Session Number: Plenary Session 1
Time: Monday, 25 August, AM

Paper Prepared for the 30th General Conference of
The International Association for Research in Income and Wealth

Portoroz, Slovenia, August 24-30, 2008

The predictive ability of poverty models. Empirical Evidence from Uganda

Astrid Mathiassen

For additional information please contact:

Astrid Mathiassen
Statistics Norway, Division for Development Cooperation
PO Box 8131 Dept, NO-0033 Oslo, Norway
E-mail: mss@ssb.no

Abstract
This paper examines the performance of a particular method for predicting poverty. The method is a supplement to the approach of measuring poverty through a fully-fledged household expenditure survey. As most developing countries cannot justify the expenses of frequent household expenditure surveys, low cost methods are of interest, and such models have been developed and used. The basic idea is a model for predicting the proportion of poor households in a population based on estimates from a total consumption regression relation, using data from a household expenditure survey. As a result, the model links the proportion of poor households to the explanatory variables of the consumption relation. These explanatory variables are fast to collect and are easy to measure. Information on the explanatory variables may be collected through annual light surveys. Several applications have shown that this information, together with the poverty model, can produce poverty estimates with confidence intervals of a similar magnitude as the poverty estimates from the household expenditure surveys. There is, however, limited evidence for how well the methods perform in predicting poverty from other surveys.

A series of seven household expenditure surveys conducted in Uganda in the period 1993-2006 are available, allowing us to test the predictive ability of the models. We have tested the poverty models by using data from one survey to predict the proportion of poor households in other surveys, and vice versa. All the models predict similar poverty trends, whereas the respective levels are predicted differently. Although in most cases the predictions are precise, some times they differ significantly from the poverty level estimated from the survey directly. A long time span between surveys may explain some of these cases, as do large and sudden changes in poverty.

Acknowledgments
I am grateful to Geir Øvensen for discussions and advice during the work process and to John Dagsvik, Bjørn Wold, Stein Terje Vikan and Ib Thomsen for useful comments and discussions. I am also thankful to Uganda Bureau of Statistics (UBOS) for providing the data and to NORAD for funding this work.

1 Introduction

The widely accepted approach for measuring poverty is through measurement of household expenditure during a twelve month survey. If policymakers demand annual estimates, and register data is not available, a country has to run a continuous household expenditure survey program. However, hardly any country can afford or justify such expenses. A cheaper alternative is to base a household survey program on a twelve month household expenditure survey every fifth or seventh year, and then supplement with light, two-month, surveys on an annual, or bi-annual, basis for the years in between. If the light surveys are combined with a poverty model, such program can give model based poverty estimates with confidence intervals of a similar magnitude as the poverty estimates from the household expenditure surveys. The main test, however, is whether the model predicts the same poverty as new twelve months household budget surveys. Hence, this paper is outlined to examine the predictive ability of a poverty correlates model.

The method applied is developed in Mathiassen (2007). The basic idea of this method is to estimate a consumption model from a household expenditure survey, linking consumption per capita and poverty to variables that are fast to collect and easy to measure. Information on the explanatory variables in the poverty model is then collected through the annual light surveys. This information, together with the estimated model, is used to predict poverty rates and their standard errors for years where there is no household expenditure survey. The method in Mathiassen (2007) as well as related ones, such as Simler, Harrower, and Massingarela (2003), Stifel and Christiansen (2007)[1] and Datt and Jolliffe, (2005), have now been applied in several countries. Empirical evidence shows that the models are able to predict poverty levels well within the sample, with standard errors at similar levels as in the traditional household expenditure survey estimates of poverty. There is, however, limited evidence for how well the methods perform in predicting poverty levels outside the sample[2].

In this paper we want to test how well the model approach performs in predicting poverty over time. The assumption that there is a stable relation between per capita expenditure and

[1] Both are modifications of the Poverty Mapping method, Elbers, Lanjouw and Lanjouw, 2003. The Poverty Mapping method is designed to combine a Census with a household expenditure survey to produce small area estimates of poverty, but can be adapted to combine a light survey with a household expenditure survey.

[2] There are two recent papers that examine the performance of the poverty mapping method, Elbers, Lanjouw and Leite, (2008) and Demombynes, Elbers, Lanjouw and Lanjouw (2007). The results from these analyses indicate that the method produce small area estimates of welfare that are in line with the actual values.

the poverty indicators is critical because one may expect that the parameters change over time, particularly in dynamic economies. In order to test the performance of the model, a comparison between predicted poverty rates and the "actual" poverty rates estimated directly from the household expenditure aggregates are necessary. Thus, to test the assumption on stable model parameters, one needs at least two household expenditure surveys with comparable household expenditure aggregates.

Fortunately, a series of seven household expenditure surveys from Uganda are available, allowing us to test the assumption of stable parameters. The surveys were undertaken in the period 1993-2006, a time period where Uganda experienced strong growth with more than twenty percentage point decrease in poverty. The surveys are well suited to our purpose as the questionnaires and sampling methods have been kept more or less unchanged over the time period in question. Moreover, significant work has been done to ensure comparable household expenditure constructs and poverty estimates from the survey (in Appleton, Emwanu, Kagugube and Muwonge (2001)). It is in fact rather unique that an African country has so many high-quality household expenditure surveys available over such a short time span.

This paper is organized as follows: The next section outlines the methodology applied. In section three we describe the data and the setting. The following sections are concerned with the empirical testing, and the paper concludes with a discussion of the results.

2 The method

In this section, we outline the main features of the methodology for predicting poverty rates. Readers looking for further references should consult Mathiassen (2007).

2.1 A predictor for the poverty 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 wish to predict the headcount ratio, i.e., the proportion of individuals with consumption below a given poverty line[3].

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,

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

i.e. when $Y_i \leq z$, and zero otherwise. The population is Ω and it consists 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. Hence, the share of poor individuals in Ω is estimated by

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

We wish to use a model to predict y for a given set of household variables (indicators). To this end we assume that:

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

where X_i is a vector of selected poverty indicators, β is a vector of unknown parameters and ε_i , *i=1,2,…*, are i.i.d. error terms with unit variance. The parameter σ represents the standard deviation of $\sigma\varepsilon_i$. 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[4].

Given model (2) which has a stochastic component in the estimated consumption level, all individuals have a non-zero probability of being poor. Thus, rather than counting the number of individuals with predicted consumption below the poverty line, we use the average probability that an individual is poor as the predicted estimator for the headcount ratio. 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[5].

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

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

[4] One should test for homoscedasticity in the empirical analyses, and if necessary apply the method allowing for heteroscedasticity as outlined in Mathiassen (2007).

[5] Other distribution function can be applied if it seems more reasonable.

It can be shown that this predictor is biased due to the errors in the estimates $\hat{\beta}$ and $\hat{\sigma}$. Hence, we will use the formula for the unbiased predictor given in (6) in the Appendix[6].

2.2 The standard error of the poverty predictor

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

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

The first term on the right-hand side in (5) is the difference between the actual and expected population poverty levels. This term 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 term 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 sample S. This term captures uncertainty due to S being 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 estimation is described in the Appendix.

[6] When calculating the standard error of the predictor below, it is however the simpler predictor in (4) that is used. This is reasonable as using the biased corrected predictor substantially increases the complexity in the calculations, and because the error caused by using the unbiased predictor is marginal.

3 The household surveys and constructs

3.1 Comparability of the household expenditure surveys

Fortunately, Uganda, as probably the only country in Sub-Sahara Africa, has had a recent large scale household survey program. It began in 1992 and covers eight households expenditure surveys up to the most recent survey in 2005.

To test the models predictive ability it is critical that the consumption aggregates are comparable between the surveys[7] and that there are sufficiently identical indicators (explanatory variables). As the Uganda household surveys rely on similar sampling procedures and questionnaires, and substantial work has been done to ensure comparability (see Appleton et al., 2001), they are suitable for our testing. The 1992 survey, however, differs too much with respect to core indicators, and it is therefore not used in the analyses[8].

Table 1 shows the period and number of households covered in each of the surveys. We will in the following refer to the survey by the year when it started.

Table 1: Survey round, 1993-2006

Survey Round	Dates	Households covered
Monitoring survey 1 (MS-1)	Aug. 1993 – Feb. 1994	4,925
Monitoring survey 2 (MS-2)	Jul. 1994 – Mar. 1995	4,925
Monitoring survey 3 (MS-3)	Sep. 1995 – Jun. 1996	5,515
Monitoring survey 4 (MS-4)	Mar. 1997 – Nov. 1997	6,654
Uganda National Household survey 1 (UNHS-1)	Aug. 1999 – Jul. 2000	10,696
Uganda National Household survey 2 (UNHS-2)	May 2002 – Apr. 2003	9,711
Uganda National Household survey 3 (UNHS-3)	May 2005 – Apr. 2006	7400

Source UBOS (2002, 2006)

As can be seen from the tables, the monitoring surveys from 1993, 1994 1995 and 1997 did not cover an entire calendar year. Because food consumption will vary over the year, the fact that it is recorded only over a short period (one week), without adjustments, implies that the consumption aggregates may be affected by seasonality. Depending on whether it is an above or below annual average season that is covered by the survey, the poverty level may be over- or under estimated, as compared to when the entire calendar year is covered. For the purpose of replicating the estimated poverty figures by using a model, we might get a biased

[7] Modelling approaches have also been used to ensure comparable poverty estimates between incomparable surveys, see Deaton, 2003 and Tarozzi, 2004. In this case one may also use expenditure variables for which the definition and question has not changed between the survey.

