the review of income and wealth

Review of Income and Wealth
Series 59, Number 1, March 2013
DOI: 10.1111/roiw.12007

TESTING PREDICTION PERFORMANCE OF POVERTY MODELS: EMPIRICAL EVIDENCE FROM UGANDA

BY ASTRID MATHIASSEN*

Statistics Norway

This paper examines the performance of a method of predicting poverty rates. Because most developing countries cannot justify the expense of frequent household budget surveys, additional low-cost methods have been developed and used. The prediction method is based on a model linking the proportion of poor households to suitable explanatory variables (consumption proxies). These consumption proxies are variables that can be collected at much lower cost through smaller annual surveys. Several applications have shown that such models can produce poverty estimates with confidence intervals of a similar magnitude to the poverty estimates from the household budget surveys. There is, however, limited evidence of how well the methods perform out-of-sample. A series of seven household budget surveys conducted in Uganda in the period 1993–2005 allows us to test the prediction performance of the model. We test the poverty models by using data from one survey to predict the proportion of poor households in other surveys, and vice versa. The results are encouraging, as all models predict similar poverty trends. Although in most cases the predictions are precise, sometimes they differ significantly from the poverty level estimated from the survey directly.

JEL Codes: C31, C42, C81, D12, D31, dI32

Keywords: household survey, money metric poverty, poverty model, poverty prediction, Uganda

1. INTRODUCTION

A widely accepted approach for assessing poverty is based on the measurement of total household expenditure obtained through surveys. In the absence of register data, a country must conduct yearly comprehensive household budget surveys to obtain poverty estimates on an annual basis. However, few countries can afford such a procedure. A cheaper way to obtain annual poverty estimates is to base a household survey program on budget surveys every fifth or seventh year, and supplement it with "light" surveys that have a limited set of questions and require a short period of fieldwork. If the light surveys are combined with a model that links the poverty rate to the explanatory variables in the light survey, a country can obtain poverty estimates with confidence intervals of a similar magnitude to the poverty estimates from the household budget surveys. In this paper we discuss whether such a method is able to produce sufficiently reliable estimates for the years between the budget surveys.

Note: I would like to thank the Norwegian Agency for Development Cooperation (NORAD) for financial support and the Uganda Bureau of Statistics (UBoS) for providing the data. I am in debt to John K. Dagsvik for advice and suggestions. I thank Bjørn Wold and Geir Øvensen for useful discussions and comments. I would also like to thank two referees of this journal for very helpful reviews.

*Correspondence to: Astrid Mathiassen, Statistics Norway, Pb 8131 Dep, 0033 Oslo, Norway (Astrid.Mathiassen@ssb.no).

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

Methods for estimating poverty rates by using models have increasingly been applied for producing both small area poverty maps and poverty predictions at aggregated levels. The so-called poverty mapping method, introduced by Elbers *et al.* (2003), is designed to combine censuses with household budget surveys to produce small area estimates of poverty. This method has been adapted to combine light surveys with household budget surveys: see the analyses by Simler *et al.* (2003) and Stifel and Christiansen (2007). The basic idea is to first estimate a consumption model from a household budget survey, linking consumption per capita and the poverty rate to variables that are fast to collect and easy to measure. From this consumption relation, it is possible to derive a model for the probability of being poor: that is, the probability that consumption is below a specified threshold. Information about explanatory variables is collected through annual light surveys and is, together with the model, used to predict consumption per capita and poverty rates including standard errors. The properties of the model for predicting poverty applied in this paper have been discussed by Mathiassen (2009). Its main difference from the methods adapted from poverty mapping is that it shows how to obtain an exact expression for the standard error term of the poverty rate, allowing for a simple and transparent estimation procedure of the corresponding confidence intervals.

The method in Mathiassen (2009), and related methods, such as those of Stifel and Christiansen (2007), Datt and Jolliffe (2005), and Simler *et al.* (2003) have now been applied in several countries. Empirical evidence shows that the models are able to predict poverty levels well in the sample, with standard errors at similar levels to those of the household budget survey estimates of poverty. However, there is little evidence of how well the methods perform in predicting poverty levels out-of-sample.[1]

The underlying modeling assumption about a stable relation between consumption and the consumption proxies is critical, because the relationship may in fact change over time. To test the performance of the model, predicted poverty rates are compared with poverty rates estimated directly from consumption per capita in the budget survey. Thus, for comparison, at least two household budget surveys with comparable household consumption aggregates are required.

A series of seven high-quality household budget surveys from Uganda is available. The surveys were undertaken in the period 1993–2005, during which time Uganda experienced strong growth and poverty fell by more than 20 percentage points. The surveys are well suited for our purpose, as the questionnaires and sampling methods have remained more or less unchanged during the time period. Thus we use this series of household budget surveys from Uganda to compare poverty predicted by a model with the poverty estimate obtained using household expenditure directly.

The paper is organized as follows. Section 2 outlines the methodology applied. Section 3 describes the survey data and Section 4 gives a short review of the poverty trends as estimated directly from the surveys. Sections 5 and 6 are

[1]There are two papers that examine the performance of the poverty mapping method: Elbers *et al.* (2007) and Demombynes *et al.* (2007). The results from these analyses indicate that the method produces small area estimates of welfare that are in line with the actual values.

concerned with the empirical strategy and results of the analyses, and Section 7 sums up the main results.

2. The Method

The headcount ratio is predicted by applying a Probit model, as a function of household specific covariates. The underlying Probit model is estimated by using data from a household survey. A closed form expression for the standard error of the variance of the predictors is obtained. The predictor is subsequently extended to allow for household specific parameters (random effects). The methodology is further discussed in the first part of this section. The second part extends the method to account for temporal variation due to shocks that affect all households. This approach can be used when at least two household budget surveys are available.

2.1. *The Poverty Predictor and Standard Error*

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: that is, the proportion of individuals with consumption below a given poverty line.

Let Y_i denote consumption per capita in household i. Consumption typically consists of aggregated consumption of food and non-food items plus user values of consumer durables and housing. Because the unit in the survey is the household, total household consumption is adjusted for the number of members in each household.[2] Let z denote the poverty line and define $y_i = 1$ if the individuals in household i are poor—that is, when $y_i \leq z$—and zero otherwise. The population is denoted by Ω and consists of N^H households. Let s_i be the number of persons in household i and let N be the number of individuals in the population. The share of poor individuals in Ω is estimated by:

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

When a household budget survey is available, the poverty status, y, is identified directly from total household consumption for each of the m^H households in the sample, B. In this case, a predictor for (1) is:

$$\hat{y} = \frac{1}{m}\sum_{i \in B} s_i y_i, \tag{2}$$

where m denotes the number of individuals in the budget survey sample. Thus (2) is the predictor for poverty headcount estimated in the traditional way.

