WPS6217

Policy Research Working Paper 6217

Poverty in Mozambique

New Evidence from Recent Household Surveys

Federica Alfani
Carlo Azzarri
Marco d'Errico
Vasco Molini

The World Bank
Africa Region
Poverty Reduction and Economic Management Department
October 2012

Policy Research Working Paper 6217

Abstract

This paper has three primary objectives: (i) to investigate potential problems regarding Mozambique's most recent nationally representative household survey on poverty dynamics; (ii) to assess the robustness and reliability of official poverty statistics; and (iii) to provide alternative estimates of poverty and welfare indicators in light of the methodological and analytical issues raised in areas (i) and (ii). It is determined that at least two significant weaknesses affect the official poverty-rate estimates: measurement errors in consumption data and flaws in the methodology used to calculate poverty lines (the cost-of-basic-needs approach based on provincial food bundles with entropy correction). A number of observations appear to be affected by substantial measurement errors, which severely distort the official poverty statistics. The paper provides methods to correct the consumption distribution by recalculating poverty lines based on a single national food basket—as opposed to the current estimates, which are based on province-specific food baskets. The revised poverty statistics differ considerably from the official estimates of poverty across provinces and are far more consistent with other poverty indicators. In addition, poverty appears to be highly concentrated in certain areas, with dramatically higher rates found in Central and Northern Mozambique, as well as in rural areas overall, compared with relatively low rates in Southern Mozambique and in the country's urban centers. These findings substantially contradict the government's official poverty figures, which appear to systematically overestimate poverty rates in Mozambique's Southern provinces and urban areas while simultaneously underestimating the prevalence of poverty in the country's Central and Northern regions and in rural areas nationwide.

This paper is a product of the Poverty Reduction and Economic Management Departyment, Africa Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at vmolini@worldbank.org.

Poverty in Mozambique: New Evidence from Recent Household Surveys[1]

Federica Alfani, Carlo Azzarri, Marco d'Errico and Vasco Molini

Keywords: Poverty lines; location specific poverty lines; measurement error; Mozambique

JEL Codes: O15, I32, POV

[1] Corresponding author, Vasco Molini, World Bank, vmolini@worldbank.org; Azzarri: International Food Policy Research Institute, C.Azzarri@cgiar.org; Alfani: Food and Agriculture Organization, Federica.Alfani@fao.org; d'Errico: Food and Agriculture Organization, Marco.DErrico@fao.org,. Comments on earlier versions by Peter Lanjouw, Kathleen Beagle, Bart van den Boom, Gero Carletto, Thomas Pave Sohnesen, Andrew Dabalen, Julio Revilla and Sean Lothrop are gratefully acknowledged. Special thanks are due to Antonio Nucifora for strongly encouraging and supporting the process of revising the poverty figures. The usual disclaimers apply. We gratefully acknowledge financial support from the World Bank. The views expressed here are those of the authors and do not necessarily reflect those of the funding organizations, and the institutions they are affiliated with.

1. Introduction

Three consecutive National Poverty Assessments (NPAs) conducted in Mozambique between 1998 and 2010 (MPD-DNEAP, 2010; MPF/IFPRI/PU, 2004; MPF/UEM/IFPRI, 1998) have provided a wealth of information on national poverty patterns and their changes over time. The most recently published NPA, based on the new Mozambique Household Budget Survey (IOF,[2] 2008/09) shows that while welfare levels have generally increased, poverty remains widespread. The total percentage of Mozambicans living below the poverty line (based on per capita consumption measured at the household level) was estimated at 54.7 percent, a slight and statistically insignificant increase from 54.1 percent observed in 2002/03, with a marginal rise in rural poverty (from 55.3 to 56.9 percent) offsetting a small decline in urban poverty (from 51.5 to 49.6 percent).

However, the poverty profile that emerged from the latest NPA (MPD-DNEAP,[3] 2010), which is computed using a consumption-based money metric as its measure of welfare, does not seem to be consistent with other poverty indicators, from child anthropometrics (see Azzarri et al., 2011) to asset ownership figures to measures of access to basic services and regional productivity estimates. These and other factors raise concerns about the robustness of the poverty statistics presented in the latest NPA, and the discrepancies in the data become even more striking at the provincial level.[4] In the most recent NPA, the rural Niassa and Cabo Delgado regions, historically considered poor according to a variety of indicators, recorded the lowest poverty rates in the country. By contrast, the country's Southern provinces and urban areas registered very high poverty rates despite their traditionally superior nonmonetary poverty indicators (e.g. child malnutrition) and higher regional productivity.

Three types of error may have contributed to producing these counterintuitive results. First, measurement errors in expenditure data could be caused by consumption underreporting (especially consumption of household produce, which is common in rural areas) or by flaws in the survey questionnaires. Second, the methodology used to compute poverty lines[5] (PLs) differs from the standard used in many other developing countries, and methodological weaknesses might result in unexpected provincial patterns. Third, the interaction between consumption underreporting and the particular methodology used to compute the PLs could magnify inaccuracies in the data.

[2] *Inquérito sobre Orçamento Familiar.*

[3] *Ministério de Planificação e Desenvolvimento – Direcção Nacional de Estudos e Análise de Políticas*: Ministry of Planning and Development – National Directorate of Policy Studies and Analysis.

[4] The Northern, Central and Southern regions are grouped according to the official country definition. *Northern* includes the provinces of Niassa, Cabo Delgado and Nampula; *Central* includes Tete, Zambezia, Manica and Sofala; and *Southern* includes Gaza, Inhambane, Maputo Province, and Maputo City. For an overview of Mozambican provinces, see Map 1 below.

[5] All the three NPAs report a different poverty line, based on local prices and consumption patterns observed in the cross-sectional rounds of, respectively, the IAF (*Inquérito aos Agregados Familiares sobre Orçamento Familiar*) 1996/97, the IAF 2002/03 and the IOF 2008/09 survey, by grouping provinces and areas according to 13 locations: Niassa and Cabo Delgado, rural and urban; Nampula, rural and urban; Sofala and Zambezia, rural and urban; Manica and Tete, rural and urban; Gaza and Inhambane, rural and urban, Maputo Province, rural and urban, and Maputo City.

This paper attempts to analyze these three potential sources of bias in Mozambique's national poverty figures and to isolate their effects on the final estimates. Revised estimates are also proposed; although tentative, the revised estimates are more consistent with a range of poverty correlates and with the historic geographical welfare distribution of the country. The paper is organized as follows: Section 2 presents an overview of potential measurement errors in the 2002/03 and 2008/09 surveys, it also describes our strategy for identifying observations affected by measurement errors and re-estimating them; Section 3 deepens the analysis of the relationship between consumption underreporting and the PLs and proposes an alternative methodology for calculating the PLs; Section 4 presents the results of using new PLs based on revised estimates of observations likely affected by underreporting and discusses the consequent differences between the official and revised poverty levels and trends over time; Section 5 concludes the analysis.

2. Measurement Errors in Consumption Estimates

2.1 Underreported consumption

This section reports the most relevant general findings regarding the apparent underestimation of consumption and describes the background and rationale for the present work. This analysis has two objectives: first, it seeks to identify a subset of potentially downward-biased observations in food-consumption reporting in the IOF 2008/09 data and, second, it attempts to examine and evaluate the possible sources of this bias. We explore the consumption data from three perspectives: overall calorie intake, dietary diversity patterns, and food-share patterns.

About 30 percent of households reported consuming an average of less than 1,000 k-calories per capita per day, which is considered to be below the minimum standard for survival by the World Health Organization (WHO 1985). Moreover, in some Southern regions of Mozambique median calorie intake is below even this basic threshold. Following the strategy described above, we run an OLS regression model in which per-capita consumption expenditure is regressed against poverty correlates. A clear pattern of error terms (not shown) becomes apparent: observations with a calorie intake of below 1000 tend to show high a concentration of negative residuals. This suggests that for households reporting average calorie intakes below 1000 per capita per day, the model predicts higher fitted expenditures, based on characteristics highly correlated to consumption, than do the original expenditure figures. In other words, the official food-consumption statistics for certain populations are conspicuously low and imply that poverty among these groups is both more prevalent and more severe than other relevant indicators would suggest.

To further assess these apparent incongruities in the data, dietary diversity is analyzed. Unexpectedly, the share of "cereals", a food category that is typically a mainstay in the diets of the poor, in fact increases with expenditures, whereas consumption of nutritionally superior (and usually much more expensive) items such as meat, fish, and dairy products does not increase with expenditures.

Finally, we find that—also contrary to normal patterns in consumption data—food consumption as a share of total household expenditures is very low in the South, despite being the region with the highest official poverty rate. Even more surprisingly, it seems that Engel's law[6] does not hold in rural areas, where the expected negative correlation between food consumption shares and expenditure levels is hardly detectable.

When looking at the causes of consumption underreporting, three elements emerge. The first is the pre-printed food-item list used in the consumption questionnaires. The 2008/09 survey included a list of 38 food items, 18 for purchased food and 20 for consumption of household produce. This list seems to be too restrictive to adequately capture dietary diversity and consumption patterns across provinces. In addition, the enumerators' apparent tendency to underreport food transactions (including food purchases, the bartering of household produce, or food received in exchange for work) could have been facilitated by an overly restrictive questionnaire, which may not have left enough room to record all food-related transactions and which did not provide extra room for reporting additional food items not included in the printed list. Also, the fact that the pre-typed list included only primary agricultural commodities, but not prepared or packaged food, could have prevented the comprehensive reporting of food consumed in urban areas, where households devote a larger share of consumption to packaged food. The 2002/03 survey was also affected by similar problems, and this appears to have biased earlier poverty figures, particularly for Southern Mozambique.

The second issue is poor management of data-collection activities. Food consumption data in the IAF[7] 2002/03 and IOF 2008/09 were collected in a different manner than initially planned. Food transactions should have been reported by the households themselves through the use of a diary, in which every two days they would record the daily food transactions that had occurred; and three enumerator visits to each household were initially envisaged. However, this procedure was not implemented. Instead, only one visit was completed in 93 percent of cases in the IOF 2008/09. It should not come as a surprise that many households showing low calorie consumption also reported fewer than ten transactions during the recall period (7 days); this is likely the result of limited supervision during the survey fieldwork.

Finally, the lack of official conversion factors for non-standard measurement units could have significantly affected calorie calculations. Two main assumptions are used in computing calorie consumption in the 2008/09 survey, namely (i) that one piece/unit of bread is set to weigh 200 grams in every province, although it is uncertain whether this is in fact true across the entire country, and (ii) that one coconut is assumed to contain 800 grams of edible parts, regardless of its actual size, ripeness or freshness.

