poverty line.

[6] This is something of an oversimplification, as calculation of the true confidence region is more complex than is shown in Figure 1. Nevertheless, it does capture the essence of the idea and is used here to illustrate the argument.

depends on the location of the poverty line. Poverty lines that are closer to the mode of the distribution, where the CDF has its steepest slope, will also tend to generate less precise poverty estimates. It should also be noted that analogous illustrations could be employed for higher orders of the FGT poverty measures, for example, by using the poverty deficit curve (Ravallion 1994b, Deaton 1997) in place of the CDF for the estimate of the average poverty gap, P_1.

<Figure 1 about here>

3. Data and methods

This section describes our approach to incorporating the sampling error of the poverty line in estimates of standard errors of poverty measures using household data from Mozambique as a case study. Before describing the approach to calculating standard errors specifically, we first describe the data collection process, the definition of the welfare metric, and the setting of poverty lines. This is presented in some detail because the individual steps determine not only the point estimates of the poverty lines and poverty measures, but also the bootstrapped estimates of their standard errors.

Data collection

We use data from the 2002–03 national Household Budget Survey in Mozambique, also known by its Portuguese abbreviation IAF (for Inquérito aos

Agregados Familiares sobre Orçamento Familiar).[7] The survey was carried out from July 2002 through June 2003, visiting 8,700 households throughout the country. The sample had 21 strata: separate rural and urban strata for each of Mozambique's ten provinces, plus one for the capital city of Maputo. A three-stage procedure was used to select sample households. Within each stratum, primary sampling units were randomly selected with probability proportional to size. In the second stage, enumeration areas (already defined on the basis of the 1997 Census) were selected, again with probability proportional to size. One month before the launch of the survey, the survey teams carried out a complete listing of all households in each of the 857 selected enumeration areas (EA). In the third stage, households were randomly selected within each EA, with 12 households per urban EA and 9 households per rural EA. The survey was limited to households residing in private residences, thus excluding those living in institutions (e.g., prisons, boarding schools, military barracks), diplomatic residences, and the homeless. The complex sampling structure implies unequal probability of selection across EAs, so sampling weights were calculated as the inverse of the probability of selection.

The content of the 2002–03 IAF is similar to that of other household budget surveys conducted in low income countries. Households were visited by interviewers at least three times over a seven day period. On the first visit, the interviewer and household completed the module on general household characteristics, and collected food and nonfood consumption information with reference to the preceding day (purchases, consumption from home production, and in kind transfers received). On subsequent visits

---

[7] Additional details about the survey may be found in INE (2004).

other parts of the questionnaire were completed (monthly expenditures, annual expenditures, and income), as well as daily consumption information for the period since the previous interview.

Definition of the welfare metric

The approach used to calculate consumption follows closely the one described by Deaton and Zaidi (2002) and Deaton and Grosh (2000), drawing from several modules of the IAF. It measures the total value of consumption of food and nonfood items (including purchases, home-produced items, and gifts received), as well as imputed use-values for owner-occupied housing and household durable goods. Market purchases were valued at the price paid, whereas non-market purchases were valued at the prevailing market price in the area at that time. The only two significant omissions from the consumption measure—both because of lack of data—are consumption of commodities supplied by the public sector free of charge (or the subsidized element in such commodities) and consumption of home produced services. For example, an all-weather road, or a public market, or a public water tap, presumably enhances the well-being of the people who use those facilities. Similarly, home produced services, such as cooking and cleaning, also add to welfare. These are often not captured by household surveys.

Food prices tend to follow a seasonal pattern, which implies that the purchasing power of a given amount of money varies during the year. For example, to acquire the same amount of food, a given household might have to spend twice as much in January as it spends in June. If the household consumed the same amount in real (quantity) terms in those months, it would appear to have a higher standard of living in January in nominal monetary terms. To avoid this kind of inconsistency, an intra-survey temporal food price index was developed from the survey data, and all nominal values of food consumption were adjusted by the index to take into account these price fluctuations.

