Discussion Papers No. 415, December 2006
Statistics Norway, Research Department

Astrid Mathiassen

A simple poverty predictor model with assessment of the uncertainty

A revised version of DP 415

Abstract:
The primus inter pares of the UN-approved Millennium Development Goals is to reduce poverty. The only internationally accepted method of estimating poverty requires a measurement of total consumption based on a time-consuming and resource-demanding measure of household expenditure in an integrated survey over 12 months. Rather than measuring poverty, say, only every fifth year, a model is presented to predict poverty based on a small set of household variables to be collected annually between two 12-monthly household surveys. Information obtained from these "light" surveys might then be used to predict poverty rates. The key question is whether the inaccuracy in these predictions is acceptable. It is recommended that these models be tested at a country level and if the test results are similar to those found here, that this approach be adopted.

Keywords: Stochastic model, Poverty measurement, Money metric poverty, Survey methods

JEL classification: C31, C42, C81, D12, D31, I32

Acknowledgement: John K. Dagsvik contributed substantially to the methodology presented in this report and throughout the work process, offering advice and valuable comments. Bjørn K. Wold and Stein Opdahl initiated the work and provided comments at various stages. Thanks to Pham Dinh Quang for help in programming, to Johan Heldal for commenting on the methodology and to Geir Øvensen for comments and input. The participants in seminars in Malawi, Mozambique, and Statistics Norway discussed and gave feedback on methodology and findings, as did participants at the IAOS–IASS conference "Poverty, Social Exclusion and Development: A Statistical Perspective" in Amman December 2004. The Norwegian Agency for Development Cooperation, Norad, funded the work.

Address: Astrid Mathiassen, Statistics Norway, Division for Development Cooperation.
E-mail: mss@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

The need for developing methods for regular and frequent monitoring of poverty has become urgent with the increased focus on the need to "eradicate extreme poverty and hunger". This is the first goal of the Millennium Declaration that was adopted in September 2000 by 189 nations. Six variables focusing on individual consumption and nutritional status were identified for measuring progress towards this goal. The two most important of these are the proportion of population below one US dollar (PPP) per day and the proportion of population below the national poverty line. A widely recognized approach for measuring the proportion of individuals below a poverty line is through a fully fledged household consumption and expenditure survey. For implementation, a proper sample and diaries or recollections of consumption of food and nonfood items extrapolated to 12-months is needed. However, few countries can justify spending resources on an annual household consumption and expenditure survey. Consequently, proper expenditure measures are often collected only every 5th or even 10th year. Because of this state of affairs, there is a need for methods that provide annual poverty estimates at a more reasonable cost.

In this paper, we discuss a statistical procedure for estimating the proportion of a population with consumption below a given poverty line, without the requirement of a full-fledged expenditure survey. The basic idea is to utilize the information in a consumption and expenditure survey to identify a small set of household variables (indicators) that can be collected annually between two such surveys. This is done by estimating a relation that links consumption and poverty to the set of indicators through a statistical model. The indicators should be fast to collect and easy to measure. Hence, they may be compiled through so-called "light surveys", such as the CWIQ[1] surveys. The information obtained from the light surveys and the estimated model is then used to predict poverty rates.

The challenge of predicting poverty is not a new one. Fofack (2000) developed a method for ranking the households in the CWIQ survey into expenditure quintiles based on the number of individuals with predicted consumption within each quintile. This method has been applied in, for example, Ghana (Fofack, 2000) and Uganda (McKay, 2001).

Our contribution is to show how these methods may be taken a step further by using a simplified approach to the calculation of the predictor of the headcount ratio as well as its standard error and bias. The methodological approach presented is inspired by statistical modeling in the

---

[1] Core Welfare Indicators Questionnaires, jointly developed by the World Bank with UNDP and UNICEF. These surveys are not designed to measure expenditure or consumption but to obtain indicators of welfare and use of and access to public services.

adjacent area of poverty mapping, cf. Elbers, Lanjouw and Lanjouw (2003) and Hentschel, Lanjouw, Lanjouw and Poggi (2000). Poverty mapping produces low-level predictions of poverty and inequality and respective standard errors by combining a household expenditure survey with a census. The poverty mapping method has also been extended to predict poverty by combining two surveys, for example in Minot (2005). In this case, the researcher adds an additional element to the standard error to account for sampling in the survey (as opposed to the use of a census, with no sampling error).

The poverty-mapping method outlines a general method for predicting various poverty and inequality measures; however, this paper spells out a simplified method for predicting the poverty headcount. A similar procedure can be used to predict other poverty measures. The poverty mapping methodology applies a Taylor approximation to derive the standard error, whereas the procedure proposed here derives an exact expression for the standard error. This allows for a simple and transparent estimation procedure. In its simplest form, we assume homoskedastic error terms, which is reasonable in the empirical applications tested here. We also outline how the method can be extended to allow for heteroskedasticity. Like similar methods for predicting poverty, it relies on the critical assumption that the relation between expenditure per capita and the poverty indicators is stable over time.

The paper is organized as follows: In Section 2, we present the model for predicting poverty rates as a function of the indicators. In Section 3, we discuss the bias and variance of a predictor and obtain formulas for these. In Section 4, we describe the process for selecting the set of indicators. In Section 5, we apply the methodology by combining the National Household Survey for Mozambique 2002–03 with a labor force survey from 2004–05, and the results are used to exemplify and discuss the approach. In Section 6, we present our conclusions.

2 Models for predicting poverty

In this section, we propose a methodology for predicting poverty rates and assessing the uncertainty. A consumption model is used to predict the expenditure for each household in the light survey sample, using parameters estimated from the expenditure survey. The headcount ratio is in the next instance computed by summing the individual probability that the predicted consumption falls below a poverty treshold. Alternatively, one could start by directly estimating the probability of being poor, by applying a binary model (e.g. logit or probit). The advantage with the binary approach is that it allows for an unknown transformation of the dependent variable, but this again implies that one cannot change the poverty threshold after the model is estimated. If, however, the modeling assumptions made in the two approaches are correct the two methods should produce the same predictions[2].

An individual is considered poor if his/her consumption[3] falls below a certain threshold. This threshold is called the poverty line[4]. We want to predict the headcount ratio, i.e. the proportion of individuals that have a level of consumption below the poverty line.

Let Y_i denote consumption for individual *i.* Individual consumption is household consumption obtained from the household expenditure survey, adjusted for the number of household members[5]. Whether consumption per capita or per adult equivalent is chosen will not affect the methodology described here. We will refer to Y_i as consumption per capita. Let *z* denote the poverty line. Let $y_i = 1$ if individual *i* is poor; $Y_i \leq z$, and zero otherwise. Since the unit for the measurement of consumption is the household, one needs to adjust for the number of members in each household to calculate individual consumption. Let s_i be the number of members in household *i,* and *N* be the number of individuals in a population Ω, consisting of N^H households. The population can, for example, refer either to countries or to regions within a country. Hence, the proportion of poor people is given by

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

We wish to use a model to predict y, given a set of household variables (indicators), and assume next that

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

where X_i is the vector of poverty indicators, β is a vector of unknown parameters and ε_i is an i.i.d. error term with known cumulative distribution function *F*, which has zero mean and unit variance. Assume further that ε_i and X_i are uncorrelated. The logarithmic transformation of the consumption variable serves to produce a symmetric distribution of the error term that usually approximates a

---

[2] In the empirical application, section 5.2, we also present predictions produced by the probit approach.

