Statistics Netherlands

Discussion Paper

Estimation of household income based on asset ownership

2014 | 01

Nino Mushkudiani

4 februari 2014

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
verkoop@cbs.nl
Fax +31 45 570 62 68
ISSN 1572-0314

60083 201401 X-10

Within the collaboration between Statistics Netherlands and the National Statistics Office of Georgia (GeoStat) a pilot study was carried out. The study aims the improvement of quality of the Integrated Household Survey (IHS) of GeoStat. One of the issues is that the income of the households obtained from the IHS is not reliable and not representative for the population. GeoStat does not have any administrative or other source with information on household income. In order to obtain information related to household income we developed an asset ownership questionnaire. The questionnaire was applied to a small group of households. Using a linear regression model based on this questionnaire we estimated the household income for the IHS data frame in the Tbilisi (Georgia) area. In this paper we recommend GeoStat to use these estimates for stratification of the IHS according to household income. The stratification can also be used for improving the weighting of the IHS and make the income distribution of households more representative. The pilot study was applied only on households in the Tbilisi area.

1 Introduction

In 2010 a bilateral cooperation between Statistics Netherlands (SN) and the National Statistics Office of Georgia (GeoStat) was initiated by the Dutch Ministry of foreign affairs, by assigning a budget to support research and process development of GeoStat. GeoStat has expressed the need of methodological and intrinsic support for improving the quality of the Integrated Household Survey (IHS) design. With joint efforts of two teams the goals of this bilateral collaboration were defined (see Woensdrecht and van Teeffelen [2011]). The initial list of the topics was very extensive. It included over 20 different problems related to the IHS. Later in April 2012, GeoStat and SN teams made a choice of three directions for this collaboration. These are:

- Data reliability (editing and control rules);
- Sampling scheme and sampling design;
- Pilot study on pre-listing of IHS sample population frame;

These areas identified the project goals and project steps (see Mushkudiani [2013]).

In this paper we deal with the third goal of this collaboration; a pilot study on pre-listing of the IHS sample population frame. Pre-listing is as a sampling technique, when a complete enumeration of auxiliary variables of each unit in the population frame is carried out (see e.g. Hauser and Hansen [1944]). In this paper we describe the design of the pilot study and results obtained from it. The goal of the study was to investigate the possibility of stratifying the IHS data frame according to the household income. One of the main issues of the IHS is that the household income obtained from this questionnaire is not reliable and not representative for the total population. GeoStat does not have any administrative or other source for household income. Therefore we want to estimate the household income for the IHS sample frame. We restricted the pilot study for the Tbilisi area. It was carried out in two stages. At first, GeoStat developed an asset ownership questionnaire and applied it to a small group (in total 790) of households from the Tbilisi area. For these households the income was known to GeoStat from the earlier IHS questionnaires. Using a linear regression model we identified the significant asset ownership variables for household income and further improved the questionnaire. At the second stage the improved questionnaire was applied to a IHS sample frame in the Tbilisi area (8170 households). Following the linear regression model we estimated the income for these households. Availability of household income information for sampling/weighting purposes can potentially bring very significant benefits to the IHS by making the income distribution of households more representative.

This paper is organized as follows: Section 2 describes the data frame of IHS. In Section 3 we explain the setup and the design of the pilot study. The design of the asset ownership questionnaire is given in Section 4. Section 5 deals with the linear regression model and Section 6 with the stratification of the IHS data frame. Finally, conclusions can be found in Section 7.

2 Integrated household survey data frame

The population frame of the IHS is out of date and needs an update. The frame is based on an address database. The data in this database were collected during the last census in Georgia, which took place in 2003. After 2003, a large number of new housing constructions took place especially in and around the Tbilisi area. The new constructions are mainly owned or bought by households with a high income. In order to include these high income households in the data frame, we need to include the addresses of the new buildings. However, there is no official register or other information of these new constructions and buildings available to GeoStat. In the frame and henceforth in the IHS sample this part of the new constructions are not included. Therefore we suspect that the distribution of income in the sample may not not be representative.

Another important issue of the IHS is the lack of information on income of the respondents. There is no register information available for GeoStat on household income. In the Netherlands we have information from several registers that are linked to households, for example house/apartment values from the land register, household income from the tax office register and social benefits from the social services office register. Based on these variables Statistics Netherlands can stratify the sampling data frame. However, for GeoStat it is impossible to stratify the sample data frame based on these variables. The IHS includes questions about household income, but it is very sensitive information and respondents are not inclined to fill in these questions truthfully. The information obtained from the IHS about household income is therefore partly unreliable and cannot be used to post-stratify the sample. To resolve this problem we want to define an income score for each household in the data frame of the IHS. Based on this income score we can stratify the data frame before sampling. In the pilot study, carried out in the Tbilisi area, we define the household income score based on information highly correlated with income. In the next section we describe the design of the pilot study.

