2007/4

Documents

Statistics Norway

Statistisk sentralbyrå

Documents

Astrid Mathiassen and Dag Roll-Hansen

Predicting poverty for Mozambique 2000 to 2005

How robust are the models?

Statistics Norway/Division for Development Cooperation

Executive Summary

There is an increasing demand for more frequent measurements of the poverty situation. Statistics Norway has developed a method for monitoring the development in the time periods between full-fledged Household Income and Expenditure Surveys. The goal of this report is to test the method.

The predictions resulting from these analyses indicate that the poverty rate in Mozambique is declining. However, the analyses also demonstrate that it is difficult to draw conclusions as long as we are not sure that we have included good and relevant explanatory variables.

The report shows that it is important to include data about issues that are frequently seen as related to poverty in the analysis. Examples of such data to include are the number of meals the family had on a given day or if the household has acquired goods like clothes or building materials. These variables add to the quality of the predictions in a substantial manner. This effect is particularly large for the rural domains. This is not surprising, as in general it is more difficult to identify good predictors in the rural than in the urban models. It implies, however, that the urban-rural-division ought to be taken into account in poverty modeling.

The analyses are based on the latest household expenditure survey for Mozambique (IAF2002/03) and two surveys that do not measure consumption; the Core Welfare Indicators Questionnaires (QUIBB 2000/01) and the labor force survey (IFTRAB, 2004/05).

1 Introduction

The increasing demand for more frequent measurements of poverty, typically for *annual* poverty estimates has generated the need for a supplement to the costly, full-fledged Household Income and Expenditure Surveys, which are usually only conducted every 5th year. Statistics Norway has developed one such method for predicting poverty and assessing the uncertainty in years when no comprehensive household expenditure is available see Mathiassen (2005). The basic idea is to utilize the information in an expenditure survey to identify a smaller set of household variables (indicators) that can be collected annually between two budget surveys. This is done by estimating a relation that links consumption and poverty to the set of indicators through a statistical model, i.e., by constructing a 'consumption model'. The indicators should be fast to collect and easy to measure. Hence, they may be compiled through so-called light surveys without collecting expenditure data. The information obtained from the light survey and the estimated model is used to predict poverty rates.

The methodological approach adopted is inspired by statistical modeling in the adjacent area of poverty mapping, cf. Elbers, Lanjouw and Lanjouw (2003) and Hentschel, Lanjouw, Lanjouw and Poggi (2000). The poverty-mapping method outlines a general method for predicting various poverty and inequality measures; however, Mathiassen (2004) outlines a simplified approach to the calculation of the predictor of the headcount ratio as well as its standard error and bias. While the poverty mapping methodology applies a Taylor approximation to derive the standard error, the procedure proposed here derives an exact expression for the standard error. This allows for a simple and transparent estimation procedure. In its simplest form, we assume homoskedastic error terms, which is reasonable in the empirical applications tested here. However, the method can also allow for heteroskedasticity.

In this report we predict poverty by applying the two latest light surveys for Mozambique, in 2000/01 and 2004/05[1]. The models were estimated using the latest household expenditure survey, IAF2002/03. The predictions are used to discuss changes in poverty in Mozambique from 2000-2005, as well as to discuss effects on the prediction of choosing different sets of indicators in the consumption model. However, before returning to the results we will briefly outline the methodology applied.

[1] The empirical analyses for predicting by use of the IFTRAB 2004/05 survey were conducted by Fátima Zacarias, Cassiano Soda Chipembe, Cristóvão Muahio, Elisio Mazive, Xadreque Maunze and Maria Mazive from INE Mozambique, and Geir Øvensen and Astrid Mathiassen from Statistics Norway, while the analyses on the QUIBB 2000/01 were carried out by Astrid Mathiassen.

2 A summary of the method

In this section, we outline the main features of the methodology for predicting poverty rates with limited reference to the statistical methods. Readers looking for further references should consult Mathiassen (2005).

1.0 A predictor for the headcount ratio

An individual is considered poor if his or her consumption or income falls below a certain threshold. This threshold defines the poverty line. We want to predict the headcount ratio, i.e., the proportion of individuals with consumption below a given poverty line[2].

Let Y_i denote the consumption for individual *i*. We refer to Y_i as household consumption per capita or the adult equivalent. Let *z* denote the poverty line. Let $y_i = 1$ if individual *i* is poor where $Y_i \leq z$, and zero otherwise. We are interested in predicting the headcount ratio, y, i.e., the share of poor individuals in a population Ω consisting of N^H households. The population can, for example, refer to a region within a country. Because the unit in the survey is the household, one needs to adjust for the number of members in each household. Let s_i be the number of members in household *i*, and let N be the number of individuals in the population. In our case, an individual is considered poor if his or her household's per capita consumption is at, or below the poverty line. Hence:

(1) $$y = \frac{1}{N} \sum_{i \in \Omega} s_i y_i \ .$$

As indicated above, we wish to use a model to predict y for a given set of household variables (indicators). We next assume that:

(2) $$\ln Y_i = X_i \beta + \sigma \varepsilon_i$$

where X_i is the vector of selected poverty indicators, β is a vector of unknown parameters and ε_i is an error term that is assumed to be distributed according to the standard normal distribution. The parameter σ therefore represents the standard deviation of $\sigma\varepsilon_i$. The assumption on normality is, as shown later, used in the step below; however, other distribution functions can be applied. Assume further that the ε and X are uncorrelated. The logarithmic transformation of the consumption variable serves to reduce the usual asymmetry in the distribution of the error term and stabilizes the variance. The assumption on homoskedasticity and normality of the error term will be further discussed and tested in the empirical section.