[3] Consumption rather than income is used since consumption, particularly in poor economies, is likely to be more precisely measured. In addition, income may vary considerably over the year; for example, a farmer receives most of his/her income at harvest time, but the household has a more stable consumption pattern over the year (see, for example, Johnson, McKay and Round, 1990). The method described here is, however, not affected by whether income is used instead of consumption in the analysis.

[4] There are two main approaches to deriving the consumption poverty line: the absolute and the relative. The absolute poverty line is derived on the basis of the cost of satisfying basic needs. The relative approach derives the poverty line as a share of the consumption estimated on the basis of the entire population. See, for example, Ravallion (1992) or Ravallion (1998) for a comprehensive discussion of poverty lines.

[5] We do not consider intrahousehold distribution effects in this framework, such as women systematically being paid less than men. See, for example, Deaton (1997) for a review.

normal distribution, and stabilizes the variance of the error term. The validity of the assumptions on homoskedasticity of the error term will be discussed in relation to the empirical analyses later.

The probability that individual *i*'s consumption falls below a poverty line, *z*, is found by inserting the regression model in a probability function:

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

The predictor for the headcount ratio is the average probability of being poor over all individuals in the survey. Thus, a predictor for the headcount ratio, y, in (1) is given by

(4)
$$P^* = \frac{1}{n}\sum_{i\in S} s_i P_i = \frac{1}{n}\sum_{i\in S} s_i F\left(\frac{\ln z - X_i\beta}{\sigma}\right),$$

where *S* denotes the light survey sample, $S \subseteq \Omega$, consisting of n^H households and *n* individuals.

3 Bias and variance of the estimator

3.1 Correcting for bias in the predictor

If the modeling assumptions underlying (2) are correct the estimator for P^* given in (4) will be an unbiased estimator for y. 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 nonlinear way. In the appendix section 7.1.1, we show how we can derive a closed form expression for this bias, which can be used to correct our estimate, in the case when *F* is a standard normal cumulative distribution function. The formula for correcting this bias is given in (13).

3.2 The variance of the prediction error

We will now discuss the problem of assessing the uncertainty of the predictor for the poverty rate by estimating the model sketched above. The prediction error is the deviation between the poverty level predicted by our model and the actual poverty level in the population. The prediction error we discuss here is computed by inserting the predictor in (10) without adjusting for the bias discussed in the previous section. This facilitates the derivation of an expression of the variance of the prediction error, and the error it produces is negligible as long as the bias in (13) is small.

One way to decompose the prediction error is as follows:

(5)

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

It is demonstrated in 7.1.4 that the three terms on the right hand side of (5) are uncorrelated, which implies that the total variance of the prediction error equals the sum of the variances of the three terms in brackets on the right side of (5).

The first term on the right hand side in (5) is the difference between the actual poverty level in the population and the expected poverty level. It captures how the headcount ratio in the population deviates from its mean expected value. In (18) in section 7.1.4, we show that the variance of this term can be expressed as

(6)
$$\operatorname{var}\frac{1}{N}\sum_{i\in\Omega}\left(s_i\left(y_i - P_i\right)\right) = \left(\frac{1}{N}\right)^2 \sum_{i\in\Omega} s_i^2\left(P_i - P_i^2\right).$$

The variance in (6) is usually small when the target population is large.

The second term in (5) is the variance of the mean error in prediction due to the uncertainty of the estimate $\hat{\beta}$. It also depends on the variance of the *X*-vector:

(7)
$$\operatorname{var}\left(\frac{1}{N}\sum_{i\in\Omega} s_i P_i - \frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i\right) = \operatorname{var}\left(\frac{1}{N}\sum_{i\in\Omega} s_i\left(P_i - \hat{P}_i\right)\right) \quad .$$

Although one can develop the calculations in (7) further, the expression is sufficiently explicit for our purpose, which is to compute the variance by simulation.

The last term in (5) is the difference between the predicted poverty level in the population Ω, and the predicted poverty level in the sample S. It is due to the uncertainty that stems from the fact that *S* is a finite random sample, and that it also depends on an uncertain estimate, $\hat{\beta}$. This term depends on the variance of the *X*-vector in the sample, the sampling fraction n^H / N^H and the size of the light survey. If S is a randomly selected sample it follows readily (see (19) in 7.1.4) that:

(8)
$$\operatorname{var}\left(\frac{1}{N}\sum_{i\in\Omega} s_i \hat{P}_i - \frac{1}{n}\sum_{i\in S} s_i \hat{P}_i\right) = \left(1 - \frac{n^H}{N^H}\right)\frac{n^H}{n^2} E \operatorname{var}\left(s_i \hat{P}_i \middle| \hat{\beta}\right) \quad .$$

As a result, the variance of the prediction error in (5) is found by summing the variance components:

(9)

$$\operatorname{var}\left(\frac{1}{N}\sum_{i\in\Omega}s_i y_i - \frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\right) =$$

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

We can allow for other sampling designs by adjusting the last term of the right hand side of (9) to the applied sample design.

In Appendix 7.1.5 we show how one by using Monte Carlo simulations can compute the variance given in (9). The point is that since we know that the parameter vector, $\hat{\beta}$, is multinormally distributed it follows that $X_i\hat{\beta}/\sigma$ is normally distributed. Thus, we estimate the variance of $X_i\hat{\beta}/\sigma$, and for each individual with the given characteristics, X_i, we compute *M* independent probabilities of being poor. Then we have sufficient variation for computing the variance, see (25).

4 Selecting the set of poverty indicators

The poverty indicators may be selected by comparing estimated models with various combinations of potential poverty indicators. Based on statistical criteria, one may choose the set of indicators that constitute the "best" model for predicting the poverty headcount ratio[6]. Since these variables will be collected in a light survey there should not be too many indicators, although sufficient variables should be selected to ensure that the marginal gain of including additional variables is low.

Before models can be compared there are some criteria that limit the pool of potential candidates. First, the set of potential poverty indicators are, of course, limited by the information in the household budget survey. In addition, the poverty indicators should, (i) be easy to measure, (ii) be easy to obtain information about, and (iii) have a high correlation with household consumption per capita. Points (i) and (ii) refer to the reliability of the poverty indicators. Point (iii) ensures that the set of indicators is limited to those that are significant in predicting household consumption and therefore also the poverty level.

To account for possible cluster effects in the expenditure survey, variables such as access to markets, rainfall and availability of electricity may be included.

Expenditure variables naturally have a high correlation with per capita expenditure. They are, however, not necessarily available for inclusion in a light survey. Questions about the value of consumption are in general quite complex and cover information on purchases, gifts as well as consumption from own production. In particular, the latter is problematic since the value of own

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

production is hard to measure. Thus, such variables can hardly be said to fulfill the criteria of being fast, easy and reliably measurable. However, it may be desirable to include information on a few consumption variables outside the category of "own production" (e.g., cooking oil and soap) as these tend to increase substantially the explanatory power of the model. Furthermore, dichotomous variables about whether particular goods (e.g., meat and transport) are consumed often increase explanatory power considerably.

The effect of obtaining the same type of information from different types of questionnaire is not well documented. There is, however, no reason to expect this to be a particular problem with the light survey compared with the expenditure survey. However, there will generally be an effect from whether the information about consumption in the expenditure survey is based on a diary[7], bounded recall (recall from one meeting to the next) or unbounded recall; i.e. recall often implies that people forget small purchases and purchases that are beyond the ordinary pattern, Deaton and Grosh (2000). There are, however, no a priori reasons to believe that the dichotomous information on whether the subjects have consumed important food, such as meat, should be influenced by whether recall or diaries were used.

