IESE Instituto de Estudos Sociais e Económicos

On the Robustness of Poverty Predictors

Patricia Silva
University of Copenhagen
Virgulino Nhate

Conference Paper nº 37

Conferência Inaugural do IESE
"Desafios para a investigação social e económica em Moçambique"
19 de Setembro de 2007

On the Robustness of Poverty Predictors

Patricia Silva
University of Copenhagen

Virgulino Nhate
Mozambique Ministry of Planning and Development

DRAFT: September 2007

Abstract:

1. Introduction

The demand for data to inform policy and monitor poverty is increasing in developing countries. Goal one of the Millennium Development Goals (MDGs)—to halve the number of people in extreme poverty—can only be measured and monitored using household budget survey data. These surveys contain detailed consumption and expenditure information, from which income poverty statistics can be obtained. Income poverty indicators are also frequently embedded in Poverty Reduction Strategy Papers (PRSPs) and timely household budget data are thus important for the evaluation of the success of poverty reduction policies.

The interval between household budget surveys is long, frequently five or more years. This makes monitoring the impact of public policy on poverty more difficult. To overcome this problem, less extensive household surveys, without consumption and expenditure information, have been developed to monitor other poverty indicators. The Core Welfare Indicator Questionnaire (CWIQ), developed by the World Bank in the mid 1990's, is one example of a non-monetary poverty monitoring survey. These "light" monitoring surveys focus on non-monetary poverty indicators, such as school attendance and literacy rates, access to health and other services, employment, household ownership of assets, etc, and are thus quicker and relatively less expensive to implement than household budget surveys. Obtaining a precise measurement of how many households fall below the poverty line, however, is not directly possible from such surveys.

Some of the household information obtained in these light surveys, however, overlaps with information available from household budget surveys. Recently, analytical techniques have been developed to predict household consumption levels using these light surveys, which are then used to estimate poverty rates. One method, employed by Datt and Jolliffe (2005), estimates poverty by using a weighted average of household probability of being poor based on the predicted consumption. We refer to this method as the "analytical method". A second method, developed by Elbers, Lanjouw, and Lanjouw (2003), uses small area estimation techniques. We refer to this method as the "simulation method". This estimation method is based on poverty mapping techniques, which combine detailed data from household budget surveys with larger population census surveys. The population census surveys provide limited information about households, but generally cover a much larger number of households than consumption and expenditure surveys, and are thus representative at smaller geographical units. Poverty mapping techniques combine these different data sources to obtain poverty estimates a lower level of spatial aggregation than household surveys are designed to be representative. Poverty mapping techniques have also recently been used to combine data from surveys in different years to obtain poverty estimates overtime, for the time periods for which no household budget survey data are available.

A critical element in this exercise of predicting poverty overtime is the stability of the parameters that determine household consumption. In order to predict poverty in future years, one must assume that the determinants of consumption have not changed. This becomes a strenuous assumption the more dynamic the economy is and the longer the time span between surveys. In this paper, we compare two alternative poverty prediction methods in order to assess the robustness of the resulting poverty estimates. Because we use data from two household budget surveys, this allows us to evaluate the poverty predictions against actual poverty figures for the later time period. It also allows us to test whether the determinants of household consumption changed between the two household budget surveys carried out in Mozambique

.

We find that the assumption of stable consumption determinants does not hold for Mozambique during the time period examined. When comparing the two poverty prediction methods, we find that their relative performance appears to be highly dependent on how far into the future the predictions are carried out. The paper then considers the policy implications of these findings for Mozambique and other developing countries with regards to support for different types of household welfare surveys.

The paper is structured as follows. Section 2 provides a brief review of the literature on poverty prediction methodology, paying particular attention to previous analysis of poverty trends in Mozambique. Section 3 describes the model specification and empirical approach employed in the analysis. Section 4 describes the data used for the analysis. Section 5 discusses the results of the analysis and Section 6 concludes.

2. Literature Review—Poverty Trends in Mozambique

Mozambique emerged from a prolonged civil war which ended in 1992, and was unarguably one of the poorest countries in the world—with an estimated GDP per capita of US$80 in 1995. In 1996, the first nationally representative household budget survey—the *Inquerito Nacional aos Agregados Familiares* (IAF)—was carried out and analysis of the survey data indicated a poverty headcount of 69 percent. As Table 1 shows, poverty was higher in rural areas (71 percent) compared to urban areas (62 percent), and in some provinces the poverty headcount exceeded 80 percent. A second nationally representative household budget survey was carried out in 2002 to measure the progress in poverty reduction efforts. The IAF 2002 survey showed that poverty declined considerably in the intervening years, with 54 percent of the population falling below the poverty line. Although the gap between rural and urban communities narrowed, poverty remained higher in rural areas, at 55 percent of the population. The estimated decline in poverty rates was consistent with overall economic growth development in the post war period (Ministry of Planning and Finance, 2004).