For welfare comparisons, the metric used is total household consumption divided by household size, or consumption per capita. A sensitivity analysis was conducted using an adult equivalence scale that gives greater weight to working age adults and lower weight to children, based on nutritional requirements. Although this led to considerable re-ranking of households, the sensitivity analysis showed no appreciable difference in the aggregate poverty measures at the national and provincial levels. As all equivalence scales are arbitrary to some degree, we opt for the simplest approach and retain the per capita normalization.

Setting poverty lines

Poverty lines were set using the CBN approach (Ravallion 1994b). Mozambique is a large country with poorly developed infrastructure and markets. High transactions costs, combined with wide variation in agro-climatic conditions and production costs, lead to wide spatial and temporal variation in the prices of basic goods. In particular, differences in relative prices across space and time affect not only the total cost of acquiring basic needs, but also the composition of the basic needs bundle, as households adjust their consumption patterns in response to differences in relative prices.

As absolute poverty lines are supposed to represent the cost of achieving the same standard of living across the domain of comparisons, it is necessary to establish region-specific poverty lines. To define the poverty lines, the country was divided into 13 regions, based on an aggregation of the 21 survey strata that preserved the distinction between rural and urban areas, but grouping adjacent strata with similar characteristics if they had relatively few observations.

For each poverty line region, the food poverty line is constructed by determining the caloric content of the typical diet of the poor in that region, the average cost (at local prices) of a calorie when consuming that diet, and the food energy intake requirements for the reference population (the poor). The food poverty line—expressed in monetary cost per person per day—is the region-specific cost of meeting the caloric requirements when consuming a food bundle comprised of goods that the poor in the region actually consume.[8] It bears emphasizing that the food bundle is not determined by an externally imposed least-cost diet, but rather by the food consumption characteristics of poor households as recorded in the survey, which contributes to the sampling error of the poverty line.

The decision to allow the basic needs food bundles to vary by region was driven by the large differences in relative food prices across the 13 poverty line regions, and corresponding consumer behavior consistent with cost minimization. Ravallion (1998)

---

[8] The typical food bundle of the poor may contain more or less calories than the requirement for that region. This bundle is then proportionally scaled up or down until it yields exactly the pre-established caloric requirement, and the cost of this rescaled bundle at region-specific prices determines the food poverty line for that region. Also, it is recognized that food energy is only one facet of human nutrition, and that adequate consumption of other nutrients, such as protein, iron, vitamin A, and so forth, is also essential for a healthy and active life. However, like most multipurpose household surveys, the information on food consumption in the IAF data set is not sufficiently detailed to permit estimation of the intake and absorption of other nutrients. Use of energy requirements alone is also well established in the poverty measurement literature (Greer and Thorbecke 1986; Ravallion 1994a, 1998).

and Tarp et al. (2002) present arguments that allowing the food bundle to vary by region can result in more consistent poverty comparisons than using a fixed national bundle. Recent poverty studies that use region-specific poverty bundles and prices include Tarp et al. (2002), Mukherjee and Benson (2003), Gibson and Rozelle (2003), Ravallion and Lokshin (2003), Datt and Jolliffe (2005), and Arndt and Simler (2005). Note that the same arguments in favor of allowing the bundle to vary over space can also be applied to comparisons over time.

The relevant food bundles and associated prices were estimated for relatively poor households using the iterative procedure described by Ravallion (1998). Households were ranked by nominal consumption per capita, with the bottom X percent identified as the relatively poor. The value for "X" may be considered as a preliminary estimate of the poverty headcount. Preliminary poverty line calculations were made, and the nominal consumption values converted to real terms (i.e., taking into account region-specific differences in the cost of acquiring the basic needs bundle). Households were then re-ranked using this first approximation of consumption per capita in real terms, and the bottom X percent of this ranking identified as the relatively poor. Observed consumption patterns and prices in this sub-sample were calculated, producing a second estimate of food poverty lines, by which the households were re-ranked again. The iterative process continues until it converges, meaning that the same, or nearly the same, sub-sample of households appears in the poorest X percent. We experimented with several starting values for X, and found that for any reasonable value of X the process tended to converge on 48 percent, with convergence occurring after four or five iterations.