3 The pilot study design

In this section we define a setup of the pilot study. The pilot study was carried out in two stages. In the first stage we designed the asset ownership questionnaire (AOQ), including questions on the variables that we thought would be related to household income. We applied the AOQ questionnaire to a small group of households (in total 790) from the Tbilisi area. We interviewed these households by a face to face interview mode. From these 790 households in total 618 (78%) completed interviews. These records have no item non-response. We analyzed the response data using a generalized linear regression model and identified the variables in the questionnaire that are significant for estimating the household income. These significant variables were used to define the improved AOQ questionnaire (see Appendix V) and the income score function.

During the second stage of the pilot study we applied the AOQ to a larger group of households (in total 8170 households). This group represents the data frame for the IHS in the Tbilisi area. From the response data of these households we estimated the income score based on the linear regression model defined at the first stage.

The income score of the households in the data frame will be used for stratification of households in the IHS frame according to their income. The following sections deal with each stage of the pilot study separately.

4 Asset ownership questionnaire

The asset ownership questionnaire consists of questions on household assets: real estate property, dwelling, rural estate and the state of the main dwelling. We also included questions about other assets, such as a car and the state of the household equipments: TV, fridge, computer, washing machine, etc. Some of these questions are taken from the Social Service's questionnaire, that is used for identifying households below the poverty line. A GeoStat working team designed the questionnaire and the Statistics Netherlands team members tested it and gave suggestions for improvement (see Kruiskamp [2013]).

Table 1 Income statistics of 790 households

Min.	1st Quantile	Median	Mean	3rd Quantile	Max.
36.0	159.2	236.5	338.0	387.0	6840.0

Next the questionnaire was applied to a small group of households, in total 790 households. The group was selected carefully. All households were from the Tbilisi area. These were the IHS respondents from 2012. The households were chosen so that their income was known from their response and assumed to be reliable. By comparing household income with expenditure it can be detected if the income is filled in incorrectly. In addition the household income is assumed to be reliable if it is almost equal to the expenditure and it is 'reasonable' for the specific household structure, such as total number of persons, number of working persons, number of retired persons, etc. In this sample of 790 households the distribution of income is representative, meaning that the data includes households with very low income and with very high income. There were in total 40 households with income less than 100 GEL's and 28 households with income greater than 1000 GEL's. The Georgian Lari (GEL) is the official currency of Georgia. Table 1 gives the summary for income of these households.

In Figure 1 the distribution of the reported income data corrected for the number of persons per household is presented. Because of the big differences in income, we have groups of households with very low and with very high income. To visualize the income distribution we split the income in two groups. The first figure displays the frequency bar chart of the household incomes up to 600 GEL and the second figure shows the bar chart of the income above 600 GEL (these were in total 94 households). The household income data in this sample has very skew density, especially on the right hand side.

GeoStat interviewed these 790 households in January and February of 2013, using the AOQ-questionnaire. As we already have mentioned above the face to face interview mode was applied and in total of 618 (78%) households completed the interview. In Tables 2 and 3 the frequencies of the real estate property variables and of the household equipment of the respondents are given.

We analyzed the sample data using a generalized linear (GL) regression model and identified the variables in the questionnaire that were significant for estimating the household income.

Figure 1 Income of 790 households

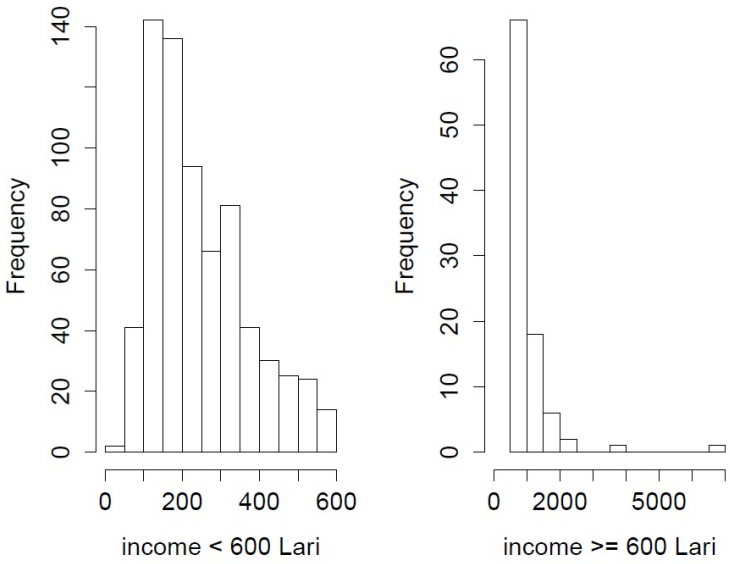


Table 2 Frequencies of real estate variables of 618 households

Variables	Frequency	Percentage
Main dwelling	618	100.00
Home/apartment 1	87	14.08
Home/apartment 2	5	0.81
Home/apartment 3	0	0.00
Summer house 1	34	5.50
Summer house 2	2	0.32
Summer house 3	0	0.00
Commercial property	5	0.81
Garage 1	63	10.19
Garage 2	1	0.16
Garage 3	0	0.00
Dwelling under construction	1	0.16
Nonagricultural land	28	4.53
other building	0	0.00