A fundamental modeling assumption is that the conditional distribution of consumption per capita *Y*, given the set of indicators, *X*, is stable over time; in other words, that the parameter vector β has not changed between the time of the expenditure survey and the light survey. If this assumption does not hold, then the model is no longer valid and will provide biased estimates. The assumption about stability is not testable if one does not observe both Y and X in both surveys[8]. As the assumption of stability in the estimated parameters are generally hard to verify, a researcher should be careful not to predict too far into the future or the past, and should compare the conditional probability of being poor when a new household survey becomes available.

5 An empirical application

In this section, we apply the multivariate modeling approach outlined above[9]. We identify poverty indicators and estimate a model on the basis of the national household expenditure survey for Mozambique: IAF 2002–03, and predict for the labor force survey: IFTRAB, 2004–05. We had no

---

[7] No food items are specified in the questionnaire, rather the household keeps a record of what they consume.

[8] Tarozzi (2005) provides indirect empirical support for this assumption when estimating the conditional probability of being poor over four rounds of expenditure surveys in India. This is possible as the data under investigation consists of several rounds of expenditure surveys over a period when the expenditure questionnaire underwent considerable changes.

[9] The empirical analyses were conducted by Fátima Zacarias, Cassiano Soda Chipembe, Cristóvão Muahio, Elisio Mazive, Xadreque Maunze and Maria Mazive from INE Mozambique, and Geir Øvensen and Astrid Mathiassen from Statistics Norway.

influence over the variables included in the IFTRAB, and thus no influence over the common set of variables in the two surveys. Since our focus is on methodology, we have chosen to present the rural and urban models for only one region, central Mozambique.

5.1 Data

The integrated household survey (IAF) for Mozambique, 2002–2003, was a comprehensive socio-economic survey of the living standard in Mozambique and consisted of about 8700 households. The data were prepared by INE, the national statistical office of Mozambique, and important variables such as total household consumption were derived and the poverty line was established; see the National Directorate of Planning and Budget et al. (2004) for documentation. The welfare measure is given by total daily per capita consumption and expenditure. The poverty line is based on the cost of basic needs.

The light survey for which we predict poverty is the labor force survey, IFRAB, 2004–05. The sample consisted of about 17,500 households. This is an extended CWIQ[1]-survey including an additional section on economic activities for all members in the households.

The range of potential indicators that were examined represents characteristics that were available from both surveys. About 150 variables were tested, and they comprise the following groups: literacy, education, employment, assets, housing, energy and water use. Information on typical poverty indicators, given by a separate section in the IAF 2002–03, was not included in the IFTRAB and therefore not available for the model.

5.2 Results

The estimates and standard errors for the parameters in the multivariate models are given in Table 2 and Table 3, section 7.3. The R-squared for the urban model are much higher than the rural model (compare 62 to 39 percent). This is a general finding in these types of models, as the urban population tend to be more heterogeneous with are a larger spread in the expenditure per capita, compared to the more homogenous group of rural households. Thus, it is easier to identify correlates explaining the variation among the urban population. Further, one can expect that location specific characteristics are more important in explaining the welfare in rural areas, and the rural model might therefore benefit from including variables as rainfall or distance to market. However, such indicators were not available from both these surveys. Note also that, if we could include indicators from all the available indicators in the IAF survey rather than relying on only those that also are included in the IFTRAB, the R-squares would increase considerably, from 39 to 48 percent in rural areas and from 62 to 69 percent in urban areas.

Plots of the residuals against the predicted value do not indicate the presence of heteroskedasticity[10]. Normal PP-plots support the assumption of a normally distributed error term; thus, the standard cumulative distribution function was applied, as may be seen from Figure 1-Figure 4 in 7.2.

Table 1 summarizes the weighted prediction results for the rural and urban central region of Mozambique. The third column shows the actual poverty level calculated from the expenditure survey in 2002–03. The last column shows the predicted headcount ratio based on the labor force survey in 2004–05. Standard errors are given in parentheses, and all standard errors are corrected for the sampling design.

The biases of the predictor as given in (13) are negligible, for both the rural and the urban domain they are about 0.01 percent of the level of the prediction. Thus, neglecting the bias in calculating the standard error for the model predictions should not be a problem.

Table 1. Poverty headcounts with standard errors

	Number of observations		Actual poverty level, expenditure survey, 2002/03	Predicted poverty level, labor force survey, 2004/05
	Expenditure survey	Labor force survey		
Rural	1924	3535	45.2 (2.9)	40.9 (3.0)
Urban	1176	2853	46.7 (3.1)	40.1 (3.1)

1000 random draws were used in the simulations

The poverty headcount ratio for the region of interest was 45 and 47 percent in rural and urban areas in 2002/03. According to predictions, poverty fell by four and seven percentage points in the rural and urban central domain, respectively, in the two-year period between the surveys[11]. The declines in poverty levels are not statistically significant[12]. Note, however, that the decrease in poverty level follows a trend with considerable decline in the poverty headcount in Mozambique as reported in the two subsequent household surveys in 1996/97 and 2002/03. In this period, the poverty headcount in the central region fell from 74 to 46 percent[13, 14].

[10] Formal tests of constant variance are ambiguous. In the urban model, neither the Breusch Pagan nor the White test rejects the hypothesis about constant variance at a 5 percent significance level. In the rural model, the Breusch Pagan and White tests reject the hypothesis about homoskedasticity. 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. When we estimate the model by using a smaller, randomly drawn sample of the expenditure survey (for example, 1000 rather than 1900 observations) none of the tests of homoskedasticity is rejected.

[11] The within sample predictions of the headcount ratio in 2002/03 is 47.4 for rural and 45.5 percent for urban. The within sample (here: the expenditure survey) predictions ensures a directly comparable reference for the out of sample (here: the labor force survey) predictions, as they are calculated by using the same method.

[12] We have also predicted the headcount ratio for 2004/05 by directly estimating a binary model (probit model). The headcount ratios in respectively rural and urban predicted by the probit model come to respectively 39.4 and 38.8 percent. Thus, the two modeling approaches produce fairly similar prediction levels.

[13] The numbers are not split on urban/rural as the definition of urban/rural changed between the two surveys.

The sampling errors of the actual poverty rates in 2002/03, column three, are about the same as the standard errors of the model prediction in 2004/05, column four. Since the labor force survey consists of a larger sample, we have also estimated the standard deviation based on a sample the same size as the expenditure survey. We have done this by estimating the variance of the predictor from a randomly drawn sample of the light survey, repeating this procedure 50 times and computing the average standard error. Comparing these numbers to the standard errors in column three gives a picture of the effect on the uncertainty by applying a model approach compared to the traditional approach with conducting a fully-fledged expenditure survey. The standard error of the actual poverty headcount for the rural domain is 2.9, compared to 3.3 for the standard error of the model-based prediction of same sample size. The corresponding figures are 3.1 and 3.4 for the urban domain. Thus, the standard error of the predictor is only slightly higher when applying the model approach. However, as sampling contributes the entire variance of the predictor in the traditional approach, only a modest share of the variance of the model approach is due to sampling. When estimating separately each of the variance components in the model approach, given in (9), we find that in both the rural and urban case, about 20 percent of the variance is due to sampling uncertainty. By using a model, we make assumptions about the expected expenditure level and thus fewer observations are required to obtain the same level of precision as when no prior assumptions are made. The gain of reduced sampling uncertainty, however, is offset by the uncertainty of estimating the parameters in the model. About 80 percent of the variance in the examples here, are due to model error. Finally, less than one percent is due to the idiosyncratic component. This component is low because we predict for large populations.