[2] We will return to the data requirement and definitions of these concepts in the next section.

Because of the stochastic component in the estimated consumption level, all individuals have a nonzero probability of being poor. Thus, rather than counting the number of individuals with predicted consumption below the poverty line to find an estimator for the headcount ratio, we use the average probability that an individual is poor as the predictor. The probability that individual i's consumption falls below the poverty line, z, is found by inserting the regression model in a probability function:

(3) $$P_i = P(Y_i < z) = P(\ln Y_i < \ln z) = P(X_i\beta + \sigma\varepsilon_i < \ln z) = \Phi\left(\frac{\ln z - X_i\beta}{\sigma}\right)$$

where $\Phi(\ )$ denotes the standard cumulative normal distribution function (but other distribution functions could be applied if it seems more reasonable).

One predictor for the headcount ratio in (1) is then given by:

(4) $$\hat{P} = \frac{1}{n}\sum_{i\in S} s_i \Phi\left(\frac{\ln z - X_i\hat{\beta}}{\hat{\sigma}}\right).$$

It can be shown that this predictor is biased. Hence, we will use the formula for the unbiased predictor given in (6) in the Appendix, section 0. However, for calculating the standard error of the predictor below, it is the simpler predictor in (4) that is used, because using the biased corrected predictor substantially increases the complexity of the calculations, and the error caused by using the unbiased predictor is marginal.

2.0 The standard error of the predictor

The prediction error is the deviation between the poverty level predicted by our model and the actual poverty level in the population. One way to decompose the prediction error is:

(5) $$\frac{1}{N}\sum_{i\in\Omega} s_i y_i - \frac{1}{n}\sum_{i\in S} s_i \hat{P}_i = \left[\frac{1}{N}\sum_{i\in\Omega} s_i y_i - \frac{1}{N}\sum_{i\in\Omega} s_i P_i\right] + \left[\frac{1}{N}\sum_{i\in\Omega} s_i P_i - \frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i\right] + \left[\frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i - \frac{1}{n}\sum_{i\in S} s_i \hat{P}_i\right] .$$

The first term on the right-hand side in (5) is the difference between the actual and expected population poverty levels. This captures how the headcount ratio in the population deviates from its expected value. This component will generally be very small when we provide predictions for large samples.

The second term in (5) is the difference between the expected poverty level and the poverty level predicted by the estimated model for the entire population, Ω. This captures uncertainty from the error in the estimate, $\hat{\beta}$.

The last term in (5) is the difference between the predicted poverty level in the population Ω and the predicted poverty level in the sample S. This is the result of uncertainty because S is a finite random sample. All error components are also affected by the variation of the X-vector in the sample.

The expression of the variance of the error in (5) and the procedure for estimating this variance are described in the Appendix, section 0.

There are other errors that we are not able to measure and that are thus not included in (5). The most critical is stability of the model parameters. Even if the model relation is true at a given time, the regression coefficients may change over time. When the economy changes, the relation between poverty predictors and expenditure may change as well. The more dynamic the economy, and the more time that passes between the surveys, the more likely it is that the model parameters are unstable. To test this assumption, two budget surveys are required to estimate the two consumption models and to test whether the parameters have changed. A short-form measure of consumption could also help to verify the assumption as one could estimate models based on this information and compare the model coefficients.

3 Data

1.0 The surveys

The analyses are based on the latest household expenditure survey for Mozambique, the IAF2002/03, two light surveys; the QUIBB[3] 2000/01 and the labor force survey: IFTRAB, 2004/05.

The consumption model is estimated on the basis of the expenditure survey, IAF 2002/2003. This is a comprehensive socio-economic survey of the living standard in Mozambique and consists of about 8 700 households. The data were prepared by the national statistical office of Mozambique (INE), and important variables such as total household consumption were derived and the poverty line was defined; see the National Directorate of Planning and Budget et al. (2004) for documentation. The welfare measure is given by total daily per capita consumption and expenditure. The poverty line is based on the cost of basic needs.

[3] QUIBB is the Portuguese abbreviation corresponding to CWIQ in English; Core Welfare Indicators Questionnaires. The CWIQ was jointly developed by the World Bank with the UNDP and UNICEF. These surveys are not designed to measure expenditure or consumption but to obtain indicators of welfare and use of and access to public services.

The QUIBB, 2000/01 had a sample of 13 800 households. It contained questions about education, health, employment and characteristics of the households and their house among other things. In addition, it contained some questions about issues that are often seen as related to poverty, such as the number of meals the family had on a given day. As we will see, these poverty predictors are potentially important in constructing a model to predict the development of poverty.

The labor force survey, IFTRAB, 2004/05, consisted of about 17 500 households. It contained elements from the QUIBB survey as well as a substantial number of questions about labor participation and other economic activities for all members in the households. Unfortunately it does not contain the questions from the QUIBB particularly related to poverty. The analytical challenge this represents will be discussed further later in this report.

2.0 The poverty indicators

The range of potential indicators that were examined represents characteristics that were available from both surveys (the expenditure survey and the light survey). In addition, a first criteria for selecting and constructing indicators are reliability, easy measurement and available information. About 150 variables were tested, and they comprise the following groups: literacy, education, employment, assets, housing, energy and water use. Information on typical poverty indicators, given by a separate section in the IAF 2002–03, was included in the QUIBB, but not in the IFTRAB and could therefore not be used for predicting poverty for 2004/05. There are also some other discrepancies between the remaining set of variables available in the two light surveys. The full list of indicators is presented in 0.