[8] In particular the household consumption expenditure on food is based on a 30 days recall period compared to 7 days in the other surveys.

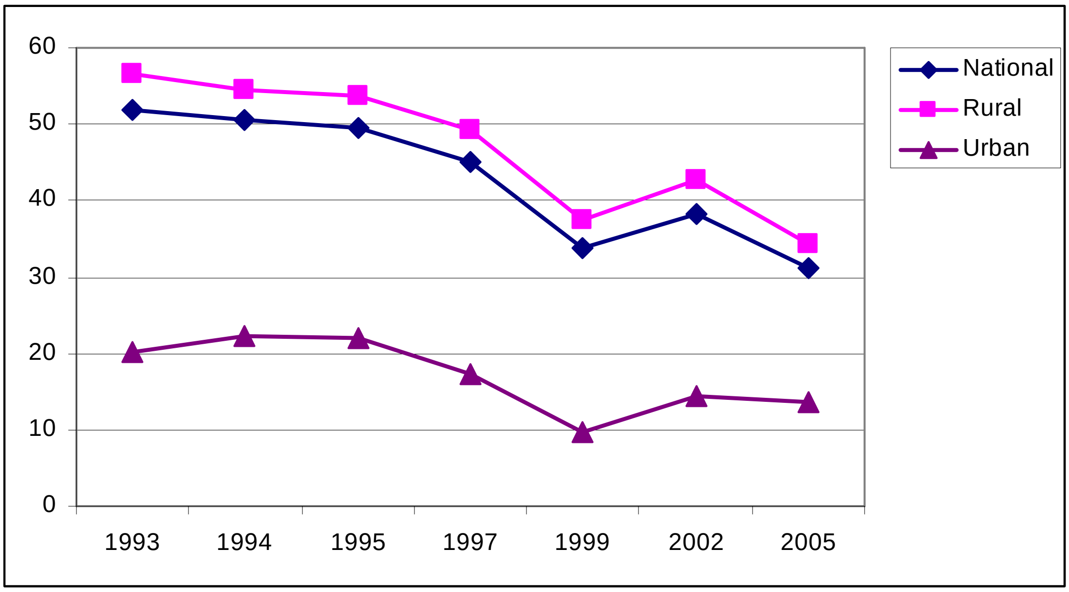
prediction if we do not adjust for the months covered. We will return to this in the empirical section.

The poverty line was computed on the basis of the 1992 survey, and it has remained fixed in real terms up to 2005, allowing for comparison of changes in poverty. It is an absolute poverty line anchored in a minimum required caloric intake; see Appleton et al. (2001) for details on the construction of the poverty line.

3.2 Poverty estimates from the household expenditure surveys

To familiarize the reader with the Ugandan setting we will briefly discuss the trends in the actual[9] poverty levels. Figure 1 shows the national and rural/urban poverty estimated from the surveys in the period[10]. Uganda experienced a substantial decrease in poverty in the period and the national headcount ratio fell from about 52 percent in 1993 to 31 percent in 2005.

Figure 1 Poverty estimated from household expenditure surveys. National and Rural/Urban.



The rural poverty trend follows the national trend closely as the major share of the population lives in these areas (85 percent in 2005, UBOS (2005)). Poverty fell more in rural than in urban areas, both in absolute and relative terms.

[9] We will throughout the paper refer to the poverty level estimated in the traditional way by using the per capita consumption from the household expenditure surveys as the actual poverty level.

[10] Due to security problems parts of districts in Northern and Western region were excluded in some of the surveys (Bundibugyo, Kasese, Gulu, Kitgum and Pader). Thus, for comparability these districts are also excluded from the other surveys.

Some papers have addressed the poverty trends up to 2002 (see Kappel, Lay and Steiner, (2004), Obwona, Okidi and Ssewanyana (2006) and Okidi, Ssewanyana, Bategeka and Muhumuza, (2005)), and the brief summary of the key factors behind the development in poverty given below, refers to these papers. As the 2005 survey is still fresh there is as yet (to the author's knowledge) no publication addressing the further development in poverty.

The decrease in national poverty from 1992 to 1999, took the form of a period with strong growth in average per capita consumption, and only small changes in the income distribution (at least up to 1997) (see Table 2 on Gini indices). From 1999 to 2002 growth slowed down, and at the same time inequality widened, resulting in an increase in poverty. From 2002-2005 poverty reduction was again on the right track, due to reduction in poverty in rural areas.

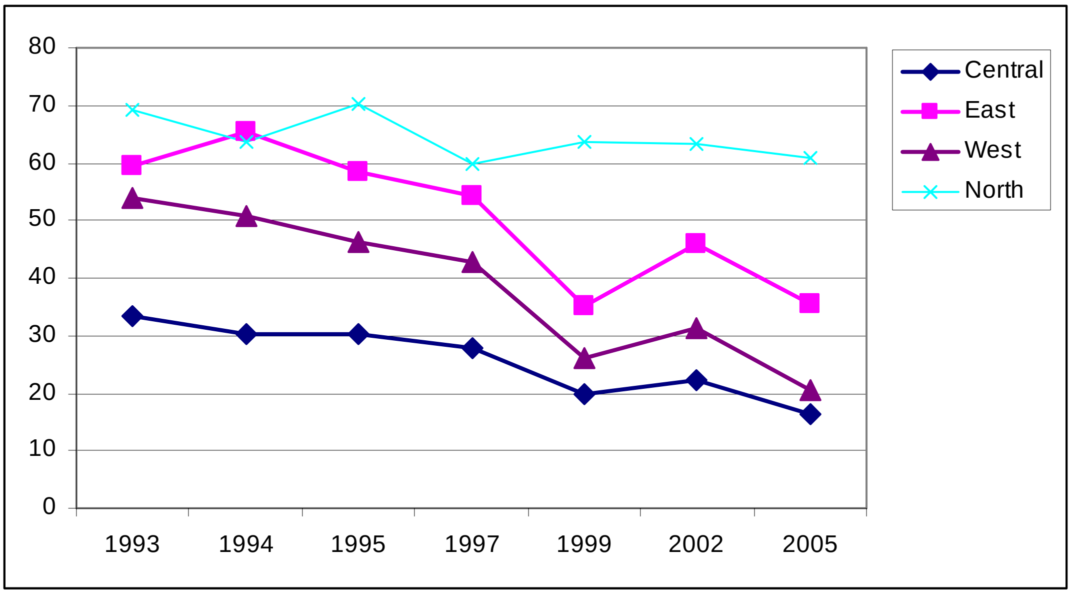
Agriculture, being the most important sector for the rural poor, has played an important role in the poverty reduction in Uganda in the 1990s. The reduction in poverty up to 1999 has been associated with a structural shift towards increased cash crops production. Trade liberalization in the early 90s and increased world market price on coffee, the main cash crop in Uganda, are important explanations for this shift. Diversification into non-agricultural activities with high growth rates further benefited many poor households. From 1999 till 2002, however, poverty among subsistence farmers increased considerably. The price of coffee started to fall after 1994, and in 2001 reached a level of only ten percent of that of 1994. Also prices on other important cash crops; cotton, tobacco and tea fell in this period, which together with slow growth in the subsistent agriculture are likely to be key factors behind the increase in poverty. UBOS (2006) suggests the recovery of coffee prices[11] to be important for the improvements of poverty from 2002 to 2005.

Figure 2 shows that similar poverty trends are found in all regions, except in the Northern region. The war-affected Northern region continues to suffer from long-term instability. The poverty level was initially high and has only slightly decreased from about 70 to 60 percent in the period. The Central region has the lowest poverty level both at the beginning and the end of the period, and faced a reduction in poverty of about 15 percentage points. This is the region with the highest level of urbanization. The Western part of Uganda has experienced the largest improvements in poverty with a decrease at nearly 35 percentage points. East, the

[11] from \$0.56/kg in 2002/03 to \$1.38/kg in 2005/06

second poorest region, has also seen large improvements with a reduction in poverty at nearly 25 percentage points.

Figure 2 Regional Poverty in Uganda



3.3 The explanatory variables

Since the aim is to test whether a model could be used in years where no household expenditure survey is available, we focus on the types of indicators that are feasible to include in a light survey. That is indicators that are easy to measure and collect. It is important that the indicators are exactly identical in the two surveys. Thus, it is both necessary that the question is identical; that the method for getting the information is the same (for example recall versus diary), and that the reference period is the same. In the case of the Uganda dataset, the recall method was always used to obtain the information about food consumption and the report period for consumption variables were the same in all questionnaires.