Table 3 Household equipment in sample data

Covariate	Frequency	Percentage
Refrigerator 1	559	91.34
Washing machine 1	466	76.14
Gas stove 1	534	87.25
Kitchen machine	57	9.31
Microwave stove	68	11.11
Foreign vacuum cleaner	200	32.68
Water heater 1	363	59.31
Personal computer 1	301	49.18
Laptop 1	150	24.51
Mini system	47	7.68
Video camera 1	21	3.43
Piano/Grand piano	202	33.01
Satellite TV antenna	48	7.84
Air-conditioner 1	46	7.52
LCD or plasma TV-set 1	159	25.98
Individual heating system 1	354	57.84

Based on these variables we defined a GL-model and the income score function. In the next section we discuss the GL-model in more detail. We also shortened the AOQ questionnaire by excluding the questions that had very low positive response frequency, e.g. in Table 2 the variable "Garage 3" has the frequency equal to 0. These variables were insignificant for the score function. The improved version of the AOQ questionnaire is presented in Appendix V.

5 The generalized linear model

We applied a generalized linear regression model on 618 household response data. Recall that these households were chosen so that we know the "Income" and "Expenditures" for each household from their response on IHS questionnaire of 2012. We consider these two variables, since for example, if for a certain quarter the household spent its savings, then the household can have expenditure bigger than the income. We would like to estimate the household income, in terms of the household assets and since we can not identify the household savings, we consider the maximum of income and expenditure. On the other hand, since income/expenditure has a very skew distribution, we define the response variable as follows:

$$Y = \log(\max(\text{Income}, \text{Expenditures})). \tag{1}$$

For some questions in the AOQ there were very few or no cases at all recorded. In Appendix A we give the list of these variables and the corresponding frequencies. We did not include the variables from this table in our model.

For all other variables we calculated the covariance matrix. We found that few were highly correlated with each other. In Table 4 we give the list of these variables and their correlations (Pearson's). It is clear that the three variables "Bedrooms", "Rooms" and "Total Space" contain almost the same information. We considered univariate linear models with each of these variables. From these models the one with "Bedrooms" has the lowest AIC score and the highest Nagelkerke R^2, even though the difference between these models is very small. If we

Table 4 High correlations

Variables	Correlation
Bedrooms, Rooms	0.93
Total Space, Rooms	0.74
Total Space, Bedrooms	0.73
Holidays Home, Number Holidays Home	0.96
Holidays Abroad, Number Holidays Abroad	0.95

include all three variables in one model, the scores, AIC and Nagelkerke R^2 hardly improve, see Table 5. Note that in all models the response variable is Y. From these three covariates we include only the variable "Bedrooms" in our model.

Table 5 Model quality scores

Model	AIC	Nagelkerke R^2
Bedrooms	1422.2	0.163
Booms	1424.4	0.158
Total Space	1444.1	0.127
Bedrooms+Rooms+Total Space	1422.7	0.172

For the variables "Holidays Home" and "Number Holidays Home", and for "Holidays Abroad" and "Number Holidays Abroad", we carried out a similar analysis. For these pairs of variables it was also obvious that the second variable did not add much information to the model and did not improve the model variation. Model scores were almost the same. Hence we do not include the variables "Number Holidays Home" and "Number Holidays Abroad" in our linear model.

We have the following covariates left in our GL-model: "Bedrooms", "Kitchen", "Bath", "Cellar", "Holiday Home", "Holiday Abroad", "Family Size", "Ownership Main Dwelling", "Score Main Dwelling", "Built Year Main Dwelling", "Last Repair Year Main Dwelling" "Home/Apartment (other than main dwelling)", "Summer House", "Garage", "Score Car", "Land", "Fridge", "Washing Machine", "Foreign Vacuum Cleaner", "Water Heater", "Kitchen Machine", "LCD or Plasma TV set", "Laptop1". For each categorie of each covariate, we included a $(0, 1)$ dummy variable in the model. The response variable for this model is defined in (1). In Appendix II the results of this GL-model are presented. The model evaluation scores are AIC= 1008, $R^2 = 0.59$, Nagelkerke $R^2 = 0.66$.