On the basis of the budget survey, it is possible to estimate a consumption model that can be used for predicting consumption per capita for the households

[2]The simplest solution is to adjust for the number of individuals living in the household. Another approach is to adjust for the number of adult equivalents in the household: i.e., apply a system of weights that depends on size of household and age and sex of the individual household members. For simplicity, we refer to Y_i as household consumption per capita.

in a light survey. To this end, we identify a vector of variables X_i for household i (consumption proxies) that are supposed to "explain" consumption Y_i for the household, through the relation:

$$\ln Y_i = a + X_i\beta + \sigma\varepsilon_i, \tag{3}$$

where a is the constant term, β is a vector of unknown parameters, and ε_i ($i = 1, 2, \ldots$) is an i.i.d. error term with unit variance. The parameter σ represents the standard deviation of $\sigma\varepsilon_i$. The log transformation of the consumption variable serves to reduce the usual asymmetry in the distribution of the error term and stabilizes the variance. Thus, rather than collecting detailed information on all consumption items, it is possible to collect information on the small set of consumption proxies, X_i, that can be used to predict poverty.

The model in (3) postulates that log consumption per capita depends on a systematic and a stochastic component. Because of the stochastic component, all individuals have a non-zero probability of being poor.[3] The probability that per capita consumption of household i falls below the poverty line, z, is found by inserting the regression model in a probability function:

$$P_i = P(Y_i < z) = P(\ln Y_i < \ln z) = P(a + X_i\beta + \sigma\varepsilon_i < \ln z) = \Phi\left(\frac{\ln z - a - X_i\beta}{\sigma}\right), \tag{4}$$

where $\Phi()$ denotes the standard cumulative normal distribution function.[4]

Let n denote the number of individuals in the light survey sample, S. Then a predictor for the headcount ratio in (1) is given by:

$$\hat{y} = \frac{1}{n}\sum_{i\in S} s_i \Phi\left(\frac{\ln z - \hat{a} - X_i\hat{\beta}}{\hat{\sigma}}\right), \tag{5}$$

where $\hat{a}$, $\hat{\beta}$, and $\hat{\sigma}$ denote the least squares estimates of the corresponding parameters. 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. It can be shown that the predictor in (5) is biased because of the errors in the estimates $\hat{\beta}$ and $\hat{\sigma}$.[5] A formula for the unbiased predictor is derived in Mathiassen (2009) and given by:

$$\hat{y} = \frac{1}{n}\sum_{i\in S} s_i \Phi\left(\frac{\ln z - \hat{a} - X_i\hat{\beta}}{\sqrt{\hat{\sigma}^2 + \mathrm{var}\left(X_i\hat{\beta}\right)}}\right). \tag{6}$$

In (6), the variance in the denominator is taken with respect to $\hat{\beta}$, conditional on the explanatory variable vector X_i.

[3]For instance, households with per capita consumption far above the poverty line will have a probability of being poor near 0, while those far below the poverty line will have a probability of being poor near 1, and those near the poverty line will have a probability of being poor near 0.5.

[4]A different distribution function can be applied if it seems more reasonable.

[5]Even if we obtain the unbiased estimate $(\hat{\beta}, \hat{\sigma})$ of (β, σ), the estimator for the probability of being poor is biased because it depends on $(\hat{\beta}, \hat{\sigma})$ in a non-linear way.

Mathiassen (2009) derives a closed form expression for the standard error of the prediction:

$$(7)\qquad \frac{1}{N}\sum_{i\in\Omega} s_i y_i - \frac{1}{n}\sum_{i\in S} s_i \hat{P}_i = \left[\frac{1}{N}\sum_{i\in\Omega} s_i y_i - \frac{1}{N}\sum_{i\in\Omega} s_i P_i\right] + \left[\frac{1}{N}\sum_{i\in\Omega} s_i P_i - \frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i\right] + \left[\frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i - \frac{1}{n}\sum_{i\in S} s_i \hat{P}_i\right].$$

The error is broken down into three components: error due to idiosyncrasy, error due to the estimated parameters, and error due to the fact that the light survey is based on a finite sample.

The model can be extended to account for heteroscedasticity owing to individual specific model parameters (see Mathiassen, 2009). This is done by applying a random coefficient model for estimating consumption per capita. If the error term, ε_i, does not have constant variance, our prediction in (5) may be biased. The predictor for the headcount ratio with the random coefficient model is given by:

$$(8)\qquad \hat{y} = \frac{1}{n}\sum_{i\in S} s_i \Phi\left(\frac{\ln z - \hat{a} - X_i\hat{\beta}}{\sqrt{\hat{\sigma}^2 + \sum_k X_{ik}^2 \hat{\omega}_i^2}}\right),$$

where ω_i^2 is the variance due to the individual specific model parameters.[6] From (8), we can see that even in the presence of heteroscedasticity, the effect on the prediction may be insignificant if the individual specific component of the variance is relatively small.

2.2. *Accounting for Prediction Uncertainty due to Temporal Variation in Unobservable Variables*

In what follows, we will discuss how the model can account for variation in latent variables in the predictions, given that we have at least two household budget surveys to hand. Let the consumption model now be expressed as:

$$(9)\qquad \ln Y_{it} = Z_t\gamma + V_{it}\delta + X_{it}\beta + \eta_{it},\ \eta_{it} \sim N(0,\sigma),$$

where Z_t is a vector of unobserved macro variables, representing the effect of weather and prices, etc., and V_{it} is a vector of unobserved micro variables, such as theft, or sickness or death of a household member. The model in (9) is supposed to represent the "ideal" model that incorporates all relevant variables. We can rewrite (9) as in (3), namely:

$$(10)\qquad \ln Y_{it} = a_t + X_{it}\beta + \sigma\varepsilon_{it},$$

where now the intercept and the error term have the interpretation

$$a_t = Z_t\gamma + \bar{V}_t\delta \quad \text{and} \quad \sigma\varepsilon_{it} = (V_{it} - \bar{V}_t)\delta + \eta_{it}.$$

[6]Note that this formula does not account for the bias. This should not pose any problem as the bias tends to be very small.

TABLE 1

Uganda Surveys, 1993–2005

Survey Round	Dates	Households Covered
Monitoring Survey 1 (MS-1)	Aug 1993–Feb 1994	5,040
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	7,400

Source: Muwonge (2006) and UBoS (2006).

From this we realize that the intercept, which is now allowed to be time-dependent, can be interpreted as a function of the latent macro variables and the mean of the latent micro variables at time t.[7]

Consider next out-of-sample predictions of the poverty level in the population. To deal with the unobserved variables represented by $\{a_t\}$, we treat a_t as an independent, identically normally distributed variable. This is of course a rather crude simplified representation, because of the fact that we have very little data on the temporal variation of the intercept. As in Mathiassen (2009), it can be shown that the mean poverty prediction for an individual in a household, i, is then given by:

$$E(\hat{y}_i|X_i) = \Phi\left(\frac{\ln z - X_i\beta - \bar{a}}{\sqrt{\sigma^2 + \operatorname{var} X_i\hat{\beta} + \operatorname{var} a_t}}\right), \tag{11}$$

where the expectation in (10) is taken with respect to the parameter vector β and the random variable a_t, and $\bar{a} = Ea_t$. As we can see from (11), both the average and the variance of the omitted macro variable will affect the point prediction.

3. The Data

Our empirical analyses rely on a series of seven household budget surveys from Uganda in the period 1993–2005. The fact that the Uganda data consists of several household budget surveys within such a short time-span makes it suitable for testing its performance for predicting poverty. However, for this purpose it is critical that the consumption aggregates are comparable between the surveys and that there are sufficiently identical variables to be used as proxies for consumption.