We identify the common set of indicators in pairs of household surveys and select among these a smaller set of indicators by comparing estimated models with various combinations of poverty indicators, including square and log-linear transformation of variables. Based on statistical criteria, we choose the set of indicators that constitute the "best" model for predicting the poverty headcount ratio[12]. Since the poverty indicators normally are collected in light surveys there should not be too many indicators, although a sufficient number of

[12] One may do this through automated stepwise procedures.

variables should be selected to ensure that the marginal gain of including additional variables is low. Typically, around twenty indicators enter into a model.

In the case of Uganda the following groups of indicators were included;

- Demography: Indicators comprising variables like dependency ratio, number of members in household, marital status, and age distribution among members in household.
- Education: Indicators comprising education level and literacy.
- Labour market: Indicators capturing type of work done by head defined by industry and whether head is employed or self-employed[13].
- Housing: for example type of roof and lightening. In the available data material for 1994 and 1995 housing variables are not available.
- Consumption of food: Binary "yes/no" variables on household consumption of food like for example meat, sugar and rice the last seven days.
- Consumption of non-durables: "yes/no" variables on consumption of non-durables like for example energy, transport and bathing soap the last thirty days.
- Consumption of semi-durables: "yes/no" variables on household consumption on semi-durables like for example clothes, shoes and furniture the last year.
- Welfare indicators: for example number of meals per week and ownership of shoes.
- Regional control: Dummies capturing regional differences in the aggregated (national and rural/urban) models.
- Seasonal adjustment: Dummies accounting for seasonal variability in food consumption. Good periods will normally be after harvest while food will be scarcer prior to harvest[14]. These control variables can only be included when we estimate models based on surveys covering the whole calendar year, thus not for 1993, 1994, 1995[15] and 1997.
- Community variables: Indicators accounting for location specific effects, such as access to markets, infrastructure and availability of electricity may be included. Due to various problems in data, we have only been able to link up the community section for two household surveys (2002 and 2005).

The explanatory variables are the driving forces behind the prediction results. They are correlated with the household consumption, and one should expect them to change with changes in the poverty level. For example, lower poverty levels are associated with higher education levels, better housing standards and a larger share of population consuming the non-inferior goods. As people may switch to more expensive calories as they become wealthier, one could also expect that consumption of some, inferior, goods decrease.

---

[13] In the 2005-survey the definitions of the labour market variables are slightly different compared to the other surveys and could therefore not be used in our analyses.

[14] "There are two agricultural seasons in a year in most part of the country. The first one is between January and June, while the second one lasts from July to December. Harvesting of the crop planted in one season usually extends to another season. For example the crop planted during the first season may be harvested in July-September. On the other hand, the crop planted during the second season may be harvested in January-March of another agricultural year. Agriculture in Uganda is rain-fed which determines whether to plant early or late depending on when the rain starts. This in turn dictates the extent to which harvesting will be pushed into the following season." UBOS (2004). We have added one month to the harvest time as one may expect food to be relatively plentiful shortly after harvest., and accordingly divided the calendar year into the following four periods; January-April, May-June, July-October, and November –December. Thus, we expect the first and third period listed to be better than the two others.

[15] For 1995-survey we lack information on date of interview.

The types of variables have different “roles” in the model. Demographic and education variables tend to change slowly, and will thus reflect long-term improvements. Consumption variables, to the contrary, are able to reflect sudden changes, or shocks to the household, for example due to a household head loosing his work and poor weather affecting the harvest. Housing variables are in-between, and change as conditions improve, but are normally not able to capture short term fluctuations in income.

One potentially important group of variables, assets, was not available for the analyses, because households were asked about the total value of a large group of assets rather than availability and value of specified assets. Thus, we were neither able to include indicators for single assets[16], nor to construct an asset index. This is unfortunate as changes in the asset stocks may capture coping strategies for the poor as they may sell off assets in bad times, while building up the stock in good times.

Thus, reasons that one survey model do not predict well for another survey, given that the modelling assumptions are reasonable, may be that important variables are omitted from the model and/or that some model parameters have changed. The parameters could change as a population become wealthier, for example at a certain level of welfare in a country whether people eat meat or not are not a result of income but rather whether one is a vegetarian. Also, one can expect that the parameters change as new food varieties and technologies are introduced leading to shifts in the demand curves.

4 Empirical Results

After identifying the joint set of variables in two surveys, we estimate a consumption model for one of the surveys. The model is then used to predict consumption per adult equivalent and thus the poverty level in the other survey. We use a t-test to compare model-predicted poverty to the actual poverty estimates. It is not feasible to include the estimation results for all models here. Therefore, we will rather discuss some general findings, before moving on to the prediction results.

[16] Except for ownership of bicycle included in the three last surveys (1999, 2002 and 2005).

4.1 Some general modelling results

The estimated models

We estimate models for urban and rural areas separately, as the underlying economic structures in these domains may differ substantially. R-square adjusted is about 0.6-0.7 for the urban models and about 0.5-0.6 for the rural models. The models were inspected for heteroscedasticity by visual interpretation of plots of the residual versus predicted expenditure per capita as well as by formal tests [17]. Because the pattern looks reasonably random, we did not correct for heteroscedasticity[18]. After inspecting the distribution of the residuals, we chose to apply the normal distribution function for estimating the poverty predictor.
In general, the models predicts precisely within sample. The within sample prediction for some of the models are shown in the Appendix tables.

Indicators from each group of the explanatory variables; demography; education; labour; housing; consumption of food; consumption of non-durables; consumption of semi-durables and welfare indicators, entered into almost every model. Community indicators, from two surveys were available, but were not selected in any of these survey-models. This may have to do with too little variation in these, or that we do not have the relevant community variables at hand. Seasonal adjustment did not seem to have an effect. It could be that seasonality is captured by other variables (for example the binary variables for food consumption).

Some surveys did not cover the entire calendar year. It would thus have been formally correct to limit the samples to the joint field work months covered by pairs of surveys. For the sake of presentation we have used the full sample in all cases[19]. We have, however, in some cases predicted poverty also when adjusting the sample size to correct for differences in coverage, but it does not seem to be important. The three last surveys cover the entire year and thus, no adjustments of the samples are necessary in comparisons among these surveys.

Standard errors of the predictions

---

[17] Formal tests (White test and Breusch Pagan) reject the assumption on constant variance of the error term (homoscedasticity) for most of the models. These tests are, however, sensitive to the number of observations as with a large number of observations a small deviation leads to rejection of the hypothesis. Thus, when we use a smaller randomly drawn sample (down to about 1000 observations for some models) the hypothesis on constant variance is no longer rejected.

[18] We have tested the impact on the prediction and the standard error of adjusting for heteroscedasticity in some of the models, but this has very little impact on the prediction results. We will return to this below.

[19] If the sample sizes for a given survey is reduced, the actual poverty estimate that is compared will change as well, and thus the presentation of the results will be rather messy.

All standard errors incorporate the two-stage sampling design, and for the survey poverty estimates, sampling is the only source for the standard error. For the model predictions, however, the standard error comprises three components[20]. At these sample sizes the largest share is due to the estimated parameters. As could be expected, total standard errors are larger for the model predictions than for the survey estimates. However, the sub-component of the model standard error due to sampling is smaller than the sampling error of the survey estimate[21]. For some predictions at the sub-regional level, the standard errors of the model-based estimates are actually lower than of the actual ones. Thus, if one accepts the confidence interval of the survey based estimates, there is no need to reject the model based predictions due to their confidence interval. However, the critical question whether the model is valid from one survey to another remains, because it is not reflected in the magnitude of the standard errors. We use t-tests for the change in the poverty level to judge whether the model prediction and the survey based estimates are statistically different.

4.2 Predicted poverty trends

For each survey, we estimate models that are used to predict rural and urban poverty for other surveys. This is repeated for all pairs of surveys, which yields the seven predicted poverty trends in . . For example, the solid line labelled the Rural 93-model shows the predictions made by models from 1993 onto each other survey from 1994 to 2005. It also includes the actual poverty level in 1993[22]. The thick lines show the actual poverty trends for rural and urban domain in the same period.

A first glance, . shows that the models are able to predict the poverty level quite well. No rural model predicts poverty at the urban level and vice versa. Although poverty predictions for urban areas, seem to fit better than for rural areas, the urban poverty level is much lower, so deviations are not smaller in relative terms. The predicted poverty trends for urban areas, however, follow more closely the actual trend than in rural areas.

[20] Uncertainty due to sampling of the indicators; uncertainty due to the estimated model parameters; as well as uncertainty due to an idiosyncratic component (which will be relatively small as we predict for large domains). See also equation (5).