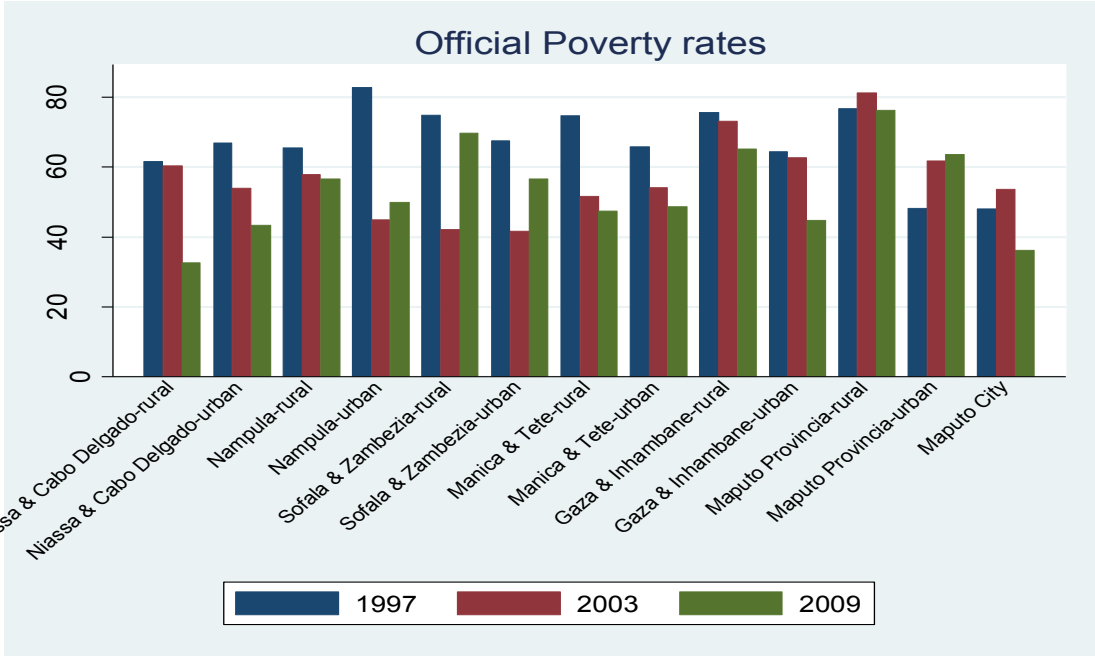
Problems related to caloric intake, and the considerable disparities between wellbeing indicators and poverty figures, also appear to affect the 2002/03 data. These flaws in data collection and analysis

[6] "Engel's law" refers to the typical observation that as income rises, the share of food consumption in total expenditures falls, or, in other words, that food is a normal good, with an income elasticity of demand between 0 and 1.

[7] *Inquérito aos Agregados Familiares sobre Orçamento Familiar*: "Household Survey of Family Budgets", the precursor to the IOF.

produce an extremely volatile and time-inconsistent poverty profile, resulting in a continuous re-ranking of provinces in terms of poverty rates (
Figure 1). However, the different (and arguably superior) design of the IAF 1996/97 produced higher statistical reliability and, as a consequence, greater coherence between poverty figures and other welfare indicators.[8] The more complete food list in the IAF 1996/97, which included 50 items, most likely led to better consumption measurements, reduced underreporting, and more limited concentration of reported food purchases in the pre-printed items.

Figure 1: Official Poverty Rates, by Province, by Survey Year



Official: 2008/09 PLs (MPD-DNEAP, 2010); **Simulated:** 1996/97 official PLs deflated by the consumer price index from World Development Indicators (2011)

The authors of the NPA (DNEAP, 2010), after carefully examining the problems in the design and implementation of household surveys, arrive at similar conclusions regarding the magnitude of underreporting and its spatial concentration. They propose, as a preliminary solution, the re-estimation of expenditures for households whose calorie intake falls below a certain threshold. These adjustments produce a substantial downward revision of poverty figures at the regional level in the South and a 3 percent lowering of the poverty rate at the national level. However, since the 2002/03 data also experienced similar problems in estimating poverty in the South, the NPA's authors argue that even though there is some regional variation, 'the trend between the two surveys results is very stable'.

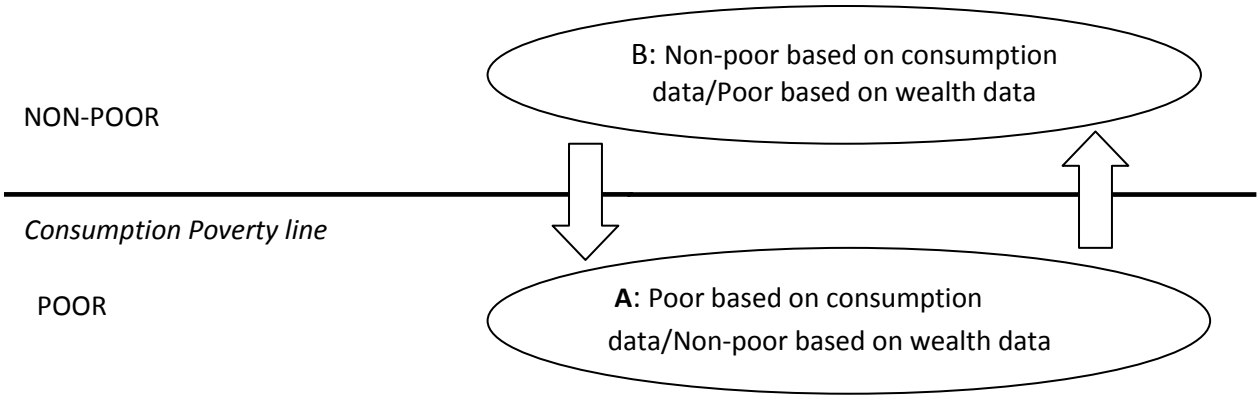
[8] In the IAF 1996/97 provincial poverty rates show a significant negative correlation with all welfare and education indicators.

2.2 Identifying unreliable observations: 2002/03 and 2008/09

In order to cope with measurement errors in consumption and with the 'error-poverty line interaction' (described in Section 3, below), we utilize a method that preserves most of the original information while correcting for only manifestly inaccurate observations. Our approach is to carefully screen all consumption data and define specific criteria by which to determine their degree of reliability.[9] In a second stage, the unreliable observations are re-estimated using out-of-sample predictions from the 1996/97 coefficients model (see Section 2.3, below). Since we also identified clear signs of measurement errors in 2002/03, after using this methodology for the 2008/09 data we apply it to the earlier survey as well.

The original consumption data are ranked and compared to a wealth index.[10] Observations that are below the poverty line (the 54th percentile of the income distribution), but above the 70th wealth-index percentile, or observations above the poverty line but below the 30th wealth-index percentile, are singled out as 'suspicious'.[11] Figure 2 illustrates the definition of these two observation sets: one set of households (A) is considered poor based on the original consumption data, but is classified as non-poor based on the wealth index; the other set of households (B) is considered non-poor based on the original consumption data, but is classified as poor based on the wealth index.

Figure 2: Observation Sets A and B



[9] The first-best approach to identifying consumption measurement errors would be to administer a small consumption-based survey questionnaire in the areas where underreporting and over-reporting are deemed most serious, revisiting the same households sampled in the IOF 2008/09. A comparison with the original data would indicate the extent to which questionnaire design and fieldwork operations may have impacted consumption data quality, diet-composition and caloric intake indicators. This information could then be used to adjust the consumption data in the IOF 2008/9.

[10] With this asset index, suggested by Filmer and Pritchett (2001), principal-components analysis is used to calculate the weights of the index. The first principal component, the linear combination capturing the greatest variation among the set of variables, can be converted into factor scores, which serve as weights. The rationale for using this index is that it captures the household's permanent welfare dimension better than simple consumption data, and can provide more reliable rankings between households.

[11] The use of the 30th and 70th percentiles is not arbitrary: choosing extreme and symmetrical values (+/- 20 percentiles from the median) of the wealth distribution allows us to screen consumption observations that were reasonably inconsistent with their corresponding wealth-index value.

Further, we control for whether the general characteristics[12] of the two sets were significantly different from those of other observations: Set A compared with all other consumption poor; set B compared with all other consumption non-poor (see Appendix); Set A compared to Set B; and finally Set A compared to the consumption non-poor. This cross-comparison substantially validates our hypothesis. The general characteristics of Set A are significantly superior (higher education levels, greater access to services, more valuable assets, etc.) to those of other consumption poor, whereas they closely resemble those of the consumption non-poor if we exclude Set B.

Set A observations also show a clear regional pattern (Table 1). In both survey rounds, about 70 percent of Set A households were located in the South (Gaza, Inhambane and the Maputo area); this concentration is not consistent with population distribution, as these regions account for only 25 percent of the national population (see Map 1 in Appendix). This suggests that, in the South, the interaction between location-specific PLs and predicted consumption might significantly affect the calculation of poverty figures.

Table 1: Set A and Set B Observations and Shares (IAF 2002/03-IOF 2008/09)

	2002/03				2008/09			
13 provinces	Set A		Set B		Set A		Set B	
	Obs.	%	Obs.	%	Obs.	%	Obs.	%
Niassa & Cabo Delgado-rural	5	1	231	24	2	0.1	128	13
Niassa & Cabo Delgado-urban	59	10	33	6	39	9	18	3
Nampula-rural	1	0.1	117	22	8	2	232	23
Nampula-urban	29	13	12	6	56	12	35	6
Sofala & Zambezia-rural	10	1	310	30	17	2	291	19
Sofala & Zambezia-urban	87	17	11	2	168	25	27	3
Manica & Tete-rural	19	2	228	25	21	2	233	22
Manica & Tete-urbana	134	20	25	4	95	21	22	4
Gaza & Inhambane-rural	142	14	89	9	126	19	64	7
Gaza & Inhambane-urban	171	31	9	2	151	27	12	2
Maputo Province-rural	84	29	3	1	53	43	1	1
Maputo Province-urban	261	48	0	0	332	57	1	0.1
Maputo City	381	41	0	0	313	35	0	0
Total	**1383**	**15**	**1068**	**12**	**1381**	**12**	**1064**	**9**

Set A: Consumption poor/ Wealth index non-poor; **Set B:** Consumption non-poor/ Wealth index non-poor;

Regarding Set B, i.e. the consumption non-poor/wealth poor set, around 80 percent of observations are concentrated in Central and Northern rural areas in both surveys (see Table 1). The general

[12] These are: the non-food share of household consumption, the head-of-household's total years of education, the highest level of education attained by any household member, the household's total asset ownership, access to a protected water source, electricity and sanitation, housing conditions and dietary variety.

characteristics of these households are significantly correlated with worse poverty indicators than those of other non-poor households, and they bear a strong resemblance to poor households not belonging to Set A (see Appendix). These results suggest that the consumption data systematically overestimate poverty in the South and in urban areas, with a corresponding underestimation of poverty in the Central and Northern regions and in rural areas. This pattern appears to be mostly due to price differentials, which are magnified by the fact that the share of consumption from household production is higher in the Central and Northern provinces and in rural areas than it is in the rest of the country.

For households in Set B, auto-consumption[13] prices for key items (maize [corn], rice, other cereals, cassava, groundnuts [peanuts], chestnuts, bananas and tomatoes) in the rural Northern and Central regions exceed both the purchase prices for these commodities and the auto-consumption prices defined for farmers in other areas;

In sum, both of these surveys present significant evidence of measurement error: in 2002/03 and 2008/09, respectively, 27 percent and 21 percent of observations record a poverty status that is inconsistent with all other household characteristics. Among these, households in Set A are likely affected by underreporting of consumption, while the results for Set B appear to be skewed by the overestimation of prices for household production. In Section 4 we re-estimate these observations by using sample estimates based on the 1996/97 survey.

2.3 Using out-of-sample estimates to re-estimate unreliable observations

The fact that the 1996/97 survey seems to be less affected by consumption underestimation and more consistent with other welfare indicators suggests that consumption-prediction models based upon it can produce a less biased fitted measure. Thus, for each survey (1996/97, 2002/03, and 2008/09) an identical consumption model has been estimated (Equation 1), the dependent variable for which is household per capita consumption $(\mathrm{y_i})$ deflated using the national consumer price index (CPI). Then, out-of-sample predictions are made for both 2002/03 and 2008/09 by replacing the original coefficients with coefficients obtained by applying the 1996/97 model parameters. Normally distributed errors are randomly assigned (Equation 2).