6 Conclusions and recommendations

Our objective in this paper is to develop an alternative approach for calculating the standard deviation of the predictor of the poverty rates by means of selected indicators, and to illustrate how this method can be used for annual poverty estimates. The method can be applied in the years between the fully-fledged consumption and expenditure surveys by obtaining indicators from the light surveys that are consequently used to predict poverty.

The poverty prediction based on the expenditure surveys every 5th year is affected by the sample inaccuracy due to the small size of the survey sample. The poverty predictions based on the method proposed in this paper suffer from the sampling inaccuracy due to the small size of the light survey sample plus the uncertainty of the model parameters.

---

[14] A considerable share of the decline in poverty in Mozambique in this period is associated with a change of the food bundle in the poverty line. The poverty lined changed as the economy opened up and with that followed a change in the relative

Test results indicate that there does not seem to be any serious deviation from the assumption of homoskedasticity error term, and this assumption does not seem to pose a problem for our analyses[15]. Thus, a fairly simple and straightforward simulation procedure can be applied to compute an estimate for the exact standard error. However, the assumptions must be tested and evaluated for every analysis undertaken. In 7.1.7 we have outlined a procedure that may be used if heteroskedasticity is a problem.

In our empirical application on data from Mozambique, the standard errors of the predictions based on the model are higher than the standard errors of the poverty head counts estimated on the basis of the fully fledged expenditure survey. They are about 10 and 15 percent higher in the rural and urban samples, respectively, given the same sample size for the light survey as for the expenditure survey. As the light surveys tend to be larger than the expenditure survey, the difference in the precision of the predictors is even smaller. Thus, if one accepts the assumption about stability of the model parameters, the additional uncertainty from the proposed method is acceptable. The problem with unstable model parameters should, however, be taken seriously and should be tested when a new expenditure survey is available[16]. For the same reason, one should be careful when forecasting a long time ahead or extrapolating back in time.

---

prices and the consumption pattern. However, even when using the same consumption bundle in the two surveys the poverty headcount fell by 15 percentage point in the Central region, see National Directorate of Planning and Budget et al. (2004).

[15] Neither does there seem to be serious deviations from these assumptions in the analyses of the remaining regions in Mozambique and in corresponding analyses from Malawi that was based on the Integrated Household Survey 2004.

[16] Inferences about stability in the model parameters can be provided if expenditure variables on some sub-group are available in both the expenditure and the light survey. If this is the case one can estimate one model from each survey and compare the coefficient estimates, or one can estimate one model from one survey, predict for the other survey and compare it to the actual prediction.

7 Appendix

7.1 Mathematical appendix

7.1.1 Correcting for bias in the predictor

In this section, we will derive a closed form expression for this bias, which can be used to correct our estimate, in the case when *F* is a standard normal cumulative distribution function.

Let $\hat{\beta}$ be the OLS estimate of β. By replacing β by $\hat{\beta}$ in (3) we get a predictor for the probability that an individual is poor:

(10) $$\hat{P}_i = F\left(\frac{\ln z - X_i\hat{\beta}}{\sigma}\right).$$

In the following, we assume that the error in the OLS estimate $\hat{\sigma}$ of σ is negligible. We can do this because we deal with large samples; refer to the formal motivation in 7.1.2 below. The error in the estimates $\hat{\beta}$ of β, will in general not be negligible as one may allow variables to be included in the model as long as they are significant at e.g. a five percent level. Let $\Phi(x)$ denote the standard cumulative normal distribution function. Then, it can be shown (see Wooldridge, 2002, p. 470) that the expectation of the predictor is:

(11) $$E(\hat{P}_i | X_i) = E\Phi\left(\frac{\ln z - X_i\hat{\beta}}{\sigma}\right) = \Phi\left(\frac{\ln z - X_i\beta}{\sigma\sqrt{\operatorname{var}\left(\frac{X_i\hat{\beta}}{\sigma}\right)+1}}\right).$$

We know that $\Phi(x)$ is an increasing function. Since $\operatorname{var}(X_i\hat{\beta}/\sigma) > 0$, it follows from (11) that the probability rate tends to be overestimated if $\ln z > X_i\beta$, and too low if $\ln z < X_i\beta$. Moreover, we have that:

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

where $\tilde{X}$ is the matrix given by $\tilde{X}' = (\tilde{X}_1', \tilde{X}_2', ..., \tilde{X}_n')$ used in estimation of β, and *X* is the matrix given by $X' = (X_1', X_2', ..., X_n')$ from the light survey (see proof in 7.1.3 below).

Hence, an estimate for the bias is given by:

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

7.1.2 Motivation of the approximation $\hat{\sigma} \approx \sigma$.

We know that $\left((n-k)\hat{\sigma}^2\right)/\sigma^2 \sim \chi^2_{n-k}$. Then, when n is large, it can be shown that

$$\operatorname{var}\left(\frac{\hat{\sigma}}{\sigma}\right) = 1 - \left(1 - \frac{1}{n-k}\right)\exp\left(\frac{1}{2(n-k)}\right) \approx \frac{1}{2(n-k)}$$

(see Sverdrup (1967)), where n is the number of observations and k is the number of parameters. When n is large this variance is close to zero. For example if n = 1000, the relative variance is about 0.0005.

Q.E.D.

7.1.3 Proof of (12):

Let ε_i be an i.i.d. error term with zero mean and standard deviation σ. We have that:

$X_i\hat{\beta} = X_i\beta + X_i\left(\hat{\beta} - \beta\right)$ and $\hat{\beta} = \beta + \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\sigma\varepsilon$ gives $X_i\hat{\beta} = X_i\beta + X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\sigma\varepsilon$ (see Greene (2003) p. 44). Since X and ε are uncorrelated, then

$$\left[E\left(X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\varepsilon\right)\right]^2 = 0,$$

and we get:

$$\operatorname{var}\left(\frac{X_i\hat{\beta}}{\sigma}\right) = \operatorname{var}\left[\frac{X_i\beta + X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\sigma\varepsilon}{\sigma}\right] = \operatorname{var}\left[X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\varepsilon\right] =$$

$$E\left(X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\varepsilon\right)\left(X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\varepsilon\right)' - \left[E\left(X_i\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\varepsilon\right)\right]^2 = X_i\left(\tilde{X}'\tilde{X}\right)^{-1}X_i'$$

Q.E.D.

7.1.4 The derivation of equation (9)

Let

(14) $$M = \left[\frac{1}{N}\sum_{i\in\Omega} s_i y_i - \frac{1}{N}\sum_{i\in\Omega} s_i P_i\right]$$

(15) $$K=\left[\frac{1}{N}\sum_{i\in\Omega}s_iP_i-\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i\right]$$

and

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

We wish to calculate the variance of M+K+R.