[21] This is because information about the dependent variable is a priori given by the model, and for a given level of sampling uncertainty one needs fewer observations when using a model compared to when one estimates poverty by using the consumption aggregates directly.

[22] We will refer to a model estimated on data for a given survey by the survey year. For example the 93-model refers to a model estimated on data for the 1993-survey.

Figure 3 Rural and urban poverty trends for Uganda, actual and predicted by seven models

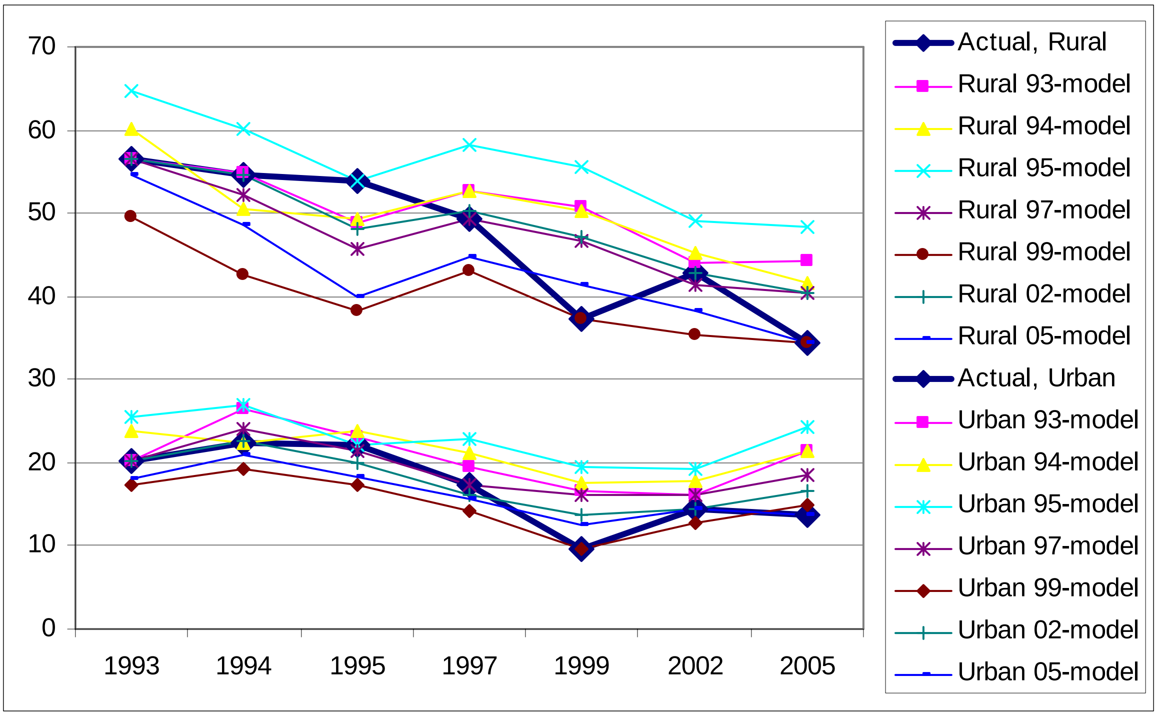


Figure 4 and Figure 5 show rural and urban predictions, respectively. We observe that all rural models are able to capture the decline in poverty over the period[23]. Even though none of the models predicts the entire 25 percentage points fall in actual poverty, the predicted fall in poverty is near to 20 percent for most survey-models, and lowest for the 1999 survey-model predicting a 15 percentage points fall in poverty from the beginning to the end of the period.

[23] Poverty fell from about 57 to 34 percent in the period

Figure 4 Rural poverty, actual and predicted by seven models

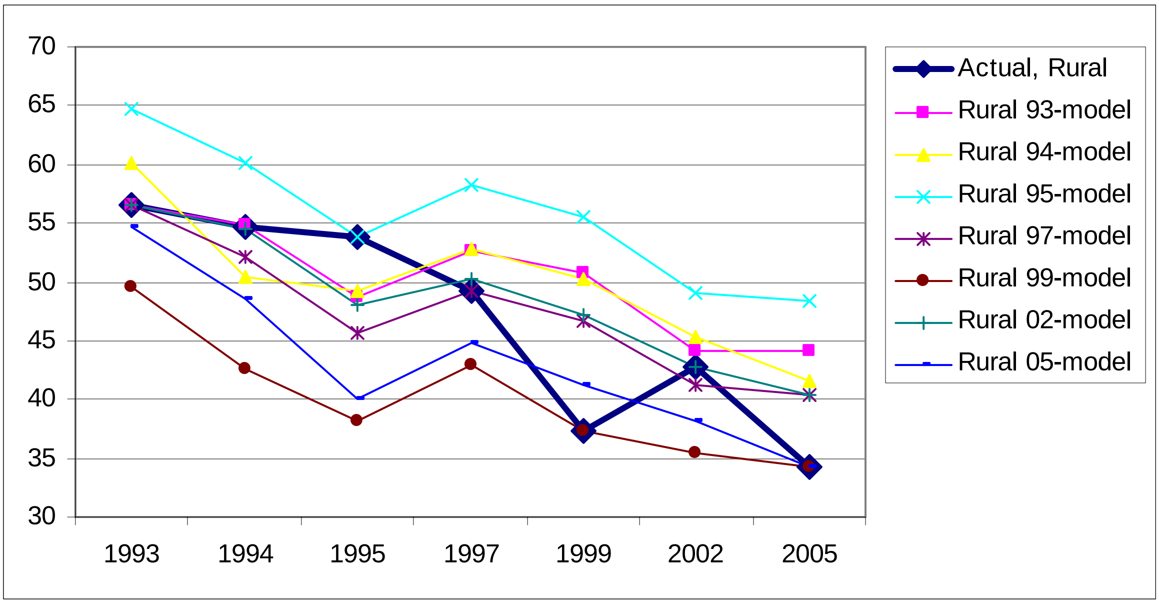
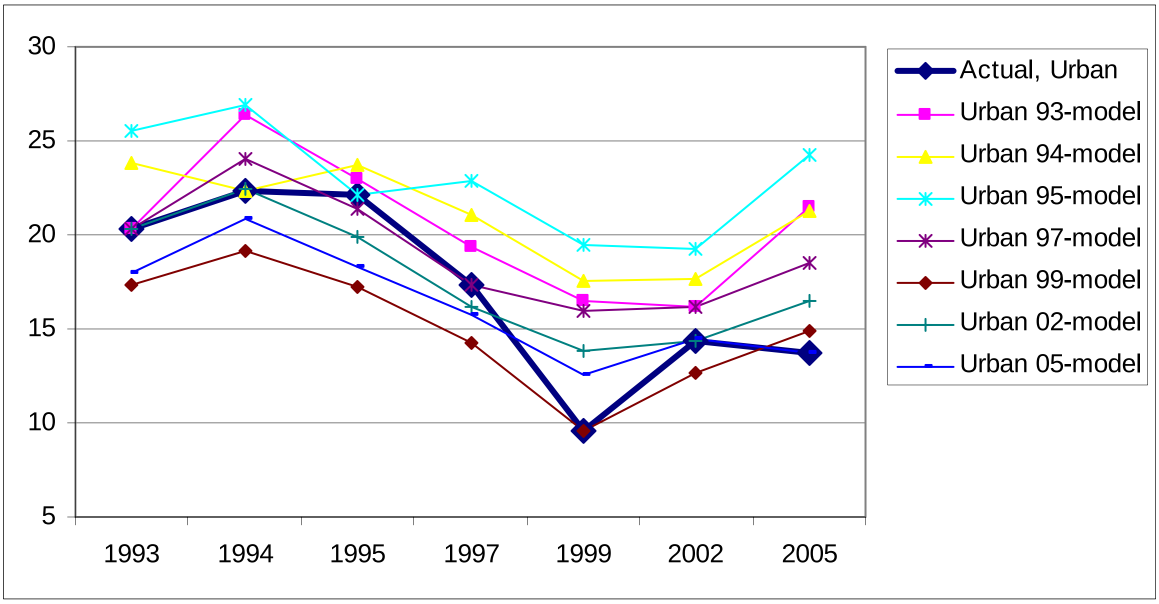


Figure 5 Urban poverty, actual and predicted by seven models



The predicted poverty trends are fairly similar for each model. Thus, focusing on changes in poverty over the period, all survey-models give nearly the same result. This suggests that the relation between the consumption aggregates and the set of explanatory variables are consistent in predicting poverty. The bias may be due to omitted variables or conditions that are more important for some years than other.

Some models estimated from surveys with relatively low poverty, like 1999 and 2005, tend to predict lower poverty for a given year compared to predictions from models based on surveys with higher poverty levels. This feature is particularly visible for urban models. For the rural domain the models estimated on data for 1993, 1994, 1997 and 2002 produce fairly similar predictions over the entire period, even though the actual poverty levels in these surveys differ substantially.