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

The preparation and the comparability of the survey data for welfare and poverty analyses have been discussed in Appleton *et al.* (2001) and in Appleton and Ssewanyana (2003). The authors discuss three aspects of sampling that may be

[7]To account for possible temporal variations in model parameters, a random coefficient formulation could also be applied, similar to the one discussed previously but where the parameters are allowed to vary over time, t, rather than over households, i.

a source of incomparability. The first is geographic coverage (security problems led some Monitoring Surveys to exclude four districts). We have followed Appleton's approach to adjust for this by excluding the four districts from all subsequent calculations.

The second sampling issue is a panel, as the Monitoring Surveys were intended to have a panel element. In practice, however, there was very high attrition of households and the panel element was abandoned in its entirety in the 1995 survey. Appleton *et al.* (2001) have investigated effects of the panel and find that there is little need to adjust for any biases that may have arisen from the attrition and replacement process.

The third sampling issue is seasonality, as the Monitoring Surveys from 1993, 1994, 1995, and 1997 did not involve fieldwork throughout an entire calendar year. Because consumption will vary over the year, the fact that most consumption goods are recorded only over a short period (food is reported for one week) implies that the inflated annual consumption aggregates may be affected by seasonality. Appleton and Ssewanyana (2003) have tested for seasonality and find that it does not seems to pose a problem.[8]

All surveys have similar socioeconomic core modules and the expenditure component was maintained in all surveys. The various surveys include additional modules on labor force, enterprise. and agriculture: see, for example, Muwonge (2006). There are some comparability issues due to smaller differences in the core modules of the questionnaire.[9] Appleton *et al.* (2001) explored the possible effects of such differences, comparing shares of total expenditure in the 1993–99 survey. The idea is that if the share of certain item groups changes markedly between the surveys it might be due to differences in the questionnaires. They find, however, that the composition of expenditure is fairly similar across the surveys.

The poverty line is a Cost of Basic Needs poverty line. It was computed on the basis of the 1992 survey and it has remained fixed in real terms up to 2005 (see Appleton *et al.*, 2001, for details).

The variable of main interest for the analyses in this paper is the consumption aggregate, aiming at measuring current welfare in the household. The consumption aggregate consists of the monetary value of all consumption items reported in the surveys. The items were aggregated into subgroups of similar items and some subgroups were re-evaluated. In particular, items received for free were evaluated and recorded at the current prices; rent for owner-occupied houses was imputed at current market prices; and food consumption from own production as well as free collection/gifts was re-evaluated into market prices. Adjustments for intertemporal and interspatial price variations were made. Finally, the total household consumption was adjusted for the number of individuals in the household by using an adult equivalence scale.

[8]To investigate the effect of seasonality, Appleton and Ssewanyana (2003) pooled the surveys (except for the 2005 survey) together and regressed the log of consumption per adult equivalent on dummy variables for location (e.g., central urban) for month of the year and for the survey.

[9]The Monitoring Surveys (MSs) from 1993 to 1997 have almost identical consumption sections. The more recent surveys have more item codes, but, unlike the MSs, did not print them on the questionnaire, instead leaving blanks for the interviewer to fill in. Also the three last surveys ask about health and education expenditure at an individual level, whereas the MSs ask about such expenses at the household level.

A study has, however, shown that there are some indications that the consumption aggregates are not comparable (see Luoto, 2007). The paper uses a method to assess comparability between consumption aggregate in the 1999 and 2002 surveys by analyzing subgroups of consumption separately.[10] The findings suggest that the subgroup of starches is not comparable between the two surveys and that there may be differences in the re-evaluation of home-produced starches into market prices in the two surveys. Unfortunately, in the datasets to hand we do not have the derived variables for subgroups of consumption (only for aggregated household consumption). Such variables could have been useful in several ways in the analyses: for assessing comparability between surveys and computing additional welfare indicators to be used for discussing trends, as well as assessing the importance of variables omitted on purpose.[11]

4. Poverty in Uganda

To familiarize the reader with the Ugandan setting, we will briefly discuss the trends in official/actual poverty levels.[12] Figure 1 shows the official poverty figures estimated directly from the surveys during the period 1993–2005. From this it can be seen that Uganda experienced a substantial decrease in poverty, with a fall in the national headcount ratio from about 52 percent in 1993 to 31 percent in 2005. Rural poverty follows the national trend closely, as the major share of the population lives in these areas—85 percent in 2005, according to UBoS (2005). Poverty fell more in rural than in urban areas, in both absolute and relative terms.

Agriculture, being the most important sector for the rural poor, played an important role in reducing poverty in Uganda during the 1990s. Trade liberalization in the early 1990s, together with a high world market price for coffee, the main cash crop in Uganda, was an important driving force in the economy (see Deininger and Okidi, 2003; Kappel *et al.*, 2005; Okidi *et al.*, 2007). An important factor for growth has been the strong general performance of the agricultural sector, with a shift toward increased cash crop production[13] (see also Bussolo *et al.*, 2007). Furthermore, diversification into non-agricultural activities, like trade and

[10]Luoto (2007) uses an adaptation of the methodology developed in Elbers *et al.* (2003). The method is similar to the one used in this paper, but it is applied differently. A relationship between total consumption and a single explanatory index representing subgroups of consumption that are restricted to the consumption items that are strictly comparable between the surveys is assessed. This relationship is used to reveal the impact of each subgroup on the prediction by predicting consumption and poverty for the other survey as one systematically aggregates new subgroups to the explanatory index. Luoto finds that, with one exception, all models predict poverty levels for 2002 that are lower or equal to the official numbers for 1999. The one exception is the model including consumption of home-produced starches (sweet potatoes, cassava, maize, sorghum, and millet), which when added to the explanatory index gives a prediction for 2002 that is almost identical to the official number in 2002.

[11]To arrive at such variables we would need to go back to the raw data, which would have to be re-evaluated and cleaned for outliers and adjusted for spatial and temporal price changes in order to arrive at comparable measures. This has not been done, as we have neither sufficient information on all the price indices used nor access to the syntaxes for reproducing the cleaning and the re-evaluation that took place.

[12]We refer to the poverty level estimated in the traditional way (using household consumption obtained directly from the household expenditure surveys) as the actual poverty level.

[13]The main cash crops next to coffee are cotton, tobacco, and tea (see Appleton and Ssewanyana, 2003).