Equation 1

$$\frac{y_i^{97}}{CPI_{97-j}} = \alpha + \beta * X_i^{97} + \varepsilon_i \qquad for \;\; j = 2002/03, 2008/09$$

Equation 2

$$\widehat{y_i^j} = a + \hat{\beta} X_i^j + f(\hat{\epsilon}_i) \qquad for \;\; j = 2002/03, 2008/09;\; f(\hat{\epsilon}_i) \sim N(0,1)$$

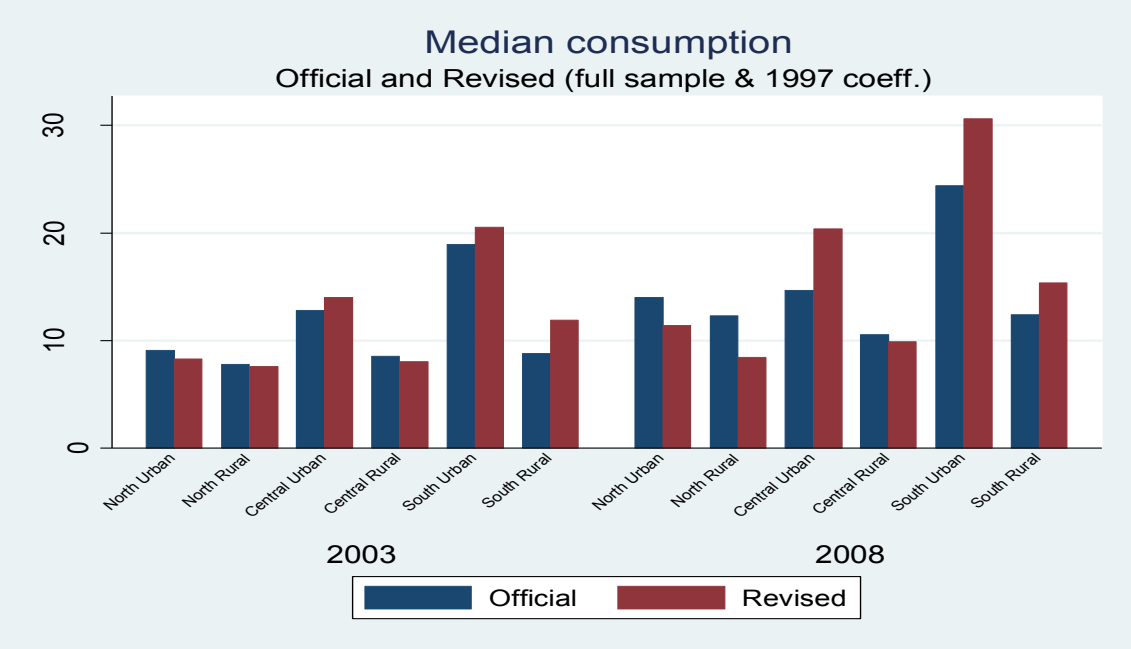
[13] "Auto-consumption" refers to the consumption of goods (in this case food) produced by the household. Auto-consumption prices should, in principle, reflect the opportunity cost of consuming household produce rather than selling it.

The results illustrate what poverty trends would have looked like had the underlying relationship between consumption and its correlates remained constant from 1996/97 to 2008/09. Coefficients' time invariance is arguably a significant assumption.[14] However, bearing in mind the evidence that the model based on the IAF 1996/97 is less affected by measurement error, the out-of-sample predictions for 2002/03 and 2008/09 produce revised consumption estimates that are demonstrably less biased than the original figures. Furthermore, using the 1996/97 parameters, variations in consumption figures over time are fully explained by variation in the independent variables (see Appendix).

Figure 3, below, shows median consumption per capita in 2002/03 and 2008/09 using both the official and revised consumption data. Whereas in 2002/03 the difference between the original and revised figures seems to be remarkable only in the South, the 2008/09 model estimate shows both higher median consumption in the Southern provinces and urban areas and much lower consumption in the Northern and Central provinces and rural areas. For example, in the revised estimates for the urban South median consumption is only slightly higher than in the official figures for 2002/03, but the gap widens dramatically in 2008/09. By contrast, in rural Nampula and Niassa-Cabo Delgado the revised estimates for median consumption are significantly lower than the official figures for both periods.

[14] An alternative could be to re-estimate the inaccurately measured observations using a so-called 'hybrid model' (see Appendix). The model is estimated by using a sample of arguably reliable observations, and the estimated β coefficients are then applied to the whole sample; the fitted consumption estimate is assigned only to the biased observations, and errors are estimated by the random assignment of normally-distributed errors from the equation. Both methodologies come with their drawbacks; estimates based on a reduced sample are not regionally representative (due to the regional pattern of dropped observations), while estimates based on 1996/97 coefficients assume no variance over time. As the results for the two methodologies are quite similar, we opted for the latter (the 1996/97-based model), judging that the lack of regional representation posed a much bigger problem than time-inconsistency.

Figure 3: Median Consumption per Capita in 2002/03 and 2008/09, Official and Revised Estimates



Official: 2008/09 PLs (MPD-DNEAP, 2010); **Simulated:** 1996/97 official PLs deflated by the consumer price index from World Development Indicators (2011)

The full-sample re-estimation for the 2002/03 and 2008/09 data confirms these findings. For 2008/09, and to a lesser extent in 2002/03, the official data are compromised by the underestimation of consumption in Southern provinces and urban areas and by the substantial overestimation of consumption in some Central and Northern provinces and in rural areas. In both cases the model indicates that general household characteristics would predict lower consumption per capita in the Central and Northern provinces and in rural areas and, keeping the official PLs unchanged, higher poverty rates in those regions.

Although it provides useful insights into the temporal dynamics of consumption and poverty, the full re-estimation of the two surveys is clearly an extreme measure, as not all consumption data are affected by measurement error. In Equation 3, using the out-of-sample estimates and combining them with the observation sets described in Section 2.2, the fitted consumption is assigned only to observations in Sets A and B for both 2002/03 and 2008/09.

Equation 3

$$\overline{y_i^j} = \widehat{y_i^j} \qquad for\ i \ \in \ A \cup B;\ for \ \ j = 2002/03, 2008/09;$$

A partial prediction using the 1996/97 estimates possesses a number of important advantages. First, the revised overall provincial poverty rates show a similar pattern to those obtained using other correction

techniques (see Footnote 9 and Appendix). Second, even when using bootstrapped error terms in Equation 2 the results remain statistically unchanged, confirming the robustness of results for the sample selected. Third, the model specification is relatively parsimonious, with consequently low noise in the fitted variable.

3. Poverty Lines and Measurement Error Interaction

3.1 The official poverty lines

A national poverty rate is typically estimated using a single poverty line and food basket, pricing that basket and adding to it the cost of basic non-food items. The PL represents a minimum threshold for consumption expenditures (or cost-of-basic-needs [CBN]) below which an individual (or household) is considered poor, and a single PL is applied to all individuals (or households) in the country.

The three NPAs use information on local prices obtained from reported household values as well as community markets, and attempt to account for dynamic local consumption patterns—using so-called "adjusted flexible bundles" (AFBs)—to create 13 location-specific PLs, all of which vary over time (Tarp et al., 2002; Arndt and Simler, 2010). Location-specific flexible PLs are designed to more accurately reflect local consumer behaviors and preferences under the assumption that consumer baskets, as well as commodity prices, are subject to significant temporal and geographic variation.

This methodology aims at capturing the substitution effect, a dimension that single-basket PLs cannot adequately address: households facing different prices can adjust their consumption baskets in order to maintain the same level of utility. This is valid both spatially and inter-temporally; in the AFB methodology, baskets used to calculate PLs can change over time to reflect changes in consumer preferences or adjustments in response to price variations. However, moving from a single CBN-basket to multiple ones raises the possibility that the resulting bundles will be utility-inconsistent, thus violating the fundamental premise that all regional PLs should represent the same level of welfare. To address this, Ravallion and Lokshin (2006) propose a scalar correction that could equalize baskets in terms of utility, but warn that "there is no guarantee that such a scalar adjustment exists".

In an effort to create a valid scalar adjustment Arndt and Simler (2010), after calculating different provincial CBN baskets, develop a methodology to correct their value in order to ensure that all PL bundles produce the same level of welfare. The correction is established by means of a constrained minimization, in which the objective is the distance between the adjusted and original budget shares and the constraints are represented by the revealed preference condition[15] and a minimum calorie requirement (around 2150 calories per capita per day). The only variable parameter is the quantity of each item in the bundles, while the price vector, obtained from the survey, is fixed.

[15] Revealed preference conditions rely on the principle of non-satiation: consumers always prefer consuming more rather than less. Applying this condition implies that the cost of the bundle for the location-specific PLs must be the same, taking relative prices into account, so that each bundle delivers the same utility level.

A comprehensive theoretical discussion of this methodology is beyond the scope of the present paper. In principle, the entropy correction methodology seems to provide a valid tool to produce utility-consistent bundles and might thus enable the establishment of consistent location-specific PLs. Our purpose, however, is different. We question the validity of using location-specific poverty lines—with or without entropy correction—in the presence of substantial measurement errors that conform to a clear geographical pattern.[16]

Any location-specific PL, being constructed based on a smaller sample than a national poverty line, has the potential to be more strongly influenced by measurement errors and to produce inconsistent, invalid, or counterintuitive results. There is a clear trade-off between the advantages of creating more locally-representative PLs and their relative robustness and consistency. There are also specific concerns regarding the entropy correction methodology, since to our knowledge the procedure doesn't include any correction mechanism for the measurement errors described above; if the information used to construct the location-specific bundles is itself flawed, the entropy-adjusted bundles are bound to be affected by the same inaccuracy. As discussed in detail in Section 3.3, constructing location-specific PLs without entropy correction produce results that are very similar to the official PLs, which include entropy correction.[17] It follows that the impact of entropy correction is minimal, but that the problems in the raw data used to create these PLs can significantly compromise the resulting poverty figures.

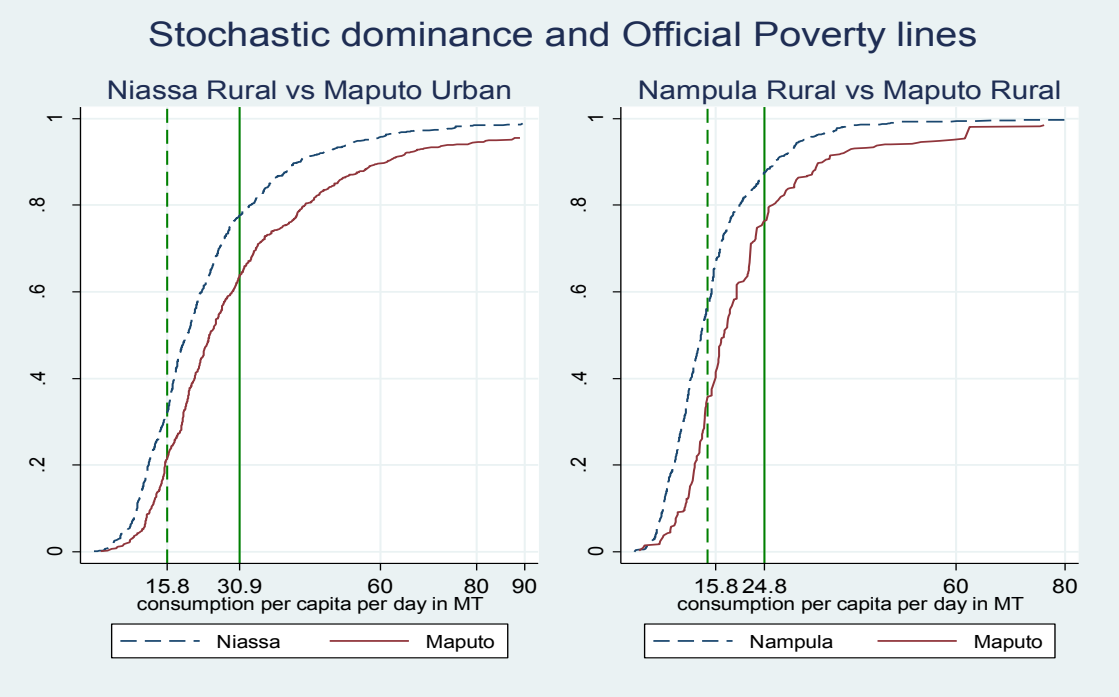
3.1 Why focus on poverty lines?