The set of indicators used in the model should jointly have a high correlation with household consumption per capita, thus the set of indicators is limited to those that are significant in predicting household consumption and therefore also the poverty level. The final set of poverty indicators included in each of the models is selected by comparing estimated models with various combinations of indicators. Based on statistical criteria, automated through stepwise procedures, we chose the set that constituted the "best" model for predicting the poverty headcount ratio. We estimated separate urban/rural models for each region, each comprising between 800 and 1 900 households, see Table 3 in the Appendix, section 0[4]. We have also estimated a national urban and a national rural model, as well as a full-coverage national model comprising all cases.

[4] There were three regional-rural models, and four regional-urban models, the latter category including a separate model for the national capital, Maputo.

4 Results

In this section we discuss the results of the poverty prediction analyses based on the expenditure survey for 2002/03 and the light surveys for 2000/01 and 2004/05. As there are some essential differences between the indicators available in the light surveys, we have estimated two sets of models used for prediction for each light survey. First, we have estimated the consumption model from the expenditure survey, where we have selected the common indicators available in the expenditure survey and the relevant light survey (referred to as the unrestricted set of indicators). Second, we have estimated the consumption model from the expenditure survey, where we have selected among the common indicators available in all the three surveys, i.e. the expenditure survey and the joint set of indicators between the light surveys (referred to as the restricted set of indicators). The latter set of indicators is apriori better when comparing the poverty levels between the years as we use identical models, while the first approach will produce better predictions as we select from a larger set of indicators. However, consistent comparison is of little help if the models produce poor predictions. Therefore, the models estimated on the basis of the restricted set of indicators are mainly included to discuss the importance on the predictions of including different sets of explanatory variables in the models.

Thus, we estimate models from the expenditure survey where we choose among three different sets of variables as described below:

1. Indicators available in the IAF and the QUIBB. The QUIBB contains a section of indicators that are labeled "poverty predictors" in the questionnaire. This is information on whether or not the household

- has consumed important food groups,
- has acquired goods like clothes, building material or made use of transport,
- has a member working as a casual agricultural worker,
- has kept chicken or
- has received regular remises[5].

These are variables that are assumed to be highly correlated with poverty.

2. Indicators available in the IAF and the IFTRAB. In the IFTRAB dataset there are no such "poverty indicators". However, the IFTRAB contains a larger list on assets, as well as number of assets (while in the QUIBB dataset only information on whether or not the household owns an asset or not). The IFTRAB/IAF set also includes agro-ecological zones.

[5] These variables were included as a result of similar analyses for the expenditure survey for Mozambique IAF96/97 and QUIBB200/01, see Simler, Harrower and Massingarela (2004). Unfortunately, these variables were not included in the IFTRAB04/05.

3. Indicators available in the IAF, the QUIBB and the IFTRAB. This is the most limited set, as it includes only the common set between 1 and 2.

Table 1 below shows the prediction results when we can choose among the unrestricted set of variables in each survey. The standard errors are given in parentheses. In addition to the actual predictions of poverty based on the expenditure survey, we have included the predictions within the sample, i.e. we have predicted the poverty level for 2002-03 by using the respective estimated model from the same sample. The within sample predictions ensure a directly comparable reference for the out of sample predictions, as they are calculated using the same method. The within sample predictions deviate with about one to two percentage points from the actual prediction within the sample, except for Urban North where the within sample based on the IFTRAB-model predicts four and a half percentage points higher poverty than the actual prediction. This region also has a considerably smaller sample of households, see Table 3 in the Appendix, section 0, which is likely to produce less precise predictions.

Table 1 Predictions based on separated models for the two light surveys

	QUIBB 2000-01	IAF 2002-03			IFTRAB 2004-05
		Prediction within IAF sample, QUIBB model	Actual prediction	Prediction within IAF sample, IFTRAB model	
All Mozambique	58.3 (2.1)	55.5	54.1 (1.4)	54.7	49.3 (2.1)
Urban	54.7 (2.6)	50.9	51.5 (2.3)	51.2	45.0 (2.2)
Rural	58.4 (2.3)	55.1	55.3 (1.7)	55.6	51.0 (2.3)
Rural North	63.3 (3.1)	59.1	59.1 (2.4)	60.4	58.6 (2.7)
Urban North	43.8 (4.0)	48.3	47.4 (5.6)	52.0	45.3 (3.6)
Rural Central	46.8 (3.4)	46.6	45.2 (2.9)	47.5	41.0 (2.9)
Urban Central	50.5 (2.3)	46.5	46.6 (5.6)	45.5	40.1 (3.1)
Rural South	77.8 (3.3)	72.4	74.4 (1.9)	72.8	68.6 (3.2)
Urban South	65.3 (3.8)	59.4	62.3 (2.4)	59.9	52.7 (2.9)
Maputo City	49.9 (4.5)	52.1	53.6 (3.1)	51.3	41.8 (3.4)

The predictions indicate that there has been a steady decline in poverty from 2000 to 2005. The decline in poverty is not statistically significant. Note, however, that the decrease in poverty level follows a trend with considerable decline in the poverty headcount in Mozambique as reported in the two subsequent household surveys in 1996/97 and 2002/03. In this period, the poverty headcount in Mozambique fell from 69 to 54 percent. The predicted fall in poverty is most pronounced in the southern region, which is also the region that experienced the lowest decline in poverty between 1996/97 and 2002/03, with a fall in poverty in this period of one percentage point.