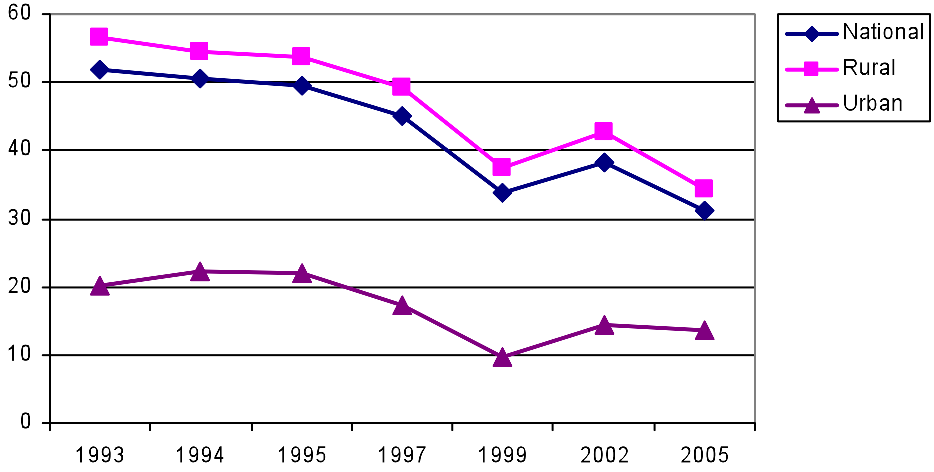

Figure 1. Official Poverty Incidence Estimated from Household Budget Surveys, %
Source: Official figures based on data from the household budget surveys.

transport with high growth rates, benefited many poor households (see Kappel *et al.*, 2005). From 1999 to 2002, poverty increased, mainly because of increased poverty among subsistence farmers (see Appleton and Ssewanyana, 2003). Okidi *et al.* (2007) attributed the increase to a structural stagnation of the economy. From 2002 to 2005 poverty fell, which has been attributed to a return of relative peace in the northern and some parts of the eastern regions and to the recovery in coffee prices (see UBoS, 2006; Ssewanyana and Okidi, 2007).

Figure 2 shows that similar poverty trends are found in all regions except the northern one. This war-affected area continues to suffer from long-term instability and the poverty level has decreased only slightly, from about 70 to 60 percent, during the period. The central region had the lowest poverty level, both at the beginning and at the end of the period, and experienced a reduction in poverty of about 15 percentage points. This is the region with most coffee-producing districts[14] and it has the advantage of including Kampala, the main market. The western part of Uganda has experienced the greatest improvement in poverty, with a decrease of nearly 35 percentage points, and seems to be catching up with the central region. It has plenty of rain, good farming conditions, and several urban marketplaces for its agricultural products. The eastern region, the second poorest, has also seen great improvements, with a reduction in poverty of nearly 25 percentage points. The northern part of this region has been affected by instability and war. The land here is dry, whereas the southern part of the eastern region has better conditions for agricultural production. Detailed poverty maps show that the southern part indeed has much lower poverty levels than the northern part of this region (see UBoS, 2003).

[14]Coffee districts as defined in Bussolo *et al.* (2007): namely Kalangala, Kapchorwa, Kiboga, Luwero, Masaka, Mpigi, Mubende, Mukono, Rakai, Mbale, Kamuli, Iganga, Bushenyi, and Jinja.

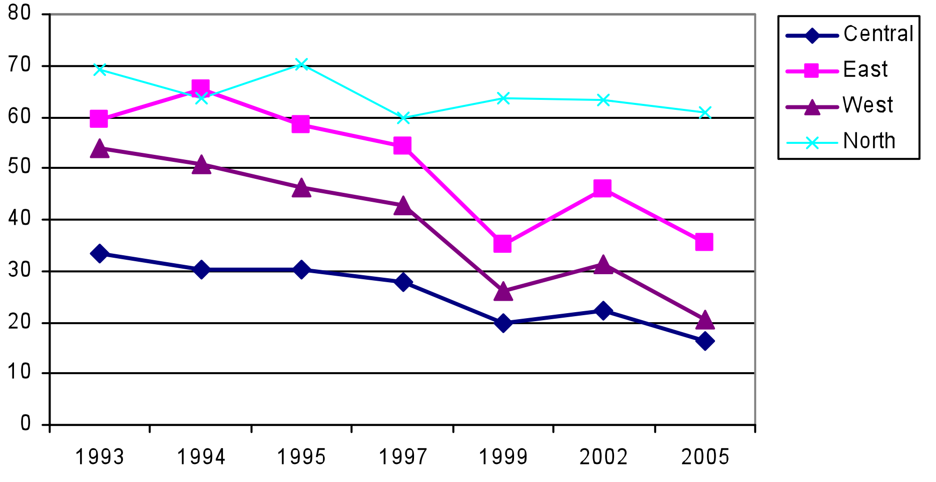

Figure 2. Official Poverty Incidence Estimated from Household Budget Surveys (Region), %
Source: Official figures based on data from the household budget surveys.

5. Empirical Strategy

This section will explain the empirical strategy followed with respect to selecting explanatory variables for the models and dealing with parameter heterogeneity.

5.1. *Explanatory Variables*

As the explanatory variables need to be suitable for including in a light survey, they should be easy to measure and collect. The full set of such variables available from the surveys includes variables of the following types: demography; education; occupation (by head); housing; consumption of food; expenditure on non-durables; expenditure on semi-durables; welfare indicators; regional dummies; community variables; and seasonal dummies. This last group is represented by dummies accounting for seasonal variability in consumption.[15] Good periods will normally be after harvest, while food will be scarcer prior to harvest.[16]

Note that all consumption and expenditure variables are binary, yes/no variables, as actual expenditure questions are not suitable for inclusion in light surveys. The dummies simply indicate whether the household consumed the various items or not.

The different types of variables have different "roles" in the model in (3). Demographic and education variables tend to change slowly, and will thus reflect long-term

[15]These control variables can be included only when we estimate models based on surveys covering the whole calendar year, so not for 1993, 1994, 1995, and 1997.

[16]Most parts of Uganda have two agricultural seasons: roughly from January to June and from July to December. The harvesting period, when food is more plentiful, usually extends to the next season (see UBoS, 2005). As our purpose is to distinguish between good and less good seasons in terms of food security, we divide the calendar year into the following four periods: January–April, May–June, July–October, and November–December. Thus we have added one month to the harvest time, as food may be expected to be relatively plentiful shortly after harvest, and accordingly we expect the first and third periods listed to be better than the remaining two.

improvements. Consumption variables, on the contrary, are able to reflect seasonality, sudden changes, or shocks to the household that will affect the immediate consumption. Housing variables are in between and change as conditions improve, but are normally not able to capture short-term fluctuations in income.

There are some limitations on availability of comparable consumption proxies in the surveys. The modules on welfare indicators in the questionnaires have typically changed much in the period and identical welfare indicators, which are potentially important explanatory variables, were not available for many models. In addition, because of various problems with the data, we have been able to link up the community information for only two household surveys, 2002 and 2005. Housing variables are not available for all models, as there was no such information in the 1994 and 1995 datasets.

One potentially important group of variables, assets, was not available for the analyses. Households were asked about the total value of a group of assets rather than the numbers and values of specified assets. Thus we were able neither to include indicators for single assets,[17] nor to construct an asset index based on the number of assets in the household. This is unfortunate, because changes in the stock of assets may be useful for capturing changes in welfare. A common strategy among poor people is to sell assets as a means of coping with shocks, while building up stocks in good times.

5.2. *Stepwise Selection of Variables*

Given the full joint set of identical explanatory variables in the two surveys, we estimate the consumption model in (3) on the basis of one of the surveys. Because there are a large number of potential explanatory variables, each consumption model is estimated by selecting variables using an automated, so-called stepwise procedure, including square and log-linear transformations of the explanatory variables.[18] The estimated model is used to predict consumption per capita and the poverty level, using (6), in the other survey.