Some measurement error is always present in household survey data, and the IAF and IOF are no exception. Recalling in the analysis presented in Section 2, we identify 27 percent (2002/03) and 21 percent (2008/09) of observations (Sets A and B) as substantially inconsistent with other data and therefore likely to be inaccurate. This was a very conservative conclusion: early attempts to circumscribe the problem produced even broader sets of questionable observations. However, in all these cases the pure re-estimation of these observations didn't produce the expected results, in particular regarding the regional distribution of poverty. After our re-estimation the Southern provinces still appeared poorer than the Central and Northern provinces, though all other indicators would suggest the opposite. This counterintuitive spatial distribution of poverty figures prompted us to examine the PLs themselves and their high degree of spatial variation.

[16] Here we implicitly assume that measurement error in each location occurs in only one direction, either negative or positive. Section 2 confirms that this is the case for Mozambique. In the South and in urban areas we observe considerable underestimation of food consumption, while in some areas of the rural North we observe a less severe, but still significant, incidence of consumption overestimation.

[17] The re-construction of official PLs with entropy correction is a sophisticated iterative procedure that we were not able to fully reproduce. However, when constructing the 13 provincial-level bundles, we followed the same procedure used in the poverty assessment, with the same calorie intake value and procedure to identify most-consumed food items, as well as the same outliers' treatment for prices and quantities consumed. As the results show, the variation from the official figures is minimal, and the geographical pattern is fully reproduced.

Figure 4: Stochastic Dominance and official poverty lines



High PL differentials significantly impact the calculation of poverty figures, and two specific examples corroborate this conclusion. The left-hand graph in Figure 34 above, compares the per capita consumption distribution of urban Maputo Province and rural Niassa and Cabo Delgado[18] (N-CD) Provinces, while the right-hand graph in Figure 3 compares rural Maputo to rural Nampula.[19] Per capita consumption in both urban Maputo and rural Maputo is stochastically dominant throughout the distribution, even though in these areas underestimation affects about 50 percent of the observations. Yet the official poverty figures imply the opposite: while urban Maputo registers a poverty rate of 63.7 percent, rural N-CD only registers a rate of 32.7 percent; rural Maputo, meanwhile, has an official poverty rate of 76.3 percent, while rural Nampula's rate is 56.7 percent (see Table 2, column 2, below).

In both cases the cause of the difference is the differing PLs (the vertical lines in the two graphs). The PL for urban Maputo is 93 percent higher than that of rural N-CD, while the PL for rural Maputo exceeds the PL for Nampula by 73 percent (see Table 4, column 6, in Appendix). As shown in Figure 3 the magnitude of the PL differentials completely offsets the dominance of per capita consumption in Maputo, both for urban and rural areas. The key question, therefore, is whether these sizeable

[18] The location-specific PLs used in the official poverty statistics group homogeneous urban and rural areas into 13 regions, sometimes combining areas from different provinces, as is the case with Niassa and Cabo Delgado Provinces (see Table 1).
[19] In both cases, we truncated the upper side of the distribution in order to close up on the lower percentiles: this way we simply disregard the upper 1st percentile. What concerns us is the rest of the distribution, in particular the part where the poor are identified.

differentials between PLs accurately reflect regional price differences for goods in the CBN basket. Based on our assessment this does not appear to be the case.

We began by analyzing price differences between urban Maputo and rural N-CD using a Paasche index[20] constructed based on survey prices, which indicates a differential of 18 percent; a Paashe index based on SIMA[21] prices (which does not differentiate between rural and urban Maputo) produces a 30 percent differential. Similar calculations performed elsewhere (Sohnesen, 2011) put the price-index ratio between 5 and 20 percent.22 As a further robustness check, in computing the Paasche indexes we used only those observations that were 20 percent below or above the PL, in order to see whether the differential increased the closer we moved to the PL. The Paasche index based on both food and non-food items from the household survey registers a differential of 30 percent provinces; based only on food the differential is just 5 percent, and the Paasche index based on SIMA prices produces a differential of 16 percent. Similar results are obtained when only observations below the 55th percentile[23] are used. In all cases, the observed price differentials are dwarfed by the vast differences in regional poverty lines.

The rural areas of Maputo and Nampula are more directly comparable, yet here as well the considerable PL spread (73 percent) is not justified by any similar difference in prices. Sohnesen's simulation indicates a price differential of between 31 and 40 percent, while our deflator (whether based on survey prices or SIMA prices) indicates little or no difference between the two areas. As in the previous case, we compute the Paasche index 'for the poor', and here the differences are minimal: 13 percent for the food and non-food index, 20 percent when using only food items, and a negative differential of 2 percent when using SIMA prices.

To summarize, these findings suggest that the spatial distribution of prices ostensibly reflected in the official PLs is highly questionable, and does not appear to represent the actual prices faced by the poor. Since poverty headcounts are calculated using these PL's, our results suggest that the spatial distribution of poverty recorded in the official figures is a consequence of methodological weaknesses in the construction of the PLs themselves and does not accurately describe the prevalence or disposition of poverty in Mozambique.

[20] The Paasche index has been computed for 1060 enumeration areas and 49 food and non-food items using the national price for every item as a base price.

[21] SIMA is the Portuguese acronym for the Agricultural Marketing System of Mozambique. SIMA is a database of agricultural market prices, updated monthly, with reasonably comprehensive coverage.

[22] A recent policy note by Sohnesen (2011), focuses on Mozambican spatial price distribution. The note examines price distribution at the local level by calculating product prices and spatial Paasche indices under different scenarios. The basic rationale is that all the scenarios proposed to compute unit prices should not impact overall prices considerably. The fact that some scenarios show a high impact on prices leads the author to argue that price data are not stable enough to support the relatively high number of PLs used by MPD-DNEAP (2010).

[23] All these results are available upon request.

3.2 Sensitivity of PLs to measurement error

As noted in previous sections, measurement error is highly concentrated in certain parts of the country. In this section we use the 2008/09 data[24] to show how the construction of location-specific PLs is likely to be affected by measurement errors and how this might produce, in the areas where measurement error is more concentrated, PLs that do not reflect the baskets consumed by the poor (see also Maia and van den Berg, 2010). Southern Mozambique and urban areas nationwide appear to be particularly affected by observations in which the food consumption component is excessively low (due to the systematic underestimation of food consumption in urban areas described above) and whose non-food consumption share is consequently overestimated.[25] These observations, we posit, have a significant impact on both the food and non-food components of the location-specific PLs.

The large share of non-food consumption in these observations makes the non-food component of the PLs inordinately influential (see Appendix). In the Maputo area, for instance, the non-food component accounts for about 1/3 of the total PL, causing the Maputo poverty line to be roughly twice as high as all other PLs in the country. This differential is not justified by corresponding differences in any spatial price index (see Appendix).

Regarding the food component, the direct impact of underestimated consumption (beyond the consequent overestimation of non-food shares) is less clear. Broadly speaking, relying on these observations to construct 'local baskets' may produce counterintuitive results. Two types of measurement errors typically occur.

The first type of error results from the fact that many food transactions are not reported in the survey; this causes the aggregate level of food expenditures to be lower than expected. The consequences of this error are clear in terms of food/non-food shares but less so in terms of their impact of the food component of the food poverty line. This is because calculating the food PL relies on the pattern of consumption, not the level.[26] Baskets relying on underreported consumption may imply a very different dietary pattern than would more precise consumption data. It is uncertain, a priori, whether the resulting food PLs will be biased upward or downward: this depends on which types of items are underreported. If relatively inexpensive items are underreported, the resulting basket will be biased upward (as will the food PL); if more expensive foods are underreported, the opposite will be true.

[24] As an additional check, we reproduced the test and procedures for 2002/2003, obtaining similar results. These findings are available upon request.

[25] Food consumption is calculated by aggregating food expenditures. Non-food consumption expenditures, instead, are constructed by using information on education, health and some utilities fees/expenditures by calculating an implicit value, such as for housing or the use value of durable goods (see Deaton and Zaidi, 2002). In the specific situation discussed, we argue, the level of non-food consumption of urban households was correctly estimated. The problem arises when we merge food with non-food consumption and we look at their shares. Being food consumption dramatically under-estimated, non-food consumption share is relatively overestimated; hence it looks like these households spend a disproportionate amount in non-food items -a typical consumption behavior of better-off households- while at the same time they cannot meet a minimal dietary requirement. Indeed, a clear contradiction.

[26] This is because PLs are rescaled to meet the minimum caloric requirement (2150 calories per capita per day).

The second type of error involves discrepancies between recorded food expenditures and actual food quantities consumed. A household may correctly report its expenditures on an item, but may not be able to accurately report the specific quantity of that item, particularly given the indefinite measures (heaps, sacks, etc.) used by the survey. Such an error would clearly affect the unitary price of the item.

In order to compare the relative error sensitivity of PLs for urban and rural areas, we constructed a single national-basket PL, as well as national-level urban and rural PLs (not reported), the latter deflated with a location-based Paasche index, and then compared these with the location-specific PLs used to determine the official poverty statistics. It is important to recall that "location" in this context means each of the 13 regions listed in Table 1, a division of the country that groups homogenous urban and rural areas, sometimes from different provinces, into the same region. Our hypothesis is that location-based PLs are more strongly affected by measurement errors in consumption data than are the overall national and national-level urban and rural PLs. Sensitivity tests on consumption data classified as unreliable (Set A and Set B observations, described above) have been constructed. Two tests are reported: (i) a test of the non-food component of the PLs; and (ii) a test of the final (food and non-food) PLs.

Table 2 illustrates our first sensitivity test of the non-food component of the PLs, constructed exactly like those in the three NPAs.[27] Column 4 of Table 2 shows, for each of the 13 regions, the non-food share of observations used to calculate the non-food PLs. The amount that these households spend on non-food items becomes the non-food poverty line; their share of non-food consumption is the value reported. As noted above, the non-food PLs rely heavily on observations in which the share of non-food consumption is very large due to the underestimation of food consumption.[28] To what extent these observations condition the construction of non-food PLs is shown below.

Table 2: Sensitivity of the Official Non-Food Component of Poverty Lines to Set A Observations

13 regions	**Official**			**Simulation eliminating underestimated observations (Set A)**		
	Non-Food Comp of PL	**Poverty Line**	**Non-food share**	**Non-Food Comp of PL**	**Poverty Line**	**Non-food share**
Niassa & Cabo Delgado-rural	3.43	15.95	0.30	3.75	16.25	0.30
Niassa & Cabo Delgado-urban	5.01	18.91	0.39	5.17	19.17	0.37
Nampula-rural	3.21	14.33	0.33	3.63	14.73	0.33
Nampula-urban	4.22	16.72	0.39	4.06	16.56	0.33

[27] We took a simple weighted average of non-food expenditures for households with expenditures between 80-120% of the food-poverty line. A triangular weighting scheme is then used: weights are higher the closer a household's expenditure is to the poverty line.

[28] As discussed in Section 2, the non-food share of these observations is significantly higher than that of all other observations.