Note also that estimating a national model in the case of QUIBB, gives a fairly different prediction compared to predictions from the rural and urban model separately. The within sample predictions for the IAF 2002-03 also indicate that the model fits the consumption-based results best when used separately for urban and rural households, pointing to the importance of applying separate urban and rural models.

A comparison of the results presented in Table 1 and Table 2 shows that the IFTRAB predictions are very similar when using the unrestricted and restricted set of indicators available for this survey. For the QUIBB, however, there are large deviations between the predictions produced by the two models. Recall that the two indicator sets included in the QUIBB-based model include fairly different types of variables sets, as only one of them contains the typical poverty predictors. Particularly for the rural domains the predictions change substantially when poverty predictors are included. This is not surprising as in general it is more difficult to identify good predictors in rural than in urban models. The overall prediction for the rural domain changes from 58 to 51 when the "poverty predictors" are removed from the initial set of indicators.

IFTRAB does not contain these kinds of poverty predictors, a fact that may make the model less precise. However, the survey contains other poverty related questions that to some degree compensate for this to some extent.

Table 2 Predictions based on the same models for the two light surveys

	QUIBB 2000-01	IAF 2002-03		IFTRAB 2004-05
		Prediction within sample	Actual poverty prediction	
All	51.0 (1.5)	55.3	54.1 (1.4)	50.2 (1.6)
Urban	51.1 (2.3)	51.5	51.5 (2.3)	46.2 (2.0)
Rural	51.0 (2.0)	55.5	55.3 (1.7)	51.0 (2.1)
Rural North	53.2 (3.0)	59.8	59.1 (2.4)	58.9 (3.0)
Urban North	45.0 (3.9)	49.9	47.4 (5.6)	43.9 (3.9)
Rural Central	39.5 (2.4)	47.3	45.2 (2.9)	41.3 (2.6)
Urban Central	46.5 (3.4)	45.4	46.6 (5.6)	41.6 (2.9)
Rural South	71.6 (3.2)	71.7	74.4 (1.9)	68.6 (3.0)
Urban South	61.2 (3.4)	59.4	62.3 (2.4)	52.6 (3.0)
Maputo City	41.8 (3.6)	50.9	53.6 (3.1)	41.4 (3.0)

The importance of including the "poverty predictors" (from the poverty indicators section in the QUIBB) is emphasized in Table 4 in the Appendix, section 0 and shows that R-square increases substantially when allowing for these variables in the QUIBB model and is particularly important for the rural model, which has a relatively low adjusted R-square.

5 Concluding Remarks

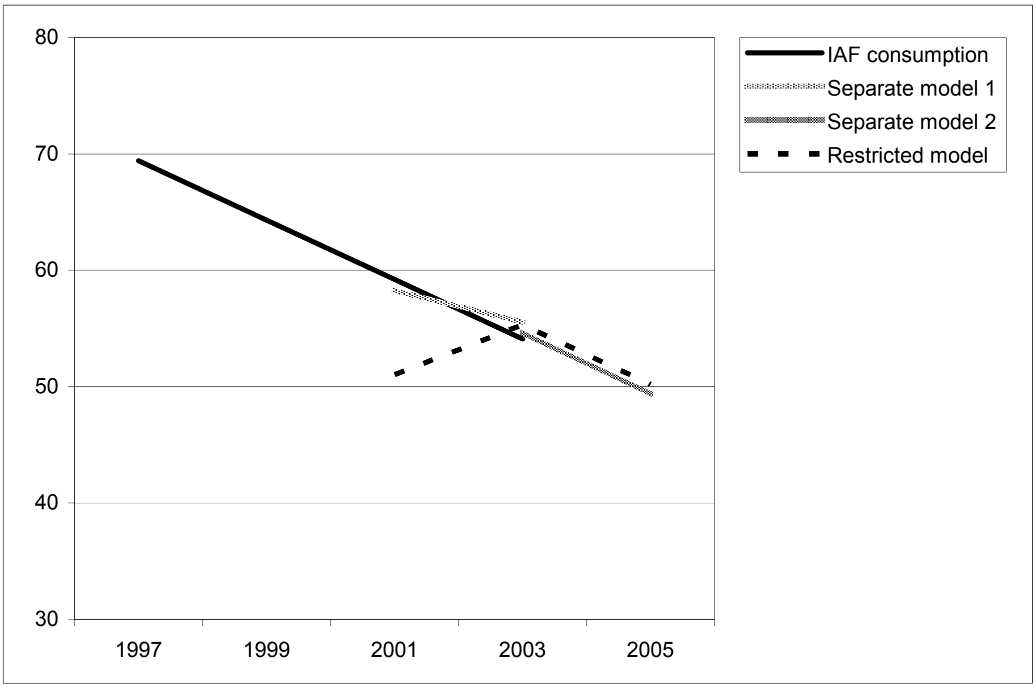
The predictions resulting from these analyses indicate that the share of poverty in Mozambique is declining. However, the analyses also demonstrate the need to be careful drawing too strong conclusions as long as we are not assured having included valid and reliable explanatory variables. The analyses have shown that it is important to include data about issues that are often seen as related to poverty. Examples of such data are the number of meals the family had on a given day or whether the household has acquired goods like clothes or building materials. These variables add substantially to the quality of the predictions. This is particularly the case for the rural domains.

As is the case for similar methods for predicting poverty, this method relies on the critical assumption that the relation between expenditure per capita and poverty indicators is stable over time. However, to test this assumption, two budget surveys are required for estimating the two consumption models and testing whether the parameters have changed.

We have looked into some important aspects of predicting poverty. We have argued that the most sustainable approach is to develop models that best fit the underlying data, rather than using the same model for making an estimate for many time periods.