Predictions for 2005 seem to indicate that the time which has elapsed from the model base survey to the prediction is important. All surveys predict too high poverty level for 2005, and older surveys tend to predict farther off from the actual prediction for 2005 than newer ones. Time, however, is not important when predicting for 2002. For example models estimated on data for 1993 predicts as well for 2002 as the models estimated on data for 2005. Rather, the combination of time elapsed and large fall in poverty level may be factors contributing to break down of the models.

Even though the rural models capture the overall trend of decreasing poverty, they do not capture the variability within the overall trend. In particular, none of the models are able to capture the strong fall in actual poverty from 1997-1999 with the following increase to 2002. The large deviations from the actual values are mainly due to predictions made by and for the surveys in 1995 and 1999. While the 1999 model produces the lowest predicted poverty levels for all the surveys, the 1995 survey predicts a substantially higher poverty level than the other models. Correspondingly, all models predict too low poverty level for 1995. Also the 1999-survey is problematic when it comes to predicting poverty levels for the urban areas.

Neither adjustment for seasonality nor accountancy for heteroscedasticity improved on the bad predictions for, and by, 1995 and 1999. Reducing the samples in 1999, 2002 and 2005 to allow for the correct comparison with 1995 does not improve on the results[24]. Similarly, reducing the sample in 1999 to account for the difference in monthly coverage compared to 1993 and 1997 does not improve on the results[25]. The degree of heteroscedasticity seems to be so small that it hardly has any impact on the predictions and the standard error in the case

[24] July and August, normally two good months in terms of food availability, was not included in the 1995 survey. Taking out these months from the samples in 1999, 2002 and 2005 to allow for the correct comparison, give only minor changes in the difference between the prediction and estimates: predicting rural and urban poverty for 1995, reduce the difference in predicted and actual poverty rates by about a half percentage points or less in the six cases tested.

[25] We have not done any more analyses to account for differences in monthly coverage in two surveys, as it does not seem to play any role for our results.

tested[26]. Figure 6 shows that when the problematic 1995 and 1999 surveys are taken out, the predictions are more in line with the actual poverty trend.

Figure 6 Poverty trends, predicted and actual for 5 surveys

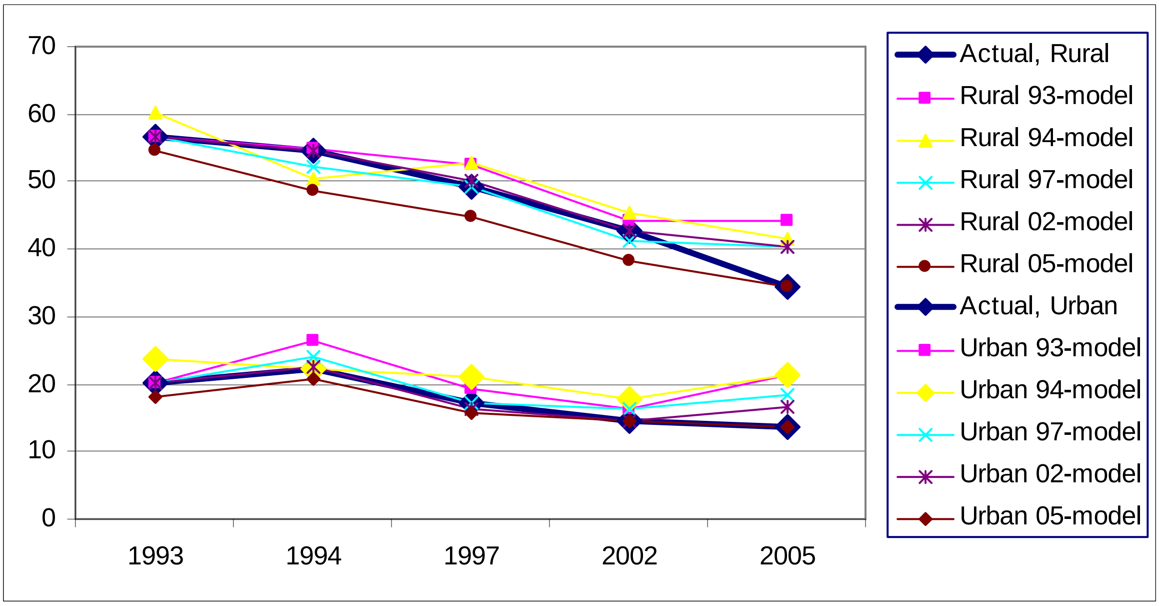


Table 3 shows t-values for the test on difference in model-based and actual rural poverty levels. 22 of the 42 predictions are unfortunately significantly different, but most of these have to do with predicting for, or by, 1995 and 1999. Also, it seems to be rather problematic to predict for 2005, and it is only the 1999-model that predict non-significantly different poverty rate in 2005. All other predictions made by and for the 1993, 1994, 1997 and 2002 surveys are not significantly different than the actual ones.

The urban models seem to do better (Table 4). Seven out of the 42 predictions are significantly different than the actual estimate. This applies to when using models from 1993, 1994 and 1995 to predict for 1999 and 2005, as well as when using the 1997 model to predict for 1999. No other predictions differ significantly from the actual value.

So why is it that the modelling approach seems to work for most of the surveys but not for all? By gathering more evidence at the sub-regional level we will explore further to which

[26] The predictions by the 1999-models for the rural and urban domain in 2002 are respectively 35,0 (1,6) and 11,4 (1,6) when applying the heteroscedasticity model compared to 35,4 (1,6) and 12,7 (1,8) with no correction for heteroscedasticity (standard errors in parenthesis). The predictions by the 1995-models for the rural and urban domain in 1997 are respectively 58,4 (2,8) and 22,7 (2,5) when applying the heteroscedastic model compared to 58,2 (2,7) and 22,9 (2,6) with no correction for heteroscedasticity. And, finally the predictions by the 2002-models for the rural and urban domain in 2005 are respectively 41,1 (2,0) and 18,0 (2,2) when applying the heteroscedastic model compared to 40,4 (1,9) and 16,5 (2,2) with no correction for heteroscedasticity.

extent this can be explained by the time elapsed between surveys, and/or large changes in poverty levels. And, in particular explore whether the models are able to capture sudden changes in the poverty level. Further, one may suspect that the reason that a model fit well at high level of aggregation is that errors at sub-regional neutralize each other. Alternatively, one could also assume that low prediction capability is caused by off predictions for only one sub-regional rather than for all.

4.3 Sub-regional level predictions

Is the time which had elapsed important?

From Figure 3 we have seen some indication that time elapsed between surveys are important as early models had lower predictive power for 2005 than for 2002. The 1993-model predicts almost perfectly for 2002, while neither the predictions at urban nor rural domain are significant when applying the 1993-model for 2005. This picture is confirmed by the predictions at the sub-regional level: the 1993 model reproduce the actual poverty estimates very well for 2002, while the predictions significantly differ for most sub-regions in 2005[27] (Table 5 and Table 6). Note also that the sub-regional predictions confirm the earlier finding that the models tend to produce rather conservative estimates, underestimating the changes in poverty.

Predictions at the sub-regional level for 2005 by the 1997- and 2002-models give some further support to the “time” hypothesis. The reason that the 2002-model predicts significantly higher rural poverty rate than the actual one in 2005, is due to one single prediction (Table 7). The problem is to predict for Rural West, which also is the rural region that had the largest fall in poverty between the two surveys[28]. Thus, from the sub-regional predictions we can conclude that the 2002-model do not predict too badly for 2005. The sub-regional figures, when using the 1997-model to predict for 2005, show the same pattern (Table 8). The problem is again to predict for the Western region. The differences in the predicted values and the actual one are, however, larger when using the 1997-models compared to using the 2002-models. This pertains to also the urban models.

---

[27] Except for Northern region for both domains as well as for urban Central (Table 6). The Northern region is the region with lowest fall in poverty.

[28] When we re-run the rural model and predict for all rural regions except West, we find that the prediction are no longer significantly different from the actual one. The predicted poverty level for the three remaining regions is 42,4 percent as compared to the actual poverty level at 39,3 percent.