Sofala & Zambezia-rural	3.30	14.35	0.29	3.12	14.52	0.27
Sofala & Zambezia-urban	5.39	19.07	0.42	4.15	17.85	0.31
Manica & Tete-rural	4.21	19.39	0.29	4.34	19.54	0.28
Manica & Tete-urban	5.93	21.47	0.40	4.97	20.57	0.32
Gaza & Inhambane-rural	5.31	18.37	0.43	5.07	18.17	0.39
Gaza & Inhambane-urban	6.32	20.31	0.47	5.78	19.88	0.41
Maputo Province-rural	7.45	24.84	0.43	6.62	24.52	0.39
Maputo Province-urban	10.25	30.86	0.50	6.25	26.95	0.31
Maputo City	12.34	33.14	0.59	7.05	27.95	0.35

Official: 2008/09 PLs and shares (MPD-DNEAP, 2010)**; Set A:** Consumption poor/ Wealth index non-poor households.

Columns 5, 6 and 7 of Table 2 report the results obtained by eliminating the underestimated observations identified above (Set A) from the calculation of the non-food PL. After eliminating observations that were singled out as suspicious we ran the estimation procedure, and once the final results were obtained we compared these with the baseline scenario, including the results obtained using the full sample.[29] As expected, both the revised non-food shares and non-food PLs were much lower than the baseline in Southern Mozambique and urban areas, which suggests that the construction of the non-food PLs was seriously affected by underestimated observations.

The direction of the bias is quite straightforward. The elimination of observations characterized by excessively high non-food shares produces lower non-food PLs and, consequently, total PLs tend to be lower. This effect becomes even clearer when comparing columns 2 and 3 of Table 2 with columns 5 and 6. Due to our simple modification of the non-food component, the total PLs decrease 5 percent and 12 percent, respectively, in the greater Maputo urban area and in Maputo City.[30] We then turn to the question of whether our alternative PLs minimize this effect. Our simulations indicate that non-food PLs calculated at the national or nationwide urban/rural levels are relatively insensitive to underestimated observations. The number of observations used to calculate the poverty line is crucial to its robustness: the more specific the PL, the higher the risk of magnifying measurement errors.

In Table 3 we extend the sensitivity test to the full PL methodology and compare the results with a national-basket alternative. Table 3 describes the revised PLs and their variation from the official figures. As noted above, whereas for the non-food PLs the impact of underestimation is easily detectable, for the food PLs the impact can be ambiguous: the bias can run in either or both directions depending on what type of information is missing.

The findings in Table 3 support our hypothesis. Location-specific PLs tend to be more strongly affected by the presence of underestimated observations in the data. The elimination of these observations

[29] At this stage of the analysis we limit our simulation to a rather simple and mechanical procedure. We reiterate that a generalization of the results is beyond the scope of this analysis.

[30] Similar results are reported in a recent paper on Mali (Delarue *et al.*, 2010). The use of provincial level PLs, also calculated based on a limited number of observations, leads to very high PLs in areas of the country traditionally regarded as relatively better-off, such as the Sikasso Province. It follows that official poverty rates tend to be higher in Sikasso than in less developed areas of the country.

produces especially large variations in the PLs in the South, with Maputo City and the other urban areas of Maputo Province recording the highest differentials, and urban areas nationwide exhibiting far greater variation than their rural counterparts. With the notable exception of urban Gaza and Inhambane, where the baseline values differ from the official more than other PLs, the direction of change in the areas most affected by underestimation (i.e. the Northern and Central regions) is towards lower PLs and thus towards lower poverty figures.

Table 3: Sensitivity of Official and Revised Poverty Lines to Underestimated Observations (Set A)

13 regions	Set A		National basket P.L			Location sp. basket P. L		
	Obs.	Share	Full sample	Without set A	**Diff.**	Full sample	Without Set A	**Diff.**
Niassa & C. Delgado-rural	2	0.00	18.79	18.11	**0.68**	18.13	18.14	**-0.01**
Niassa & C. Delgado-urban	39	0.09	20.61	19.87	**0.74**	22.89	22.16	**0.73**
Nampula-rural	8	0.02	14.57	14.05	**0.52**	10.36	10.30	**0.06**
Nampula-urban	56	0.12	16.55	15.96	**0.59**	13.16	12.50	**0.66**
Sofala & Zambezia-rural	17	0.02	15.81	15.25	**0.56**	13.36	13.21	**0.15**
Sofala & Zambezia-urban	168	0.25	17.62	16.99	**0.63**	19.13	17.68	**1.45**
Manica & Tete-rural	21	0.02	20.87	20.12	**0.75**	19.11	19.05	**0.06**
Manica & Tete-urban	95	0.21	20.19	19.46	**0.73**	21.40	20.00	**1.40**
Gaza & Inhambane-rural	126	0.19	15.99	15.42	**0.57**	19.18	18.77	**0.41**
Gaza & Inhambane-urban	151	0.27	17.04	16.43	**0.61**	15.87	16.52	**-0.65**
Maputo Province-rural	53	0.43	18.74	18.07	**0.67**	28.37	27.75	**0.62**
Maputo Province-urban	332	0.57	22.59	21.78	**0.81**	32.07	28.02	**4.05**
Maputo City	313	0.35	22.31	21.51	**0.80**	33.78	28.79	**4.99**
National	**1381**	**0.12**	**17.80**	**17.20**	**0.60**	**19.20**	**18.40**	**0.80**

National: single national PL and 13 provincial price deflators; **Location specific**: 13 PLs based on local food basket, without entropy correction.

In conclusion, underestimated observations tend to bias the location-specific PLs through both their food and non-food components. By contrast, a single national PL seems to be far less affected by consumption underestimation. The sensitivity of PLs to measurement error seems to be directly correlated to the number of excluded observation sets: the larger the sample of consumption observations, the lower the effect of underreporting. In results not reported even the national-level urban and rural PLs were more affected by underestimation than the single, national PL.

4. Poverty trends: Official and Revised

4.1 Revised poverty-line construction

In a recent assessment of possible alternatives to the official poverty figures, van den Boom (2010) compares the official figures with figures obtained by averaging the region-specific poverty lines and creating a unique aggregate national poverty line. The simulation leads to striking changes in poverty patterns in Mozambique. Under the aggregate PL, poverty appears significantly higher among the rural population (with a 65 percent poverty rate for 2008/09 compared to the official 57 percent), while the urban population appears considerably less poor (39 percent for 2008/09 compared to the official rate of just under 50 percent). These poverty dynamics would seem to be more consistent with other data sources, including regional economic growth rates and anthropometric indicators, both at the regional and national levels. In addition, van den Boom's methodology produces dramatic improvements in the stability of poverty trends over time, as well as the near-complete elimination of the puzzling poverty-rate swings described above.

1. Using the information contained in the 1996/97 survey we construct a single-basket poverty line:

Equation 4

$$\underline{q}_{97} = \sum_i w_{97i}\, q_{97i} \qquad for\ i\ \in P = \{y_i : y_i \leq \bar{y}\}$$

where $\underline{q}_{97}$ is a bundle typically consumed by the poorest segment of the population in 1996/97 (see Appendix for a list of items and food shares). This bundle is developed by taking the population weighted average $(\sum_i w_{97i})$ of bundle q_{97i} consumed by household below a certain consumption threshold $\bar{y}$ (in our case the 60th percentile) and scaled up to the standard threshold of 2150 calories per capita per day. The bundle $\underline{q}_{97}$ is then multiplied by a vector of prices $\underline{p}_t$

Equation 5

$$\underline{p}_t = \sum_i w_{ti}\, p_{ti} \qquad for\ i\ \in P = \{y_i : y_i \leq \bar{y}\}\ ;\ for\ t = 1996/97, 2002/03, 2008/09$$

Combining Equation 4 and Equation 5 and adding $\partial \in (0,1)$ to account for the non-food share and p_{ts}, a spatial Paasche index for the 13 regions, we obtain

Equation 6

$$z_{ts} = p_{ts}(1 + \partial_t)\, \underline{p}_t \underline{q}_{97} \quad for\ t = 1996/97, 2002/03, 2008/09; for\ s = 1, \ldots 13$$

Equation 6 produces a national, fixed-basket, spatially-deflated poverty line z_{ts} for each of the three survey periods. The results are presented in Table 4 (see Appendix), along with the official figures.

4.2 Comparison between official and revised PLs

A comparison of the official and revised PLs confirms a number of hypotheses discussed in previous sections. The revised PLs are more stable over time, reducing the frequent reshuffling of regional poverty rankings in the official data. The coefficient of variation (CV)[31] calculated for the revised PLs ranges from 0.19 in 1996/97 to 0.20 in 2002/03 to 0.14 in 2008/09, whereas the CVs for the official PLs are 0.29, 0.46, and 0.28, respectively. When looking at the ratio of the highest PL (Maputo City in all cases) over the lowest PL, the results are more stable for the revised figures (around 1.6). Conversely, the ratios for the official PLs are 2.54 for 1996/97, a remarkable 3.57 for 2002/03, and 2.31 for 2008/09.

In addition, a fixed national basket produces a substantial reduction in the PLs for Southern Mozambique and urban areas, both in 2002/03 and in 2008/09, the survey periods in which we identified the greatest impact of measurement error. In the 1996/97 survey, which is less affected by measurement error, the gap is smaller. Interestingly, the gap between the revised and official PLs widens between 1996/97 and 2002/03 and then either stabilizes at the 2002/03 level or slightly decreases in 2008/09. This appears to provide additional evidence of the fundamental problem in the more recent surveys: the gap widens when the error is introduced (2002/03) and stabilizes thereafter, as the error distribution does not change much between 2002/03 and 2008/09 (see Table 1).

One possible objection to these findings is that differences between 1996/97 and the two later surveys are not necessarily related to measurement error and might be interpreted as merely a consequence of fixing the basket in 1996/97, with the gap between the revised and official PLs widening because of changes in the baskets themselves, i.e. substitution.[32] However, this argument is not entirely convincing: when constructing national-basket PLs with baskets that change every year (see Section 3.2 and Appendix) the resulting values are still closer to our revised PLs, suggesting that the inter-temporal effect of basket substitution is minimal. The gap is therefore either capturing some spatial basket-substitution effect or, as we argue, is a product of the interaction between location-specific PLs and systematic measurement errors in the data.

Finally, the comparison reveals another interesting phenomenon: the growth rate of the poverty lines. This provides, for each of the 13 regions, a measure of the change in the cost of living exactly at the poverty line. By design, this particular measure of inflation is most relevant for the poor, as the contents of the food bundle reflect items that are consumed by the poor and regarded as necessary for basic survival. Although the PLs themselves appear to be significantly impacted by measurement error, their rate of change over time is consistent with the available price data. While the rates of PL growth

[31] $CV = \sigma/\mu$, where σ = the standard deviation and μ = the mean. Lower CV values mean less variability.

[32] Revised PLs are constructed by fixing the basket at national level in 1997, therefore the whole inter-temporal variation is explained by price changes only.

between 1996/97 and 2002/03 are different in different regions, leading to an increased spatial disparity in 2002/03, both the revised and official PLs for 2002/03 and 2008/09 show a similar geographical pattern of growth (see Figure 4, in Appendix) and a high and significant rank correlation (0.74).

As is extensively discussed in the most recent NPA, the pattern of price increases observed between 2002/2003 and 2008/09 helps to explain regional poverty dynamics over the period (DNEAP, 2010). Notably, sharp price increases combined with stagnant agricultural productivity contributed to the bad poverty outcomes of the Central provinces. Likewise, according to the NPA the relatively slower growth of PLs in urban areas nationwide triggered a rapid reduction in observed poverty; this effect was consistent throughout the country (see Figure 4 and the last column of Table 7, both in Appendix).