5.3. *Parameter Heterogeneity*

A model may not yield sufficiently reliable predictions out-of-sample if important variables are omitted from the model or model parameters have changed. Structural changes are likely to take place as a population becomes wealthier. Furthermore, the parameters may in fact change as new food varieties and technologies are introduced, leading to shifts in demand. The overall strategy for evaluating the performance of a model is to compare poverty predicted by a model with the actual poverty estimate, and use a t-test to judge whether the two differ significantly. Thus we do not test whether the estimated model parameters have

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

[18]The forward stepwise procedure applied begins with no variables in the model. Variables are sequentially included according to the magnitude of the F-value. New variables are included provided that they have a low p-value, in this case 0.01. When new variables are included, others may add less information and will then be removed. One problem with this method is that it may produce standard errors that are smaller than is actually the case (see, e.g., Altman and Andersen, 1989). Note that the models used for the predictions have been re-estimated in a standard regression analysis and the tables in the Appendix shows results from a "normal" regression output.

changed from one survey to the other. If one or several model parameters change, the effect is not necessarily so large that it has any significant impact on the prediction. As time passes and the economy changes, we suspect that the effect of parameter changes and omitted variables becomes larger and that eventually the model no longer produces accurate poverty predictions.

Parameter heterogeneity can also pose problems across space. For example, use of charcoal may be a sign of poverty in urban areas, whereas it may be a sign of welfare in rural areas. To handle differences in the underlying economic structures between urban and rural areas, we estimate models for urban and rural areas separately.

In addition, we test whether heterogeneity across individuals seems to be an issue that needs to be considered in our analyses. This is done by comparing poverty predictions using the consumption model with individual specific parameters as given in equation (8) with poverty predictions using the consumption model with constant model parameters over space, equation (6).

6. Empirical Analyses

This section is divided into three main parts. We will first discuss results when predicting at the urban/rural level. Next we turn to subregional analyses for further validation of the method. The third part discusses the overall findings in light of other evidence, such as the effect of including variables omitted on purpose and trends in other welfare indicators.

6.1. *Predicting Rural and Urban Poverty*

6.1.1. Brief Review of Some Main Modeling Results

In total 84 rural and urban models were estimated. As it is not feasible to present all model results, just a selection is included (see Tables A1–A4 in the Supplementary Appendix).

The adjusted R-square is about 0.5–0.7 and generally higher for the urban models than for the rural models. One or more variables from each group—demography, education, occupation (by head), housing, consumption of food, expenditure on non-durables, expenditure on semi-durables, welfare indicators—was selected into almost every model when available. Community variables from two surveys were available but were not significant in any of these survey models. The regional dummies are important explanatory variables while seasonal dummies are generally not included into the model.

The models were inspected for heteroscedasticity by visual interpretation of plots of the residual versus predicted consumption and by formal tests, which did not lead us to correct for heteroscedasticity in the first place.[19] We applied the

[19]Formal tests—White and Breusch–Pagan tests (see, e.g., Wooldridge, 2002)—reject the assumption of constant variance of the error term (homoscedasticity) for most of the models. However, these tests are sensitive to the number of observations, because with a large number of observations a small deviation leads to rejection of the hypothesis. Thus when we use a smaller randomly drawn sample (of about 1,000 observations), the hypothesis of constant variance is no longer rejected.

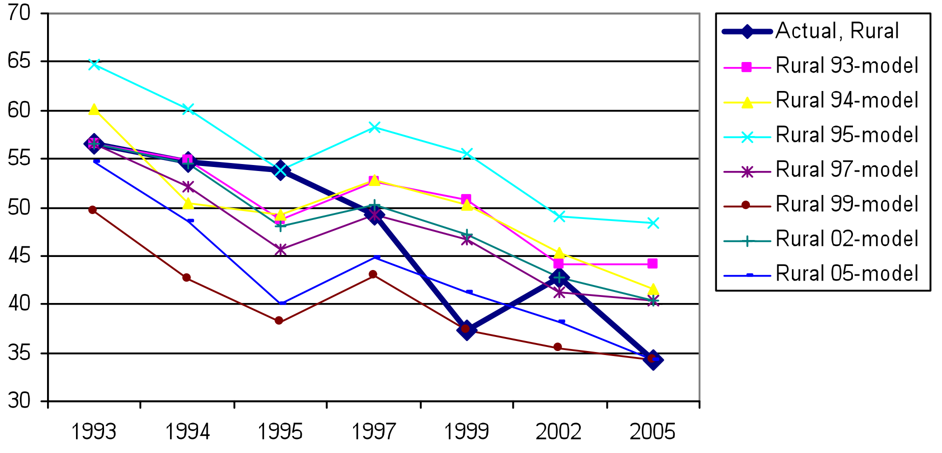

Figure 3. Actual and Predicted Poverty by Seven Models (Rural), %
Source: Author's calculations based on data from the household budget surveys.

normal distribution function for the probability function in (6), to estimate the poverty predictor, which seems reasonable according to the distribution of the residuals.

Sampling error is the only source of error in the standard errors of the actual poverty estimates. The standard error of the model predictions comprises three components: see equation (7). The idiosyncratic component is small and contributes to less than 1 percent of the variance in all our cases. The largest part of the variance is due to the estimated model parameters, accounting for about 60–80 percent of the total variance of the predictions. The subcomponent of the model standard error due to sampling is, however, smaller than the variance of the survey estimate. As would be expected, total standard errors are larger for the model predictions than for the actual estimates. However, the difference is not so large that there is reason to reject the model-based predictions because of their confidence interval.

6.1.2. The Predicted Poverty Trends

For each survey we estimate one rural model and one urban model that are then used to predict rural and urban poverty for another survey. This is repeated for all pairs of surveys and yields the seven predicted poverty trends. We will refer to a model estimated from data for a given survey by the survey year. For example, the solid line labeled the Rural 93-model in Figure 3 shows the predictions made by models from 1993 on to each survey from 1994 to 2005. It also includes the actual poverty level in 1993. The thick lines show the actual poverty in the same period.

Figures 3 and 4 show rural and urban predictions, respectively. A striking feature is that the predicted poverty trends are very similar for each survey model. Thus, focusing on changes in poverty over the period, all survey models give nearly

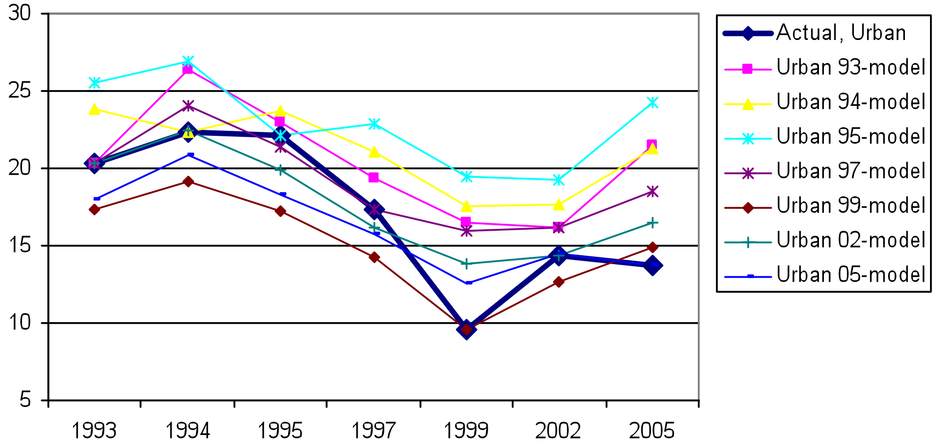

Figure 4. Actual and Predicted Poverty by Seven Models (Urban), %

Source: Author's calculations based on data from the household budget surveys.

the same results. This suggests that the relation between the consumption aggregates and the set of explanatory variables is consistent for predicting poverty trends.

