Session Number: 2B
Session Title: Improving Estimates from Survey Data
Session Organizer(s): Stephen Jenkins, University of Essex, Colchester, UK, and Holly Sutherland, University of Essex, Colchester, UK
Session Chair: Holly Sutherland, University of Essex, Colchester, UK

Paper Prepared for the 29th General Conference of
The International Association for Research in Income and Wealth

Joensuu, Finland, August 20 – 26, 2006

Improving inequality and poverty estimates in the Moldovan Household Budget Survey

Ludovico Carraro (OPM), Lilian Galer and Ala Negruta (National Bureau of Statistics)

For additional information please contact:

Author Name(s) : Ludovico Carraro
Author Address(es) : Oxford Policy Management
6 St Aldates Courtyard
38 St Aldates
Oxford OX1 1BN
United Kingdom
Author E-Mail(s) : ludovico.carraro@opml.co.uk
Author FAX(es) : + 44 (0) 1865 250 580
Author Telephone(s) : + 44 (0) 1865 207 300

Table of Contents

List of Tables and Figures

Abbreviations

CPI	Consumer Price Index
EU	European Union
GDP	Gross Domestic Product
HBS	Household Budget Survey
NBS	National Bureau of Statistics, Moldova
NIPA	National Income and Product Accounts
OECD	Organization for Economic Cooperation and Development
OPM	Oxford Policy Management
PSU	Primary Sampling Unit

Abstract

This paper describes how the problem of survey non-response was addressed in the Moldovan Household Budget Survey (HBS) to correct for biases on poverty and inequality.

Non-response rate in the HBS is about 12%, but it is much higher in cities (35%) than in rural areas (3%) and it is a common finding that non-compliance is usually selective. Therefore, the practice of substitution and simple geographical re-weighting correction can generate biases on key estimates of this survey. To investigate such issues, this paper studies the reasons for non-response as well as the main characteristics of non-respondents. Moreover, it explores a number of techniques to correct for non-response biases and assess the sensitivity of their impact. These techniques use information from “the response to basic questions”, interviewers’ external observation, geographical characteristics as well as unique auxiliary information available for all addresses in Moldova: electricity consumption.

This dataset (kindly provided by the electricity companies operating in Moldova) is used to analyse differences in the distribution of electricity consumption in the population and in the sample, and to study the probability of non-response in different population sub-groups. Using such information the paper is able to assess the distortion determined by the practice of substitution and geographical re-weighting on various welfare estimates and proposes different methods to correct for non-response.

1. Introduction

This paper aims at investigating the problem of unit non-response in the Moldovan Household Budget Survey (HBS) with the specific objective of reducing bias in poverty and inequality estimates.

In 2005 the overall non-response rate was about 12%, but it is much higher in cities (35%) than in rural areas (3%). Although these rates are still relatively low compared to those of most European countries[1], non-response in the HBS has shown a clear increasing trend since 1997. In fact, when the first HBS took place, non-response in cities was only 23%. Moreover, the expectation is that, as the economy keeps developing, non-response will also tend to increase.

Up to now the issue of non-response has been largely neglected in Moldova, and limited to approximate estimates of the overall non-response rates in major surveys. The practical correction adopted in the HBS is substitution of non-respondent households, but there is no study of the actual impact of such correction on the non-response bias. On a purely theoretical level the effect of substitution on means is unlikely to be different from that of simple geographical re-weighting, and if non-response is selective and correlated to the main variables under investigation (namely income), it is expected that these methods poorly correct for non-response. Indeed, they are likely to leave a substantial bias on measures of mean income and possibly on distributional analysis, in particular measures of poverty and inequality, although for the latter ex-ante theoretical arguments cannot establish the actual impact (see Deaton 2004 and Ravallion et. al 2005). Therefore, the task of this paper is to study alternative methods of correction and assess their impact on key measures of poverty and inequality so that better solutions can be implemented.

The lack of information on the effects of non-response in the case of the HBS in Moldova is not an exception. In fact in general there is little knowledge on how survey non-response affects survey estimates, and in particular how it affects poverty and inequality. Survey non-response is said to be a bigger problem in developed economies, although to some extent its effects have been neglected in many developing countries and, sometimes reported high response rates, are the consequence of inappropriate practices of substitution that occur in the field without proper documentation. Moreover, in many CIS and eastern European countries non-response rates are relatively high (see for instance Eurostat 2004).

The problem of non-response has received much more scrutiny in countries where non-response rates are and have been particularly high since the 1980s. Various studies tried to identify the main characteristics of non-respondents with the use of some auxiliary information (population registers and Census). In these studies it is a common finding that lower response rates are expected in urbanised areas, among single persons, childless households, older persons, divorced or widowed, people with lower educational attainment, and the self-employed. Such studies often informed the correction method to be adopted. For instance, in northern European countries statistical agencies have studied in some detail how to correct for non-response, though such correction has not generally focused on some specific variable, but as a treatment for the problem of non-response in general. The prevalent approach is some re-weighting based on the probability of response of certain population sub-groups, and in most cases post-stratification and calibration are used to correct for non-response as well as for sampling errors. In the UK the probability of response is studied using auxiliary information that comes from matching non-respondent

[1] In European countries non-response rates in similar surveys are commonly between 30 and 40% (Eurostat 2003, pag. 56).

households with census information about the same households[2] (see Foster 1998, and the same was done in the US as documented in Groves and Couper 1998). In many northern European countries auxiliary information contained in population registers is used in one step to correct both for non-response and sampling errors (see Lundström and Sändarl 1999).

However such method of correction is often simply based on demographic characteristics, which cannot fully explain the reason for non-response. The same adjustment is made for groups of people (adjustment cells) that are essentially formed ad hoc (Ravallion et al 2005). If such auxiliary information is a good predictor of the variable under analysis the procedure is satisfactory in correcting the non-response bias, otherwise it is a relatively inefficient method.

Ravallion et al (2005) recently suggest a different method of adjustment based on the estimation of a direct relationship between non-response and income obtained from non-response rates recorded in the various survey strata. In their application to the Current Population Survey in the US they found that their correction increases mean income as well as inequality measures.

In the context of international comparisons and calculation of world poverty and inequality, some authors have also partly addressed the problem of non-response by substituting survey means with those of the National Accounts (NIPA estimates of consumption and even GDP) (Borguignon and Morrisson (2002), Sala-i-Martin (2002) and Bhalla (2002)). However, the method of substituting survey means with those of the NIPA makes two strong assumptions: 1) the reasons for the discrepancies between NIPA and survey do not have effects on the income distribution, and 2) NIPA measures of mean income/consumption are more accurate than those of the surveys. The first assumption implies that non-response and under-reporting occur proportionally in the same way throughout the income distribution. However, this goes against the available evidence on survey non-compliance in western countries. On the second hypothesis Deaton (2004) points out the limitations of the NIPA in measuring consumption and income. Although the available evidence suggests that the substitution of survey means with estimates from the NIPA is incorrect, the methodology remains defendable due to the lack of evidence of the effect of non-response in developing countries. Therefore, NIPA based poverty estimates remain provoking and point to a serious mismatch between survey estimates and those of the national accounts. Deaton (2004) argues that more research is needed on the effects of survey non-compliance and this could be an area that reconciles some of the existing differences between national account estimates of final household consumption and survey estimates.

The approach used to investigate the effect of non-response on income and distributional analysis ultimately depends on the specific information at our disposal, which in the case of the HBS is particularly rich and allows us to see the effect of alternative methods, providing evidence on the best approaches to use in the future.

The paper begins with a presentation of our data sources and then provides some descriptive analysis about trends of non-response over time, and their causes. The main characteristics of non-respondent are compared both with those of the respondents and the substitute households. Subsequently we implement four different ways to correct for non-response: substitution, simple geographical re-weighting, re-weighting based on the inverse probability of response in different mutually exclusive population sub-groups, econometric imputation of income of non-respondents. A final section offers some conclusions.

---

[2] For instance in the 2004-05 Family Expenditure Survey, firstly weights are used to compensate for non-response using 10 weighting classes (based on a study matching Census households with those of the FES in 1991); secondly a post-stratification exercise that matches population totals by age groups and gender.

2. Data

The main data source used in our analysis is the household budget survey (HBS), but also some auxiliary information: responses to 'basic questions', interviewers' external observations, and electricity consumption for all addresses of Moldova. The next two sub-sections describe the main characteristics of such data.

2.1 The Household Budget Survey

The HBS is the main quantitative survey used for poverty estimates and poverty analysis in the country. Besides its use for welfare analysis (both income and non-income aspects of living standards), the HBS provides weights for the consumer price index, and other information used for the National Accounts. Because of its comprehensive questionnaire, the HBS can also be described as a multi-purpose survey; it contains sections on household demographic characteristics, employment, housing characteristics, education, health, income and expenditure data. The survey is nationally representative and the sample has a two-stage stratified cluster design. The sample frame is based on the 1996 electoral lists[3], which were partly updated with the 2000 electoral lists. Clusters or primary sampling units (PSU) correspond to electoral sectors of 1500-3000 voters. When electoral sectors were smaller than the targeted population they were merged with other geographically adjacent sectors in order to reach the minimum number of 1500 voters. Similarly, if electoral sectors were bigger than 3000 voters, they were split into segments of approximate equal size. The selection of clusters was done in 1997, when the first HBS took place, and such primary sampling units have not changed until 2005[4]. The survey is a permanent exercise of the NBS in the sense that 12 households are interviewed each month in each PSU. In total there are four strata (Chisinau, Balți, towns and the rural areas[5]), 45 PSUs and 6480 interviews each year. Moreover, the sample is designed to provide reliable estimates for three domains: cities, towns and rural areas. However, although the sample size is relatively large the precision of the estimates is limited by the comparatively small number of PSUs.

Within strata the number of PSUs is allocated approximately in proportion to the size of the population of each stratum, and PSUs are selected with equal probability. However, the sample is not self-weighted and sampling weights are used to generate population estimates[6]. Within PSUs households are selected as a simple random sample without replacement. Half of the sample is part of a panel survey in which the household is interviewed in the same month of the year for four consecutive years (there are actually two overlapping waves of panel data: the first wave started in the second quarter of 1997, the second wave in the second quarter of 1999, and both lasted for 4 years and constituted each a quarter of the interviewed households, new waves started

---

[3] These were preferred to the 1989 Census, which occurred before the country independence.

[4] A new sample design has commenced in 2006 based on the new sample frame provided by the 2004 Census.

[5] The PSUs are divided within four strata as follows: 9 PSUs in Chisinau, 2 PSUs in Balți, 8 PSUs in towns, and 26 PSUs in rural areas.

[6] The sampling weights are computed in a standard way, representing the inverse of the probability of selection ($w_{ik} = \frac{1}{p_{ik}}$ and $p_{ik} = \frac{m_k}{M_k} \cdot \frac{n_{ik}}{N_{ik}}$).

respectively in the second quarter of 2001 and 2003). Moreover, all the other households are interviewed a second time after three months[7].

In each interview, households are visited three times: at the beginning of the month, in the middle and at the end of the month. During the first visit the interviewer asks about household composition, housing characteristics, employment and some infrequent expenditure on non-food items (recall modules with different reference periods)[8]. The interviewer also introduces and explains the use of the 'diary' to the household, who will record income and expenditure transactions during the whole month. In the second visit the enumerators check that the diary is properly kept and answer any doubt the household may have, and take with them the first half of the diary. At the end of the month a final interview gathers information about education, health, durable goods and subjective perceptions of household welfare, and the second part of the diary is checked and collected.

Data is considered of good quality given the extensive checks performed by the enumerators, supervisors and the central office. The data entry software contains a number of consistency checks that alert the enumerator of any possible problem and eventual uncertainties are resolved with the enumerator and sometime with further clarifications from the household.

Poverty and inequality are measured using consumption expenditure as the main welfare indicator. The consumption aggregate is computed in a comprehensive manner including food consumption, education and health expenditure, housing (including imputed rents for dwelling owners), consumption flow from durable items, and other expenditure for non-food items. Household consumption expenditure is also corrected by regional and over time price differences using a monthly PSU Paasche price index. Finally household consumption is corrected by household size using the 'old OECD equivalence scales'[9]. The old OECD equivalence scales are preferred to the 'modified OECD scale' because of the consumption patterns prevailing in the country (a large share of consumption is spent on food, while expenditure on 'quasi-public goods' -housing and consumption flow from durable items- is relatively low, compared to levels observed in OECD countries).

Some type of non-response is reduced in three different ways: 1) the interviewer's visit is anticipated by a letter to the selected household presenting the survey and explaining its benefits; 2) each household is given a monetary contribution of 25 lei (approximately 1.7 euros); 3) non-contacts are reduced because the interviewer has the duty to try to contact the household at least three times (three visits), which take place in a period of one week.

More generally the NBS is also reviewing the questionnaire to make it easier to fill in as well as to reduce the burden for the household whenever possible. Furthermore, it is also planning to provide more specialised training on required attitudes of the interviewer towards respondents, especially in difficult environments of urban areas.

---

[7] Although theoretical interviews are 6480 per year, the number of interviewed households per year should be 5265, since 1215 households are interviewed twice in the same year (other 405 are also interviewed twice but across two years, fourth quarter of previous year and first quarter of the year under analysis).

[8] Since the first survey the questionnaire has changed in order to improve its ability to monitor poverty. While at its beginning, the survey mainly aimed at generating average national estimates (for the CPI and the National Accounts), the questionnaire gradually introduced expenditure recall modules and other changes that make more robust intra-household comparison of consumption expenditure.

[9] Such equivalence scales consider as 1 the head of the household, 0.7 any other adult members and 0.5 all children aged 15 and less (this is because compulsory education is until this age).

2.2 Auxiliary data

Whenever the interviewer is unable to establish a contact with the selected household or the household refuses to participate in the survey activities, the enumerator has the duty of filling in a "non-response sheet". This one-page questionnaire collects some key information: the reason for non-response, the often called 'responses to basic questions', and interviewers' external observations. Basic questions and interviewers' observations include information about household size, whether there are children in the household, occupational status, type of dwelling, agricultural land, ownership of car, and interviewer's judgement of household's welfare.

Moreover, we have data on electricity consumption available for all addresses in Moldova (more than 1.2 million addresses). Such data was kindly provided by the three electricity companies operating in Moldova: RED-North, RED-North-West and Union Fenosa. The data has been assembled and consistently organised so that for each address we have data on monthly consumption (kWh) from January to May 2005.

The quality of supplier data is considered of high standards especially for two of the electricity suppliers, Union Fenosa and RED-North-West, whereas for RED-North the dataset does have some problems because in some cases it contains not only the current month of consumption, but also arrears and low consumption values. However, such inconsistencies are likely to be relatively small. In fact, from the HBS, which also collects information about arrears, we know that in 2005 arrears of more than two months represented only 0.15% of cases, and all arrears represent about 1.8% of cases. Moreover, RED-North supplies electricity to less than 25% of all addresses in Moldova (Union Fenosa serves about 60% of addresses and RED-North-West the remaining 15%).

3. Non-response rates and characteristics of non-respondents

3.1 Non-response and sample frame imperfections

In the non-response sheet reasons for non-interview are coded in 10 different categories, which for the purpose of this study are divided in non-responses and sample frame imperfections.

Non-responses are re-classified in three different categories:

1) Refusals: they consist of non-response due to three reasons: 'lack of time', 'don't consider it useful', and 'don't believe information will remain confidential'.

2) Non-contacts: they occur when no one can be found at the selected address (after three visits)

3) Other non-interviews: they represent cases where household members are not able to participate in the survey (due for instance to their old age, health problems, or illiteracy) and other reasons (household is on holiday or does not want to open the door without providing any explanation[10]).

A different problem is the one of sample frame imperfections in which the selected addresses constitute 'ineligible' households: the enumerator was unable to interview the selected household because information in the sample frame was not correct, so that the address provided does not exist or no one lives there[11]. In 2005 such cases were 8.2% of the households initially selected, though if we differentiate between first time selected households and panel households, sample frame imperfections were respectively 12.6% and 2.3%. The rate for sample imperfections for first time selected households is therefore very high and represents a relevant problem for the NBS: not only it implies a waste of resources, but also has implications on sampling probabilities. Moreover, on a theoretical ground, over-coverage, implicit in sample frame imperfections, also suggests a possible problem of under-coverage: there could be many households that are not in the sample frame because formed between the creation of the sample frame and the interview. Nevertheless, a convincing explanation for sampling frame imperfections is the high emigration abroad (a 2004 survey estimated that between 12% and 18% of the population is working abroad), and, although this is not the only reason of sample frame imperfections, it suggests that under-coverage is a relatively smaller problem (household formation and internal migration are not significant as external migration).

In the case of sampling frame imperfections, where the selected address does not exist or no one lives there, it could be appropriate to substitute the case with another address because such cases should not have been in the sample frame. Alternatively such cases could be simply dropped from

---

[10] The classification of 'other reasons' into non-interviews is problematic because it could contain both refusals and non-contacts, but overall it seems more likely to capture reasons that are not related to the reluctance to be interviewed.

[11] Such cases also include duplicate listings and wrong address (students' hostel, orphanage, hospice, etc.), but 80% of reasons for sample frame imperfections is because no one lives at the selected address.

the analysis[12]. We opted for this second approach, and we computed non-response rates as follows:

$$Nr = \frac{\text{number of non} - \text{responses}}{(\text{completed interviews} + \text{non} - \text{responses})}$$

Table 3.1 reports un-weighted non-response rates for the three domains of analysis between 1997 and 2005. The overall non-response rate increases slightly from 8.6 to 11.6%, but in cities the increase is much more pronounced. Indeed, while in 1997 non-response was only 22.7%, it became 35.1% in 2005. However, the interpretation of these non-response rates over time is not straightforward, because each year contains different proportions of first time interviews and panel interviews. Therefore in the same table we also report non-response rates differentiating between panel and first time interviews. Among first time interviews non-response doubled increasing from 8.6 to 16%, and the increase of non-response in cities went up from 22.7% to 45%.

As suggested by Groves and Couper (1998), we believe it is also important to analyse separately the three types of non-response and table 3.2 shows the percentages of the three types of non-response in 2005 in the three domains of analysis and by type of interview (panel and first-time interview). The majority of non-responses are refusals, which constitute about 60% of all non-responses, while both non-contacts and other non-interviews represent each about 20% of non-interviews, and there are not significant differences in the composition of non-responses between first-time and panel interviews, though within first-time interviews refusals are slightly more prevalent. It is also important to note the different composition of non-response in cities as opposed to towns and rural areas. In fact, it is in cities that non-contacts and non-interviews are relatively more important. This finding is in agreement with what we did expect, since the literature suggests that such types of non-response are often a result of urbanisation, and obstacles in contacting households posed by the type of dwelling structures (apartment blocks).

Table 3.1 Non-response rates by strata (un-weighted, %), 1997-2005

	1997	1998	1999	2000	2001	2002	2003	2004	2005
Cities	22.7	25.7	28.6	27.6	29.0	29.2	36.3	37.9	35.1
Towns	7.8	9.0	5.7	6.9	5.2	5.7	8.5	7.8	6.4
Rural areas	2.6	2.3	3.0	2.0	2.9	3.1	3.0	2.3	2.7
Overall	8.5	9.2	10.0	9.4	10.0	10.2	12.7	12.6	11.6
Panel	-	5.7	6.5	6.1	5.1	6.8	7.5	6.8	6.2
First contact	8.5	13.1	12.2	14.5	14.0	15.2	16.9	21.0	16.0
% of panel	0.0	54.6	40.1	62.0	47.9	62.7	47.9	63.3	48.0

Source: Calculations of the authors based on HBS data.

The composition of different types of non-response over-time does not seem to have changed significantly, though we can observe an increase of non-interviews at the expenses of a relative

[12] The two approaches should theoretically provide the same results, since when addresses are substituted we would again have some non-responses, which would add up to the earlier non-responses, and similarly new cases of sample frame imperfections could be again substituted until all addresses are either non-responses or completed interviews.

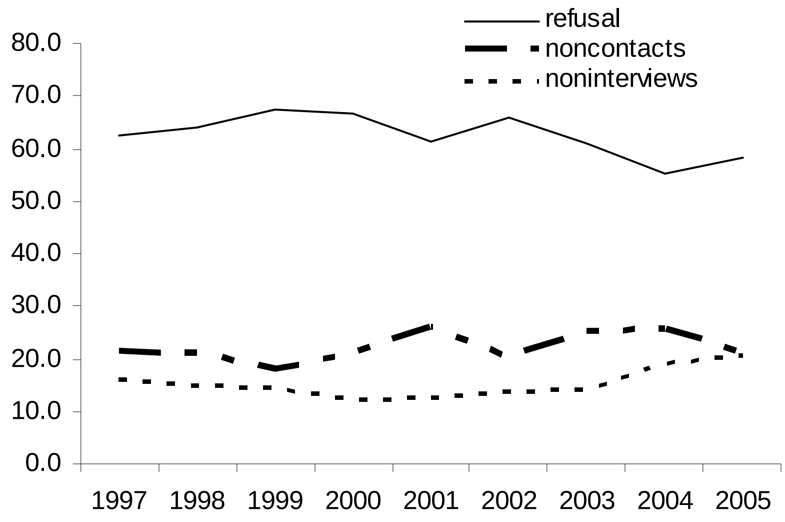
reduction in refusals. Non-interviews increased from about 15% in 1997 to 21% in 2005, while refusal went down from 63 to 58% (see figure 3.1).

Table 3.2 Reasons for non-response by strata and type of interview (%), 2005

	Cities	Towns	Rural areas	Overall
Overall	**35.1**	**6.4**	**2.7**	**11.6**
Type of non-response:	100.0	100.0	100.0	100.0
Refusal	52.6	80.4	76.3	58.3
Noncontacts	24.9	5.9	9.2	21.1
Noninterviews	22.4	13.7	14.5	20.6
First-time interview	**45.0**	**10.1**	**2.8**	**16.0**
Type of non-response:	100.0	100.0	100.0	100.0
Refusal	54.2	80.5	76.2	58.9
Noncontacts	24.1	4.9	11.9	21.0
Noninterviews	21.7	14.6	11.9	20.1
Panel interviews	**19.3**	**2.5**	**2.5**	**6.2**
Type of non-response:	100.0	100.0	100.0	100.0
Refusal	46.7	80.0	76.5	56.6
Noncontacts	28.3	10.0	5.9	21.3
Noninterviews	25.0	10.0	17.7	22.1

Source: Calculations of the authors based on HBS data.

Figure 3.1 Composition of non-response (%), 1997-2005



Source: Calculations of the authors based on HBS data.

3.2 Main characteristics of non-respondents

In order to study the characteristics of non-respondents we make use firstly of the dataset on electricity consumption and secondly of data captured in the “non-response sheet”, as well as information on some non-respondents obtained because the same households were interviewed either one year or three months earlier.

3.2.1 Comparing electricity consumption in HBS and electricity suppliers’ data

Before looking at the level of electricity consumption of non-respondent households we can compare the overall distribution of electricity consumption in the HBS and for the actual population of Moldova.

It is first of all interesting to observe that for about 8.5% of addresses, data from the electricity suppliers shows no consumption between January and May. It is likely that these represent cases of ‘abandoned addresses’, and indeed when we matched addresses of ineligible households with data on electricity consumption from suppliers we found that in many cases these households had zero consumption. This percentage is very close to the percentage of cases in the HBS that could not be interviewed because no one was living at the selected address in the case of first contacts (10.1%).

Therefore, in comparing electricity consumption in the two dataset we excluded cases in which consumption was zero for five consecutive months. We are aware that this is an imperfect adjustment since there might be some genuine cases of no consumption, but the choice of dropping them is preferable to leaving them in the dataset[13].

Figure 3.2 plots the cumulative probability functions of monthly electricity consumption in kWh as obtained from the HBS and the three electricity companies. For electricity suppliers’ data we used the average value of the five months for which we have data (January to May 2005)[14]. While for HBS, since values are reported in terms of expenditure rather than consumption of kWh, we transformed expenditure into kWh using prices of the different suppliers (0.72 lei per kWh for RED-North and RED-North West, and 0.78 Lei per kWh for Union Fenosa). It is also important to note that in the HBS data there is not difference in the distribution of electricity consumption in the first five months of the year and the one for the whole year.

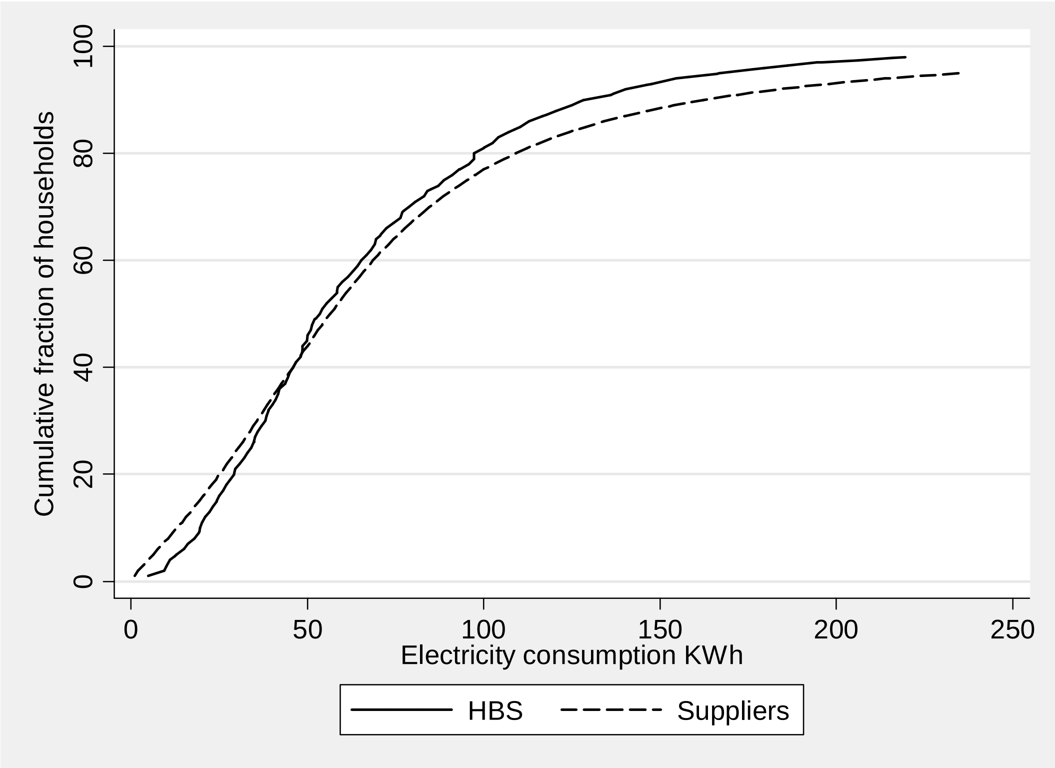
From figure 3.2 we can see that median consumption in the two datasets are very similar and the two distributions overlap each other for a good part of the central distribution, but the suppliers’ curve is above HBS’s one for lower consumption values and it then falls below HBS’s curve for high consumption values. This implies that the HBS underestimates both the proportion of households with low and high electricity consumption.

---

[13] According to the HBS data only 0.5% of households have zero electricity consumption, but it is important to remember that this is the case for one month observation and in the electricity dataset we do maintain cases where in some of the five months we do have one observation with zero consumption.

[14] Various adjustments to the electricity data were also attempted, but the overall distribution did not change significantly. These are the various corrections that have been experimented: we excluded cases where consumption was above 1000 kWh, we corrected cases where one single observation with high consumption was responsible of very high variance in the five observations, we excluded from the calculation of mean consumption cases with zero consumption in consecutive months and more than three months with zero consumption, we excluded cases with negative values. Most of these problems occurred with data from RED-North.

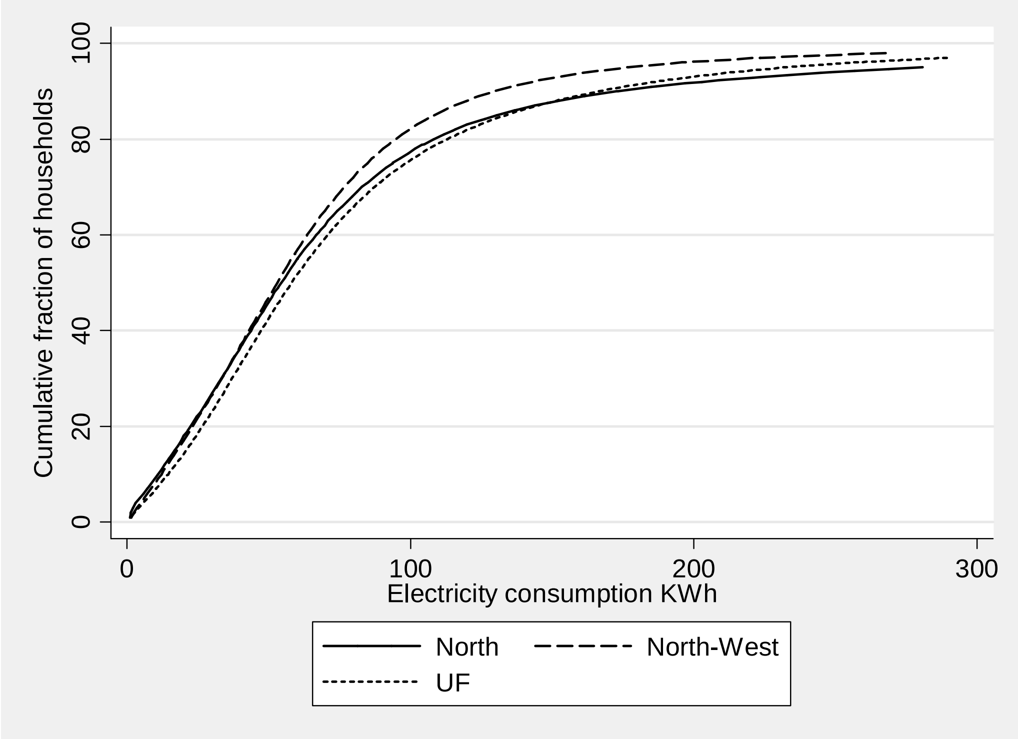
Figure 3.2 Comparison of cumulative probability functions of monthly electricity consumption (kWh) between HBS and electricity supplier, 2005



Source: Calculations of the authors based on the electricity database and HBS data.

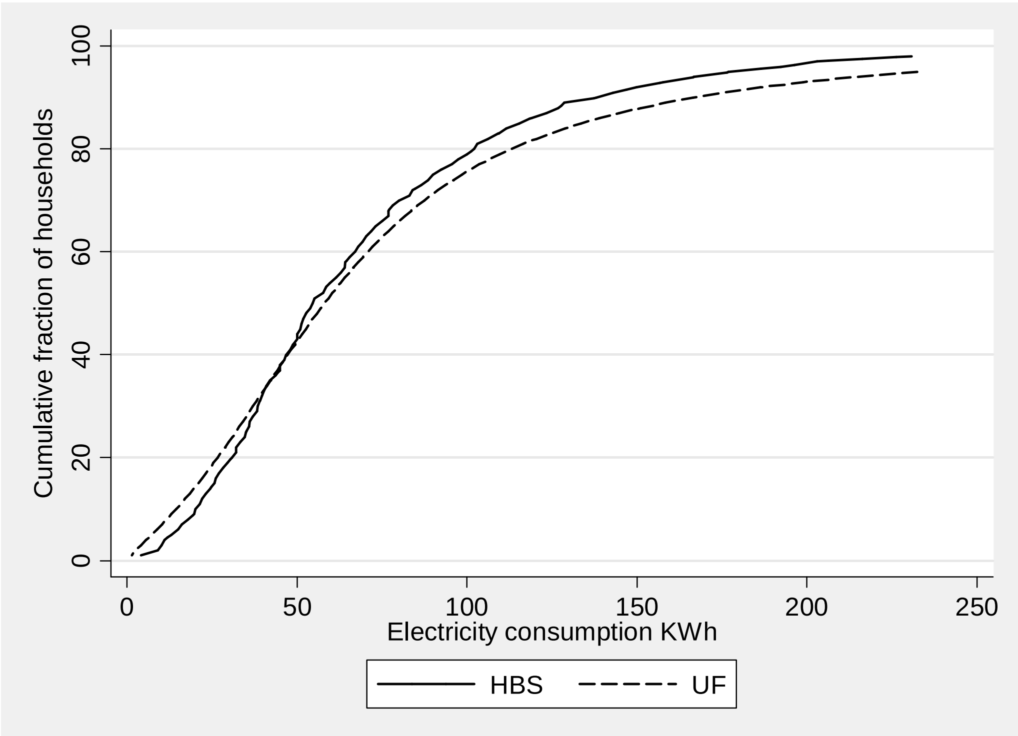
Because of doubts on the quality of data of one of the electricity supplier (RED-North) we also looked at the distribution per supplier (see figure 3.3). We expected to find higher consumption for costumers of Union Fenosa, since they serve the capital, Chisinau, but it is suspicious that RED-North has both the lowest and highest values of electricity consumption. To see whether this has an impact on the comparison between HBS and suppliers data we also compared data for Union Fenosa and HBS data in areas served only by Union Fenosa. This comparison confirms the existence of a HBS bias both for low and high consumption cases, though the differences in the lower part of the distribution are now somewhat smaller (see figure 3.4).

Figure 3.3 Cumulative probability functions of monthly electricity consumption (kWh) for the three electricity suppliers, 2005



Source: Calculations of the authors based on the electricity database.

Figure 3.4 Comparison of cumulative probability functions of monthly electricity consumption (kWh) between Union Fenosa and HBS, 2005



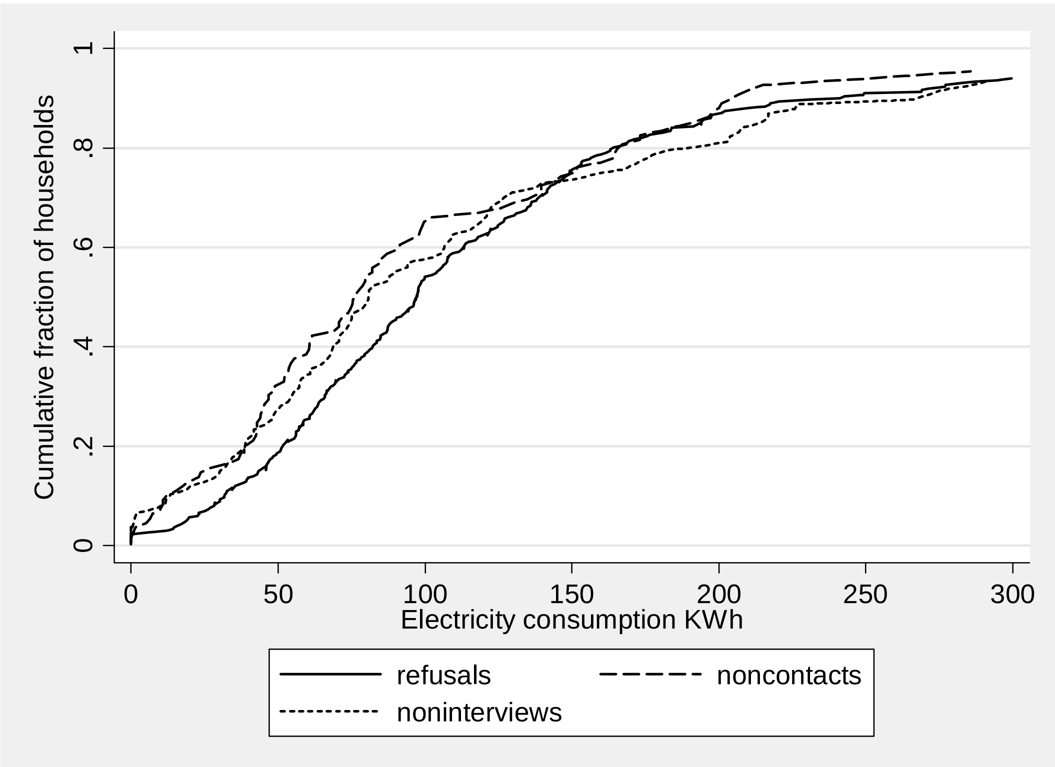
Source: Calculations of the authors based on the electricity database and HBS.

This initial analysis suggests that the HBS sample underestimates both low and high consumption of electricity and given that there is a positive correlation between electricity consumption and economic welfare we think that this might have some implications on measures of poverty and inequality.

However, this first analysis has not yet addressed whether and to what extent this problem is a consequence of non-response. The observed difference between the actual distribution of electricity and the one measured by the HBS could be due to sampling errors, or inaccuracies on the way consumption is reported in the HBS. We therefore matched addresses of non-respondent households with those in the electricity dataset to see whether their consumption levels are either particularly high or very low. We managed to match most addresses (92%), although in some cases, especially in rural areas (24%), where there are localities without street names and numbered houses, we could not match non-respondents and information in the electricity dataset.

Figure 3.5 shows the cumulative distribution functions of electricity consumption for non-respondents, divided into refusals, non-contacts and other non-interviews. The graph shows that refusals tend to have higher electricity consumption than non-contacts and other non-interviews. It also clearly appears that overall non-respondents have substantially higher consumption levels. In fact 60% of non-respondents have consumption levels that are higher than 80 kWh per month, while only 30% of respondents had such consumption. Moreover, among non-contacts and other non-interviews there is also a substantial proportion of households with very low consumption: the percentage of households with consumption lower than 10 kWh per month is double the one we observed among respondents.

Figure 3.5 Cumulative probability functions of monthly electricity consumption (kWh) for non-respondent households, 2005



Source: Calculations of the authors based on the electricity database and HBS.

However since non-respondents are mainly found in cities, where consumption is generally higher, it is important to compare electricity use in the three different geographical areas. Therefore, table 3.3 compares mean and median values of electricity consumption between respondents and non-respondents by geographical areas, showing that non-respondents have consistently higher consumption than respondents. It also differentiates between types of non-responses: either at the first interview or follow-up interview (at least one earlier interview occurred in 2004), revealing that first contact non-respondents tend to have a higher consumption.

Table 3.3 Mean electricity consumption for respondents and non-respondents by strata (kWh), 2005 (median reported in parentheses)

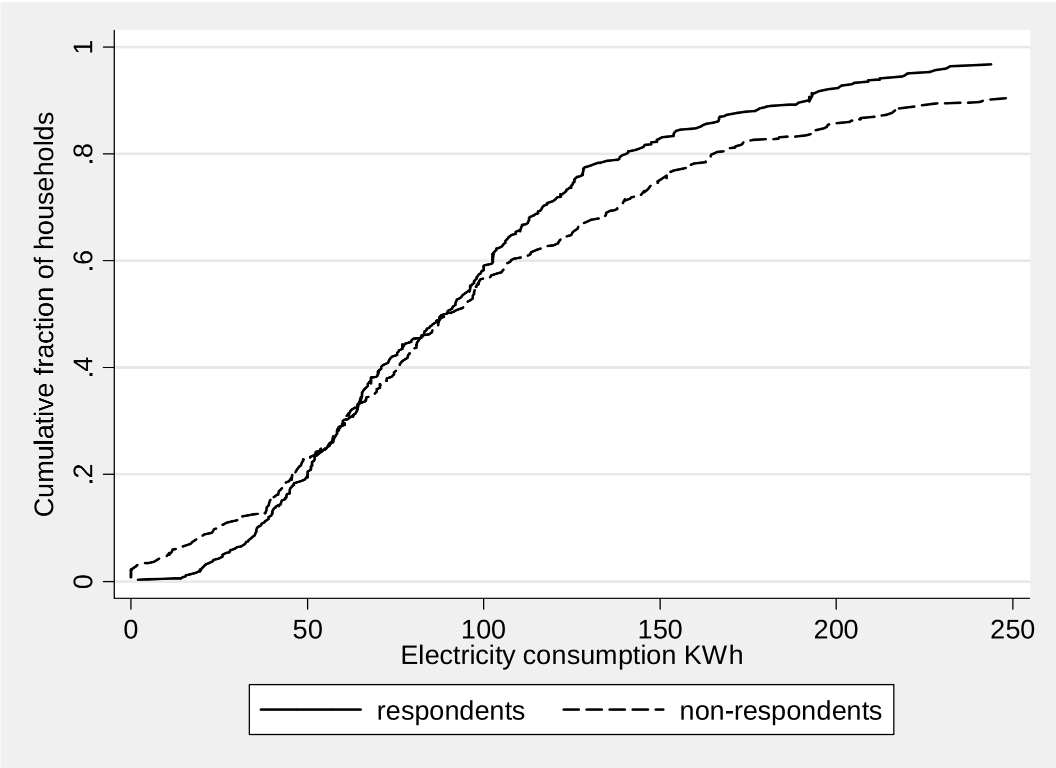
Geographical area	Interviews	Non-responses (un-weighted)		
		Overall	Panel type	
			First contact	Panel
rural area	54.7 (45.8)	75.0 (55.4)	97.4 (74.8)	55.5 (46.0)
towns	68.7 (55.6)	96.7 (75.4)	99.1 (80.2)	88.8 (75.2)
cities	101.4 (87.5)	125.9 (97.6)	123.8 (96.8)	133.2 (104.0)
Overall	68.0 (53.5)	117.7 (90.0)	119.9 (91.4)	111.3 (84.3)
% of missing		*8.3*	*9.6*	*4.4*
% of panel		*25.2*	*0.0*	*100.0*

Source: Calculations of the authors based on HBS and electricity suppliers data.

However, differences in mean and median values could depress the actual differences between respondents and non-respondents if non-respondents have both very low and very high electricity consumption. In order to consider this we look at the distribution of electricity consumption among HBS respondents and HBS non-respondents. This is done using only first-time selected households. Figure 3.6 reports the respective cumulative probability functions in cities. It is immediate to see that among non-respondents there is both a higher share of households with low consumption (below 40 kWh per month) and a substantially higher share of households with high consumption (above 100 kWh). Similar graphs can also be generated for towns and rural areas, and indeed provide a very similar result, but the number of non-respondents in these other domains is relatively small and therefore the cumulative functions of non-respondents are less reliable (see figure 3.7).

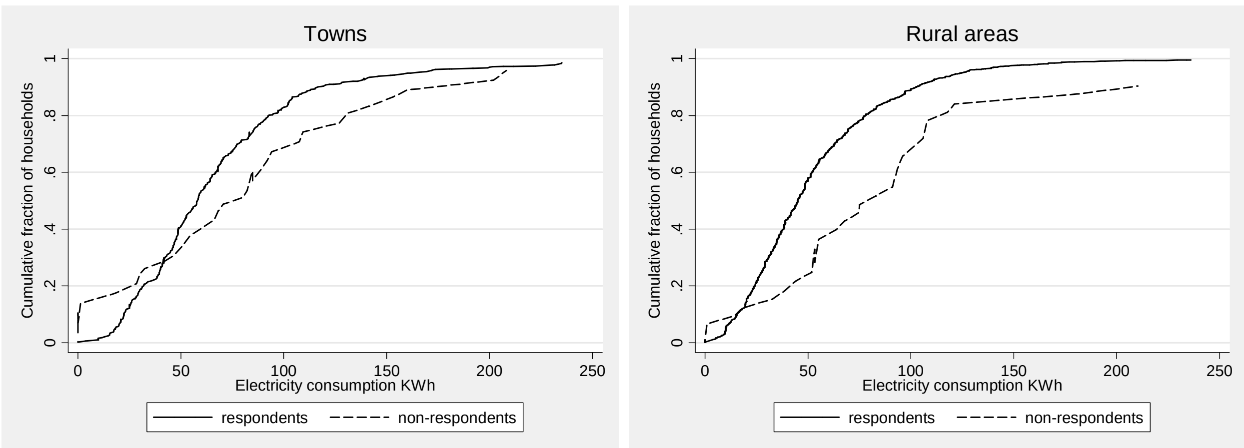
Finally, we compute the difference between electricity consumption of non-respondents and consumption of substitute households. The result of such comparison is reported in table 3.4, where we present data by reason of non-response. It is significant to note that the difference is negative for other non-interviews (especially those unable to participate - due to old age, illiteracy, etc.) where non-response seems to cause an underestimation of low electricity consumption levels.

Figure 3.6 Cumulative probability functions of monthly electricity consumption (kWh) for HBS first-contact respondent and non-respondent households in cities, 2005



Source: Calculations of the authors based on the electricity database and HBS.

Figure 3.7 Cumulative probability functions of monthly electricity consumption (kWh) for HBS first-contact respondent and non-respondent households in towns and cities, 2005



Source: Calculations of the authors based on the electricity database and HBS.

Table 3.4 Difference between electricity consumption of non-respondents and their substitute households (kWh, un-weighted), 2005

Reason of non-response	Mean	Median	Number of observations
Lack of time	43.6	27.8	67
Not useful	27.0	15.9	74
No confidential	109.5	109.5	1
Refusals	*35.4*	*21.0*	*142*
Non-contacts	*21.2*	*8.2*	*70*
Unable to participate	-41.6	-16.4	19
Other	-15.7	-0.8	37
Other non-interviews	*-24.5*	*-1.6*	*56*
Overall	*19.2*	*13.8*	*268*

Source: Calculations of the authors based on HBS and electricity suppliers data.

3.2.2 Main characteristics of non-respondents

In analysing the main characteristics of non-respondents we want to take into consideration two main attributes of non-responses:

1) Panel element: first contact non-interview and non-response of households that had been successfully interviewed at least once in the past (three months earlier or one year earlier).
2) Reason of non-response: Refusals, non-contacts and other non-interviews.

Indeed, we believe that these types of non-responses are likely to have different characteristics. However, we do not have enough observations to analyse separately reasons of non-responses by first contact and panel households.

As shown earlier, non-response rates are higher with first-contacts than within panel households, and in 2005 the rates were respectively 16% and 6%.

Table 3.5 compares the distribution by geographical area among interviews and non-responses (which are also disaggregated by type of non-responses).

Table 3.5 Interviews and non-responses by strata (%), 2005

	Interviews	Non-response (unweighted)					
		Overall	Panel status		Reasons		
			First contact	Panel	Refusals	Non-contacts	Other non-interviews
Rural areas	60.1	13.5	9.8	25.0	17.6	5.9	9.5
Towns	16.3	9.0	9.6	7.4	12.5	2.5	6.0
Cities	23.6	77.5	80.6	67.7	69.9	91.6	84.5
Overall	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% of panel	*48.0*	*24.1*	*0.0*	*100.0*	*23.4*	*24.4*	*25.9*

Source: Calculations of the authors based on HBS data.

Apart from observing that non-responses mainly occur in cities, something we knew already, it is relevant to note that among panel households there is a much larger percentage of non-responses from rural areas, whereas, looking at the reason of non-response, among refusals, in cities there are relatively lower non-responses than in non-contacts and other non-interviews.

Table 3.6 shows how non-respondent households' welfare was assessed by the enumerators. The possible classifications are three: bad, satisfactory and good. Most cases were considered 'satisfactory', though more households were judged to be in a 'bad state' than 'good' state.

However, such information was not completed for all households, since in some cases enumerators were not in a position to make a judgement. Overall we do not have information for 40% of non-respondents, and this percentage is higher among non-contacts and other non-interviews than for refusals. First time interviews are slightly better off than panel interviews, and refusals appear to be much better off than 'other non-interviews' and also non-contacts.

Table 3.6 Judgement on non-respondents' welfare status (%), 2005

Household welfare status	Overall	Panel type		Reason		
		First contact	Panel	Refusals	Non-contacts	Other non-interviews
bad	27.7	27.3	28.7	19.0	26.6	59.0
satisfactory	63.1	62.0	66.0	70.1	65.6	36.1
good	9.2	10.7	5.3	10.9	7.8	4.9
Overall	100.0	100.0	100.0	100.0	100.0	100.0
% of missing	*40.4*	*45.4*	*30.9*	*35.9*	*46.2*	*47.4*
% of panel	*28.0*	*0.0*	*100.0*	*26.5*	*32.8*	*27.9*

Source: Calculations of the authors based on 'non-response sheet'.

We now try to use information from interviews conducted in 2004 and information collected in the 'non-response sheet' to investigate to what extent non-respondents differ from those who participated in the survey[15].

Table 3.7 looks at some basic characteristics: household size, presence of children in the household and car ownership. For each of these variables we report the percentage of missing cases, and the percentage of observation that come from panel interviews.

In order to compare estimates from cases of non-response with those who completed the interview we present both estimates for the whole sample and for cities. This is because non-responses occur mainly in cities. From table 3.5 it emerges that non-respondents have a larger household size, higher not only than that in cities, but also higher than the overall country mean. Such mean is higher among panel households than in first time interviews, and more importantly there are large differences by reason of non-response. In fact, in other non-interviews and non-contacts household size is significantly lower than in refusals. Partly the larger household size among refusals is explained by the high percentage of households with children, about 45% against 35% in interviews in cities. Among non-contacts and other non-interviews, households with children is

[15] For households that were interviewed in 2004 we actually have information both in the non-response sheet and from previous interviews. Such cases offered the opportunity to verify that information collected in the non-response sheet matches closely the one from previous interviews, and where information in the non-response sheet was missing it was replaced with that of previous interviews.

lower than among successful interviews. Finally, non-respondents are more likely to own a car than households that participated in the survey. The percentage of households with cars is higher among refusals and lower among other non-interviews.

Table 3.7 Non-respondents and respondents main characteristics, 2005

	Interviews		Non-response (un-weighted)					
				Panel type		Reason		
	Overall	Cities	Overall	First contact	Panel	Refusals	Non-contacts	Other non-interviews
Mean hh size	2.5	2.3	2.6	2.5	2.8	2.8	2.4	2.1
Median hh size	2	2	2	2	2	3	2	2
% of missing			*35.6*	*47.0*	*0.0*	*31.9*	*42.0*	*39.7*
% of panel			*37.5*	*0.0*	*100.0*	*34.4*	*42.0*	*42.9*
hhs with children (%)	37.5	35.0	41.0	41.3	40.4	45.5	32.8	33.8
% of missing			*38.1*	*50.2*	*0.0*	*33.1*	*46.2*	*44.0*
% of panel			*39.0*	*0.0*	*100.0*	*35.0*	*45.3*	*46.2*
hhs with car (%)	11.3	11.3	14.2	13.2	15.4	15.9	14.3	9.5
% of missing			*47.7*	*62.9*	*0.0*	*46.5*	*52.9*	*45.7*
% of panel			*46.1*	*0.0*	*100.0*	*43.8*	*51.8*	*47.6*

Source: Calculations of the authors based on ‘non-response sheet’ and HBS data.

More comparisons between non-respondents and interviewed households are offered in table 3.8: main source of income/type of employment, age of the household head, and type of dwelling.

We can notice that among non-respondents the distribution of main source is significantly different from that of interviews. In particular the percentage of households working in the private sector, self-employed and unemployed is higher for non-respondents than for interviews. Moreover, there are significant differences between ‘refusals’ and ‘other non-interviews’. Indeed among ‘other non-interviews’ there is a majority of pensioners and a remarkably high percentage of unemployed, whereas among ‘refusals’ there are more self-employed and employee in the private sector.

Looking at the age of the household head we can observe that non-respondents show a larger concentration in middle age groups, especially so for refusals, while, in agreement with findings on sources of income, ‘other non-interviews’ have higher percentages of heads in old age.

Although results on the type of the dwelling are mainly driven by the area where the household lives (rural areas being almost entirely made of houses, cities of apartments and towns falling in between), the results show that ‘refusals’ are more likely to live in houses than other non-respondents.

Overall, these findings suggest that non-respondents have clearly different characteristics from households that participated in the survey. Moreover, we also find compelling evidence that ‘refusals’ and ‘non-contacts’ represent households that are economically better off than interviewed households, even within the same area of residence. On the other hand among ‘other non-interviews’ there are also cases of particularly poor households. In the next section we try to assess the implications of these findings when we want to measure poverty and inequality.

Table 3.8 Non-respondents and respondents main characteristics (%), 2005

	Interviews		Non-responses (un-weighted)					
				Panel type		Reason		
	Overall	Cities	Overall	First contact	Panel	Refusals	Non-contacts	Other non-interviews
Main source of income								
state employee	12.5	19.1	13.0	10.1	17.7	13.8	16.7	7.1
private employee	27.0	45.5	46.5	50.2	40.4	51.6	51.5	25.7
self-employed	2.3	2.1	4.8	3.7	6.6	6.5	0.0	4.3
farmer	19.1	0.0	1.1	0.9	1.5	1.4	1.5	0.0
pensioner	36.7	26.3	28.3	30.9	24.3	20.3	27.3	54.3
unemployed	0.1	0.1	4.5	3.2	6.6	4.6	1.5	7.1
other	2.4	6.9	1.7	0.9	2.9	1.8	1.5	1.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% of missing			*37.4*	*49.3*	*0.0*	*34.0*	*44.5*	*39.7*
% of panel			*38.5*	*0.0*	*100.0*	*35.5*	*43.9*	*42.9*
Age of household head								
<30	7.9	20.9	6.7	4.8	12.5	7.2	6.7	5.4
30-39	13.9	16.9	12.9	11.6	16.9	11.8	12.6	16.2
40-49	22.3	17.1	27.8	28.2	26.5	30.5	27.7	19.8
50-59	21.6	19.5	25.1	27.2	18.4	28.4	23.5	17.1
60-69	16.5	13.1	15.3	15.4	14.7	14.0	18.5	15.3
70+	17.8	12.6	12.3	12.8	11.0	8.1	10.9	26.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% of missing			*2.3*	*3.0*	*0.0*	*2.4*	*0.0*	*4.3*
% of panel			*24.7*	*0.0*	*100.0*	*24.0*	*24.4*	*27.0*
Type of dwelling								
apartment	29.7	91.0	78.0	82.0	65.4	72.9	88.2	81.9
house	68.2	0.6	18.5	15.0	29.4	25.6	5.9	11.2
hostel	2.1	8.5	3.6	3.0	5.2	1.5	5.9	6.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% of missing			*0.2*	*0.0*	*0.2*	*0.3*	*0.0*	*0.0*
% of panel			*24.2*	*0.0*	*100.0*	*23.5*	*24.4*	*25.9*

Source: Calculations of the authors based on 'non-response sheet' and HBS data.

4. Different methods to correct for non-response in measuring poverty and inequality

In this section we explore different ways to correct for non-response, in particular with the aim of observing the impact of such corrections on welfare measures. The following methods are implemented: substitution, geographical re-weighting, adjustment cells re-weighting based on probability of non-response, and welfare imputation of non-respondents.

4.1 Substitution and geographical re-weighting

As already mentioned earlier, the Moldovan National Bureau of Statistics replaces non-respondent households with other households in the same primary sampling units that were originally randomly selected in a reserve list[16]. It is therefore possible to use the sample of respondents and substitute households to generate welfare estimates. We compute mean and median per adult equivalent consumption, the Gini coefficient and poverty measures using three different poverty lines. These poverty lines are chosen with the only purpose to assess the effect of non-response at different levels of the distribution and are expressed in monthly consumption expenditure per adult equivalent. Their values are: 450, 580 and 830.

An alternative way to correct for non-response is simply to adjust the probability of selection at the primary sampling level according to the number of responses, so that the original sampling weights are multiplied by the ratio between the number of originally sampled households in the primary sampling unit and the actual interviews. Therefore, we exclude from our sample substitute households and compute new sampling weights that implicitly consider responding households to be representative of their primary sampling units. Table 4.1 reports welfare measures using these two approaches.

As expected, estimates are very similar, if not identical. Only within cities there exist some minor differences.

[16] It replaces both non-respondents and ineligible households, and in some cases the substitution occurs more than once because also substitutes could be ineligible or not participate in the survey.

Table 4.1 Welfare measures: substitution and geographical re-weighting, 2005

	Real per adult equivalent consumption					
	Mean		Median		Gini coefficient	
	Substitution	Geographical re-weighting	Substitution	Geographical re-weighting	Substitution	Geographical re-weighting
Cities	1310	1316	1204	1202	0.22	0.22
Towns	701	699	638	611	0.24	0.25
Rural areas	735	734	677	674	0.24	0.24
Total	857	855	748	740	0.27	0.28
	Head-count % (low poverty line)		Head-count % (middle poverty line)		Head-count % (high poverty line)	
	Substitution	Geographical re-weighting	Substitution	Geographical re-weighting	Substitution	Geographical re-weighting
Cities	1.4	0.8	3.3	2.5	15.7	15.0
Towns	19.7	20.3	41.9	45.0	75.5	76.4
Rural areas	16.7	17.4	35.7	35.8	70.1	70.3
Total	13.8	14.3	29.5	30.0	58.9	59.2

Source: Calculations of the authors based on HBS data.

4.2 Re-weighting based on probability of non-response

Based on common information on respondent and non-respondent households it is possible to try and determine the probability of non-response in different groups of the population. Unfortunately the number of variables for which we have complete or almost complete information (both for respondents and non-respondents) is very limited and consists of geographical area, electricity consumption, age of the household head, and type of dwelling.

Using such variables and their interactions/transformations we estimated the probability of response (using a logit model). Such model singles out the most important variables, which are then used to identify adjustment cells, in which the ratio of overall response (in each strata) and the probability of response for a certain group is used to correct the weights of responding households.

In order to better observe an eventual impact, we estimated the models excluding panel households, thus having a larger probability of non-response. Adjustment cells are generated in each strata using information on electricity consumption and age of the household head[17], and constructing cells of a minimum of approximately 100 observations. Correction factors range from a minimum of 0.72 and 1.64 (with both extremes occurring in cites), but most factors take values between 0.9 and 1.1.

Using this re-weighted sample we estimate again the welfare measures, but results are not different from those obtained using the simple geographical re-weighting.

[17] We also generated adjustment cells using first exclusively information on strata and electricity consumption, and afterwards on strata and age of the household head. However, the results do not change significantly.

4.3 Econometric imputations of welfare

Using variables for which we have information both for respondents and non-respondents, it is possible to estimate a consumption module using interviewed households and then predict consumption levels of non-respondent households. Once again, in order to observe eventual differences in welfare measures, we exclude panel households.

The results of the regression model are reported in table 4.2. Considering that we have relatively few variables the R-squared is high and equal to 0.34, but the explanatory power of the model is low in absolute terms. Unfortunately, we have not alternative ways to improve the model performance and we proceed with this model in estimating consumption levels of non-respondent households[18].

Table 4.2 Logarithm of real per adult equivalent consumption, 2005

Variable	Coef.	Robust Std. Err.	t-value
Electricity consumption (kWh)	0.00430	0.001	7.2
Squared electricity consumption	-0.00001	0.000	-5.1
Age of household head	0.01857	0.009	2.0
Whether the household is in cities	0.47768	0.054	8.9
Constant	6.26997	0.071	88.6

Standard errors are corrected for clustering effect and the model is estimated using sampling weights.

Source: Calculations of the authors based on HBS data.

In table 4.3 we compare welfare estimates obtained using the simple geographical re-weighting and estimates using data from the original sample, where values for non-respondents were imputed using the above model[19].

Once again welfare values for the country as a whole are not significantly different, but especially in cities and towns we now find some differences in the head-count rates for the three poverty lines.

---

[18] We assume that the error term is normally distributed (skewness is 0.1, kurtosis 3.7 and key percentiles are close to those of a normal distribution). Therefore we predict the logarithm of consumption and add random values of a normal distribution with standard error equal to the standard error of the regression and we then take the exponential values (predicted real per adult equivalent consumption = $e^{(xb+u)}$, where u is error term component).

[19] For some households we also imputed the household size.

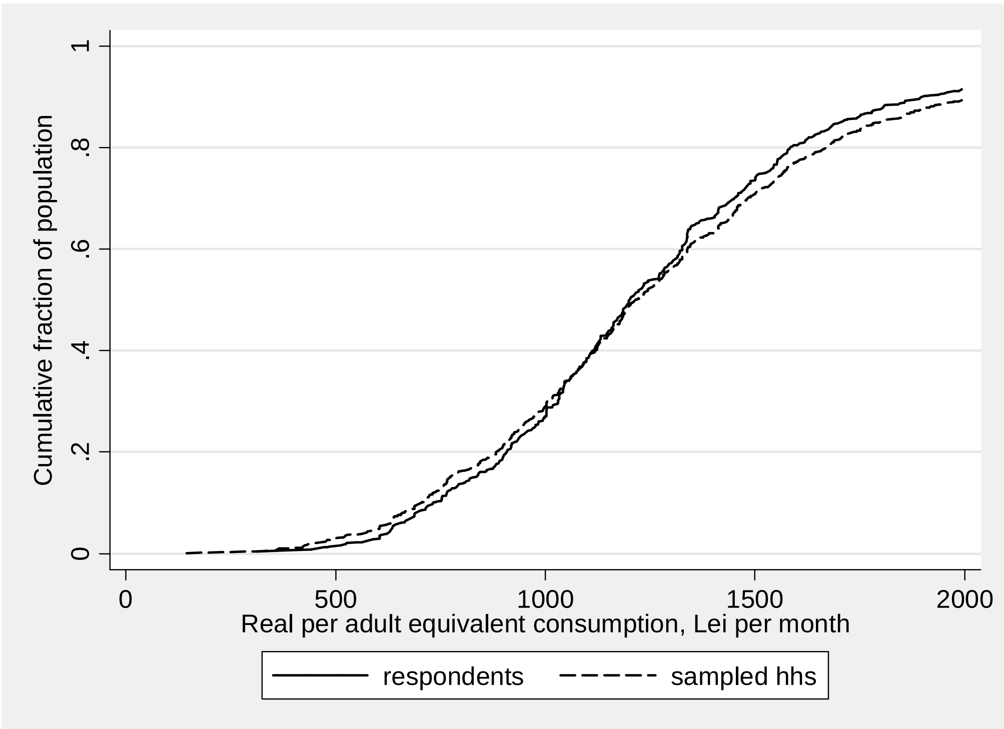
Table 4.3 Welfare estimates for sample with imputed values and respondent-only sample, 2005

	Real per adult equivalent consumption					
	Mean		Median		Gini coefficient	
	Imputation	Geographical re-weighting	Imputation	Geographical re-weighting	Imputation	Geographical re-weighting
Cities	1321	1316	1216	1202	0.23	0.22
Towns	701	699	611	611	0.25	0.25
Rural areas	737	734	674	674	0.24	0.24
Total	856	855	737	740	0.28	0.28
	Head-count % (low poverty line)		Head-count % (middle poverty line)		Head-count % (high poverty line)	
	Imputation	Geographical re-weighting	Imputation	Geographical re-weighting	Imputation	Geographical re-weighting
Cities	1.9	0.8	4.2	2.5	16.6	15.0
Towns	18.8	20.3	43.6	45.0	74.8	76.4
Rural areas	17.2	17.4	35.1	35.8	69.2	70.3
Total	14.2	14.3	29.8	30.0	58.7	59.2

Source: Calculations of the authors based on HBS data.

It is even more significant though to look at the cumulative distribution functions in the two scenarios. Figure 4.1 shows the CDFs in cities, where we can see that the adjustment with simple geographical re-weighting underestimates both the percentage of households with low and high consumption (see figure 4.1).

Figure 4.1 Cumulative probability functions of monthly per adult equivalent consumption for full sample and respondent households, 2005



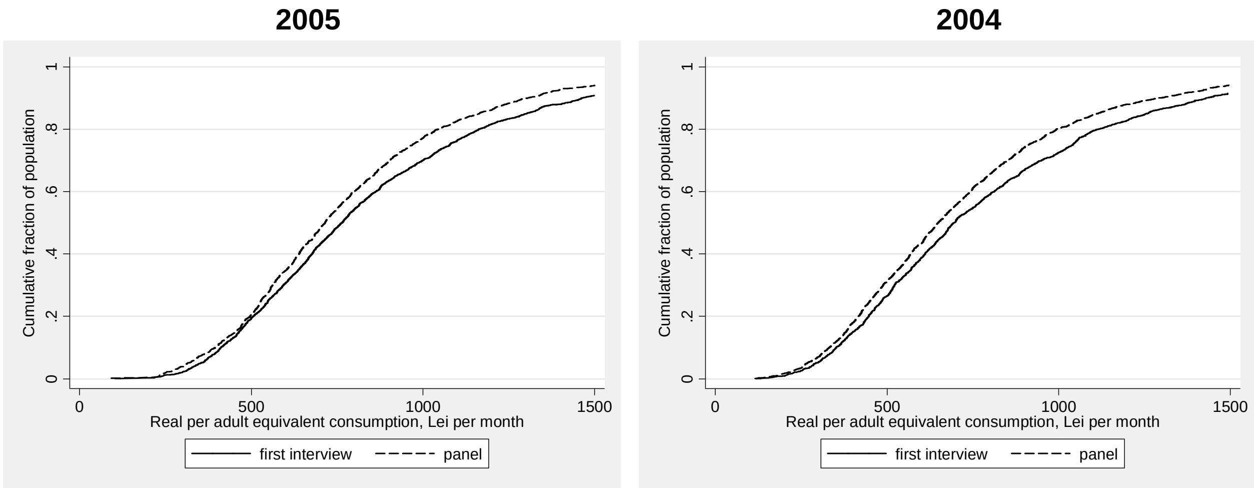
Source: Calculations of the authors based on HBS data.

4.4 Further analysis of impact of non-response on welfare estimates using panel households

In the previous analysis, investigating the effects of non-response with and without panel households, we noticed that panel households appeared to be worse off than first-interview households. We now therefore look in more detail at such differences.

A comparison between first-contact and panel households of the cumulative distribution functions of the real per adult equivalent consumption clearly shows that the sample of first-contact households is better-off than the sample of panel households. The difference is not ignorable. Moreover, we computed the same CDFs using data from 2004, and once again we found that non-panel households are better-off (see figure 4.2). After exploring the same CDFs by strata, we found that differences are once again ignorable in rural areas, but become relevant in towns and very substantial in cities.

Figure 4.2 Cumulative probability functions of monthly per adult equivalent consumption for first-interview and panel households, 2005 and 2004



In such comparisons we are observing the impact of non-response after the household agreed at least to be interviewed once. Moreover, in figure 4.2 we observe the impact on panel households of more than one year non-response. For example some of the panel households in 2004 were originally first interviewed in 2001, then in 2002, 2003 and 2004. In each of these years non-response is likely to have progressively 'selected' less well-off households.

Finally, to show that this is indeed the case we compared the CDFs in 2004 among panel households, and among only those households that we know will accept to be interviewed also in 2005. Although the differences are lower than those shown in figure 4.2, in cities the effect of non-response remains visible and confirms our hypothesis that differences between panel and first-time interview households is indeed due to non-response.

5. Conclusions

This paper studied the problem of unit non-response in the Moldovan Household Budget Survey and assessed the impact of non-response on welfare measures.

We found that in the Household Budget Survey non-response increased overall from about 8.5% to 16%, and in cities from 22.7 to 45%, and that non-response is somewhat underestimated because of the panel component in the sample. Indeed, non-response is substantially lower among panel households than among first-contact households, being respectively 6 and 16%. Most non-responses are refusals, but there are also non-contacts and other non-interviews, especially in cities.

Using information available on non-respondents we also found that their characteristics are different from those of the rest of the sample, even within the same strata. Moreover, there are also significant differences in characteristics between the different reasons of non-response. Refusals tend to be better-off than other non-interviews.

As expected, substitution is basically equivalent to correct for non-response using geographical re-weighting, since using these two methods produce almost the same welfare estimates.

Methods that were expected to show the bias of non-response were the re-weighting based on non-response probabilities for different population groups (adjustment cells) and the direct imputation of welfare levels of non-respondent households. However, the impact on welfare estimates for the first of such methods is very small and almost irrelevant. Instead, imputation of consumption levels for non-respondent households produces different welfare estimates from those obtained with the geographical re-weighting. In particular, the effect of non-response becomes particularly noticeable in cities, not so much on mean and median levels, but on the tails of the welfare distribution, since non-response reduces both the percentage of households with low and very high income. However, the overall impact remains limited and lower than what information on characteristics of non-respondents would have led us to believe. Moreover, suspicion that the actual impact on welfare might be significantly more substantial than what found in the imputation methods come from the comparison of welfare levels between panel and first-time interviews. In fact, panel households appear to be systematically worse-off than first interview households (both in 2005 and 2004), and we have clear evidence that cumulative non-response is the reason for lower welfare levels among panel households.

Therefore, although non-response corrects at least some of the non-response bias, we believe that the implemented correction methods are not fully capable to identify the actual non-response bias. Ultimately this is because we would need more comprehensive information on non-respondents. Future research should try to address this issue probably by matching non-respondent households with information collected in the latest census, which took place in October 2004.

References

Bethlehem J. and Schouten B. (2004): Nonresponse adjustment in household surveys; Discussion paper 04007; Statistics Netherland.

Bhalla, S. (2002): *Imagine there is no country: poverty, inequality, and growth in the era of globalization*. Washington DC, Institute for International Economics.

Bourguignon, F. and Morrisson C. (2002): “Inequality among world citizens: 1820-1992”, American Economic Review, Vol. 92, No. 4, pp. 727-744.

Census Bureau (2002): Current Population Survey: Design and Methodology, Technical Paper 63RV; Washington, DC.

Chen, S. and Ravallion, M. (2001): How well did the world’s poorest fare in the 1990s?; Review of Income and Wealth, Vol. 47, No. 3, pp. 283-300.

Deaton, A. (1997): *The Analysis of Household Surveys: A microeconometric approach to development policy*, published for The World Bank, The John Hopkins University Press, Balțimore and London.

Deaton, A. (2004): Measuring Poverty in a Growing World or Measuring Growth in a Poor World; NBER Working Paper No. 9822, Forthcoming in the Review of Economics and Statistics

Deaton, A. and Grosh, M. (2000): Consumption; in Grosh M and Glewwe P.: Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standard Measurement Study; Washington D.C., World Bank, Vol.1, pp. 91-133.

Eurostat (2003): Household Budget Surveys in the EU: Methodology and recommendations for harmonisation, Luxemburg, 2003.

Eurostat (2004): Household Budget Surveys in the Candidate Countries: Methodological analysis 2003, Luxemburg, 2004.

Foster J, Greer J., and Thorbecke E. (1984): “A class of decomposable poverty measures”; *Econometrica*, Vol. 52, pp. 761-765.

Foster, K. (1998): Evaluating non-response on household surveys: report of a study linked to the 1991 Census, Government Statistical Service; Methodology Series No. 8.

Groves, R and Couper M. (1998): *Nonresponse in Household Interview Surveys*. New York: John Wiley and Sons.

Holt, D. and Elliot D. (1991): Methods of Weighting for Unit Non-Response. *The Statistician*, Vol. 40, pp. 333-342.

Korinek, A. Mistiaen, J. and Ravallion, M. (2005): An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys; Forthcoming in the Journal of Econometrics.

Lanjouw J. O., and Lanjouw P. (1997): “Poverty comparisons with noncompatible data: theory and illustrations”, Policy Research Working Paper, World Bank, Washington, DC.

Lanjouw P., Milanovic B. and Paternostro S., (1998): “Poverty and Economic Transition: How Do Changes in Economies of Scale Affect Poverty Rates of Different Households?”, Policy Research Working Paper 2009, World Bank, Washington, DC.

Little, R. J. A. and Rubin, D. B. (1987): *Statistical Analysis with Missing Data*. New York: Wiley.

Little, R. J. A. (1988): Missing-Data Adjustments in Large Surveys; *Journal of Business & Economic Statistics,* Vol. 6, No. 3, pp. 287-296.

Lundström, S. and Särndal C. (1999): Calibration as a Standard Method for Treatment of Nonresponse; *Journal of Official Statistics*, Vol. 15, No. 2, 1999, pp. 305-327.

Office for National Statistics (2005): Family Spending: A report on the 2004-05 Expenditure and Food Survey.

Ravallion M. (2001): “Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree?”, Policy Research Working Paper no. 2665, World Bank, Washington, DC.

Rubin, D. B. (1987): *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Sala-i-Martin, X. (2002): The disturbing ‘rise’ of global income inequality; Cambridge, MA; NBER Working Paper No. 8904.

Sen A. K. (1976): “Poverty: An Ordinal Approach to Measurement”, *Econometrica*, Vol. 48, pp. 437-446.