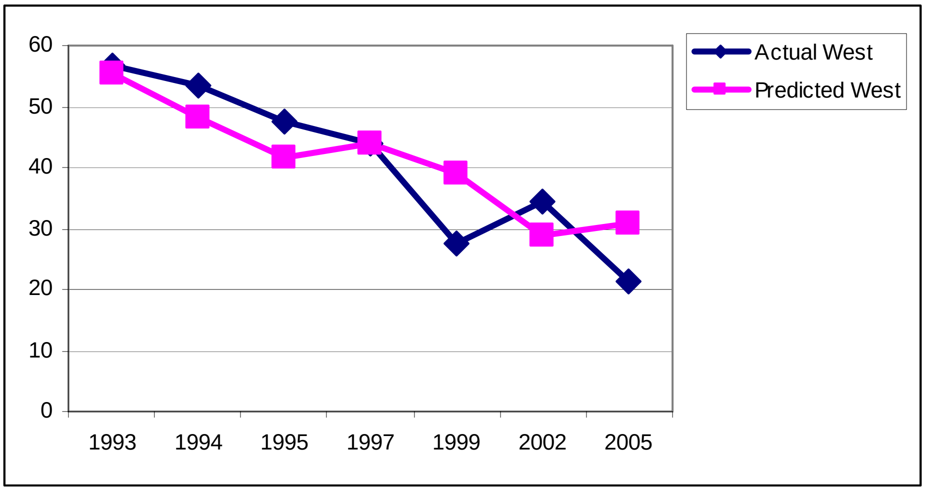
Thus, the sub-regional predictions for 2005 give some indications that time elapsed between surveys are important as 2002-models perform better than 1997-models which again perform better than the 1993-models in predicting poverty for 2005. The 1993-models do not; however, seem to be outdated in the same way when predicting for 2002. One reason may be that there is a substantially higher fall in rural poverty in 2005 compared to 2002, but this does not hold for urban areas for which the poverty level is about the same in 2002 and 2005. One may, however, expect that structural changes in the consumption pattern are introduced gradually, and that there is some time lag before it becomes visible.

Is it more difficult to predict large changes in poverty?

The above results suggest that a more important factor for explaining when the model "does not work" seems to be large and/or sudden changes in the poverty level. We will explore this further by looking at predicted poverty trends in the rural regions. We have used the 1997-survey as the base for estimating the models, because it is in the middle, and because it performs well in predicting for most surveys.

West was the most dynamic region, in the sense that it had the largest fall in poverty. In rural areas the actual poverty level fell by 34 percentage points from 1993 to 2005, while the model predicts about 25 percentage points decrease in poverty (Figure 7). The predictions are significantly different from the actual ones in the survey years 1999 and 2005 that had the most dramatic fall in poverty[29] (Table 9).

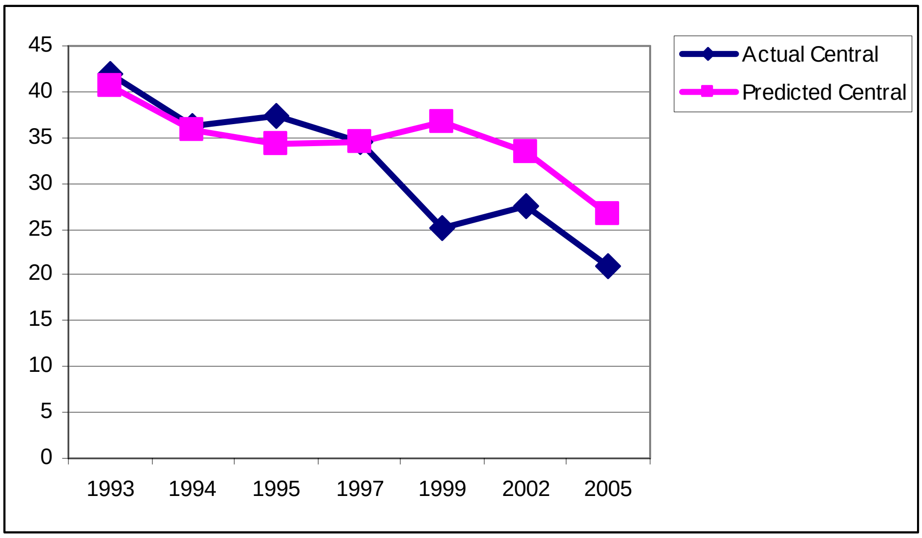
Figure 7 Poverty trend for Rural West, actual and estimated by 1997 model.



[29] 1999 and 2005, showed respectively 17 and 11 percentage point reduction in poverty from the previous survey.

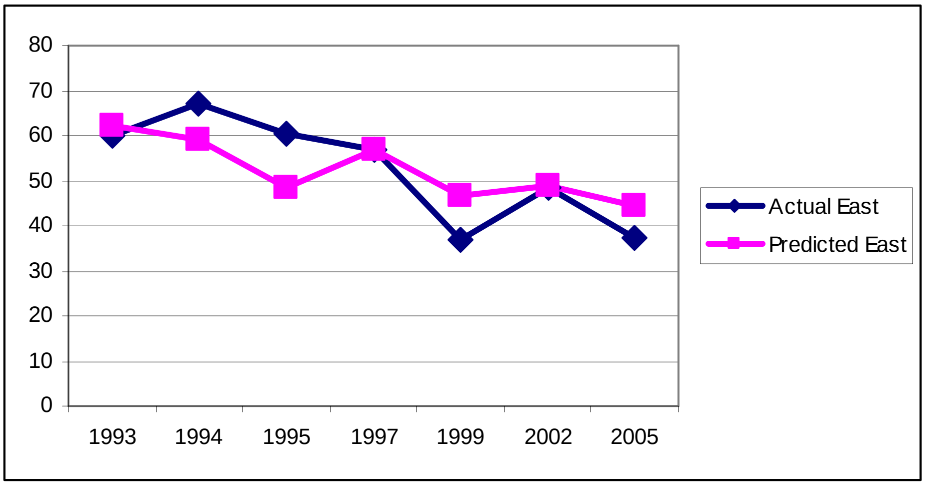
The model is not able to capture the fall in poverty level in Central regions at 10 percentage points from 1997 to 1999 (Figure 8 and Table 9). In fact, the model predicts increased poverty between these two surveys.

Figure 8 Poverty trend for Rural Central, actual and estimated by 1997 model.



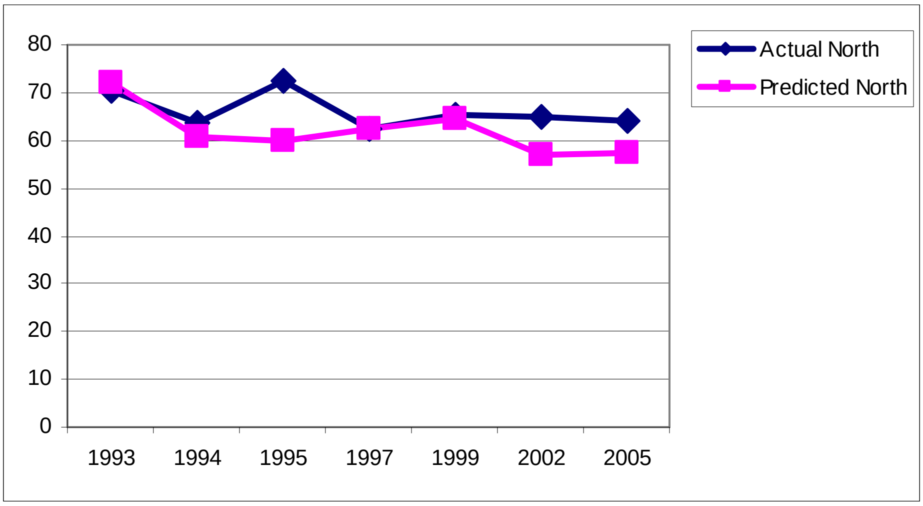
In the Eastern region poverty fell with more than 20 percentage points over the period. Here the model predicts far too low poverty in 1995, and it was not able to predict the almost 20 percentage points fall in poverty from 1997 to 1999 (Figure 9 and Table 9).

Figure 9 Poverty trend for Rural East, actual and estimated by 1997 model.



Finally, for the north the model works fine except for 1995. There was a substantial increase in poverty this year which is not captured by the 1997-model (Figure 10 and Table 9).

Figure 10 Poverty trend for Rural North, actual and estimated by 1997 model.



Overall, the predicted poverty trends for the rural regional are to a large extent in line with the actual poverty trends. Most cases when it fails coincide with large falls, at more than 10 percentage points, in poverty compared to the previous survey. Worryingly, in some of these cases we unexpectedly predict an increase in poverty. Also the 1995 survey is problematic in a couple of the regions, which do not coincide with large changes in poverty levels.

Thus, adding evidence from more disaggregated levels suggest that the elapsed time may be an important factor when the models does not predict well. Further, the models have trouble capturing sudden and large changes in the poverty level.

4.4 The variables behind the predictions

In this section we show that the poor performance of the 1995 and 1999 surveys may be attributed to some unexpected development in the explanatory variables. Examining the trends in the aggregates of these variables give a crude picture, but still provide some useful insight. Table 10 and Table 11 show the weighted average of the selected explanatory variables for each survey, for rural and urban areas. We have selected variables to represent each groups included in the model; demography, education, employment, housing, and from each group we have tried to report on indicators that typically enter into a model[30].

[30] As the questions concerning welfare indicators changed between the surveys, identical welfare variables were only identified in two or three surveys. Thus, although this group of variables is important in the models, we cannot produce trends of these variables and they are not included in the tables.