To further analyze the robustness of the revised PLs vis-à-vis the official figures we constructed a correlation matrix showing poverty rates and consumption proxies for 2002/03 and 2008/09 across the 13 regions (see Table 5 in Appendix). Using the revised PLs, the 13 regional poverty-rate rankings appear more consistent with what other poverty correlates would suggest. These correlations are both negative and statistically significant for proxies associated with lower poverty (e.g. a higher wealth index, better access to services, a higher head-of-household education level), and both positive and statistically significant for proxies associated with poverty (e.g. a higher percentage of stunted and underweight children). By contrast, the official poverty rates show a non-significant rank correlation with all other poverty indicators, raising further doubts about their validity.

In addition, a standard consumption-prediction model is presented in
Table 6 (see Appendix) and can be expressed as:

Equation 7

$$\log\left(\frac{y_i}{z_{jk}}\right) = \beta_{1k}\mathrm{HEAD}_i + \beta_{2k}\mathrm{EDU}_i + \beta_{3k}\mathrm{SHOCKS}_i + \beta_{4k}\mathrm{ASSET}_i + \beta_{5k}\mathrm{AGRIC}_i + \mathrm{MONTH} + \varepsilon_{ik}$$
$$\text{for } k = 1, 2$$

where yi refers to total official household consumption expenditures per capita per day, with zk representing the two different groups of PLs, and subscripts i and j referring to households and locations, respectively.

Since the scale of the dependent variable differs slightly in the two models due to the different PLs used (see Table 3), the simple difference in the t-statistics for each parameter is reported instead of the statistical difference in the coefficient usually reported in the seemingly unrelated regression equation (SURE) models. A higher absolute difference for each parameter can be interpreted as a better capacity of the ratio between consumption expenditures and PLs to be explained by conventional poverty correlates. The results indicate that the model using revised PLs (B) better fits the standard poverty predictions. The adjusted R-square of this model is 22 percent higher than when the official PLs are used. Moreover, the difference in t statistics for the variables associated with poverty (e.g. assets and access to services) are all positive and relatively high, with the only exception being the variable

associated with household ownership of a bicycle, as this is actually an indicator of relative poverty (Hanlon and Smart, 2008).

4.3 Revised poverty outcomes: 1996/97-2002/03-2008/09

Combining the results presented in Sections 3 and 4, we calculate revised poverty figures using the revised PLs and extrapolating from the 1996/97 model (see Section 2.3). Table 7 (see Appendix) compares the revised poverty figures with the official statistics. In Figure 5, below, we group poverty figures by consolidated urban and rural areas in the Northern, Central and Southern macro-regions.

This portion of the analysis yields three important conclusions. First, at the national level the revised figures show a different trend than that indicated by the official statistics. Rather than recording a rapid reduction in poverty during the first period (1996/97-2002/03) and stagnation in the second (2002/03-2008/09), essentially an L-shaped trend, we observe a change in the gradient of poverty reduction: from 2002/03 to 2008/09 poverty reduction slowed but did not stop (see also van den Boom, 2010). As shown by the confidence intervals in the last row of Table 7, the revised values for poverty at the national level in 2002/03 and 2008/09 are statistically different and indicate a declining trend during the period.

Second, the official and revised poverty levels across regions are similar for 1996/97 (with the South being significantly richer than the Central and Northern regions) but differ substantially in the later surveys. According to the official figures, in 2002/03 poverty was heavily concentrated in the South: rural Maputo is ranked as the poorest rural area, and urban Maputo is the second-poorest urban area. The rest of the country, with the exception of the rural North, shows poverty rates of below 50 percent. In 2008/09, the official poverty figures prompt a sharp re-ranking of regions, and the Central provinces, together with the Maputo area, become the poorest in the country.

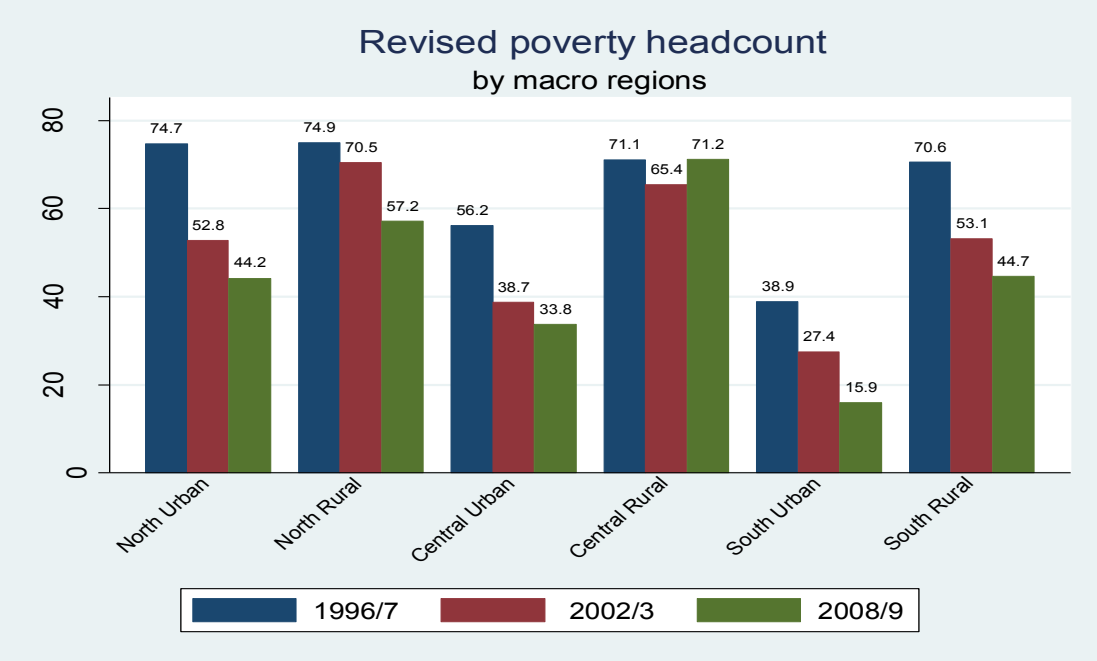
Our revised figures, by contrast, show that in 2002/03 the Southern provinces, both urban and rural, seem to be better-off than the Central and Northern regions, with urban areas consistently recording lower poverty rates than rural areas. This pattern is confirmed in the revised figures for 2008/09. Poverty is lowest in urban areas and in the South, and highest in rural Central and Northern Mozambique.

Third, the official and revised poverty rate trends also diverge quite strikingly across regions. According to the official figures the poverty ratio over the entire period (1996/96-2008/09) drops substantially only in Niassa and Cabo Delgado, both urban and rural, while there is a moderate reduction in the rural and urban South. Poverty rates stagnate in rural Nampula, while increasing in the Central region, both in urban and rural areas, and in urban Nampula.

Our revised figures indicate a very different pattern (see Figure , below). Already evincing a lower level of poverty at the outset of the period, urban areas in the South show a considerable reduction in

poverty rates, which fall to just 15 percent in 2008/09. There is also some apparent spillover from urban to rural areas, with poverty in the rural South falling from 70 percent in 1996/97 to about 45 percent in 2008/2009. As a result the rural South becomes the least poor rural area in the country and one of the regions in which poverty rates decline fastest. In the rest of the country substantial declines in urban poverty are not accompanied by corresponding improvements in adjacent rural areas. A critical feature that distinguishes Southern Mozambique, and the Maputo area in particular, from the rest of the country is the simultaneous decline of both urban and rural poverty rates. By contrast, urban areas in the Central and Northern regions remain islands of relative prosperity surrounded by rural areas mired in severe and persistent poverty. In conclusion, with the notable exception of the urban Central region, overall poverty reduction seems to be driven by consistent reductions in urban poverty nationwide and by the strong performance of the Southern region as a whole.

Figure 5: Revised Poverty Headcount, by Consolidated Regions



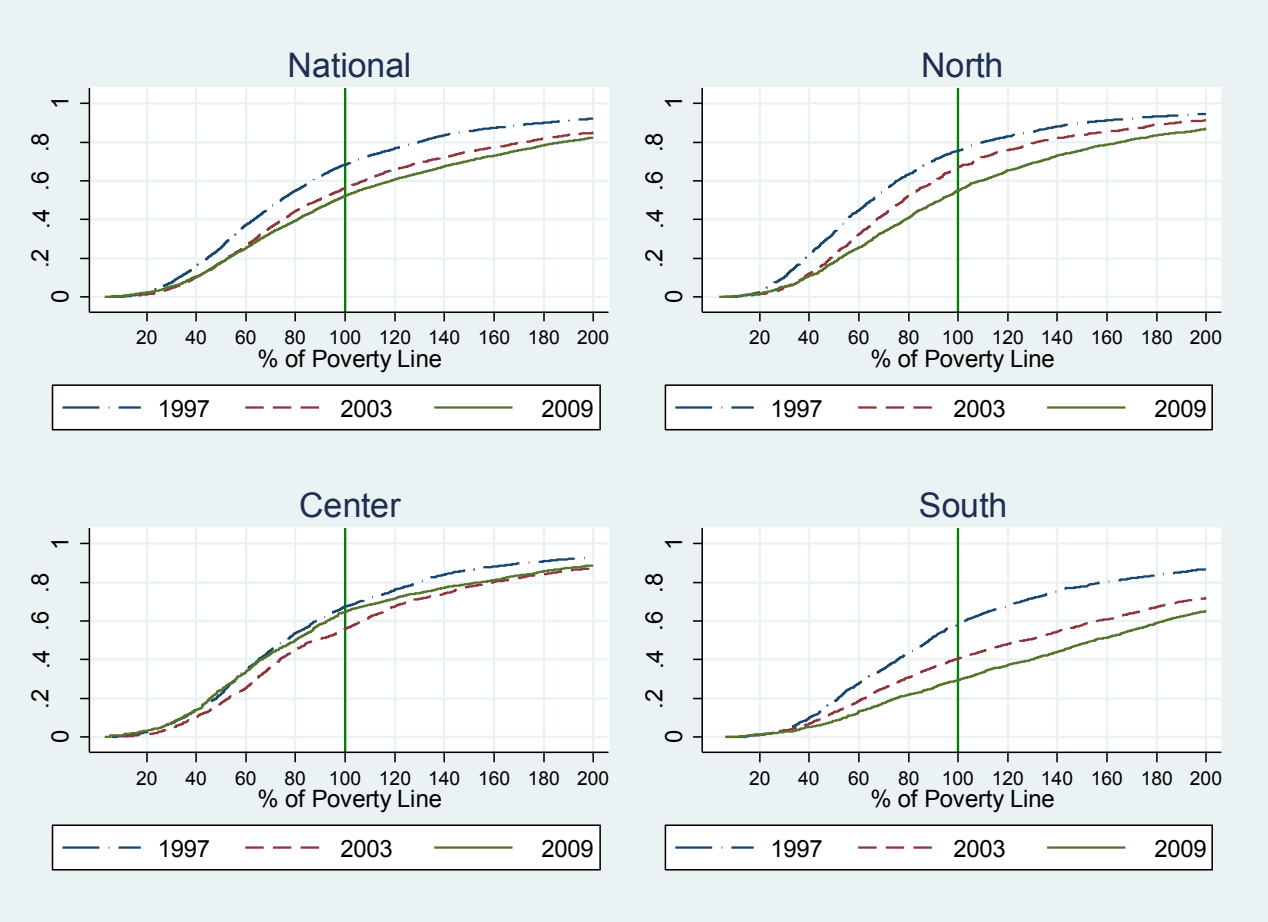
Revised: national 1996/97 based basket PL with 13 location price deflators and 2002/03 and 2008/09 data partially re-estimated (set A and B table1)

An analysis of the distribution of real consumption (see Figure 6, below) obviates the potential disadvantages of considering only one measure of welfare, the poverty line, while overlooking the rest of the consumption distribution. The horizontal axis represents revised consumption measured as a percentage of the poverty line. The vertical axis represents the population percentage. Each point on the distribution function shows the share of the population below a specified consumption level.[33] The

[33] Given our focus on the performance of the poorest segments of the population we cut off the distribution at 200% of the poverty line.

poverty levels in each of the three survey years can be read from the distribution functions at the point where the function crosses the vertical line that indicates 100 percent of the poverty line (DENEAP, 2010). The distribution can also be compared in welfare terms: any distribution appearing to the right of another is statistically dominant and can be considered a welfare improvement.