In order to analyze the developing poverty trends in Mozambique, several researchers have combined household budget data from the IAF surveys with the core welfare indicator questionnaire (QUIBB) data, which was carried out in 2000, to predict poverty rates in years between the two IAF surveys. Simler et al (2003) use data from the 1996 IAF as the basis for their prediction of poverty rates in 2000. Mathiassen and Hansen (2005), on the other hand, use data from the 2002 IAF for their prediction of poverty in 2000. The forward poverty predictions based on the 1996 IAF data seem to suggest faster poverty reduction rates than the backward poverty prediction based on the 2002 IAF data. In their analysis, Simler et al (2003) note that their predicted poverty estimates appear to consistently underestimate poverty in 1996. One could construe that as sign that the predicted poverty rates they provide are conservative estimates of poverty reduction that has actually taken place. However, this may not necessarily be the case, as generally higher predicted poverty rates for 2000 are obtained by Mathiassen and Hasen (2005).

Another significant difference between these two studies is the consumption model estimated which is used for the poverty predictions. Simler et al (2003) analysis is based on a consumption model estimated separately for each of the ten provinces in Mozambique. Mathiassen and Hansen (2005), on the other hand, estimate a consumption model that distinguishes between rural and urban areas in each of the three regions of Mozambique. Maputo city is treated separately in both estimations. The literature on poverty determinants suggests that there are significant differences in poverty determinants between rural and urban areas and that it is generally harder to obtain good predictors of consumption and poverty for rural models.

The different poverty estimates for 2000 reported by these two studies could thus be attributed to either differences in the consumption model used as a basis for poverty predictions or the use of a different time period used to estimate the consumption model. In this paper, we assess the robustness of poverty estimates by examining whether these differences derive from the use of a different consumption model or the use of a different base year for the consumption estimates. If the difference in predictions is driven primarily by the use of different base year for the consumption model estimated, this would suggest that the determinants of consumption in Mozambique were not stable during the time period in question. If this is the case, then the usefulness of the application of this poverty prediction techniques overtime must be carefully considered. We thus formally test for the stability of consumption determinants, using the two household budget surveys. Additionally, to assess the robustness of the methodology employed in the previous poverty prediction studies, we compare the poverty prediction results using two alternative prediction methods.

3. Model Specification

The basic idea behind the poverty prediction methodology is to first estimate household consumption per capita, the indicator of household welfare and poverty status, and a set of explanatory variables common to both the household budget survey and the non-budget survey. By restricting the set of explanatory variables in this way, the estimated regression coefficients from the consumption model can then be used generate estimates of consumption levels for the population represented in the non-budget survey. In the poverty mapping methodology, this method is applied to obtain poverty estimates at lower levels of spatial aggregation by combining two contemporaneous household survey data sets, a smaller and more detailed survey with a larger and more representative survey.

In this paper, we apply the poverty mapping method to generate estimates of poverty at a later point in time. We then compare the resulting predictions, obtained through simulations, with an alternative method to predict poverty, which treats predicted consumption as a stochastic variable and analytically derives the probability of the consumption measure falling below the poverty line. The weighted average of household probability of being poor gives the poverty headcount estimate (Datt and Jolliffe, 2005).

A critical assumption in both of the applied prediction methods is that the estimated parameters of the consumption model are stable overtime—in other words, the relationship between consumption and the explanatory variables used to estimate it does not change in the span of time between the two surveys. This is the assumption adopted in previous studies that make poverty predictions overtime. In some of the poverty prediction studies, the poverty predictions are only a few years away from the original household expenditure survey (Simler et al, 2003, Mathiassen and Hansen 2005), whereas in at least one study, the time span is much longer (Stifel and Christiansen, 2006). However, the more dynamic the economy and the more time that passes between the surveys, the more likely it is that the estimated model parameters are unstable (Mathiassen and Hansen, 2005). In this paper we test whether this assumption of stable consumption determinants holds for Mozambique using the two household budget surveys from 1996 and 2002.