Note that M is random only because $\{\varepsilon_i,\ i\in\Omega\}$. Also $\{y_i,\ i\in\Omega\}$ is independent of $\{\hat{P}_i,\ i\in S\}$ and $\{\hat{P}_i,\ i\in\Omega\}$ because $\hat{\beta}$ is estimated by the use of an independently conducted sample. Thus, M and (K+R) are independent. Next, we need to show that K and R are independent:

(17) $$\begin{aligned}&\operatorname{cov}(K,R)=E(KR)=EE\left[KR\middle|\hat{\beta}\right]=E\left[E\left(K\middle|\ \hat{\beta}\ \right)E\left(R\middle|\ \hat{\beta}\ \right)\right]=\\&E\left\{E\left[\frac{1}{N}\sum_{i\in\Omega}s_i\left(P_i-\hat{P}_i\right)\middle|\ \hat{\beta}\ \right]\left(E\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-\frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\right)\middle|\ \hat{\beta}\ \right)\right\}=\\&E\left\{\left[\frac{1}{N}\sum_{i\in\Omega}s_i\left(P_i-\hat{P}_i\right)\right]\left(\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-E\left(\frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\ \middle|\ \hat{\beta}\ \right)\right)\right)\right\}=\\&E\left\{\left[\frac{1}{N}\sum_{i\in\Omega}s_i\left(P_i-\hat{P}_i\right)\right]\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i\right)\right\}=0\quad.\end{aligned}$$

We have that

(18) $$\operatorname{var}(M)=\operatorname{var}\frac{1}{N}\sum_{i\in\Omega}\left(s_i\left(y_i-P_i\right)\right)=\left(\frac{1}{N}\right)^2\sum_{i\in\Omega}s_i^2\ \operatorname{var}y_i=\left(\frac{1}{N}\right)^2\sum_{i\in\Omega}s_i^2\left(P_i-P_i^2\right).$$

The variance of R can be expressed as

(19) $$\begin{aligned}&\operatorname{var}R=E\operatorname{var}\left(R\ \middle|\ \hat{\beta}\right)+\operatorname{var}E\left(R\ \middle|\ \hat{\beta}\right)=\\&E\operatorname{var}\left(\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-\frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\right)\middle|\ \hat{\beta}\right)+\operatorname{var}\left(E\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-\frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\right)\middle|\ \hat{\beta}\right)=\\&E\operatorname{var}\left(\left(\frac{1}{n}\sum_{i\in S}s_i\hat{P}_i\right)\middle|\ \hat{\beta}\right)+\operatorname{var}\left(\left(\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i-\frac{1}{N}\sum_{i\in\Omega}s_i\hat{P}_i\right)\right)=\\&\left(1-\frac{n^H}{N^H}\right)\frac{n^H}{n^2}E\operatorname{var}\left(\hat{P}_i\ \middle|\ \hat{\beta}\right)+0\end{aligned}$$

and we have proved (9).

7.1.5 Estimating the variance of the prediction error

One can use Monte Carlo simulations to compute the variance given in (9). Since the parameter vector, $\hat{\beta}$, is multinormally distributed (asymptotically) it implies that $X_i\hat{\beta}/\sigma$ is normally distributed with mean $X_i\beta/\sigma$ and variance τ_i^2 given in (12). We can therefore write

(20) $$\frac{X_i\hat{\beta}}{\sigma} = \frac{X_i\beta}{\sigma} + \tau_i\eta_i,$$

where η_i is a random term that is normally distributed *N(0,1).* Hence, we can generate random draws and compute a predictor as follows. Let

(21) $$D_{ij} = F\left(\frac{\ln z - X_i\beta}{\sigma} - \tau_i\eta_{ij}\right), \qquad \overline{D}_{i.} = \frac{1}{M}\sum_{j=1}^{M} D_{ij}$$

where η_{ij}, *j=1,2,...M,* is i.i.d. random draws from *N(0,1)* (see 7.1.6 below for a justification). Here, D_{ij} is an analogue to $\hat{P}_i$ in (10), 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. Since the error in $\hat{\beta}$ is common to all households, η_{ij} is dependent across *i* for given *j*; however, this poses no problems for our simulations. First, $\overline{D}_{i.}$, being the average over all the random draws for each household can be used as an estimator for $P_{i.}$ since each *j* is independent. In the simulation of the second term of (9), variance is taken over *j* and, since each *j* is independent, there is no problem with the interdependency over *I*; see (23). The expectation we make in the latter term of (9) does not depend on whether there is independence or not, as the variance is conditioned on $\hat{\beta}$; see (24). Thus, by means of $\{D_{ij}\}$ and inserting $(\hat{\beta},\hat{\sigma})$ for (β,σ), one can simulate:

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

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

and

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

The total variance of the prediction error can therefore be simulated by

$$\text{(25)} \qquad \begin{aligned} &\frac{1}{N}\frac{1}{n}\sum_{i\in S} s_i^2\left(\overline{D}_{i.} - \overline{D}_{i.}^2\right) + \\ &\frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{n}\sum_{i\in S} s_i\left(\overline{D}_{i.} - D_{ij}\right)\right)^2 + \\ &\left(1-\frac{n^H}{N^H}\right)\frac{n^H}{n^2}\frac{1}{M}\frac{1}{n^H}\sum_{i\in S}\sum_{j=1}^{M}\left(s_i D_{ij} - \frac{1}{n^H}\sum_{i\in S} s_i D_{ij}\right)^2 \quad . \end{aligned}$$

In the case where we do not have a simple random sample, the last term of (9) can be estimated by using the syntax for a stratified and/or clustered sample in, for example, SAS or STATA to compute the variance for each random draw j, and estimate the expectation as the average over all j.

7.1.6 Justification of (21)

Let ε_i be an i.i.d. error term with zero mean and standard deviation σ. We have that:

$X_i\hat{\beta} = X_i\beta + X_i\left(\hat{\beta} - \beta\right)$ and $\hat{\beta} = \beta + \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\sigma\varepsilon$ gives $X_i\hat{\beta} = X_i\beta + X_i\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\sigma\varepsilon$ (see Greene (2003) p. 44).

Thus, we get

$$\hat{P}_i = \Phi\left(\frac{\ln z - X_i\hat{\beta}}{\sigma}\right) = F\left(\frac{\ln z - X_i\beta}{\sigma} - X_i\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\varepsilon\right) \quad .$$

We have that $X_i\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\varepsilon \overset{d}{=} \tau_i\eta_i$, where τ_i is the standard deviation to $X_i\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\varepsilon$ and $\eta_i \sim N(0,1)$.

Q.E.D.

7.1.7 Expanding the model to allow for heteroskedasticity

Assume heteroskedasticity in the sense of a random coefficient model:

$$\textbf{(26)} \qquad Y_i = X_i\beta_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0,\sigma).$$

Define

$$\textbf{(27)} \qquad \eta_i = \beta - \beta_i .$$

Hence, we can write

$$\textbf{(28)} \qquad Y_i = X_i\beta + e_i ,$$

where e_i is given by

$$\text{(29)} \qquad e_i = \varepsilon_i + \sum_{k=1}^{K} X_{ik}\eta_{ik} , \qquad \eta_{ik} \sim N(0,\varpi_i)$$

and k refers to explanatory variable. Note that, if we substitute X with Z, where Z is a set of explanatory variables not necessarily included in X, we get a more general expression for the heteroskedastic model.