The strong fall in poverty from 1997 to 1999, followed by the increase in poverty in 2002 and a subsequent fall in poverty in 2005 are not reflected in the variables in the consumption model, in particular for the rural domain. The developments of the housing variables reflect a steady improvement. Even indicators which are able to change fast, like consumption variables, do not reflect the marked dip in poverty in 1999. Rather, the figures in Table 10 indicate that there has been a steady improvement, and 2002 seems to be a relatively good year in terms of the share of the population that consumed food, non- and semi durables goods like meat and bathing soap. This holds not only for the variables included in the table, but for all consumption variables in our dataset. This finding is also confirmed at the sub-regional level. From 1999 to 2002 the growth in mean consumption per capita was negative, and only among the upper 20 percentile was there an increase in per capita consumption, Kappel et al. (2005). The variables in Table 10, however, are not variables that one associate with the wealthiest, and thus the improvements in these are not in accordance with what one could expect from the fall in consumption per capita in this period.

Also for 1995, the other problematic survey, the share of the population that consumed the various goods is considerably higher than the previous and subsequent surveys, 1993, 1994, and 1997. This pertains to all food indicators available for the analyses, not only the ones included in the tables. This holds also for all non-durables, and for all semi-durables[31], in both 1993 and 1994. The same picture is found when examining the rural regions separately. The high level on the consumption variables relative to the poverty level in 1995 might explain why the 1995-models predict too high poverty levels for the other surveys and why the other survey models predict too low poverty levels for 1995. As the 1995 dataset, and thus the models to predict for and by 1995, does not include housing variables, the consumption variables may be assigned a higher weight than in models including housing variables.

Thus, the low poverty level in 1999 does not coincide with “high scores” on the poverty correlates, and the “high scores” on the poverty correlates in 1995 do not correspond to the poverty level in 1995. However, further investigation of the surveys to examine whether other types of indicators developed in the same manner seems necessary. In particular, it would be relevant to follow amount consumed, as a supplement to the share that consumed the goods,

[31] Except furniture.

as well as the value of the asset stocks[32]. Looking at the value of assets could have an additional purpose, namely to see whether there is some indication that the households obtained a high consumption level in 1995 by depleting their asset stocks, and vice versa in 1999.

5 Conclusion

In this paper we have tested the predictive ability of poverty models relative to poverty figures estimated directly from consumption aggregates. Altogether, we have had seven comparable household expenditure surveys from Uganda from 1993 to 2006 at our disposal. Using poverty models calculated from each of these surveys in turn, we have been cross-testing the models onto the other surveys.

In most cases this simple modelling approach produces predictions at rural/urban and sub-regional levels that are in line with the poverty levels estimated from surveys in the traditional way. The method is as good at sub-regional as at aggregate levels, and there is no tendency that good predictions at an aggregate level hide poor predictions at the sub-regional level. On the contrary, bad predictions at aggregate levels are sometimes due to a single bad prediction in one region.

The poverty trends predicted by the seven models estimated from each of the surveys are consistent. The difference in the ability to predict poverty stems from differences in levels predicted, while the models, independent of which surveys it was based on, predicts approximately similar changes in poverty level over time. The model predictions carry forward their “base poverty” level: A model based on a survey with low poverty tends to predict lower poverty than a model based on a survey with high poverty.

Some times, however, the model approach gives significantly different poverty estimates than poverty obtained directly from the survey. Such cases may be attributed to the long elapsed time between surveys. Further, the model approach does not work well with sudden and large changes in poverty. One hypothesis may be that the model parameters are able to capture most of the change, but the faster the change is, the more is explained by other factors.

[32] Unfortunately, we do not have all the information required for such analyses.

However, when scrutinising the data we find that more importantly for poor predictive ability seems to be some divergence in the data. The two surveys that are problematic seem to be at variance with the other surveys with respect to the development in the explanatory variables. This finding may suggest that the bad prediction is a result of survey issues.

Thus, even though the overall testing results are encouraging, one will not be able to evaluate the results when predicting poverty for a new, light-survey. If one had only one of the Uganda expenditure surveys at hand to serve as the base for the model, one could risk that it was one of the problematic ones, producing significantly different predictions compared to the actual level. The good news is that all models tend to predict the same changes in poverty level, and one can expect a similar predicted trend in poverty independent from which of the seven surveys were available. If two expenditure surveys are available, and there is less than ten years between each survey and the new light-survey, one should use the average of the predictions made by the two survey models. It is also important to keep in mind that this is a second-best, low-cost solution in years when no household expenditure survey is available, and one should be willing to update the poverty predictions when a new expenditure survey is available. This could be done by predicting backwards using the new survey and combining the current and the previous predictions offsetting the differences in the level predicted by each survey separately.

Finally, one should try to improve the models as much as possible. This could be done by including additional important variables in the model. In the case of Uganda, for example, that would mean to include questions in the surveys on number and type of assets. Elber and al. (2008) and Demombynes et al. (2007) find that the ability of the small area estimates approach to reproduce that actual welfare indices depends on whether locality-level explanatory variables are included or not. Even though one should not expect these variables to be as critical when predicting at aggregate levels, one should aim to identify good community variables, and if possible combine the survey data with information on climatic and location specific issues from other sources.

References

Appleton, S, Emwanu, T, Kagugube, J. and Muwonge, J. “Changes in poverty and inequality", chapter 4 of P. Collier and R. Reinnikka (eds.) “Uganda’s Recovery: the role of Farms, Firms and Government” pp.83-121, World Bank: Washington DC, 2001

Appleton, S and Ssewanyana. “Poverty estimates from the Uganda National Household Survey II, 2002/03”, 2003

Datt, G. and Jolliffe, S. ”Poverty in Egypt. Modelling and Policy Simulations” *Economic Development and Cultural Changes* pp. 557-572, 2005

Deaton, A. “Adjusted Indian Poverty Estimates for 1999-2000” Economic and Political Weekly, January 25, 2003

Demombynes, G., Elbers, C., Lanjouw, J.O. and Lanjouw, P.: “How Good a Map? Putting Small Area Estimation to the Test” *World Bank Policy Research Working Paper* 4155, 2007

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

Elbers, C., Lanjouw, P. and Leite P.G. “Brazil within Brazil: Testing the Poverty Map Methodology in Minas Gerais” *Policy Research Working Paper* 4513, The world Bank, 2007

Kappel, R., Lay, J. and Steiner, S. “The Missing Link – Uganda’s Economic Reforms and Pro-Poor Growth” Paper Presented on the International Conference on African Development and Poverty Reduction: The Macro-Micro Linkage, 2004

Mathiassen, A.”A model based approach for predicting annual poverty rates without expenditure data” *Journal of Economic Inequality*, DOI 10.1007/s10888-007-9059-7, 2007

Obwona, M., Okidi, J.A. and Ssewanyana, S. “The growth-inequality-poverty nexus in Sub-Saharan Africa: Evidence from Uganda’s micro level data” Paper presented at the AfDB/AERC Workshop on Accelerating Africa’s Development Five years into the Twenty First Century, Tunisia 22-24, 2006.

Okidi, J.A. and McKay, A. “Poverty Dynamics in Uganda: 1992 to 2000”, *CPRC Working Paper* No 27, Chronic Poverty Research Centre, 2003

Okidi, J.A., Ssewanyana, S., Bategeka, L. and Muhumuza, F. “Distributional and Poverty Impacts of Uganda’s Growth: 1999 to 2003”, *Research Series* No. 46, Economic Policy Research Centre, Kampala, 2005

Simler, K., Harrower, S. and Massingarela, C. "Estimating Poverty Indices from Simple Indicator Surveys." Mimeo. International Food Policy Research Institute: Washington, D.C., 2003

Stifel and Christiansen "Tracking Poverty Over Time in the Absence of Comparable Consumption Data" *The World Bank Economic Review*, Vol. 21 No. 2 pp. 317-341, 2007

UBOS: "Uganda National Household Survey 2005/06. Report on the Socio-Economic Module" Uganda Bureau of Statistics, 2006

UBOS "Report on the results of the pilot permanent agricultural statistics system (PASS) 2004, in five districts" Uganda Bureau of Statistics, 2005.

Appendix

Methodological appendix

In this section, we present 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) $$\tau_i^2 = \mathrm{var}\left(\frac{X_i\hat{\beta}}{\sigma} \mid X_i\right) = X_i\left(\tilde{X}'\tilde{X}\right)^{-1}X_i'$$

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

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.

One can use Monte Carlo simulations to estimate the variance given in (9). It can be shown that one 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 j[th] 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:

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

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

and:

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

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 the 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 sub-sample 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, one specifies Dij as the variable for which one wants to calculate the sampling errors and the strata, clusters and household weights as given by the survey.