We follow the consumption model specified by Simler et al (2003), using the 1996 IAF survey, and test the stability of the estimated consumption model parameters using the 2002 IAF survey. The consumption model estimated is specified, as follows:

$$\ln(y_{hc}) = X_h^{'}\beta + u_{ch} \qquad (1)$$

where y_{ch} is per capita consumption of household h in cluster c at time period t, X_h is a set of household and community characteristics that are found in both surveys, and u_{ch} is the household specific stochastic disturbance term at time t.

As Elbers et al. (2003) show, the disturbance term has three components, which account for the difference between the actual consumption value and its estimated value:

$$y - \tilde{\mu}^s = (y - \mu) + (\mu^s - \hat{\mu}^s) + (\hat{\mu}^s - \tilde{\mu}^s) \qquad (2)$$

The first component of the error term is the idiosyncratic error $(y - \mu)$, which measures how the household's expenditures deviate from their expected values. The idiosyncratic error depends on the size of the population in the target survey and the explanatory power of the model. The smaller the subgroup for which the estimates are carried, the larger the potential size of the idiosyncratic error. This source of the idiosyncratic error is less of a concern with poverty estimates overtime, since these are applied to sizable representative populations, as opposed to subgroups as in small area estimation. The idiosyncratic error also depends on careful selection of explanatory variables. Thus it in not uncommon in this literature to maximize the explanatory power of the model by estimating equation (1) through stepwise regressions. This assures that only variables which contribute to explaining the variation in consumption are included in the model to maximize efficiency.

The second error component is the model error $(\mu^s - \hat{\mu}^s)$, which is due to the variance in the first stage estimates of the parameters. Stifel and Christiansen (2006) discuss the magnitude and sources of model error. The magnitude of the model error is difficult to determine without comparable expenditure surveys overtime. Model error can also be affected by slight differences in definition of variables among the different survey instruments. Of particular importance for this analysis, is the model error due to instability in estimated coefficients overtime. Stifel And Christiansen (2006) suggest the inclusion of time varying explanatory variables in the consumption model specification (such as rain fall and price data) in order to mitigate the magnitude of this type of error. Correction for heteroskedacity in the estimated consumption model also helps increase the efficiency of the parameters estimated and to reduce model error.

The last component of the error term is a computational error $(\hat{\mu}^s - \tilde{\mu}^s)$, which is uncorrelated with the other two types of errors. The computational error depends on the computational method chosen and its asymptotic distribution can be determined based on the simulation methods chosen.

The consumption model specified in (1) is estimated separately for each of the ten provinces in Mozambique and for Maputo city, using a stepwise procedure to select the relevant explanatory variable for each provincial equation. The estimated regression does not account for all the variation in the dependent variable—the prediction $X_h^{'}\hat{\beta}$ has a smaller variance than the true y_{ch}. We thus use simulation methods, drawing from the estimated distributions of u_{ch} and β, to generate estimates of $\hat{u}_{t+k}$ and $\hat{\beta}$. These are then used to predict per capita household consumption for the later survey at time $t+k$, conditional on the values of X_{t+k} observed in the later survey, so that:

$$\ln(\hat{y}_{t+k}) = X_{t+k}^{'}\hat{\beta} + \hat{u}_{t+k} \tag{3}$$

The Foster-Greer-Thorbecke (FGT) poverty measures are then calculated based on the simulated consumption levels for the later Mozambican household survey.

4. Data

The analysis in this paper is based on household survey data from Mozambique. We use household expenditure survey data from 1996-97 and 2002-03 to test the robustness of alternative methods for poverty prediction. We also use the QUIBB 2000 data and the labor force survey, IFTRAB 2004-05. Because the 2002-03 is an expenditure survey, this allows us to evaluate the poverty predictions against actual poverty figures. The QUIBB and IFTRAB, however, are not expenditure surveys, so it is not possible to evaluate how good the predictions for 2000 and 2004 are against the actual poverty levels.

5. Results

We begin our discussion of the results by first examining how the two alternative poverty prediction methods perform in sample. Using the 1996 IAF we estimate the consumption model specified in (1) and then use the estimated coefficients to predict poverty using the same dataset. This allows us to compare the prediction results against the actual poverty rates, without introducing any disturbances due to the instability of consumption determinants. Table 2 presents the results of this in sample prediction. We find that the poverty mapping simulation method produces more reliable estimates of the poverty headcount. The predicted poverty rates for national, regional and zone levels are fairly close to the actual estimates of poverty. However, since the provincial level poverty headcount have high standard deviation, the predicted poverty rates at the provincial level are somewhat less precise.