Predicted urban poverty trends follow more closely the actual trends than is the case for rural areas. The main deviation from the actual trend in rural area is that the predicted trends do not capture the marked dip in poverty in 1999. Furthermore, the models predict a small dip in poverty in 1995, which was not to be found in the actual trend. Finally, the predicted trends do not capture much of the actual decrease in poverty from 2002 to 2005. All rural models are, however, able to capture most of the decline in poverty at 23 percentage points over the period, with a predicted fall in poverty close to 20 percentage points for most survey models.

The level predicted by the various surveys varies. Models estimated from surveys with relatively low poverty, such as 1999 and 2005, tend to predict lower poverty levels for a given year compared with predictions from models based on surveys with higher poverty levels. This pattern is particularly pronounced for urban models. For rural areas, 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.

The large deviations from the actual values in rural areas are mainly from 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 a poverty level for 1995, while all models predict too high a poverty level for 1999. In urban areas the largest deviation from the actual figures is mainly found when predicting for 1999 and 2005.

Predictions for 2005 could suggest that the time elapsed from the model base survey to the prediction is important. All surveys except 1999 predict too high a

TABLE 2
t-Values for the Difference in Actual and Predicted Poverty Level

	1993-Model	1994-Model	1995-Model	1997-Model	1999-Model	2002-Model	2005-Model
Rural							
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	
Urban							
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	0.7	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	

Notes: For t-test comparing proportions in two populations, see, e.g., Battacharyya and Johnson (1977, pp. 308–12) (note that the difference is statistically significant, at 5%, for the t-value at $|t| > 2$).
Source: Author's calculations based on data from the household budget surveys.

poverty level for 2005, and older surveys predict the highest poverty level for 2005. Time elapsed between the surveys, however, does not seem to be as important when predicting for 2002. Models estimated on data for 1993 predict as well for 2002 as the models estimated on data for 2005. Rather, the combination of length of time between surveys and a large fall in poverty levels seems to be contributing to the failure of the models.

We applied the random coefficient model using the predictor in (8) to account for possible individual heterogeneity in the model parameters. If this type of heterogeneity is prevalent, it will result in heteroscedasticity. The degree of heteroscedasticity, however, seems to have only a small impact on the predictions and the standard errors of using the random coefficient model in the case tested.[20]

Table 2 summarizes the prediction results. showing t-values for the test on differences in predicted and actual poverty levels.

Twenty-two of the 42 rural predictions are significantly different from the actual poverty level. Most of the deviations are due to the 1999 sample or involve use of the 1995 survey in the analyses. Furthermore, except for 1999, all survey models predict poverty that is significantly different from the actual poverty level in 2005. All other predictions made by, and for, the 1993, 1994, 1997, and 2002

[20]The predictions from the 1999 models for rural and urban areas in 2002 are, respectively, 35.0 (1.6) and 11.4 (1.6) when applying random coefficient models, compared with 35.4 (1.6) and 12.7 (1.8) with no correction for heteroscedasticity (standard errors in parentheses). The predictions from the 1995 models for rural and urban areas in 1997 are, respectively, 58.4 (2.8) and 22.7 (2.5) when applying the random coefficient model, compared with 58.2 (2.7) and 22.9 (2.6) with no correction for heteroscedasticity. Furthermore, the predictions from the 2002 models for rural and urban areas in 2005 are, respectively, 41.1 (2.0) and 18.0 (2.2) when applying the random coefficient model, compared with 40.4 (1.9) and 16.5 (2.2) with no correction for heteroscedasticity.

surveys are not significantly different from the actual ones. The urban models seem to do better. Seven of the 42 predictions are significantly different from the actual poverty level. This applies when using models from 1993, 1994, and 1995 to predict for 1999 and 2005, and when using the 1997 model to predict for 1999.

6.1.3. Predicted Rural and Urban Poverty Accounting for Omitted Variables and Temporal Variation

As we have seen in Figures 3 and 4, the survey models tend to produce the same trend, although the predicted level may differ substantially. This suggests that important variables have been omitted or that the models are not able to capture shocks. We will return to the effects on the predictions of including omitted variable in the models; here we will explore the effects of applying a model that captures shocks.

The empirical analyses are now based on (10) and (11), accounting for temporal variation not captured by the explanatory variables in the previous model. The temporal variation could capture shocks—due to, for example, prices—that affect all households. As discussed before, a main driving force behind the reduction of poverty in Uganda is world market prices for cash crops. Note that we still assume constant model parameters for the consumption proxies in the period, but we now allow for a time-dependent random variable in the model.

As not all potential explanatory variables are available in all surveys, using all surveys for estimating a common model reduces the set of available explanatory variables. For example, housing cannot be used, as no such variables were available in the 1994 and 1995 datasets. The time-dependent intercepts in the models are shown in Figure A1 in the Supplementary Appendix. There might be an upward trend, but the number of time periods covered is small and therefore it is hard to make inferences. A way to account for the observed trend patterns in the intercept is to estimate a time-series model for a_t, although this idea will not be pursued here.

Figure 5 shows the predictions for rural poverty using the model to account for omitted macro variables. For each survey year we use a model based on all the other surveys to predict poverty.

The 1995, 1999, and 2005 surveys are still the most troublesome, although the predictions for these are much better than for most of the annual models. The effect of correcting for the variance of omitted macro variables, however, is small. This is because the variance of a_t contributes to only a small part of the overall variance.

Figure 6 shows the predictions for urban areas using the model accounting for omitted macro variables. The predictions are generally good, except for 1999, and better than for most of the annual models in Figure 4.

6.2. *Predicting Subregional Poverty*

As we have seen in Section 4, there are considerable differences in the poverty dynamics at a subregional level in Uganda, from the northern region with almost no changes in poverty to the western region with a fall in poverty in the period of nearly 35 percentage points. From subregional predictions we aim to explore

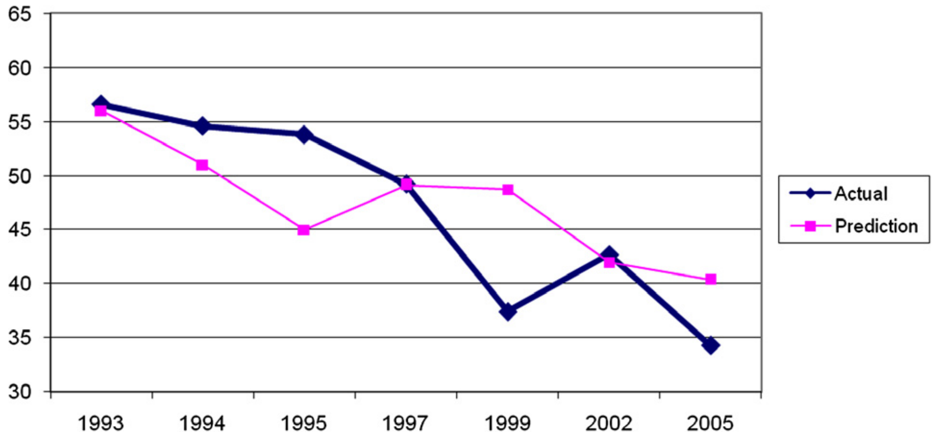

Figure 5. Predictions Using All Surveys in One Model (Rural), %
Source: Author's calculations based on data from the household budget surveys.