The variance of the error term can be written as

$$\text{(30)} \qquad \operatorname{var} e_i = \chi_i^2 = \sigma^2 + \sum_{k=1}^{K} X_{ik}^2 \omega_i^2 .$$

Thus, we have

$$\text{(31)} \qquad Y_i = X_i\beta + \chi_i u_i , \qquad u_i \sim N(0,1).$$

Hence,

$$\text{(32)} \qquad P(Y_i > 0|X) = P\left(\frac{X_i\beta}{\chi_i} + u_i > 0 \middle| X\right) = P\left(\frac{X_i\beta}{\chi_i} > -u_i \mid X\right) = \Phi\left(\frac{X_i\beta}{\chi_i}\right).$$

Since

$$\text{(33)} \qquad \operatorname{var} \hat{e}_i = E\hat{e}_i^2 - (E\hat{e}_i)^2 = E\hat{e}_i^2 \quad ,$$

we can estimate $\operatorname{var} \hat{e}_i$ by

$$\text{(34)} \qquad \hat{e}_i^2 = \hat{\sigma}^2 + \sum_k X_{ik}^2 \hat{\omega}_i^2 + v_i , \qquad v_i \sim N(0,1).$$

Then, one can replace for the (β, χ_i) by $(\hat{\beta}, \hat{\chi}_i)$ in (32) where $\hat{\chi}_i$ is given by

$$\text{(35)} \qquad \hat{\chi}_i = \sqrt{\hat{\sigma}^2 + \sum_k X_{ik}^2 \hat{\omega}_k^2} .$$

The formulas for the standard errors are analogues to (9), when replacing X_i by X_i/χ_i . We disregard the uncertainty attached to $\hat{\chi}_i$ as long as χ_i is precisely estimated.

Another simulation approach, also applied by Elbers et al. (2003), is to generate random draws of $(\hat{\beta} - \beta, \hat{\sigma} - \sigma, \hat{\omega} - \omega)$ from the multivariate normal distribution with zero mean and covariance matrix (Σ) obtained in the estimation procedure. This can be done as follows. Let A be a matrix determined by $AA' = \hat{\Sigma}$. Then, one can write

$$\left(\hat{\beta} - \beta, \hat{\sigma} - \sigma, \hat{\omega} - \omega\right) = A\theta,$$

where θ is a vector with components that are independent with standard normal distribution. Hence, when A has been obtained, one can carry out simulations of the multivariate normal vectors by drawing independent standard normal random variables.

7.2 Testing assumptions

Figure 1 Plot of residual versus predicted value, Central Rural

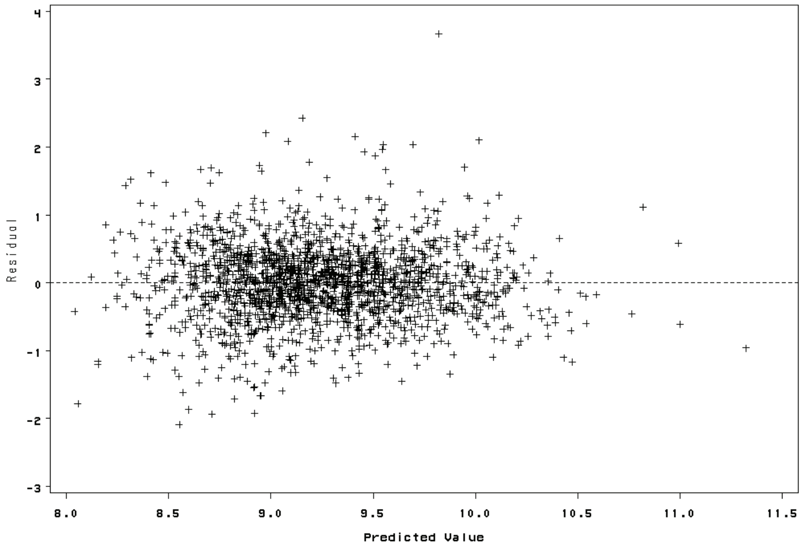


Figure 2 Plot of residual versus predicted value, Central Urban

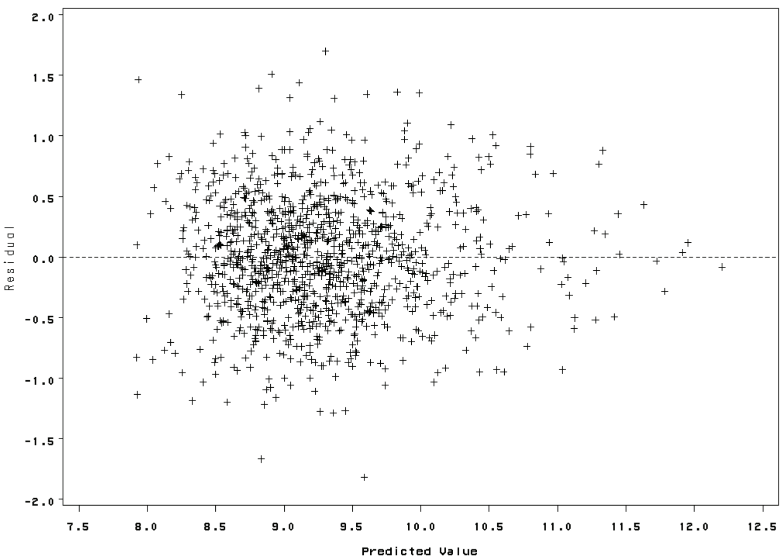


Figure 3 PP-plot for residual, Central Rural

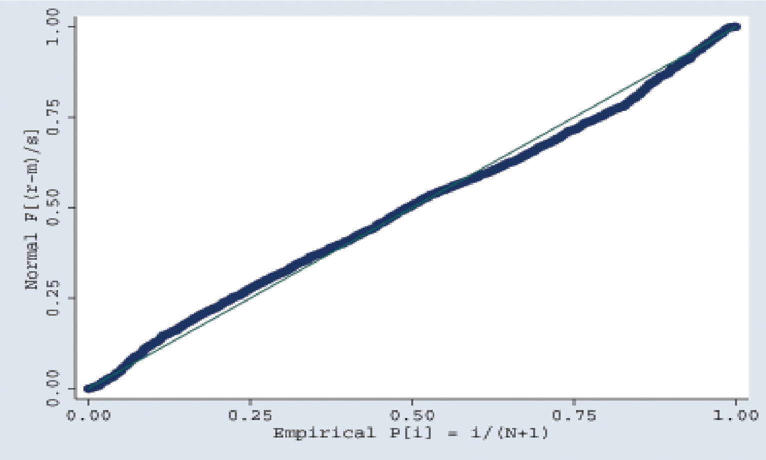
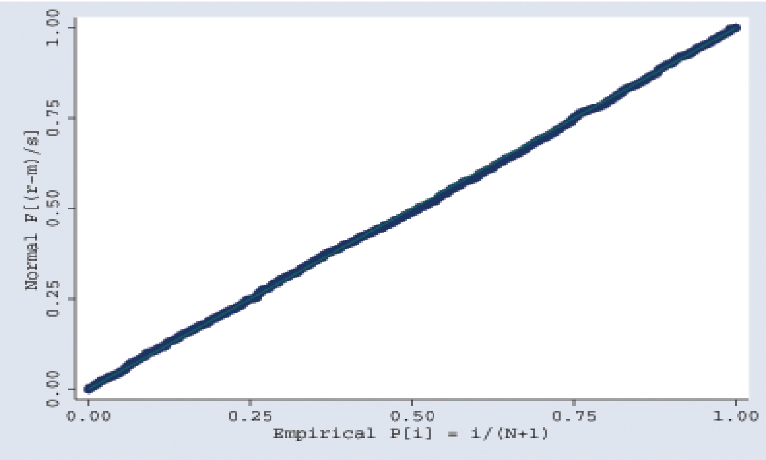


Figure 4 PP-plot for residual, Central Urban



7.3 Regression results

Table 2 Regression results, Central Rural Mozambique (OLS)

# Observations:	1917		
Root Mean Square Error:	0.56239		
Adj. R-Square:	0.3861		
Variable	**Parameter Estimate**	**Standard Error**	**t-value**
Intercept	9.8	0.07	140.2
Asset index	0.07	0.01	7.8
No. of spouses in household	0.07	0.02	2.8
If head never married	0.38	0.15	2.6
No. of members in household	-0.21	0.01	-15.0
No. of members in household, squared	0.01	0.00	8.8
If head works in tertiary sector	0.11	0.05	2.4
One or two generations and no children	0.21	0.05	4.4
If household owns bicycle	0.08	0.03	2.6
If household owns hi-fi set	0.29	0.06	5.0
If improved water	0.05	0.02	2.6
If paraffin used for lightening	0.10	0.03	3.6
If no toilet	-0.10	0.03	-2.9
Tete (Province)	-0.22	0.04	-5.9
Sofala (Province)	0.38	0.04	9.6

This model also includes dummies for agro-ecological zones

Table 3 Regression results, Central Urban Mozambique (OLS)

# Observations:	1172		
Root Mean Square Error:	0.48997		
Adj. R-Square:	0.619		

Variable	**Parameter Estimate**	**Standard Error**	**t-value**
Intercept	9.90	0.07	144.9
Members in household	-0.24	0.02	-14.3
Members in household., squared	0.01	0.001	7.7
Asset index	0.10	0.01	12.5
Asset index, squared	0.00	0.00	-5.3
if cooking with charcoal	0.15	0.03	4.5
if cooking with electricity	0.47	0.10	4.6
if cooking with paraffin	0.67	0.22	3.0
One or two generations, with children below 15	-0.08	0.03	-2.5
Single person	0.22	0.09	2.6
No. of non-relatives in household	0.15	0.06	2.4
If household owns car	0.55	0.09	6.3
If household owns a TV	0.17	0.06	3.0
If candle used for lighting	0.20	0.08	2.4
If asbestos roof	0.08	0.03	2.5
if grass roof	-0.08	0.03	-2.8
One adult illiterate	-0.09	0.03	-2.9
Mobile phones per person	0.48	0.19	2.6
Tete (Province)	-0.33	0.04	-8.3

References

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

Deaton, A. (1997): *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy* The Johns Hopkins University Press, USA.

Deaton, A. and Grosh, M. (2000): "Consumption" in Grosh, M. and Glewwe, P. (editors): *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study* Vol. 1, pp 91–133. Oxford University Press for the World Bank

Fofack, H. (2000): "Combining Light Monitoring Surveys with Integrated Surveys to Improve Targeting for Poverty Reduction: The Case of Ghana" *The World Bank Economic Review*, Vol. 14 No. 1 pp. 195–219.

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

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

Johnson, M., McKay, A. and Round, J. (1990): "Income and Expenditure in a System of Household Accounts: Concepts and Estimation" *Social Dimensions of Adjustment Working Paper* No. 10, Washington D.C: The World Bank.

McKay, A. (2001): "Report on Investigating the Development of a Poverty Correlates Model for Uganda" *Unpublished paper.*

Minot, N. (2003): "Are Poor, Remote Areas Left Behind in Agricultural Development: The Case of Tanzania" *MTID Discussion Paper*, No. 90, IFPRI.

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

Ravallion, M. (1992): "Poverty Comparison: a Guide to Concepts and Methods" *LSMS Working Paper* No. 88, The World Bank.

Ravallion, M. (1998): "Poverty Lines in Theory and Practice" *LSMS Working Paper* No. 133, The World Bank.

Ravallion, M. and Bindani, B. (1994): "How Robust is a Poverty Line?" *World Bank Economic Review,* Vol. 8, pp. 75–102.

Scott, C., Strode, M. and Chinganya, O. (2000): "Estimating Annual Expenditure from Monthly Survey Data" *Inter-Stat* No. 22, DFID, Eurostat.

Sverdrup, E. (1967): *Laws and Chance Variations* Universitetsforlaget, Oslo.

Tarozzi, A. (2004): "Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India" *Working Paper* Duke University.

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

Recent publications in the series Discussion Papers

324 E. Røed Larsen (2002): Estimating Latent Total Consumption in a Household.

325 E. Røed Larsen (2002): Consumption Inequality in Norway in the 80s and 90s.

326 H.C. Bjørnland and H. Hungnes (2002): Fundamental determinants of the long run real exchange rate:The case of Norway.

327 M. Søberg (2002): A laboratory stress-test of bid, double and offer auctions.

328 M. Søberg (2002): Voting rules and endogenous trading institutions: An experimental study.

329 M. Søberg (2002): The Duhem-Quine thesis and experimental economics: A reinterpretation.

330 A. Raknerud (2002): Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption

331 M.W. Arneberg, J.K. Dagsvik and Z. Jia (2002): Labor Market Modeling Recognizing Latent Job Attributes and Opportunity Constraints. An Empirical Analysis of Labor Market Behavior of Eritrean Women

332 M. Greaker (2002): Eco-labels, Production Related Externalities and Trade

333 J. T. Lind (2002): Small continuous surveys and the Kalman filter

334 B. Halvorsen and T. Willumsen (2002): Willingness to Pay for Dental Fear Treatment. Is Supplying Fear Treatment Social Beneficial?

335 T. O. Thoresen (2002): Reduced Tax Progressivity in Norway in the Nineties. The Effect from Tax Changes

336 M. Søberg (2002): Price formation in monopolistic markets with endogenous diffusion of trading information: An experimental approach

337 A. Bruvoll og B.M. Larsen (2002): Greenhouse gas emissions in Norway. Do carbon taxes work?

338 B. Halvorsen and R. Nesbakken (2002): A conflict of interests in electricity taxation? A micro econometric analysis of household behaviour

339 R. Aaberge and A. Langørgen (2003): Measuring the Benefits from Public Services: The Effects of Local Government Spending on the Distribution of Income in Norway

340 H. C. Bjørnland and H. Hungnes (2003): The importance of interest rates for forecasting the exchange rate

341 A. Bruvoll, T.Fæhn and Birger Strøm (2003): Quantifying Central Hypotheses on Environmental Kuznets Curves for a Rich Economy: A Computable General Equilibrium Study

342 E. Biørn, T. Skjerpen and K.R. Wangen (2003): Parametric Aggregation of Random Coefficient Cobb-Douglas Production Functions: Evidence from Manufacturing Industries

343 B. Bye, B. Strøm and T. Åvitsland (2003): Welfare effects of VAT reforms: A general equilibrium analysis

344 J.K. Dagsvik and S. Strøm (2003): Analyzing Labor Supply Behavior with Latent Job Opportunity Sets and Institutional Choice Constraints

345 A. Raknerud, T. Skjerpen and A. Rygh Swensen (2003): A linear demand system within a Seemingly Unrelated Time Series Equation framework

346 B.M. Larsen and R.Nesbakken (2003): How to quantify household electricity end-use consumption

347 B. Halvorsen, B. M. Larsen and R. Nesbakken (2003): Possibility for hedging from price increases in residential energy demand

348 S. Johansen and A. R. Swensen (2003): More on Testing Exact Rational Expectations in Cointegrated Vector Autoregressive Models: Restricted Drift Terms

349 B. Holtsmark (2003): The Kyoto Protocol without USA and Australia - with the Russian Federation as a strategic permit seller

350 J. Larsson (2003): Testing the Multiproduct Hypothesis on Norwegian Aluminium Industry Plants

351 T. Bye (2003): On the Price and Volume Effects from Green Certificates in the Energy Market

352 E. Holmøy (2003): Aggregate Industry Behaviour in a Monopolistic Competition Model with Heterogeneous Firms

353 A. O. Ervik, E.Holmøy and T. Hægeland (2003): A Theory-Based Measure of the Output of the Education Sector

354 E. Halvorsen (2003): A Cohort Analysis of Household Saving in Norway

355 I. Aslaksen and T. Synnestvedt (2003): Corporate environmental protection under uncertainty

356 S. Glomsrød and W. Taoyuan (2003): Coal cleaning: A viable strategy for reduced carbon emissions and improved environment in China?