Next we evaluate the performance of the two poverty prediction methods overtime. No previous studies have actually compared their predict poverty results against actual poverty estimates, as we do here. Table 3 presents the predicted poverty rates in 2003, using the 1996 IAF data as the basis for the prediction. In this case, we find that with both methods the predicted poverty rates fall mostly outside of the confidence interval of the actual poverty headcount. There appears to be no consistent patterns of under or over prediction with either method. This is particularly the case at the provincial level, where the results are more mixed. Neither method, therefore, appears clearly preferable in this case.

The poor performance of both prediction methods for 2003 suggests that the source of the problem may not be the prediction methodology per say, but rather the underlying assumption of stability of consumption determinants. During the 6 years between the two household budget surveys, Mozambique experience significant economic changes. We thus test whether the determinants of consumption changed between the two survey years, by estimating the same consumption model used for the prediction using the 2002 IAF data. In all but two provinces, we reject the hypothesis that the coefficients of the consumption model are equal in the two time periods. This could very well explain why so many of the predict poverty rates are outside the confidence interval of the actual poverty rates.

The question which emerges then is how far into the future can poverty predictions be confidently made? There is no such straight forward answer. Yet, it is interesting to compare how the two prediction methods compare for different time horizons. We thus predict poverty in 2004, using the 1996 IAF data and the IAF 2002 data. Table 4 shows the poverty predictions for 2004. We find substantial differences in predicted poverty rates between the two methods when the predictions are based on the 1996 IAF data. The poverty mapping simulation method generally predicts higher poverty rates than the analytical method. However, when we use the 2002 IAF as the basis for the poverty predicts, the two methods produce fairly close estimates of poverty in 2004. The closer poverty estimates obtained by the two methods over a shorter time horizon are thus reassuring. This suggests frequent household budget surveys are clearly important for close monitoring of poverty developments. Whereas the poverty prediction methods can make reasonable predictions over a short time horizon, the estimates get less accurate the further into the future the predictions are made.

6. Conclusions

Recently, analytical techniques have been developed to predict poverty using "light" monitoring surveys. These light monitoring surveys are generally less expensive to implement than household budget surveys and thus can be carried more frequently. In this paper, we assess the robustness of poverty predictions methods which combine household budget survey data with other survey data to predict poverty developments overtime in Mozambique. We do this using household budget survey data from 1996 and 2002, and data from other surveys taking place in 2000 and 2004.

Prediction of poverty overtime relies on the assumption that the determinants of consumption be stable overtime. We find that this assumption does not hold for Mozambique for the time period between 1996 and 2002. Consequently, poverty prediction using different methods can produce a wide range of poverty estimates. However, when predictions are made over a shorter time horizon, the two methods evaluated produce reasonably close estimates of poverty. This suggests that while the poverty prediction methodology is useful to fill in information gap between household budget surveys, one must not stretch the method and cast predictions far into the future. A continuing assessment of poverty through the collection of household budget surveys remains important. Our results suggest that periodic household budget surveys are thus important not just to measure poverty, but also to understand how the determinants of poverty are changing overtime.

7. References

Elbers, C., J. Lanjouw, and P. Lanjouw. 2003. Micro-Level Estimation of Poverty and Inequality. Econometrica, pp.355-364.

Datt, G. and D. Jolliffe. 1998. Poverty in Egypt: Modeling and Policy Simulations. Economic Development and Cultural Change, pp.327-346.

Datt, G., D. Jolliffe, and M. Sharma. 2004. Robust Poverty and Inequality Measurement in Egypt: Correcting for Spatial Price Variation and Sample Design Effects. Review of Development Economomics, pp.557-572.

Mathiassen, A., and D. Hansen. 2005. Predicting Poverty for Mozambique 2000 to 2005. How Robust are the Models? Statistics Norway, Division of Development Cooperation.

Ministry of Planning and Finance, International Food Policy Research Institute, and Purdue University. 2004. Poverty and Well Being in Mozambique: The Second National Assessment.

Simler, K., S. Harrower, C. Massingarela. 2003. Estimating Poverty Indices from Simple Indicator Surveys. mimeo.

Stifel, D., and L. Christiansen. 2006. Tracking Poverty Overtime in Absence of Comparable Consumption Data. World Bank Policy Research Working Paper 3810.