In the figure below, the graph describing the IAF consumption is the development of the share of poor in Mozambique, simply illustrated as the straight line between the actual poverty rates in 1996/97 and 2002/03, calculated from the Household Budget Surveys (IAF). Separate model 1 describes the change from the QUIBB 2000/01 to the IAF 2002/03 and separate model 2 is based on the IAF 2002/03 and the IFTRAB 2004/05. The restricted model is based on variables that are includes in all three data sets (QUIBB 2000/01, IAF 2002/03 and IFTRAB 2004/05).

Figure 1: Poverty headcount in Mozambique; comparing different models



The graph based on the consumption measured in IAF shows a rapid decline in the poverty headcount. The separate model describing the development from QUIBB 2000/01 to the IAF 2002/03 (separate model 1) shows a similar development. The restricted model graph describes the model using only the variables appearing in all three surveys, as given in Table 2. Looking at this graph we see a quite different development. It seems that the variables included in the restricted model do not describe the development of consumption in a good manner. The main difference between the models is that separate model 1 includes the indicators labeled "poverty predictors" in the QUIBB questionnaire.

Both models present a similar picture of the development from IAF 2002/03 and the IFTRAB 2004/05 (separate model 2 and the restricted model). Comparing these models, we see that separate model 2 does not contain the poverty predictors from the QUIBB questionnaire. It does, however, contain other variables related to poverty partially compensating for this lack. We are, however, not able to verify against the actual consumption until the next Household Budget Survey is conducted. Given the rapid economic development in Mozambique and the fact that both models give similar results, the development appears to be likely.

We have argued that the urban-rural distinction is essential in models predicting the prevalence of poverty. Figure 2 and 3 compare the consumption-based measures with the different projections for urban and rural areas.

Figure 2: Poverty headcount in urban Mozambique; comparing different models

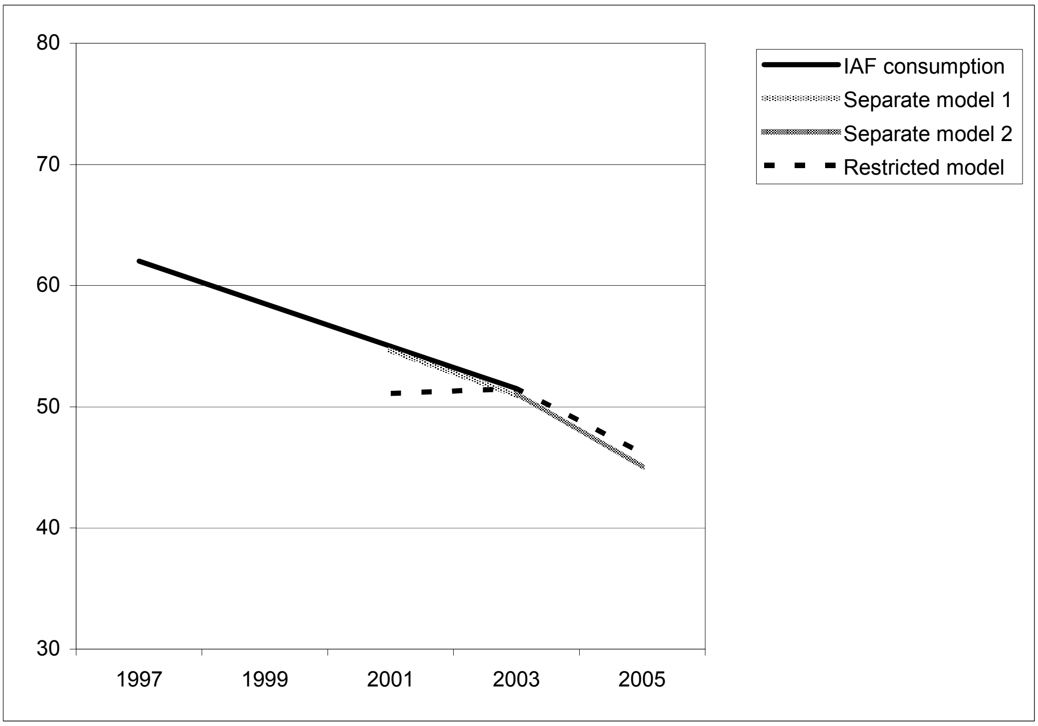
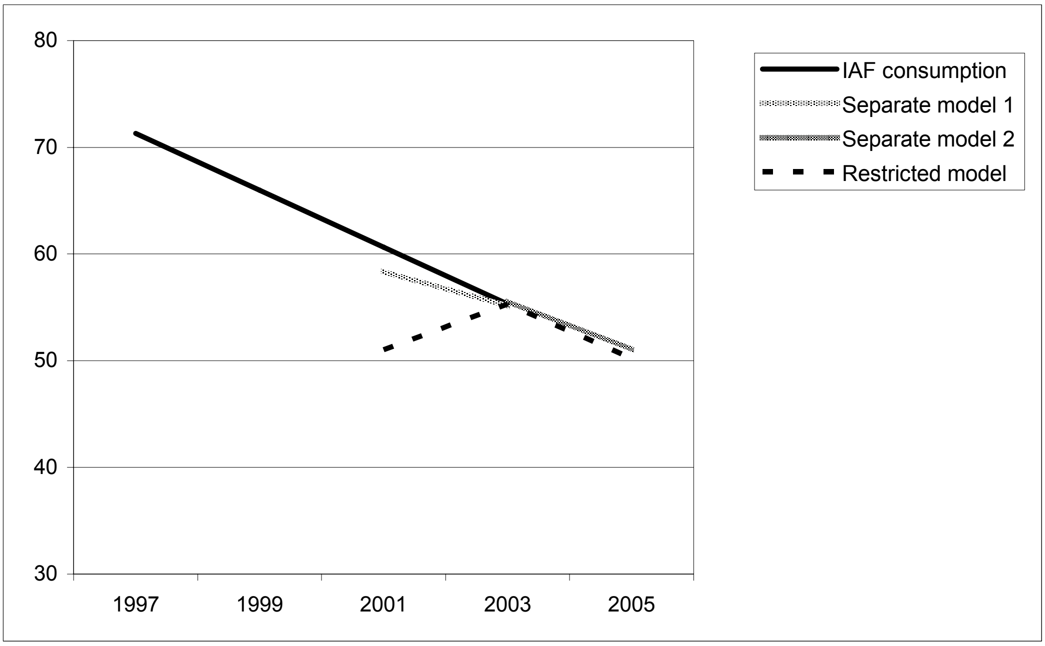