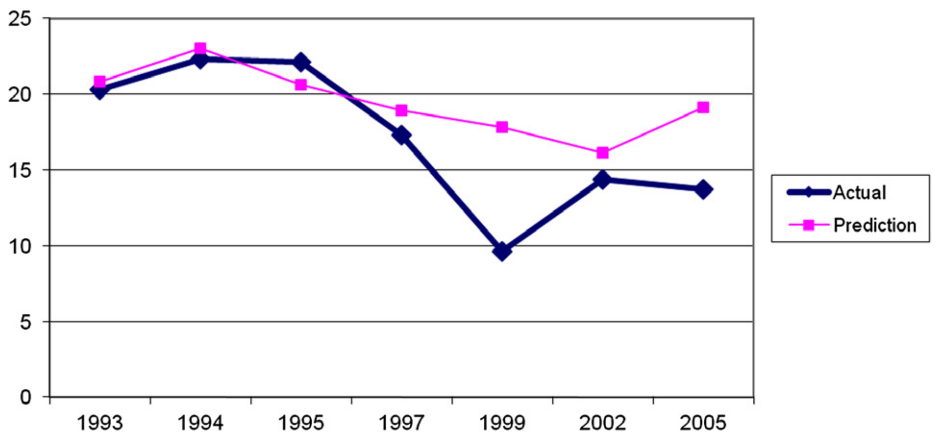

Figure 6. Predictions Using All Surveys in One Model (Urban), %
Source: Author's calculations based on data from the household budget surveys.

whether the inability to predict accurately could be associated with structural changes arising from rapid welfare changes and/or the length of time between surveys.

Of the 126 rural predictions, 51 differ significantly from the actual poverty estimate, whereas 31 of the total 126 urban predictions differ significantly from the actual poverty estimate (Table A5). Thus the proportion of "mis-predictions" increases at an urban level (from 17 to 25 percent), whereas it decreases at the subregional rural level (from 52 to 40 percent). The urban case is one where a mis-prediction at an aggregated level tends to be followed by mis-predictions for several of the subregions, and we may also find one poor prediction at a subregional level, even if the prediction at the aggregated level is good. For rural models we find the opposite: some models fail to predict the rural poverty rate because they fail to predict for a single region. For example, the 2002 rural model predicts a significantly lower poverty level than the actual level in 2005. The problem is to predict for the rural western region. The subregional figures, when using the 1997 model to predict for 2005, show the same pattern. The problem is once again predicting for the western region, which is the rural region that had the largest fall in poverty between the two surveys (see Figure 2). The subregional t-values confirm, not surprisingly, that the models tend to predict accurately when there are small changes in the poverty level: almost all models predict accurately for the north, the region with small changes in poverty levels, while the majority of mis-predictions are found in the eastern and western regions, those with the largest fall in poverty in the period. This is in accordance with the finding that most models predict accurately for their neighboring survey years (excluding the 1995 and 1999 models). This could suggest that that models become outdated.

Figures 3 and 4 show that the older surveys have lower predictive power for 2005 compared with 2002. The 1993 model predicts almost perfectly for 2002, while the predictions for both urban and rural regions differ significantly from the actual poverty levels when applying the 1993 model for 2005. This picture is confirmed by the predictions at the subregional level: the 1993 model reproduces the actual poverty estimates very well for 2002, while the predictions differ significantly for most subregions in 2005.

Thus the subregional predictions for 2005 suggest that one reason for poor performance could be the length of time between surveys; the 2002 models perform better than the 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.[21]

6.3. *Discussion of Results and Additional Evidence*

As we have seen from the previous analyses, there are three surveys—1995 for rural, and 1999 and 2005 for both urban and rural—that predict poverty significantly different from the official figures. The 2005 results may in part be attributed to the length of time between surveys, whereas the findings for 1999 could be due to comparability problems in the consumption aggregate in 1999, as shown by Luoto (2007). As discussed briefly in Section 3, the type of method applied in Luoto (2007) systematically adds expenditure groups to a consumption model in 1999 and compares the prediction to assess whether the prediction is particularly affected by certain subgroups of consumption and, if so, to assess if this seems reasonable. Using this method shows that the increased poverty level between the two surveys seems to rest on one subgroup of consumption. Luoto concludes that if there was indeed an increase in poverty, there is no plausible reason for households to shift away from home-produced starches (comprising five of the six

[21]An example in the data about how model parameters may change over time and thus contribute to the outdating of models is the parameter for charcoal usage. Use of charcoal has a positive and significant impact on the consumption level in urban areas in the "old" surveys, whereas from 1999 onwards charcoal yields a negative impact on the consumption level, although the charcoal parameters are no longer significant. Thus in the early 1990s charcoal was associated with the wealth of a household, whereas it has the opposite association from 1999 onwards.

staples consumed in Uganda); rather, one would expect to find the opposite if welfare indeed was reduced. The findings are consistent with the results here, showing that 1999 is out of line with the other surveys. It would have been useful to assess comparability between the other surveys in the same manner, in particular the 1995 and 2005 surveys, but as discussed in the data section, the sublevel consumption groups are not readily available, so we have not been able to do this type of analysis.[22]

We have, however, explored the effect of including an important variable omitted on purpose in the models. Comparable information on total value of asset holding can be found in the 1999 and 2002 surveys, and in the surveys from 1993 to 1997. Value of assets is expected to have a high correlation with welfare, but is omitted on purpose from the previous analyses as it is not suitable to collect information on asset values in a light survey. The value of assets per capita is an important variable and its parameter is significant in all models. However, its inclusion does not lead to any large change in the predictions and does not improve the predictions (see Table A6). On the contrary, including value of assets in the 1999 model to predict poverty for 2002, as well as in the 1993 model to predict for 1995, predicts lower poverty (2 percentage points) and is further away from the official figures.

In the remainder of this section we will discuss whether the trends in poverty are consistent with trends in other welfare variables in order to highlight whether one would expect poverty trends as predicted by the models or by the official poverty trends. The World Bank Poverty Assessment for Uganda in 2005, like Luoto (2007), also questioned the relatively low poverty level in 1999 compared with 2002 (see World Bank, 2005). According to this assessment, other important poverty indicators, such as values of assets and food shares, are incompatible with a higher poverty level in 2002 than in 1999.

Table A7 shows annualized growth rates in GDP per capita and private consumption as estimated from the surveys. The table shows that the growth rates predicted by the surveys in general are lower than the GDP growth rates, except for 1997–99, when it is substantially higher. The higher growth rate from 1977 to 1999 is what one would expect if the consumption aggregate in 1999 is inflated.