A challenge in the development of multiple poverty bundles involves comparability in the quality of bundles across regions (and through time). In principle, each bundle should afford the same level of utility. Arndt and Simler (2005) employ revealed preference tests to determine whether the estimated bundles satisfy revealed preference conditions. They also propose and implement a procedure, based on information theory, for adjusting the bundles such that they satisfy revealed preference conditions and retain the maximum information content inherent in the original estimated bundles.

The adjusted bundles meet revealed preference conditions and satisfy calorie requirements. Caloric requirements for moderately active individuals, disaggregated by age and sex, were obtained from the World Health Organization (WHO 1985). Average per capita requirements were allowed to vary by poverty line region, reflecting differences in the average household composition across regions. In practice, the average daily food energy requirement varies little across the 13 regions, averaging approximately 2150 kilocalories per person.

Whereas physiological needs provide the conceptual underpinning of the food poverty lines, no similar basis is readily available for defining nonfood needs. Yet, even very poor households in virtually all settings allocate a sizeable proportion of their total consumption to nonfood items, such as shelter and clothing. We estimate the nonfood poverty line by examining the proportion of total consumption allocated to nonfoods among those households whose total expenditure is approximately equal to the region-specific food poverty line (Ravallion 1994b, 1998; Ravallion and Bidani 1994). Specifically, we estimate the nonfood component of the poverty line as the average nonfood budget share of households whose total consumption is between 80 and 120 percent of the food poverty line, using a triangular kernel to give more weight to those households closer to 100 percent of the food poverty line. The rationale is that if a household's total consumption is only sufficient to purchase the minimum amount of calories using a food bundle typical for the poor, any expenditure on nonfoods is either displacing food expenditure or forcing the household to buy a food bundle that is inferior to that normally consumed by the poor, or both. In either case, the nonfood consumption of such a household displaces essential food consumption, and can be presumed to be a basic need.

Estimating poverty measures and their standard errors

After calculating the consumption variable and estimating the region-specific poverty lines, obtaining point estimates of FGT poverty measures for the population and sub-groups requires nothing more than application of equation (3) to the survey data. Obtaining consistent estimates of the standard errors of the poverty measures is less obvious.

It should be clear from the description of constructing the poverty lines that the poverty lines, as well as the welfare metric, are built from a series of estimates of population characteristics from the sample survey data. Food energy requirements are based on survey estimates of the population's age and sex distributions. The expenditure patterns that determine the basic needs food bundles are also estimates that are subject to sampling error, as are the nonfood budget shares that determine the nonfood poverty line. Similarly, the prices used to estimate the cost of the basic needs bundles come from the survey. In this light, it seems difficult to justify the common assumption that the poverty lines are not a source of sampling error in poverty estimates.

Given the complexity of the construction of the poverty lines, deriving standard errors of the poverty measures analytically is intractable, so we estimate them via a bootstrapping procedure. Bootstrapping is a general means of generating consistent estimates of an estimator's sampling distribution when an analytical solution cannot be derived or requires unreasonable assumptions (Efron 1979, Efron and Tibshirani 1993). It is based on repeated resampling (with replacement) of sample data, and requires minimal assumptions, the primary assumption being that the distribution of the estimator in the observed sample is a good approximation of its distribution in the population.

In our application, the bootstrap samples are drawn in a manner that mimics the stratified cluster sample design of the IAF survey. That is, within each stratum, c clusters are randomly drawn, with replacement, where c is the number of primary sampling units in the survey. When a cluster is drawn, all of the households in that cluster are drawn. Because the bootstrap sampling is done with replacement, each cluster (and household) may appear one or more times in a given bootstrap sample, or not at all. The estimated poverty lines, poverty headcount, poverty gap, and squared poverty gap are calculated for each bootstrap sample. The process is repeated 1,000 times. The standard deviation of a poverty measures over the 1,000 bootstrap replications is a consistent estimator of the standard error of that poverty measure. The point estimates of the poverty measures are calculated from the original, non-bootstrapped sample (Efron and Tibshirani 1993).

The process of estimating the poverty lines and poverty measures in each bootstrap replication is summarized in Table 1, which is divided into three columns. The first column lists processes that can be undertaken prior to the bootstrap loop. The calculation of *nominal* consumption per capita for each household occurs at this step as this measurement is (almost entirely) independent of the particular sample drawn.[9] The second column contains processes undertaken within the bootstrap loop. These are the steps described earlier for calculating the poverty lines and the point estimates for the FGT poverty measures for each bootstrap sample. The third column shows post-bootstrap processing, which is simply the calculation of the standard deviations of the poverty measures over the bootstrap replications.

<Table 1 about here>

4. Results

The 13 region-specific food, nonfood, and total poverty lines are shown in Table 2. The variation in the cost of basic needs is considerable across regions. Some general patterns are evident, such as the higher poverty lines in urban areas of a given province or province grouping, and the tendency for the poverty lines to increase (within urban and rural zones) as one moves down the list, which is roughly ordered from northern

---

[9] In the Mozambique case, hedonic regressions were used to impute use-values for owner-occupied housing. Obviously, the value obtained then depends upon the sample. Nevertheless, nominal use-values (rent foregone) for owner-occupied housing is in principle observable at the household level. The poverty line, in contrast, is not. Based on this distinction, we elect to treat estimates of use-value for owner occupied housing as data.

provinces to southern provinces. Table 2 also shows the estimated standard errors of the total poverty line, estimated via the bootstrap process described earlier with 1,000 replications. The poverty line standard errors range from 4 to 14 percent of the point estimates, with most of them between 4 and 8 percent.

<Table 2 about here>

Table 3 presents estimates of the poverty headcount index at the national level and for several sub-national groupings. The national headcount ratio is 54.1 percent, and ranges from 36 percent in Sofala province to 81 percent in Inhambane province. The column showing standard errors without poverty line error uses the Howes and Lanjouw (1998) method described in section 2, which includes complex sample design effects and is the method used most often in the current literature. At higher levels of aggregation, such as the national level or estimates for rural and urban areas, the standard errors are 2 to 4 percent of the point estimate. As sample size decreases with further disaggregation, the standard errors reach as high as 11 percent of the point estimates, although some of the provincial estimates are still fairly precise (e.g., Inhambane and Maputo province).

<Table 3 about here>

The next to last column of Table 3 shows the standard errors including the sampling error of the poverty lines, as estimated using the bootstrap procedure described in the preceding section. These standard errors are larger in all instances but two where the point estimated standard errors are slightly smaller. The possibility of poverty line error offsetting the error in the welfare measure that Preston (1995) described does not result in qualitatively smaller standard errors. As seen in the rightmost column, the standard error of the national headcount is 27 percent higher when poverty line sampling error is included. For other levels of aggregation, including the poverty line as a source of variation increases the standard error of the headcount estimate from effectively zero to more than 33 percent in Gaza province. On average, including the poverty line sampling error increases the estimated standard errors of the poverty headcount by about 15 percent.

Table 4 shows the same set of results for the poverty gap index. At each level of aggregation the standard errors of the poverty gap index are larger relative to the point estimate than is observed for the headcount index. This is consistent with Kakwani's (1993) observation that the precision of FGT poverty measures (measured as the standard error divided by the point estimate) tends to decrease for higher levels of α, a finding that is corroborated by the results of Howes and Lanjouw (1998). Comparing the standard errors estimated with and without poverty line sampling error we see that in Manica province, including the poverty line sampling error reduces the total standard error of the poverty gap estimate. For all other estimates the poverty line error increases the standard error of the poverty gap estimate, in some cases by as much as 34 percent. On average, the inclusion of poverty line sampling error increases the standard errors of the poverty gap estimates by about 17 percent.

<Table 4 about here>

Analogous estimates for the squared poverty gap (P_2) are shown in Table 5. These are qualitatively similar to the poverty gap results. For either set of standard errors the point estimates are less precise than the headcount or poverty gap estimates. In all cases excepting Manica, including the poverty line error increases the standard error of the P_2

estimates. On average, including poverty line error increases the standard error estimates by approximately 17 percent with a maximum increase of 36 percent.

<Table 5 about here>

How important is the increase in standard errors of the estimated poverty measures when poverty line sampling error is included? One way of assessing this is to put it in the context of the existing literature. As indicated earlier, Howes and Lanjouw (1998) found that accounting for sample stratification and clustering increased the standard errors of estimated FGT poverty measures by 26 to 33 percent in Pakistan and 45 to 64 percent in Ghana. Adding the poverty lines as a source of error increases the standard errors of the national-level poverty estimates in Mozambique by 27 to 29 percent. This suggests that accounting for poverty line sampling error may be slightly less important quantitatively than accounting for complex sample design, although results from other countries, and using alternative methods of setting the poverty lines, would be needed before drawing a firm conclusion in this regard. It should also be noted that there is no conflict between incorporating sample design and including poverty line error. Rather, it is advisable to do both, as in the present example, in which the complex sample design was also included in estimating the poverty line error. In short, there is no good reason to consistently overstate the precision of the poverty headcount by 27 percent, and the higher order FGT measures by approximately the same margin.

5. Conclusions

Poverty reduction is a fundamental objective of economic development, and reducing poverty is a major focus of governments, international financial institutions, and non-governmental and community-based organizations. The success of policies, programs, and development lending is increasingly judged in terms of poverty reduction. There has been substantial progress over the past three decades in the measurement of poverty, with the development of additively decomposable measures that reflect not only the number of poor persons, but also the depth and severity of poverty for sub-groups of the population. As most poverty estimates come from sample survey data, the statistical properties of poverty measures and appropriate inference procedures are important for evaluating the precision of poverty estimates and the statistical significance of poverty comparisons.

Studies of relative poverty have observed that there is sampling error associated with both the welfare metric and relative poverty lines calculated from the survey data. The recognition of poverty lines' sampling error has not extended to absolute poverty lines, even though they are also routinely estimated from sample survey data. This paper addresses this gap by proposing a general method for including the sampling error of poverty lines in the standard error of poverty measures, using CBN poverty lines and FGT poverty measures as one example. The approach is based on bootstrap methods that can be similarly applied to other methods of setting poverty lines (such as the FEI approach) and to other additively decomposable poverty measures.[10]

Using recent data from Mozambique, we estimate that accounting for the sampling error of poverty lines increases the standard errors of FGT poverty measures by an average of about 15 percent, with the standard errors increasing by up to 36 percent for some sub-groups. Significantly, larger scale measures were relatively strongly affected. For example, the estimated standard deviation of the poverty headcount at the national level increased by 27 percent. Thus, to be considered statistically significant, changes in poverty levels need to be larger than previously believed.

Are there circumstances in which one can safely ignore the sampling error of poverty lines, and treat them as fixed constants, without sampling error that contributes to the error of the poverty measures? In our view, the only situation would be the case of poverty lines that are determined exogenously, without reference to survey data. As absolute poverty lines are supposed to reflect the same standard of living across the domain of comparisons, and the cost of acquiring basic needs inevitably varies spatially and temporally, it is highly improbable that one could divine utility-consistent poverty lines without reference to data. Given a choice between arbitrarily specifying poverty lines that are certain to be utility-inconsistent to an unknown degree, and accepting a measurable loss in precision by estimating poverty lines from available data, the latter has clear advantages.

---

[10] The scope for applying these methods to non-additively decomposable poverty measures, such as those proposed by Sen (1976) and Kakwani (1980), is a topic for further investigation.

Despite the appeal of making poverty comparisons across a range of plausible poverty lines, stochastic dominance approaches to comparing money metric welfare distributions are not completely exempt from considering the sampling error of poverty lines when making statistical inferences. Poverty lines are not only a dividing line (admittedly artificial) between the poor and nonpoor, but also serve as cost of living indices, permitting interpersonal welfare comparisons when the cost of acquiring basic needs varies over time or space (Ravallion 1998). Stochastic dominance methods require that the same welfare metric is used for comparing populations or sub-groups, which is frequently accomplished by mapping nominal consumption to real consumption using indices derived from poverty lines, or expressing the welfare metric as a proportion of the relevant poverty line. If these poverty lines or cost of living indices are estimated from survey data, then the associated sampling error should be included in the confidence interval around the empirical cumulative distribution function, which will affect the precision of poverty comparisons.[11] Adapting the methods presented in this paper to stochastic dominance approaches to poverty comparisons is an area for future research.

---

[11] Likewise, because the dollar-a-day poverty line is based in part on statistically estimated purchasing power parity (PPP) calculations, it is not immune from the poverty line sampling error described in this paper.

References

Arndt, C. and K. Simler, Estimating Utility Consistent Poverty Lines, National Directorate of Studies and Policy Analysis discussion paper E6, Ministry of Planning and Development, Maputo, 2005.

Atkinson, A. B., On the Measurement of Poverty, *Econometrica*, 55(4): 749–764, 1987.

Bishop, J. A., K. V. Chow, and B. Zheng, Statistical Inference and Decomposable Poverty Measures, *Bulletin of Economic Research*, 47(4): 329–340, 1995.

Clark, S., R. Hemming, and D. Ulph, On Indices for the Measurement of Poverty, *Economic Journal*, 91(362): 515–526, 1981.

Davidson, R. and J.-Y. Duclos, Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica*, 68(6): 1435–1465, 2000.

Deaton, A. and M. Grosh, Consumption, in M. Grosh and P. Glewwe (eds.) Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience, World Bank, Washington, D.C., 2000.

Deaton, A. and S. Zaidi, Guidelines for Constructing Consumption Aggregates for Welfare Analysis, Living Standards Measurement Study Working Paper No. 135, World Bank, Washington, D.C., 2002.

Datt, G. and D. Jolliffe, Poverty in Egypt: Modeling and Policy Simulations, *Economic Development and Cultural Change*, 53(2): 327–346, 2005.

Deaton, A., *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore, The Johns Hopkins University Press for the World Bank, 1997.

Efron, B., Bootstrap Methods: Another Look at the Jackknife, *Annals of Statistics*, 7(1): 1–26, 1979.

Efron, B. and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.

Foster, J. E., J. Greer, and E. Thorbecke, A Class of Decomposable Poverty Measures, *Econometrica*, 52(3): 761–766, 1984.

Foster, J. E. and A. F. Shorrocks, Poverty Orderings, *Econometrica*, 56(1): 173–177, 1988.

Gibson, J., and S. Rozelle, Poverty and Access to Roads in Papua New Guinea, *Economic Development and Cultural Change*, 52(1): 159–186.

Greer, J. and E. Thorbecke, A Methodology for Measuring Food Poverty Applied to Kenya, *Journal of Development Economics*, 24(1): 59–74.

Guio, A., Income Poverty and Social Exclusion in the EU25, Luxembourg: Eurostat, 2005.

Howes, S. and J. O. Lanjouw, Does Sample Design Matter for Poverty Rate Comparisons?, *The Review of Income and Wealth*, 44(1): 99–109, 1998.