Figure 3: Poverty headcount in rural Mozambique; comparing different models



The estimated development in poverty differs somewhat between urban and rural areas, but the general impression of the relationship between the models seems to be the same as for the national model.

Appendix

1. Methodological appendix

In this section, we present the results of the mathematical derivations of the bias and the standard error of the predictor. The reader may wish to consult Mathiassen (2005) for further details, as well as Green (2003) and Wooldridge (2002) for a presentation of the econometrics used.

It can be shown that an unbiased predictor for predicting the headcount ratio is given by:

(6) $$\hat{P} = \frac{1}{n}\sum_{i \in S} s_i \Phi\left(\frac{\ln z - X_i \hat{\beta}}{\hat{\sigma}\sqrt{\tau_i^2 + 1}}\right)$$

where:

(7) $$\tau_i^2 = \operatorname{var}\left(\frac{X_i \hat{\beta}}{\sigma} \mid X_i\right) = X_i\left(\tilde{X}'\tilde{X}\right)^{-1} X_i'$$

and $\tilde{X}$ is the matrix of poverty indicators obtained from the budget survey given by $\tilde{X}' = \left(\tilde{X}_1', \tilde{X}_2', ..., \tilde{X}_n'\right)$, and X is the matrix given by $X' = \left(X_1', X_2', ..., X_n'\right)$.

Let w_i denote the sampling weight for household i. The predictor is then given by:

(8) $$\hat{P} = \frac{1}{\sum_i w_i}\sum_{i \in S} w_i s_i \Phi\left(\frac{\ln z - X_i \hat{\beta}}{\hat{\sigma}\sqrt{\tau_i^2 + 1}}\right).$$

It can be shown that the variance of the error in (5) can be written as follows:

(9) $$\operatorname{var}\left(\frac{1}{N}\sum_{i \in \Omega} s_i y_i - \frac{1}{n}\sum_{i \in S} s_i \hat{P}_i\right) = \\ \left(\frac{1}{N}\right)^2 \sum_{i \in \Omega} s_i^2\left(P_i - P_i^2\right) + \operatorname{var}\left(\frac{1}{N}\sum_{i \in \Omega} s_i\left(P_i - \hat{P}_i\right)\right) + \left(1 - \frac{n^H}{N^H}\right)\frac{n^H}{n^2} E \operatorname{var}\left(s_i \hat{P}_i \middle| \hat{\beta}\right)$$

Here N^H denotes the number of households in the target population.

In this expression, we have assumed simple random sampling. We can, however, allow for other sampling designs by adjusting the last term of the right-hand side of (9), and we will shortly return to how this should be done.

We can use Monte Carlo simulations to estimate the variance given in (9). It can be shown that we can generate random draws and compute a predictor as follows. Let:

(10) $$D_{ij} = \Phi\left(\frac{\ln z - X_i\beta}{\sigma} - \tau_i\eta_{ij}\right), \qquad \overline{D}_{i.} = \frac{1}{M}\sum_{j=1}^{M} D_{ij}, \qquad \overline{D}_{j.} = \frac{1}{n^H}\sum_{i\in S} s_i D_{ij}$$

where η_{ij}, j = 1,2,…M, is i.i.d. random draws from N(0,1). τ_i is given in (7). Here, D_{ij} is analogue to $\hat{P}_i$ in (5) and corresponds to the jth random draw of the stochastic error term. In other words, for each household with the given characteristics, X_i, we generate M independent probabilities of being poor. We use the average over these M simulated probabilities of being poor, $\overline{D}_{i.}$, as an estimator for $P_{i.}$ when computing the variance. By generating random draws, we are able to produce an estimate for the variance of the predictor, even though we initially only had one observation for each individual.

By means of $\{D_{ij}\}$, one can simulate:

$$\frac{1}{N}\sum_{i\in\Omega} s_i^2\left(P_i - P_i^2\right) \quad \text{by} \quad \frac{1}{n}\sum_{i\in S} s_i^2\left(\overline{D}_{i.} - \overline{D}_{i.}^2\right)$$

$$\operatorname{var}\left(\frac{1}{N}\sum_{i\in\Omega} s_i\left(P_i - \hat{P}_i\right)\right) \quad \text{by} \quad \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{n}\sum_{i\in S} s_i\left(\overline{D}_{i.} - D_{ij}\right)\right)^2$$

and:

$$E\operatorname{var}\left(s_i\hat{P}_i \middle| \hat{\beta}\right) \quad \text{by} \quad \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{n^H}\sum_{i\in S}\left(s_i D_{ij} - \frac{1}{n^H}\sum_{i\in S} s_i D_{ij}\right)^2\right).$$

Thus, total variance of the prediction error can be simulated by:

(11) $$\begin{aligned} &\frac{1}{N}\frac{1}{n}\sum_{i\in S} s_i^2\left(\overline{D}_{i.} - \overline{D}_{i.}^2\right) + \\ &\frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{n}\sum_{i\in S} s_i\left(\overline{D}_{i.} - D_{ij}\right)\right)^2 + \\ &\left(1 - \frac{n^H}{N^H}\right)\frac{n^H}{n^2}\frac{1}{M}\frac{1}{n^H}\sum_{j=1}^{M}\sum_{i\in S}\left(s_i D_{ij} - \frac{1}{n^H}\sum_{i\in S} s_i D_{ij}\right)^2 \quad . \end{aligned}$$

In the first term, equation (11), because of the idiosyncratic component, we replace the expected poverty level for each individual with the mean predicted probability of being poor generated by the random draws. We use the variation within sample n^H as a proxy for the variation within the population. The second term, because of uncertainty in the estimated model parameters, is the variance of the mean error in prediction. Because we only have predictions for the sample and not the entire population, we use the mean error in the subsample n^H as a proxy to calculate this variance. We calculate the mean prediction in the sample for each random draw and use these to calculate an empirical variance. The third term, because of sampling, is the expected variance of the predictor given the estimated parameters. It is computed by calculating the empirical variance of the predictor in the sample and over the random draw. The latter takes care of the fact that it is an estimate for the expected variance. In the case where we do not have a simple random sample frame, the third term of (11) can be estimated by using the syntax for estimating sampling variances as given in the packages, for example, SPSS, SAS or STATA. In this case, we specify Dij as the variable for which we want to calculate the sampling errors and the strata, clusters and household weights as given by the survey.

2. List of poverty indicators

Literacy:

- All adults illiterate
- Some adults illiterate
- One adult illiterate
- No adult illiterate
- Number of illiterate adults in household
- Head illiterate

Education:

- Education of highest educated female member
- Education of highest educated household member
- Education of highest educated male member

Employment:

- Head employed in primary sector
- Head employed in secondary sector
- Head employed in tertiary sector
- If head not employed

Assets:

- Simple additive asset index
- Simple additive expensive asset index
- Beds per person
- Bicycles per person
- Mobiles per person
- Radios per person
- Household owns air conditioner
- Household owns bed
- Household owns bicycle

- Household owns car
- Household owns oven
- Household owns computer
- Household owns electric iron
- Household owns fan
- Household owns freezer
- Household owns fridge
- Household owns hi-fi set
- Household owns mobile phone
- Household owns motorcycle
- Household owns printer
- Household owns radio
- Household owns sewing machine
- Household owns telephone
- Household owns TV
- Household owns wall watch
- Household owns washing machine

Energy, water and sanitation:

- Type of energy used for cooking
- Type of energy used for lighting
- Type of water source
- Type of toilet

Housing:

- Type of roof
- Type of toilet
- Type of walls

Demographic composition:

- Demographic dependency ratio
- Number of members in household
- Number of members younger than 15 years
- Number of persons 65 years or older
- Number of adults in household
- Number of disabled in household
- Number of daughters of head or spouse in household
- Number of sons of head or spouse in household
- Number of children of head or spouse in household

- Number of spouses in household
- Number of non-relatives in household
- Number of non-close relatives in household
- Number of heads and spouses in household
- Age of household head
- If head is divorced/separated
- If male head
- If head is married
- If head never married
- If head is widowed
- One or two generations with children younger than 15
- One or two generations with no children younger than 15
- Three generations or complex
- Single person
- Single parent with children younger than 15
- Single parent with adult sons/daughters
- Couple with children younger than 15
- Couple with adult sons/daughters
- Couple
- Extended family (outside core)

So-called poverty predictors:

- Acquired agricultural tools or inputs last 3 months
- Acquired building materials last month
- Acquired building materials last 3 months
- Acquired clothes or shoes last month
- Acquired clothes or shoes last 3 months
- Acquired domestic utensils last 3 months
- Acquired furniture last month
- Acquired furniture last 3 months
- Acquired soap last month
- Consumed bread last week
- Consumed eggs last week
- Consumed maize flour last week
- Consumed meat last week
- Consumed milk products last week
- Consumed cooking oil last week

- Consumed rice last week
- Consumed seafood last week
- Consumed sweet potato last week
- If no meals yesterday
- If one meal yesterday
- If two meals yesterday
- If three meals yesterday
- If paid for transport last month
- If usually use detergent for washing clothes
- If any household members contracted laborers last season
- If any household members did occasional agricultural work last season
- If household owns poultry
- Rooms per capita

3. Tables

Table 3 Number of cases by survey and region

Region	IAF02/03	QUIBB00/01	IFTRAB04/05
North, rural	1494	2855	2697
North, urban	816	838	1888
Central, rural	1924	3639	3535
Central, urban	1176	1216	2853
South, rural	1277	2470	2339
South, urban	1090	1427	2350
Maputo	923	1345	1699
All	8700	13790	17361

Table 4 Adjusted R-squared for the models

	Unrestricted set of indicators		Restricted set of indicators
	Quibb model	Iftrab model	
All	0.60	0.56	0.53
Urban	0.70	0.67	0.64
Rural	0.53	0.48	0.46
Rural North	0.59	0.51	0.50
Urban North	0.62	0.62	0.59
Rural Central	0.44	0.39	0.35
Urban Central	0.69	0.62	0.59
Rural South	0.57	0.50	0.49
Urban South	0.65	0.61	0.58
Maputo City	0.83	0.81	0.77

References

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003): Micro Level Estimation of Poverty and Inequality, *Econometrica*, Vol. 71, No. 1. pp.355-364.

Greene, W. (2003): *Econometric Analysis,* Prentice Hall, Englewood Cliffs.

Hentschel, J., Lanjouw, J.O., Lanjouw, P. and Poggi, J. (2000): Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador, *The World Bank Economic Review*, Vol. 14, No 1., pp.195-218.

Mathiassen, A. (2005): A Statistical Model for Simple, Fast and Reliable Measurement of Poverty, Discussion Paper No. 415, Statistics Norway.

National Directorate of Planning and Budget, Economic Research Bureau, International Food Policy Research Institute and Purdue University (2004): *Poverty and Well-being in Mozambique: The Second National Assessment*.

Simler, K.R., Harrower, S. and Massingarela, C. (2004): "Estimating poverty indices from simple surveys". http://www.csae.ox.ac.uk/conferences/2004-GPRaHDiA/papers/3p-SimlerMassingerela-CSAE2004.pdf. Accessed 12/01/2006.

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Massachusetts.

Recent publications in the series Documents

2004/18 T. Karlsen, D. Quang Pham and T. Skjerpen: Seasonal adjustment and smoothing of manufacturing investments series from the quartely Norwegian national accounts

2005/1 V. Skirbekk: The Impact of a Lower School Leaving Age and a Later Retirement on the Financing of the Norwegian Public Pension System.

2005/2 H. Utne: The Population and Housing Censushandbook 2001.

2005/3 J. L.Hass and R. Straumann: Environmental Protection Expenditure: Methodological work for the Oil and Gas Extraction Industry. Report to Eurostat.

2005/4 L. Hobbelstad Simpson: National Accounts Supply and Use Tables (SUT) in Constant Prices SNA-NT "SUT/CONSTANT"

2005/5 L. Hobbelstad Simpson: National Accounts Supply and Use Tables (SUT) in Current Prices. SNA-NT "SUT/STARTER"

2005/6 S. Todsen: SNA-NT User's Guide for Supply and Use Tables in Current and Constant Prices.

2005/7 E. Ugreninov, T.M. Normann and A. Andersen: Intermediate Quality Report EU-SILC 2003 Statistics Norway.

2005/8 H.V. Sæbø: Metadata strategy in Statistics Norway. Eurostat Metadata Working Group Luxembourg, 6-7 June 2005.

2005/9 J.L. Hass, K.Ø. Sørensen , K. Erlandsen and T. Smith: Norwegian Economic and Environment Accounts (NOREEA). Project Report 2002.

2005/10 A. Benedictow and T. Harding: Modeling Norwegian balances of financial capital.

2005/11 A.L. Mathiassen,J.B Musoke, P.Opio and P. Schøning: Energy and Poverty A feasibility study on statistics on access and use of energy in Uganda.

2005/12 E. Vinju, R. Strauman, Ø. Skullerud, J. Hass and B. K Frøyen: Statistics on pre-treatment of waste. Pilot study - Norway 2004. Report to Eurostat

2005/13 H. Skullerud, Ø. Skullerud and S. Homstvedt: Pilot study: Treatment of Hazardous Waste. Final report to Eurostat.

2005/14 H. Skiri, B. Strand, M. Talka and H. Brunborg: Selected Documents on the modernisation of the Civil Registration System in Albania Vol. II.

2006/1 O. Andersen og M. Macura: Evaluation of theproject "Modernisation of the Civil Registration System in Albania"

2006/2 T. Åvistland: The problem with a risk premium in a non-stochastic CGE model.

2006/3 Å Cappelen, R. Choudhury and T. Harding: A small macroeconomic model for Malawi.

2006/4 J. Ramm og A. Sundvoll: Translating and Testing the European Health Status Module in Norway, 2005.

2006/5 A.G. Hustoft og J. Linnerud: Statistical Metadata in Statistics Norway.

2006/6 H.V. Sæbø: Systematic Quality Work in Official Statistics - Theory and Practice

2006/7 H. Skullerud: Methane emissions from Norwegian landfills Revised calculations for waste landfilled 1945-2004.

2006/8 R. Choudhury: User's Guide for a Macroeconomic Model for Malawi.

2006/9 J. Dias Loureiro, B. K. Wold and R. Harris: Compendium from Workshop on Light Core Surveys for Policy Monitoring of National PRSPs and MDGs in Maputo, December 2005

2006/10 K. Gimming, O.M. Jacobsen and P. Løkkevik: Changes in inventories in the Norwegian National Accounts. Eurostat project report

2006/11 B. Bye, T-R. Heggedal , T.Fæhn and B. Strøm: A CGE model of induced technological change: A detailed model description

2006/12 S.E. Stave: Water Consumption in Food Processing and the Service Industries in Norway

2006/14 T. Harding and H.O. Aa Solheim: Documentation of a model on the household portfolio

2006/16 O. Ormanidhi and A. Raknerud: Productivity growth in the Norwegian manufacturing sector 1993 – 2002. A comparison of two methods

2007/1 E. J. Fløttum: Differences between SNA93 and ESA95

2007/3 T. Skjerpen: Matrix oriented formulae of dynamic multipliers in general autoregressive distributed lag and VARX models

2007/4 A. Mathiassen and D. Roll-Hansen: Predicting poverty for Mozambique 2000 to 2005. How robust are the models?