Figure 6: Distribution Function for Revised Real Consumption



Revised: national 1996/97 based basket PL with 13 location price deflators and 2002/03 and 2008/09 data partially re-estimated (Sets A and B, Table1)

These conclusions regarding poverty trends hold true at both the national and consolidated macro-region levels. The 2002/03 survey recorded a dramatic reduction in poverty and a general improvement in overall welfare. However, income growth was fastest at higher income levels, and inequality consequently increased. It is worth noting that the unequal growth of consumption has a clear spatial correlation: welfare improvements in the South between 1996/97 and 2002/03 (Figure , bottom right graph) were twice as high as those in the Central and Northern regions.

From 2002/03 to 2008/09 the rate of poverty reduction decreased and became less even, but poverty reduction itself did not stop. The consumption distribution in 2008/2009 is superior to 2003/04, but only

for those with expenditures above 60 percent of the PL; this confirms that the poorest of the poor did not benefit from growth during the period. Many of those in this category are likely located in Central Mozambique (see Figure , bottom left graph) where for most households consumption reverted to its 1996/1997 levels. Finally, the positive performance of the South is less pronounced than in the previous period, but the poverty rate nevertheless fell to about 30 percent, and clear improvements can be observed across the distribution.

Conclusions

Over the past decade there has been intense debate in Mozambique surrounding not only poverty rates and trends, but also data quality and analytical rigor. The aim of our analysis is to advance the discourse on poverty measurement by providing alternative methods to compute poverty rates. Our findings indicate that two interrelated factors have created significant weaknesses in the official reporting of poverty statistics, with the greatest degree of inaccuracy occurring between 2002/03 and 2008/09. We are presented with compelling evidence of the widespread underreporting of consumption expenditures in the survey data, and our assessment reveals serious methodological issues regarding the formulation and use of location-specific poverty lines based on those data. To address these issues we have proposed specific methods to reduce the occurrence and minimize the impact of underreporting and to construct less specific, yet more accurate, PLs based on a single national consumption basket. The national basket is developed using data from the first (and apparently most comprehensive) available survey (1996/97) combined with updated price data from each subsequent survey round. The revised figures produced by our methodology differ substantially from the official poverty lines, which are based on province-specific and time-flexible bundles with entropy correction (MPD-DNEAP, 2010; MPF/IFPRI/PU, 2004).

By examining the location-specific PLs we determined the likely cause of the excessively high regional differentials: observations affected by consumption underreporting, which are characterized by an unusually large share of non-food consumption, tend to skew local PLs upward in those areas where underreporting is most common (i.e. urban areas in general, and the South of the country in particular). As an alternative to the evidently inaccurate official PLs we constructed a single national PL and two addition PLs reflecting urban and rural divisions at the national level. Not only are these measures more consistent with standard poverty correlates, both at the regional and household levels, but the overall national and nationwide urban/rural PLs seem to be less affected by measurement error, as they rely on a larger sample in calculating their consumption baskets.

The results of our revised methodology demonstrate that the substitution of province-specific PLs with a national PL based on a 1996/97 consumption basket and combined with a partial re-estimation of observations (about 1/4 in the two years) produces poverty estimates for 2002/03 and 2008/09 that are significantly more consistent with other socioeconomic indicators and considerably more stable over time than the official statistics. In our revised estimates, poverty is substantially lower in Southern Mozambique and in urban areas overall than the official figures would indicate, and dramatically higher

in Central and Northern Mozambique and in rural areas nationwide. Compared to the official figures, this represents a complete re-ranking of provinces based on poverty incidence. In addition, our calculations do not support the ostensibly enormous spatial price variations that are the basis for the large differentials in location-specific PLs and consequently call into question the observed changes in poverty headcount rates between 2002/03 and 2008/09.

In terms of poverty dynamics, our revised estimates indicate very different trends in poverty reduction than those recorded in the official statistics. Rather than suffering economic stagnation, Mozambique is continuing to experience a general decrease in poverty, though this has been heavily concentrated in urban areas where only 30 percent of population resides; meanwhile, rural poverty has remained largely intractable, with the notable exception of rural areas in the greater Maputo region. In sum, our analysis strongly suggests that not only are the Southern and urban areas of Mozambique considerably richer than the official figures would indicate, but growth in these areas has been relatively robust and consistent over time. Meanwhile, poverty in the Central and Northern regions of the country, and in rural areas nationwide, appears to be systematically underestimated in the official statistics. Ostensibly high rates of poverty reduction in these areas, as recorded in the most recent national surveys, are likely to be the result of changes to the survey itself and its analytical methodology, and do not necessarily reflect real improvements in the economic conditions of the nation's poor. If the evident weaknesses in the official statistics are not adequately addressed they may ultimately become the foundation for ill-targeted or even potentially ineffective economic growth and poverty-reduction policies.

References

Arndt C., K. Simler (2010). Estimating Utility-Consistent Poverty Lines with Applications to Egypt and Mozambique, *Economic Development and Cultural Change*, 58(3): 449-474.

Arndt C., A. Garcia, F. Tarp and J. Thurlow (2010). Poverty reduction and economic structure: Comparative path analysis for Mozambique and Vietnam, *UNU-WIDER Working Paper*, n. 122, United Nations University, World Institute for Development Economics Research, Helsinki.

Azzarri, C., Carletto, G., Davis, B., Nucifora, A., Sohnesen T.P. (2011). Child Undernutrition in Mozambique, World Bank/UNICEF, Maputo.

Boom van den B. (2010). Analysis of poverty in Mozambique, household poverty status, child malnutrition and other indicators 1996/97, 2002/03, 2008/09, study commissioned by the Royal Dutch Embassy in Maputo, mimeo.

Cunguara, B. and Joseph Hanlon, J. (2010). Poverty is not being reduced in Mozambique, *Crisis States Working Papers Series*, no.2 (74), DESTIN, London School of Economics, London.

Delarue J, Mesple-Somps S., Naudet J.D, Robilliard A.S. (2009). The Sikasso Paradoxe: Cotton and Poverty in Mali, *Document de Travail*, n. 2009-09, Développement, Institutions et Mondialisation (DIAL), Paris.

Deaton A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy, World Bank.

Deaton A., Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates For Welfare Analysis, *Living Standards Measurement Study Working Paper*, No. 135, The World Bank.

Filmer D., and Pritchett, L. (2001). Estimating Wealth Effects Without Expenditure Data or Tears: An Application to Educational Enrollments in States of India, *Demography,* 38(1): 115–32.

Hanlon, J. and Smart, T. (2008). Do Bicycles Equal Development in Mozambique*,* Woodbridge: James Currey.

Maia C., van der Berg S. (2010). When the remedy is worse than the disease: Adjusting survey income data for price differentials, with special reference to Mozambique, *MPRA Working Paper,* n. 26571.

MPD-DNEAP (2010). Poverty and Wellbeing in Mozambique: Third National Poverty Assessment, Ministry of Planning and Development - National Directory of Studies and Policy Analysis-, Maputo.

MPF/IFPRI/PU (2004). Poverty and Wellbeing in Mozambique: Second National Poverty Assessment, Ministry of Planning and Finance/Purdue University/International Food Policy Research Institute, Maputo.

MPF/UEM/IFPRI (1998). Understanding Poverty and Well-being in Mozambique: The First National Assessment, Ministry of Planning and Finance/Eduardo Mondlane University/International Food Policy Research Institute, Maputo.

Ravallion, M. and Lokshin, M. (2006). On the consistency of poverty lines, in *Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke*, ed. A. de Janvry and R. Kanbur, Amsterdam: Kluwer Academic Publishers.

Sohnesen T. (2011). Are price data from IOF 2008/09 stable enough to support multiple poverty lines and entropy adjustment?, mimeo.

UN (2010). Millennium Development Goals Report, United Nations, New York.

World Development Indicators (2011), accessed online at http://data.worldbank.org/indicator, The World Bank, Washington, DC.

WHO (World Health Organization) (1985). Energy and protein requirements, *Technical Report Series*, n. 724. Geneva

Maps, graphs, and tables

Map 1: Population density by regions and districts

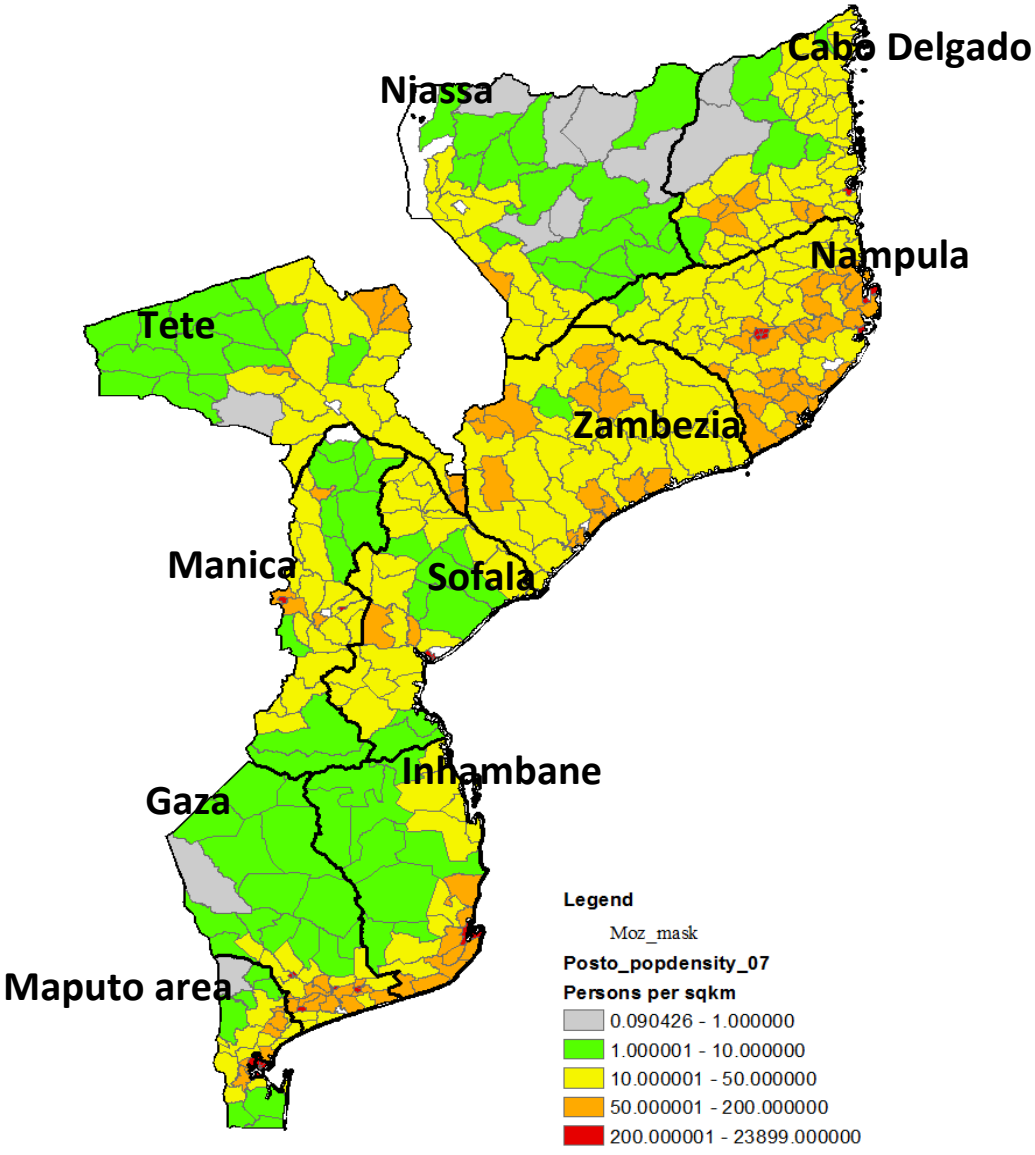


Table 4: Official and Revised PLs, 1996/97, 2002/03, 2008/09, by 13 Regions and National-Level