Table 1. Poverty Headcount in Mozambique 1996 and 2002

	1996-97		*2002-03*	
	Poverty Headcount	*Standard Error*	*Poverty Headcount*	*Standard Error*
National	69.4	1.14	54.1	1.36
Urban	62.0	2.67	51.5	2.25
Rural	71.3	1.25	55.3	1.68
North	66.3	2.28	55.3	2.57
Center	73.8	1.60	45.5	2.40
South	65.8	1.96	66.5	1.35
Niassa	70.6	3.78	52.1	5.44
Cabo Delgado	57.4	4.19	63.2	3.41
Nampula	68.9	3.29	52.6	3.82
Zambesia	68.1	2.60	44.6	4.60
Tete	82.3	3.22	59.8	4.22
Manica	62.6	5.95	43.6	4.11
Sofala	87.9	1.46	36.1	2.76
Inhambane	82.6	2.45	80.7	2.16
Gaza	64.6	3.26	60.1	2.60
Maputo Province	65.6	5.41	69.3	2.83
Maputo City	47.8	4.06	53.6	3.09

Note: Standard error of poverty headcount estimates corrected for sample design effects

Table 2. Poverty Prediction—In Sample Comparisons, 1996

	Poverty Headcount 1996-97		
	Actual	in sample prediction (analytical method)	in sample prediction (simulation method)
National	69.4	65.7*	67.2
Urban	62.0	60.1	65.1
Rural	71.3	67.1*	71.3
North	66.3	63.2	62.6
Center	73.8	70.0*	60.9*
South	65.8	61.5*	68.7
Niassa	70.6	68.5	68.3
Cabo Delgado	57.4	53.3	57.7
Nampula	68.9	66.0	67.5
Zambesia	68.1	63.7	65.4
Tete	82.3	79.2	81.1
Manica	62.6	58.5	58.7
Sofala	87.9	85.3	85.7
Inhambane	82.6	77.0*	79.3
Gaza	64.6	58.1*	57.6*
Maputo Province	65.6	63.6	64.8
Maputo City	47.8	45.6	47.0

Note: * estimated poverty rates fall outside the 95% confidence interval of actual poverty rates

Table 3. Comparison of Actual and Predicted Poverty in 2002

	Poverty Headcount 2002-03		
	Actual	Analytical prediction	Simulation prediction
National	54.1	51.7	58.1*
Urban	51.5	44.2*	47.6
Rural	55.3	55.3	63.1*
North	55.3	50.6	54.2
Center	45.5	58.2*	61.9*
South	66.5	43.1*	57.0*
Niassa	52.1	52.8	54.3
Cabo Delgado	63.2	68.4	73.7*
Nampula	52.6	42.1*	45.6
Zambesia	44.6	76.3*	78.3*
Tete	59.8	98.3*	97.7*
Manica	43.6	12.0*	18.9*
Sofala	36.1	20.7*	29.4*
Inhambane	80.7	62.7*	69.6*
Gaza	60.1	12.0*	57.0
Maputo Province	69.3	58.2*	55.2*
Maputo City	53.6	41.6*	42.9*

Note: * estimated poverty rates fall outside the 95% confidence interval of actual poverty rates

Table 4. Poverty Predictions for 2004

	Poverty Prediction (based on 96 data)		*Poverty Prediction (based on 02 data)*	
	Analytical prediction	Simulation prediction	Analytical prediction	Simulation prediction
National	38.3	43.1	38.9	38.9
Urban	33.6	46.9	33.9	33.0
Rural	41.0	36.4	41.6	42.1
North	39.8	41.0	44.0	46.4
Center	43.8	46.2	35.0	32.2
South	28.8	41.1	38.8	39.5
Niassa	46.8	42.2	41.9	45.1
Cabo Delgado	69.6	76.2	49.4	48.7
Nampula	22.0	22.3	41.9	45.7
Zambesia	51.8	54.3	23.3	19.0
Tete	83.6	81.4	52.2	54.4
Manica	9.3	8.4	85.6	81.5
Sofala	28.2	30.0	6.2	10.9
Inhambane	30.5	43.4	54.0	54.2
Gaza	36.5	43.4	25.2	26.4
Maputo Province	65.0	59.3	51.0	53.5
Maputo City	14.2	17.0	21.1	21.3

IESE

INSTITUTO DE ESTUDOS SOCIAIS E ECONÓMICOS

Av. Patrice Lumumba, 178 - Maputo
MOÇAMBIQUE

Tel. + 258 21 328894
Fax + 258 21 328895
www.iese.ac.mz