357 A. Bruvoll T. Bye, J. Larsson og K. Telle (2003): Technological changes in the pulp and paper industry and the role of uniform versus selective environmental policy.

358 J.K. Dagsvik, S. Strøm and Z. Jia (2003): A Stochastic Model for the Utility of Income.

359 M. Rege and K. Telle (2003): Indirect Social Sanctions from Monetarily Unaffected Strangers in a Public Good Game.

360 R. Aaberge (2003): Mean-Spread-Preserving Transformation.

361 E. Halvorsen (2003): Financial Deregulation and Household Saving. The Norwegian Experience Revisited

362 E. Røed Larsen (2003): Are Rich Countries Immune to the Resource Curse? Evidence from Norway's Management of Its Oil Riches

363 E. Røed Larsen and Dag Einar Sommervoll (2003): Rising Inequality of Housing? Evidence from Segmented Housing Price Indices

364 R. Bjørnstad and T. Skjerpen (2003): Technology, Trade and Inequality

365 A. Raknerud, D. Rønningen and T. Skjerpen (2003): A method for improved capital measurement by combining accounts and firm investment data

366 B.J. Holtsmark and K.H. Alfsen (2004): PPP-correction of the IPCC emission scenarios - does it matter?

367 R. Aaberge, U. Colombino, E. Holmøy, B. Strøm and T. Wennemo (2004): Population ageing and fiscal sustainability: An integrated micro-macro analysis of required tax changes

368 E. Røed Larsen (2004): Does the CPI Mirror Costs.of.Living? Engel's Law Suggests Not in Norway

369 T. Skjerpen (2004): The dynamic factor model revisited: the identification problem remains

370 J.K. Dagsvik and A.L. Mathiassen (2004): Agricultural Production with Uncertain Water Supply

371 M. Greaker (2004): Industrial Competitiveness and Diffusion of New Pollution Abatement Technology – a new look at the Porter-hypothesis

372 G. Børnes Ringlund, K.E. Rosendahl and T. Skjerpen (2004): Does oilrig activity react to oil price changes? An empirical investigation

373 G. Liu (2004) Estimating Energy Demand Elasticities for OECD Countries. A Dynamic Panel Data Approach

374 K. Telle and J. Larsson (2004): Do environmental regulations hamper productivity growth? How accounting for improvements of firms' environmental performance can change the conclusion

375 K.R. Wangen (2004): Some Fundamental Problems in Becker, Grossman and Murphy's Implementation of Rational Addiction Theory

376 B.J. Holtsmark and K.H. Alfsen (2004): Implementation of the Kyoto Protocol without Russian participation

377 E. Røed Larsen (2004): Escaping the Resource Curse and the Dutch Disease? When and Why Norway Caught up with and Forged ahead of Its Neughbors

378 L. Andreassen (2004): Mortality, fertility and old age care in a two-sex growth model

379 E. Lund Sagen and F. R. Aune (2004): The Future European Natural Gas Market - are lower gas prices attainable?

380 A. Langørgen and D. Rønningen (2004): Local government preferences, individual needs, and the allocation of social assistance

381 K. Telle (2004): Effects of inspections on plants' regulatory and environmental performance - evidence from Norwegian manufacturing industries

382 T. A. Galloway (2004): To What Extent Is a Transition into Employment Associated with an Exit from Poverty

383 J. F. Bjørnstad and E.Ytterstad (2004): Two-Stage Sampling from a Prediction Point of View

384 A. Bruvoll and T. Fæhn (2004): Transboundary environmental policy effects: Markets and emission leakages

385 P.V. Hansen and L. Lindholt (2004): The market power of OPEC 1973-2001

386 N. Keilman and D. Q. Pham (2004): Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area

387 G. H. Bjertnæs and T. Fæhn (2004): Energy Taxation in a Small, Open Economy: Efficiency Gains under Political Restraints

388 J.K. Dagsvik and S. Strøm (2004): Sectoral Labor Supply, Choice Restrictions and Functional Form

389 B. Halvorsen (2004): Effects of norms, warm-glow and time use on household recycling

390 I. Aslaksen and T. Synnestvedt (2004): Are the Dixit-Pindyck and the Arrow-Fisher-Henry-Hanemann Option Values Equivalent?

391 G. H. Bjønnes, D. Rime and H. O.Aa. Solheim (2004): Liquidity provision in the overnight foreign exchange market

392 T. Åvitsland and J. Aasness (2004): Combining CGE and microsimulation models: Effects on equality of VAT reforms

393 M. Greaker and Eirik. Sagen (2004): Explaining experience curves for LNG liquefaction costs: Competition matter more than learning

394 K. Telle, I. Aslaksen and T. Synnestvedt (2004): "It pays to be green" - a premature conclusion?

395 T. Harding, H. O. Aa. Solheim and A. Benedictow (2004). House ownership and taxes

396 E. Holmøy and B. Strøm (2004): The Social Cost of Government Spending in an Economy with Large Tax Distortions: A CGE Decomposition for Norway

397 T. Hægeland, O. Raaum and K.G. Salvanes (2004): Pupil achievement, school resources and family background

398 I. Aslaksen, B. Natvig and I. Nordal (2004): Environmental risk and the precautionary principle: "Late lessons from early warnings" applied to genetically modified plants

399 J. Møen (2004): When subsidized R&D-firms fail, do they still stimulate growth? Tracing knowledge by following employees across firms

400 B. Halvorsen and Runa Nesbakken (2004): Accounting for differences in choice opportunities in analyses of energy expenditure data

401 T.J. Klette and A. Raknerud (2004): Heterogeneity, productivity and selection: An empirical study of Norwegian manufacturing firms

402 R. Aaberge (2005): Asymptotic Distribution Theory of Empirical Rank-dependent Measures of Inequality

403 F.R. Aune, S. Kverndokk, L. Lindholt and K.E. Rosendahl (2005): Profitability of different instruments in international climate policies

404 Z. Jia (2005): Labor Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure

405 Z. Jia (2005): Retirement Behavior of Working Couples in Norway. A Dynamic Programming Approch

406 Z. Jia (2005): Spousal Influence on Early Retirement Behavior

407 P. Frenger (2005): The elasticity of substitution of superlative price indices

408 M. Mogstad, A. Langørgen and R. Aaberge (2005): Region-specific versus Country-specific Poverty Lines in Analysis of Poverty

409 J.K. Dagsvik (2005) Choice under Uncertainty and Bounded Rationality

410 T. Fæhn, A.G. Gómez-Plana and S. Kverndokk (2005): Can a carbon permit system reduce Spanish unemployment?

411 J. Larsson and K. Telle (2005): Consequences of the IPPC-directive's BAT requirements for abatement costs and emissions

412 R. Aaberge, S. Bjerve and K. Doksum (2005): Modeling Concentration and Dispersion in Multiple Regression

413 E. Holmøy and K.M. Heide (2005): Is Norway immune to Dutch Disease? CGE Estimates of Sustainable Wage Growth and De-industrialisation

414 K.R. Wangen (2005): An Expenditure Based Estimate of Britain's Black Economy Revisited

415 A. Mathiassen (2005): A Statistical Model for Simple, Fast and Reliable Measurement of Poverty