Table Appendix

Table 2 Gini Indices, Uganda

	National	Rural	Urban
1993	0.345	0.296	0.365
1994	0.365	0.320	0.396
1995	0.366	0.325	0.373
1997	0.347	0.311	0.345
1999	0.395	0.332	0.426
2002	0.428	0.363	0.483
2005	0.408	0.363	0.432

Table 3 t-values for difference in actual and predicted poverty level. Rural

	1993-model	1994-model	1995-model	1997-model	1999-model	2002-model	2005-model
1993		-0,9	-2,1	0,0	2,1	0,0	0,6
1994	-0,1		-1,8	0,9	4,7	0,0	2,2
1995	1,5	1,4		2,6	6,4	2,1	4,5
1997	-1,1	-1,2	-3,0		2,6	-0,4	1,7
1999	-4,4	-4,3	-6,4	-3,3		-3,8	-1,5
2002	-0,4	-0,9	-2,1	0,5	3,7		1,9
2005	-3,2	-2,5	-5,0	-2,2	0,0	-2,8	

Table 4 t-values for difference in actual and predicted poverty level. Urban

	1993-model	1994-model	1995-model	1997-model	1999-model	2002-model	2005-model
1993		-0,6	-0,9	0,0	0,6	0,0	0,4
1994	-1,2		-1,3	-0,6	1,0	-0,1	0,4
1995	-0,2	-0,4		0,2	1,4	-0,3	1,1
1997	-0,6	-1,2	-1,8		1,1	0,4	0,6
1999	-2,4	-2,8	-3,4	-2,3		-1,6	-1,1
2002	-0,7	-1,3	-1,9	-0,8	0,8		0,0
2005	-2,5	-2,5	-3,6	-1,7	-0,4	-1,0	

Table 5 1993 models predicting for 2002

				Rural				Urban			
	National	Rural	Urban	Central	East	West	North	Central	East	West	North
Actual Poverty 1993	52	56,6	20,3	40,7	62,1	55,4	72,1	12,7	28,8	26,2	41,5
Predicted within sample	53,8	57,7	23,6	41,2	62,3	56,5	73	16,4	31,6	27,4	46,8
Actual Poverty 2002	38,3	42,7	14,4	27,6	48,3	34,3	65	7,9	17,9	18,6	38,9
St.error of actual poverty 2002	1	1,2	1,1	2,2	2	1,9	2,4	1,3	1,9	2,2	4
Predicted Poverty 2002	40	44,1	16,2	28,9	53,8	37,5	67,1	10,8	25,3	18,4	45,3
Sterror of prediction	2,1	2,9	2,4	3,7	4,7	4	3,7	3,2	4,3	3,1	5
t-value	-0,7	-0,4	-0,7	-0,3	-1,1	-0,7	-0,5	-0,8	-1,6	0,1	-1,0

Table 6 1993 models predicting for 2005

				Rural				Urban			
	National	Rural	Urban	Central	East	West	North	Central	East	West	North
Actual Poverty 1993	52	56,6	20,3	40,7	62,1	55,4	72,1	12,7	28,8	26,2	41,5
Predicted within sample	53,8	57,7	23,7	41	62,3	56,4	73,1	16,2	31,6	27,6	46,2
Actual Poverty 2005	31,1	34,3	13,7	20,9	37,5	21,4	64,2	5,5	16,9	9,3	39,7
St.error of actual poverty 2005	0,96	1	1,6	1,8	1,6	1,9	1,9	1,3	2,7	1,2	5,2
Predicted Poverty 2005	38,7	44,2	21,5	29,5	50,4	38,2	66,9	11,5	30,3	17,1	49,9
Sterror of prediction	2,3	2,9	2,7	3,9	5	4,3	3,7	2,8	4,5	2,5	5,5
t-value	-3,0	-3,2	-2,5	-2,0	-2,5	-3,6	-0,6	-1,9	-2,6	-2,8	-1,3

Table 7 2002 models predicting for 2005

	National	Rural	Urban	Rural Central	Rural East	Rural West	Rural North	Urban Central	Urban East	Urban West	Urban North
Actual Poverty 2002	38,3	42,7	14,4	27,6	48,3	34,3	65	7,9	17,9	18,6	38,9
Predicted within sample	40,1	43,2	16,4	29,6	48,1	34,8	65,4	10,4	19,9	20,8	40,8
Actual Poverty 2005	31,1	34,3	13,7	20,9	37,5	21,4	64,2	5,5	16,9	9,3	39,7
St.error of actual poverty 2005	0,96	1	1,6	1,8	1,6	1,9	1,9	1,3	2,7	1,2	5,2
Predicted Poverty 2005	36,2	40,4	16,5	22,7	40,8	35,8	68,1	8,3	22,4	15,4	43,4
Sterror of prediction	1,9	1,9	2,2	2,9	3,8	3,2	3,6	1,9	3,7	2,5	6,1
t-value	-2,4	-2,8	-1,0	-0,5	-0,8	-3,9	-1,0	-1,2	-1,2	-2,2	-0,5

Table 8 1997 models predicting for 2005

	National	Rural	Urban	Rural Central	Rural East	Rural West	Rural North	Urban Central	Urban East	Urban West	Urban North
Actual Poverty 1997	45	49,2	17,3	34,5	56,8	44	62,6	11,8	25,2	19,7	33,3
Predicted within sample	44,5	49	18,6	35,7	56,5	43,4	62,4	13,1	25,9	22,6	37
Actual Poverty 2005	31,1	34,3	13,7	20,9	37,5	21,4	64,2	5,5	16,9	9,3	39,7
St.error of actual poverty 2005	0,96	1	1,6	1,8	1,6	1,9	1,9	1,3	2,7	1,2	5,2
Predicted Poverty 2005	36,4	40,4	18,5	26,7	44,6	30,8	57,5	11	23,3	20,2	42,6
Sterror of prediction	2,2	2,6	2,4	3,5	4	4,2	4,1	2,4	3,4	3,4	5
t-value	-2,2	-2,2	-1,7	-1,5	-1,6	-2,0	1,5	-2,0	-1,5	-3,0	-0,4

Table 9 t-values for the Rural regions when applying models from 1997

	Rural			
	Central	East	West	North
1993	0,2	-0,4	0,2	-0,2
1994	0,1	1,8	1,0	0,5
1995	0,7	2,1	1,1	2,0
1999	-2,9	-2,7	-2,8	0,2
2002	-1,5	-0,2	1,4	1,8
2005	-1,5	-1,6	-2,0	1,5

Table 10 Average, weighted, over selected indicators, Rural

	1993	1994	1995	1997	1999	2002	2005
Poverty level	*57*	*55*	*54*	*49*	*37*	*43*	*34*
Household size	4,8	4,9	5,1	5,1	5,4	5,3	5,3
Education level, head	3,9	3,9	4,1	4,1	4,2	5,0	5,0
Head works in primary sector	0,83	0,85	0,82	0,74	0,78	0,63	
Walls of burnt bricks	0,05			0,06	0,12	0,18	0,28
Floor made of earth	0,92			0,90	0,87	0,86	0,83
Cooking with wood	0,97			0,90	0,95	0,90	0,89
Ate bread last week	0,05	0,06	0,12	0,08	0,06	0,12	0,15
Ate sugar last week	0,42	0,48	0,59	0,58	0,59	0,60	0,58
Ate meat last week	0,33	0,34	0,38	0,28	0,33	0,41	0,38
Ate onions last week	0,40	0,49	0,53	0,45	0,53	0,66	0,65
Bought toothpaste last month	0,09	0,10	0,16	0,20	0,21	0,36	0,39
Bought bathsoap last month	0,09	0,14	0,22	0,26	0,24	0,29	0,32
Bought charcoal last month	0,18	0,30	0,42	0,55	0,56	0,66	0,73
Bought shoes last year	0,33	0,42	0,53	0,51	0,55	0,61	0,69

Table 11 Average, weighted, over selected indicators, Urban

Urban, weighted averge over selected poverty predictors							
	1993	1994	1995	1997	1999	2002	2005
Poverty level	*20*	*22*	*22*	*17*	*10*	*14*	*14*
Household size	4,1	4,3	4,6	4,4	4,4	4,1	4,6
Education level, head	7,9	7,4	7,4	7,6	7,9	8,0	7,9
Head works in primary sector	0,20	0,22	0,24	0,12	0,12	0,10	
Walls of burnt bricks	0,34			0,43	0,57	0,61	0,65
Floor made of earth	0,35			0,30	0,27	0,27	0,30
Cooking with wood	0,31			0,21	0,20	0,21	0,23
Ate bread last week	0,40	0,36	0,45	0,40	0,38	0,41	0,40
Ate sugar last week	0,85	0,85	0,86	0,88	0,88	0,85	0,79
Ate meat last week	0,56	0,56	0,52	0,50	0,52	0,57	0,51
Ate onions last week	0,80	0,79	0,80	0,79	0,81	0,81	0,80
Bought toothpaste last month	0,51	0,57	0,66	0,72	0,74	0,79	0,80
Bought bathsoap last month	0,29	0,35	0,42	0,52	0,47	0,51	0,58
Bought charcoal last month	0,42	0,60	0,57	0,70	0,72	0,74	0,82
Bought shoes last year	0,67	0,64	0,72	0,72	0,82	0,81	0,80