13 provinces	1996/97		2002/03		2008/09	
	Official	Revised	Official	Revised	Official	Revised
Niassa & C. Delgado-rural	4.02	4.33	7.10	7.72	15.95	17.33
Niassa & C. Delgado-urban	5.43	4.47	10.23	10.60	18.91	19.01
Nampula-rural	3.36	4.52	5.97	6.37	14.33	13.44
Nampula-urban	4.95	5.13	6.66	7.52	16.72	15.27
Sofala & Zambezia-rural	4.85	4.41	5.47	7.01	14.35	14.58
Sofala & Zambezia-urban	7.60	6.19	8.77	9.12	19.07	16.25
Manica & Tete-rural	4.71	4.45	6.93	7.99	19.39	19.25
Manica & Tete-urban	7.41	6.02	9.69	9.69	21.47	18.62
Gaza & Inhambane-rural	6.43	5.99	9.01	6.74	18.37	14.75
Gaza & Inhambane-urban	7.83	5.78	10.72	8.00	20.31	15.71
Maputo Provincia-rural	7.32	6.10	16.76	9.40	24.84	17.28
Maputo Provincia-urban	8.71	7.19	18.30	11.33	30.86	20.83
Maputo City	8.54	7.21	19.52	11.77	33.14	20.57
National	**5.27**	**5.03**	**8.47**	**7.94**	**18.41**	**16.41**

Official: 13 flexible basket PLs based on location specific baskets, with entropy correction **Revised**: national 1996/97 based basket PL with 13 location price deflators;

Figure 4: Poverty Line Ratios, Official and Revised, by Region

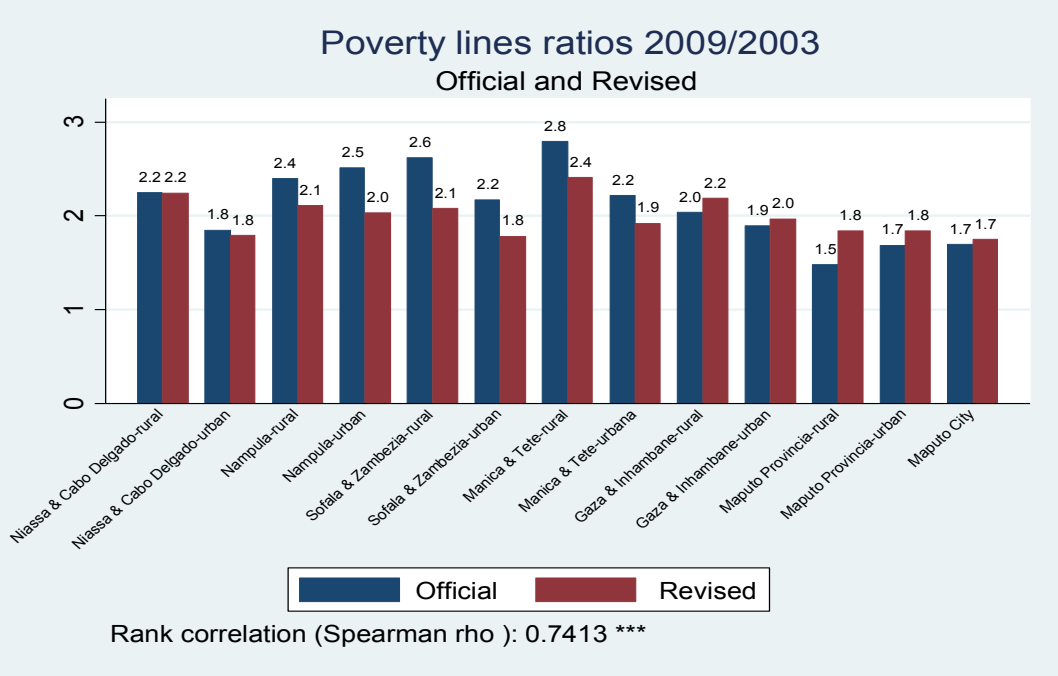


Table 5: 13-Regions Rank Correlation (Spearman rho) between the Local Poverty Rate and its Correlates, Official and Revised PLs (IAF 2002/03 and IOF 2008/09)

Year	2002/03		2008/09	
	Official	Revised	Official	Revised
Wealth Index	-0.06	-0.95**	-0.17	-0.70**
Piped water (%)	-0.27	-0.70**	-0.24	-0.66**
Toilet (%)	-0.13	-0.92**	0.0	-0.68**
Electricity (%)	-0.20	-0.89**	-0.21	-0.64**
Concrete walls (%)	0.09	-0.95**	0.05	-0.59**
Household head education (years)	-0.31	-0.49	-0.36	-0.73**
Kids stunted (%)	-	-	-0.19	0.56**
Kids underweight (%)	-	-	-0.12	0.54*
note: *** p<0.01, ** p<0.05, * p<0.1				

Official: 13 location-specific PLs (MPD-DNEAP, 2010); **Revised:** national 1996/97 based basket PL with 13 location price deflators;

Table 6: Official and Revised Consumption/PL OLS Regression Models (dependent variable: ratio between total household consumption per-capita/day and PLs; difference=national - official)

	Model with official PLs		Model with revised PL		Difference of t- values
	Coef	t	Coef	T	
Household characteristics					
Household head age	0.009***	5.500	0.010***	6.342	**0.842**
Hh head age squared	-0.000***	-5.837	-0.000***	-6.668	**0.832**
Head is female	0.006	0.398	0.016	1.007	**0.609**
Head is widow/er	-0.081***	-3.654	-0.072***	-3.327	**-0.327**
Hh head is polygamous	0.106***	5.205	0.090***	4.554	**-0.651**
Household size	-0.123***	-35.800	-0.123***	-36.599	**0.799**
Avg adult yrs of age	0.002*	1.852	0.002**	2.517	**0.664**
Dependency ratio	-0.087***	-11.925	-0.089***	-12.543	**0.618**
Education and work					
Unemployment ratio + discouraged and seasonal	-0.076***	-2.769	-0.098***	-3.658	**0.889**
Education of household head	0.032***	14.901	0.029***	13.711	**-1.190**
Hh not engaged in agr. Activity	0.024	1.067	0.094***	4.305	**3.238**
Shocks					
Hh had a relevant natural shock during past 5 years	-0.025*	-1.760	-0.023*	-1.665	**-0.095**
Hh had a relevant death-related shock during past 5 years	-0.010	-0.660	-0.027*	-1.734	**1.074**

Assets and access to services					
Hh has piped or protected water source	0.033*	1.827	0.085***	4.824	**2.997**
is the make of roof good?	-0.042**	-2.372	0.075***	4.352	**1.980**
is the make of floor good?	0.069***	3.871	0.119***	6.799	**2.928**
Hh has any toilet	0.310***	10.787	0.333***	11.881	**1.094**
Hh has electricity?	0.186***	7.833	0.178***	7.686	**-0.147**
Hh has cell phone?	0.201***	11.289	0.250***	14.433	**3.144**
Hh owns tv	0.198***	8.564	0.213***	9.420	**0.856**
Hh has any bike	0.167***	12.197	0.119***	8.893	**-3.304**
Hh has car?	0.849***	22.328	0.856***	23.071	**0.744**
Agricultural assets					
Tropical Livestock Units: cattle	0.005	0.839	0.005	1.017	**0.178**
Tropical Livestock Units: sheep	0.315***	3.038	0.314***	3.102	**0.064**
Tropical Livestock Units: goats	0.065***	3.545	0.062***	3.501	**-0.044**
Tropical Livestock Units: pigs	0.069***	3.424	0.077***	3.895	**0.471**
Tropical Livestock Units: chicken	0.092**	2.434	0.089**	2.410	**-0.024**
Tropical Livestock Units: ducks	0.310	1.528	0.379*	1.916	**0.388**
Tropical Livestock Units: other animals	0.598***	4.293	0.501***	3.686	**-0.607**
Hh sold maize	0.126***	5.461	0.102***	4.564	**-0.898**
Hh cultivated maize	0.175***	11.129	0.165***	10.776	**-0.353**
Hh sold cassava	0.171***	5.820	0.168***	5.844	**0.024**
Hh cultivated cassava	-0.016	-1.121	-0.004	-0.260	**-0.861**
Hh sold potatoes	0.055	0.952	0.034	0.608	**-0.345**
Hh cultivated potatoes	-0.005	-0.199	-0.008	-0.342	**0.143**
R2	0.427		0.532		
Adj-R2	0.425		0.53		
F-statistics (46, 10715)	173.54		264.73		
Observations	10762		10762		
note: *** p<0.01, ** p<0.05, * p<0.1					

*Monthly dummies are not reported

Official: 13 location-specific PLs (MPD-DNEAP, 2010); **Revised :** single 1996/97 basket national PL and 13 provincial price deflators

Table 7: Poverty Headcounts in 1996/97, 2002/03, 2008/09, by 13 Regions and National-Level

13 Regions	1996/97		2002/03		2008/09	
	Official	Revised	Official	Revised	Official	Revised
Niassa & C. Delgado-rural	61.63	66.34	60.37	71.74	32.70	46.44
Niassa & C. Delgado-urban	66.88	56.43	53.93	51.10	43.38	40.06
Nampula-rural	65.44	81.16	57.81	69.32	56.67	65.68
Nampula-urban	82.76	82.90	44.91	53.47	49.90	46.38
Sofala & Zambezia-rural	74.83	70.69	42.07	64.28	69.69	76.70
Sofala & Zambezia-urban	67.18	58.04	41.72	37.72	56.66	35.67
Manica & Tete-rural	74.60	71.98	51.62	67.68	47.51	62.27

Manica & Tete-urban	65.57	53.46	54.09	40.27	48.67	30.01
Gaza & Inhambane-rural	75.31	71.68	73.14	54.49	65.19	45.76
Gaza & Inhambane-urban	63.77	42.75	62.72	31.61	44.86	21.17
Maputo Provincia-rural	76.81	66.39	81.18	46.12	76.33	39.26
Maputo Provincia-urban	48.24	41.35	61.83	30.40	63.66	19.96
Maputo City	47.84	37.14	53.60	23.20	36.15	9.53
National	**69.38**	**68.38**	**54.07**	**56.40**	**54.69**	**52.08**
Confidence intervals: National	**68.4-70.46**	**67.4/ 69.4**	**53.0/ 55.1**	**55.4/ 57.5**	**53.7/ 55.6**	**51.1/ 53.0**

Official: 13 flexible basket PLs based on location specific baskets, with entropy correction (MPD-DNEAP, 2010)**; Revised:** national 1996/97 based basket PL with 13 location price deflators and 2002/03 and 2008/09 data partially re-estimated (set A and B table1).