Tables A8 and A9 show, respectively, rural and urban trends in variables collected in the surveys that typically enter into a model and may be used as "free-standing" welfare indicators.[23] The tables confirm the picture drawn by the World Bank with an improvement in welfare from 1999 and 2002. The changes in the housing variables reflect a steady improvement. Even indicators that are prone to change quickly, such as consumption variables, do not reflect the marked dip in poverty in 1999. It seems that 2002 was a relatively good year in terms of the

[22]Without doing a considerable amount of work, including making assumptions on the procedures for cleaning of outliers and re-evaluation.

[23]Note that there is a separate module in most of the questionnaires (except in 1993) on "welfare" indicators: e.g. the number of meals taken and if each household member has at least two pairs of clothes. Such variables could be useful for tracking changes in welfare in the period. However, we are not able to produce trends, as the variables included in the module changed between the surveys, and identical indicators can be identified in only two or three surveys.

proportion of the population that consumed food, including non- and semi-durable goods such as meat and bathing soap. This holds not only for the variables included in these tables, but also for all consumption variables in our dataset. This finding is confirmed when looking at the welfare quintiles separately and at the subregional level.

Focusing on the rural domain of the 1995 survey, which is the other troublesome part of our dataset, we see that even though poverty gradually declined from 1993 to 1997, the shares of the population that consumed food and non-food commodities listed are considerably higher in 1995 than in the previous and subsequent surveys, 1993, 1994, and 1997. This pertains to all food indicators available for the analyses, not only the ones shown in the tables. This holds also for all non-durables, and for all semi-durable indicators,[24] in both 1993 and 1994. The same picture emerges when examining the wealth quintiles and the rural regions separately. The high levels on the consumption variables relative to the poverty level in 1995 might explain why the 1995 models predict poverty levels that are too high for the other surveys and why the other survey models predict poverty levels too low compared to the official figures for 1995. Because 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 that include housing variables.

Tables A8 and A9 show a mixed picture when it comes to changes in welfare indicators from 2002 to 2005. Slightly fewer households reported consumption of the food items bread, sugar, and meat in 2005 than in 2002, except in rural areas, where the proportion consuming bread increased slightly. Most of the other variables indicate a little improvement, including some additional welfare indicators comparable only in 2002 and 2005 (see Table A10).

7. Conclusion

This paper has examined the performance of a survey to survey-based method for predicting the headcount poverty ratio by comparing predicted poverty figures with the official poverty figures estimated directly from the consumption aggregates. Altogether, seven household budget surveys for Uganda, from 1993 to 2005, have been used.

The models have been estimated and poverty has been predicted at both urban and rural levels, as well as at the subregional level. In total more than 200 predictions have been evaluated. All models tend to predict similar changes in poverty levels and we get similar trends in poverty, irrespective of which of the seven surveys we base our model on. In most cases this simple modeling approach produces predictions at rural, urban, and subregional levels that are in line with the official poverty figures. However, there are also many cases where the predictions differ significantly from the official poverty figures. It is interesting to note that this is particularly the case when using the 1999 survey, which has been questioned and tested by other authors, as to whether this survey is indeed comparable with the

[24]Except furniture.

subsequent survey in 2002. The results here give further support to the suspicion that the marked dip in poverty in 1999 does not reflect real improvements in welfare.

A model should be used with caution if too much time has passed between the surveys, and in particular if the region or country is in a phase of large changes in welfare. If more than one survey is available, the samples should be joined in order to create a model accounting for variations in year-specific factors, thus improving the predictive performance.

References

Altman, D. G. and P. K. Andersen, "Bootstrap Investigation of the Stability of a Cox Regression Model," *Statistics in Medicine*, 8, 771–83, 1989.

Appleton, S. and S. Ssewanyana, "Poverty Estimates from the Uganda National Household Survey II, 2002/03," Unpublished Paper, University if Nottingham and Economic Policy Research Center, 2003.

Appleton, S., T. Emwanu,, J. Kagugube, and J. Muwonge, "Changes in Poverty and Inequality," in P. Collier and R. Reinnikka (eds), *Uganda's Recovery: The Role of Farms, Firms and Government*, World Bank, Washington DC, 83–121, 2001.

Battacharyya, G. K. and R. A. Johnson, *Statistical Concepts and Methods*, Wiley, New York, 1977.

Bussolo, M., O. Godart, J. Lay, and R. Thiele, "The Impact of Coffee Price Changes on Rural Households in Uganda," *Journal of Agricultural Economics*, 37, 293–303, 2007.

Datt, G. and S. Jolliffe, "Poverty in Egypt: Modelling and Policy Simulations," *Economic Development and Cultural Changes*, 53, 327–46, 2005.

Deininger, K. and J. Okidi, "Growth and Poverty Reduction in Uganda 1992–2000: Panel Data Evidence," *Development Policy Review*, 21, 481–509, 2003.

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

Elbers, C., J. O. Lanjouw, and P. Lanjouw, "Micro Level Estimation of Poverty and Inequality," *Econometrica*, 71, 355–64, 2003.

Elbers, C., P. Lanjouw, and P. G. Leite, "Brazil Within Brazil: Testing the Poverty Map Methodology in Minas Gerais," World Bank Policy Research Working Paper 4513, World Bank, Washington DC, 2007.

Kappel, R., J. Lay, and S. Steiner, "Uganda: No More Pro-Poor Growth," *Development Policy Review*, 23(1), 27–53, 2005.

Luoto, J., "What Happened in the UNHS-2? Re-estimating Uganda's Poverty Numbers," Unpublished Paper, University of California, Berkeley, 2007.

Mathiassen, A., "A Model Based Approach for Predicting Annual Poverty Rates Without Expenditure Data," *Journal of Economic Inequality*, 7, 117–35, 2009.

Muwonge, J., "The Role of Household Surveys in Poverty Reduction Efforts: A Case of the Uganda National Household Survey Programme," *African Statistical Journal*, 6, 111–34, 2006.

Okidi, J. A., S. Ssewanyana, L. Bategeka, and F. Muhumuza, "Uganda's Experience with Operationalizing Pro-Poor Growth, 1992 to 2003," in T. Besley and L. J. Cord (eds), *Delivering on the Promise of Pro-Poor Growth: Insights and Lessons from Country Experience*, World Bank, Washington DC, and Palgrave Macmillan, Basingstoke, UK and New York, 169–98, 2007.

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

Ssewanyana, S. and J. A. Okidi, "Poverty Estimates from the Uganda National Households Survey III, 2005/06," Economic Policy Research Center, Occasional Paper No. 34, 2007.

Stifel, D. and L. Christiansen, "Tracking Poverty Over Time in the Absence of Comparable Consumption Data," *World Bank Economic Review*, 21, 317–41, 2007.

UBoS, "Where Are the Poor? Mapping Patterns of Well-Being in Uganda," Uganda Bureau of Statistics, Entebbe, 2003.

———, "Report on the Results of the Pilot Permanent Agricultural Statistics System (PASS) 2004, in Five Districts," Uganda Bureau of Statistics, Entebbe, 2005.

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

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA, 2002.
World Bank, "Uganda Poverty Assessment Report," World Bank, Washington DC, 2005.

Supporting Information

Additional Supporting Information may be found in the online version of this article: