A Discussion Note: Constructing Comparable Household Survey Data for the Analysis of Poverty in South Africa (1995–2000)

Statistics South Africa

<u>**Important Note:**</u>
This discussion note reflects 'work-in-progress'; a final report will be published during the third quarter of 2007.

Introduction

In 2000, Statistics South Africa (Stats SA) published the report '*Measuring Poverty in South Africa*' as a first step in presenting a multi-dimensional approach to poverty analysis in South Africa. By 2002, additional data had become available making it possible to generate 'time series type' datasets for the analysis of poverty. However, the creation of such datasets presented two major challenges.

First, in terms of income-poverty it became clear that a poverty threshold (in the form of single or multiple poverty lines) was needed to monitor trends in this dimension of poverty. The second challenge was the construction of comparable datasets to facilitate poverty comparisons over time.

This discussion note reflects 'work-in-progress' by Stats SA and collaborating partners – including the World Bank – in addressing these challenges. The work presented in this report covers the period 1995 to 2000, and a final report covering the period 1995 to 2006 will be published during the third quarter of 2007.

Data sources and methodology

The data to construct the nominal consumption aggregates come from four Stats SA surveys that were conducted in 1995 and 2000. Two of the surveys are the 1995 October Household Survey (OHS) and the second round of the bi-annual Labour Force Survey (LFS) September 2000, while the other two are the Income and Expenditure Surveys (IES) of 1995 and 2000. The IES survey was conducted every five years

among participants of the concurrent OHS or LFS.[1] While the size of the sample was the same in both years, the units differed. Combined, these surveys provide information on household income and expenditure, along with information on other household characteristics, such as household size and composition, work, access to services and housing.

The data we utilize to transform nominal household expenditure into real expenditure comes from the price surveys Stats SA conducts on a monthly basis. The Consumer Price Survey is conducted to collect information for the production of the monthly Consumer Price Index (CPI) using a reference set of goods and services called the CPI basket. The CPI reflects changes in the price levels of consumer goods and services. The basket and weights are re-designed every five years based on the Income and Expenditure Survey. From January 1997, the coverage of the CPI was increased   to cover 14 metropolitan areas and 39 other urban areas throughout all nine provinces of South Africa. Rural areas are not directly covered.

Building a comparable consumption aggregate

Comparable consumption aggregates over time usually refer to the building blocks of the consumption aggregate being the same in all reference periods. This is because, due to improvements in survey design or to capture changes in the economy, statistical offices change the way questions are asked, discard certain questions, add new ones, change the recall periods, etc. The literature shows that even small changes in survey design that look harmless to the researcher can lead to significant biases in poverty

---

[1] In 1995, the household survey was the annual OHS, and in 2000 it was the second round of the bi-annual LFS because the OHS series was discontinued after 1999.

comparisons.[2] In the case of South Africa, we are fortunate in this regard. The income and expenditure module has hardly changed between 1995 and 2000, and it is possible to build a consumption aggregate based on the same items for both 1995 and 2000.

However, there is another issue of comparability that is often not discussed, as it is usually a non-issue: the sampling frame. Usually, one is assured that the household survey from which the welfare indicators are derived is representative of the population. However, the sampling frame for the 1995 OHS and IES was the 1991 population census, which had several flaws. The 1991 census was carried out under the apartheid regime and had marked coverage problems, which carried over to the sampling frame of the 1995 survey. Drawing a sample that was representative of the population in South Africa in 1995 was difficult, especially in the absence of maps of enumeration areas for various parts of the country, as described by Hirschowitz in *Living in South Africa* (1996). Discrepancies in the distribution of population groups could cause overstatement (or understatement) of overall poverty, as well as for certain sub-groups, causing significant biases in any assessment of changes in welfare between 1995 and 2000.

Stats SA   recently recalculated the household weights for the IES 1995 sample, using information from the 1996 Population Census. Application of the new weights results in better population distribution estimates at provincial level (compared to 1996 Population Census figures) than previous OHS/IES 1995 population estimates. Similarly for 2000 surveys, we utilize a new set of sampling weights benchmarked to the 2001 Population estimates.

---

[2] See, e.g. Lanjouw and Lanjouw (1995).

Another issue regarding the comparability of the consumption aggregates has to do with the way a household was defined in each survey. Although the size of the sample was the same in both years (3000 Enumeration Areas in 1995; 3000 Primary Sampling Units in 2000),   the units for selecting sample households differed. The unit for sampling households in 1995 was the visiting point which usually meant selection of one household per visiting point, however, by 2000, the survey methodology had changed to enumeration of all households found at a sampled dwelling unit. Therefore, the 2000 surveys interviewed slightly more households than in 1995.

Building blocks of the consumption aggregate

We have closely followed the recent literature on building consumption aggregates, in particular the guidelines put forth by Deaton and Zaidi (2002), Ravallion (2001), and Lanjouw et al (1996).[3] The consumption aggregate includes the following expenditure categories: food, beverages, and cigarettes (excluding home production); housing (imputed rental value of residence and utilities); compensation for domestic workers; personal care, household services, and other household consumer goods; fuel (excluding firewood and dung); clothing and footwear; transport (excluding cost of purchased vehicles); communication; education; reading matter, cost of licences and other rental charges; and cost of insurance.

Important categories of expenditure we have excluded from the consumption aggregate are: water, firewood and dung; health; imputed value of household durables; food consumption from home production; lobola/dowry, funerals, religious or traditional ceremonies, gambling; and 'lumpy' expenditures, such as furniture, appliances, vehicles,

---

[3] The literature on using income vs. expenditure to measure household welfare is well-established and we abstract from that debate here. We believe, mainly for empirical reasons, that using data on household expenditures is preferable to using data on incomes.

sound and video equipment, etc. Below, we explain in some detail the construction (or the reason for exclusion) of some important components of the consumption expenditure.

Water, firewood, and dung

Water, firewood and dung are items for which a significant percentage of households have no expenditure. However, we also know that households collect water and firewood, and hence may consume them while reporting no expenditure on these items. Quality concerns aside, if two households consume the same amount of water (or firewood), but one household pays for it while the other collects it, the inclusion of expenditure on these items would bias the consumption aggregate in favor of people who are paying for them. A preferable option would be to impute the value of water, firewood, and dung for households who collect it, but there is no satisfactory way in which this could be done. So, we have opted to leave out such expenditures in the consumption aggregate.

Health, lobola/dowry, funerals, religious or traditional ceremonies, and other 'lumpy' expenditures

Although 'lumpy' expenditures are relatively large, they are usually infrequent, i.e. only a small percentage of the households incur such expenditures during the recall period in a household survey. Including such expenditure would possibly result in overstating the welfare of households which incurred such expenditure during the recall period. Ideally, one would like to 'smooth' these lumpy expenditures, but unfortunately we do not have the means to do this with the available data, and hence we leave them out of the consumption aggregate.

Household durables, and home production of food

Purchases of household durables also fall into the 'lumpy' expenditures category, but in most welfare surveys it is possible to construct a 'rental equivalent' for durable goods that are reported to be in the household's possession. However, to make such calculations we need to have information not only on the household assets, but also on when they were purchased, the value of the item when purchased, etc., and unfortunately these data are not available to us.

Consumption from home production of food should also be ideally included in the consumption aggregate; however, the quality of the data was not deemed satisfactory to calculate the value of this consumption. While this may bias downwards the welfare of poorer households (as they are more likely to rely on home production), Deaton and Zaidi (2002) report that consumption of food from home production in South Africa, constituted only 2% of total household expenditures in 1993. Barring major changes in the consumption patterns in South Africa, the exclusion of consumption from home production should not significantly affect the profile of poverty in our analysis.

Housing

To derive the monetary value of the flow of services a household receives from occupying its dwelling, the obvious choice is to include the rent paid for that dwelling in the consumption aggregate (Deaton & Zaidi, 2002, pg. 35). However, many households own the dwelling they live in, or are given subsidized or freehousing that is subsidized or free of charge by their employer or the government, and hence do not pay rent. In most questionnaires such households are asked how much they think they would be paying if they were renting their house. This implicit rental value can be used in consumption

aggregates, if the information is deemed reliable. However, while this question was asked in the IES 1995, the module on housing changed slightly in IES 2000 and this question was omitted. In the absence of such data, we have decided to impute the rental value of houses based on housing characteristics (dwelling type, number of rooms, materials for roof and walls, etc.), access to services (electricity, water, sanitation, etc.), and type of area (urban, farm, etc.). We ran a separate regression for each province in each survey.[4] The explanatory power in these regressions was satisfactory - the R-squared ranged from 0.47 to 0.66 in 2000, and 0.29 to 0.48 in 1995. Using these regressions, rental expenditures were predicted for those households that did not pay rent for the dwellings in which they resided.

Poverty lines, Spatial and Inter-temporal price adjustments

To draw normative poverty lines for analysis, we use the 'cost-of-basic-needs' method. In this method, a basic consumption bundle is derived from household consumption data and cost estimates obtained from prices data (Ravallion, 2001). The food basket we have selected consists of the mean per capita quantities of each food item consumed by the third quintile of the (nominal) expenditure distribution in 2000. Using the nutritional value information for each food item obtained from the Medical Research Council (MRC) in South Africa, we calculated that this bundle provides households with roughly 1927 kilocalories per capita per day.

The IES 1995 and IES 2000 asked households for their expenditure on a large number of food and non-food items for the past month or year, but did not collect information on quantity purchased, to allow one to construct 'unit values'. Nor did these surveys collect

---

[4] As there were not enough renters in our sample for Limpopo, we performed one regression for Limpopo and Mpumalanga provinces. For considerations of space, we do not include all the regression results here, but they are available from the authors upon request.

information on prices of various food items in the markets in the communities sampled. However, Stats SA has been collecting monthly price data for practically all the items in the food module of the IES surveys in metropolitan and urban areas for some time for the CPI. We used the prices from January 2001 to construct an average representative food bundle in the IES 2000 and to price this bundle in each province.[5] With Western Cape as the '*numeraire*' (i.e. the price index in 2000 in Western Cape is 100), the ratio of the cost of the food bundle in each province to the cost in Western Cape gave us the Laspeyres food price index.

To derive the overall price index for 2000, we followed the methodology employed in Lanjouw et al. (1996). We derived a housing price index by predicting the "rental value" of a house in an urban area that has 4 rooms, brick walls, a flush toilet, and access to electricity and street lighting, in each province.[6] Again selecting Western Cape as the numeraire, we built a consumer housing price index by dividing the rental value of this house in each province by that in Western Cape. Finally, to derive a consumer price index for the remaining consumption items (non-food, non-housing), we took a weighted average of the food and the housing price indices, where the weights are the budget shares of food and housing. A weighted average of these three price indices – food, housing, and other non-food – were then used to construct an overall spatial price index for the 2000 IES.

The 'average' representative bundle of the third quintile costs 180 Rand in real 2000 prices, and as mentioned above it provides 1927 kilocalories per person. Using the

---

[5] The IES 2000 was conducted in October 2000, so the prices used are from only a few months after the survey was conducted. The changes in prices between September 2000 and January 2001 were small.
[6] These characteristics were not ad hoc selections, they are the median values for these measures for South Africa as a whole.

average energy allowances recommended by the Medical Research Council (MRC), we calculated that the consumption in kilocalories recommended for an average South African household per capita is 2261   Linearly adjusting the 180 Rand figure by 2261/1927, we arrive at a food poverty line (Z^f) of 211 Rand – the amount necessary to purchase enough food to meet the basic daily food-energy requirements.

To derive the overall poverty line, we set a lower bound and an upper bound for cost-of-basic-needs poverty lines in South Africa, following Ravallion (1994).[7] We calculate the mean non-food expenditure of those households whose total consumption lies in small but increasing intervals around the food poverty line (e.g. between 0.99 Z^f and 1.01 Z^f, 0.98 Z^f and 1.02 Z^f, …, 0.95 Z^f and 1.05 Z^f). The simple average of these mean non-food expenditures plus the food poverty line gives a lower bound poverty line of 322 Rand. The basic premise here is that if a household's total expenditure is equal to the food poverty line, then any non-food expenditure for that household must be absolutely necessary, as the household is giving up basic food needs for that consumption. Using the same technique, but this time calculating the mean total expenditure of households whose food consumption is equal to the food poverty line, we derive an upper bound poverty line of 593 Rand. If the basic needs norms that are anchored to food-energy requirements of South African households are deemed reasonable, then a per capita poverty line for South Africa may lie between 322 and 593 Rand in 2000 prices.

Finally, to incorporate the changes in price over time, we inflate the consumption of the households in the 1995 IES, using the provincial price indices published monthly by Stats SA in CPI statistical releases. However, the provincial indices only extend back to January 1997, and do not cover changes in overall prices by province for the year 1996.

---

[7] Ravallion (1994), pages 119-125.

To derive a spatial price index for 1995, we assumed that the change in prices in each province between January 1996 and January 1997 was the same as the annual change in the average prices in the four-year period of January 1997 to January 2001 in that province. Then, using the spatial price index in 2000 and the changes in prices between January 1996 and January 2001 in each province, we calculated the spatial price index for households in the 1995 IES. Finally, the real consumption of each household in 1995 prices was then converted to real consumption in 2000 prices by scaling it up by the change in prices in the province.

The prices we utilized to construct a food basket, to derive price indices, and to draw poverty lines have one possible drawback: they do not reflect urban-rural price differences. As mentioned earlier, Stats SA does not collect prices in rural areas, but only in metropolitan and some other urban areas. As we have deflated the consumption of each household by a provincial price index, regardless of the type of area they live in, our results comparing urban and rural expenditures may be somewhat biased. This possible bias can also affect provincial comparisons, as the urban/rural composition can differ vastly across provinces.[8] If prices in rural areas are significantly lower than those in urban areas, then one would expect our estimates of rural poverty to be biased upward. However, South Africa has a relatively good road network, so the price differential between urban and rural areas within each province is probably not as high as in countries with poor infrastructure. Furthermore, according to information collected on the area of purchase of goods and services in the 2000 IES, households in rural areas buy a significant amount of food items and most of their non-food items in nearby

---

[8] For example 12% of the population in Limpopo lived in urban areas in 2000, while that figure is 95% in Gauteng.

urban areas.[9] These two factors are likely to attenuate the size of the possible bias in our analysis. Furthermore, if the urban/rural price differential is constant over time, the bias should not affect inter-temporal poverty comparisons.

Next Steps

This discussion report has presented part of on-going work at Stats SA. Current tasks include several sensitivity tests using new data from the 2005/06 Income and Expenditure Survey, and from the recently re-engineered Consumer Price Index series. A final detailed report will be published during September – October 2007.

---

[9] For example, approximately 60% of rural households reported purchasing grain products in nearby urban areas.

References

Central Statistical Service (1996). *Living in South Africa*. Pretoria: Central Statistical Service.

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

Deaton Angus, and Salman Zaidi (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. LSMS Working Paper No. 135, World Bank, Washington.

Food and Nutrition Board, National Research Council (1989) *Recommended Dietary Allowances, 10th ed*. Washington DC: National Academy Press

Lanjouw, Jean Olson, and Peter Lanjouw (1995). How to Compare Apples and Oranges: Poverty Measurement based on Different Definitions of Consumption. *Review of Income and Wealth.* 47(1), pages 25-42.

Lanjouw, Peter, Giovanna Prennushi, and Salman Zaidi (1996*). Building Blocks for a Consumption-Based Analysis of Poverty in Nepal*. The World Bank. Unpublished manuscript.

Medical Research Council (1991) South African Food Composition Database Version 1.2. Tygerberg: Medical Research Council.

Ravallion, Martin (1994). *Poverty Comparisons*. Harwood Academic Publishers.

Ravallion, Martin (2001). *Poverty Lines: Economic Foundations of Current Practices*. The World Bank. Unpublished manuscript.

World Development Indicators (2002). The World Bank.