In this model there are quite a few covariates that are not significant. We want to exclude these covariates. To do it in an optimal way we decided to choose the model based on the AIC score. We want to find the combination of covariates that result in a model with the lowest AIC score.

In order to find the 'best' model we could consider all combinations (subsets) of the covariates and based on the model evaluation scores identify the best model. However there are

$$\sum_{k=1}^{23} \binom{23}{k} = 2^{23} - 1 = 8\,388\,607$$

combinations of the covariates from the set of 23. To make the model calculations easier, we defined the core model. In the core model we included the 10 most significant variables, based on the P-values in Appendix II. We included these variables as default covariates in our model and add all possible combinations from the rest of the covariates. The ten covariates included in our core model are: "Holiday Home", "Holiday Abroad", "Score Car", "Family size", "Score Main Dwelling" , "Ownership", "Fridge", "Foreign vacuum cleaner", "LCD or PlasmaTVset1",

"Laptop1". In Appendix III the results of the GL model with these ten covariates are presented. The model validation scores of the core model are AIC= 982.45, $R^2 = 0.58$, Nagelkerke $R^2 = 0.65$.

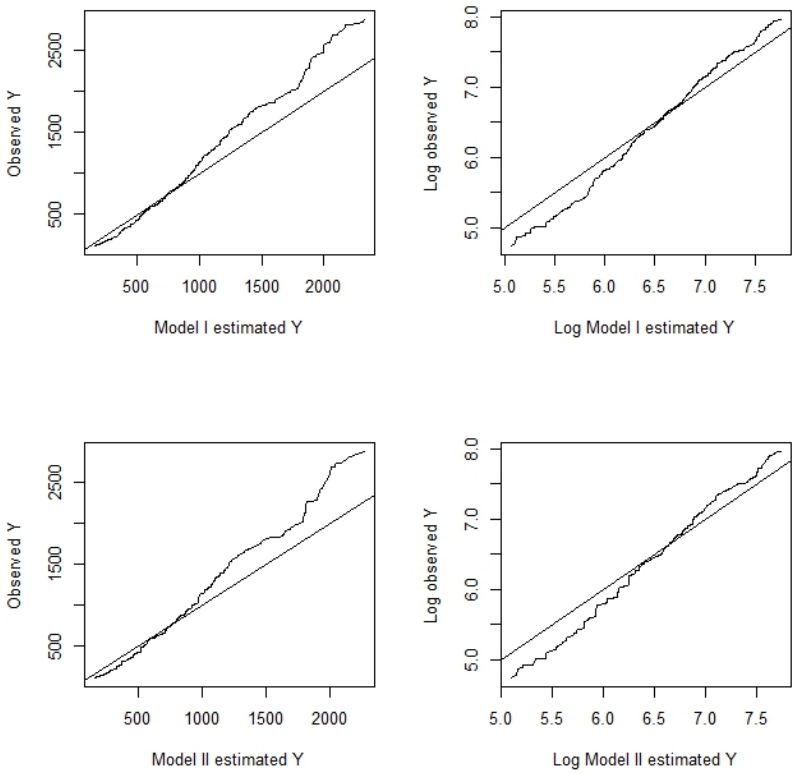
Table 6 Additional covariates for the models with the best AIC score

Covariates	AIC	R^2
Washing machine, Water Heater	982.45	0.653
HomeApartment1, Washing machine, Water Heater	982.80	0.654
Garage, Washing machine, Water Heater	983.62	0.654
Built Year, Washing machine, Water Heater	983.65	0.655
HomeApartment1, Garage, Washing machine, Water Heater	984.05	0.655
Washing machine, kitchen machine, Water Heater	984.08	0.653
HomeApartment1, Water Heater	984.12	0.652
Built Year, HomeApartment1, Washing machine, Water Heater	984.12	0.656
Water Heater	984.17	0.651
Land, Washing machine, Water Heater	984.19	0.653
SummerHouse1, Washing machine, Water Heater	984.33	0.653
HomeApartment1, Washing machine, kitchen machine, Water Heater	984.41	0.654
HomeApartment1, Land, Washing machine, Water Heater	984.74	0.654
HomeApartment1, SummerHouse1, Washing machine, Water Heater	984.76	0.654
Bath, Washing machine, Water Heater	984.77	0.654
Built Year, Garage, Washing machine, Water Heater	984.78	0.656
Washing machine	985.03	0.650
Bath, HomeApartment1, Washing machine, Water Heater	985.04	0.655
Garage, Washing machine, kitchen machine, Water Heater	985.22	0.654
HomeApartment1, Washing machine	985.30	0.651

We chose all possible combinations of the rest of the 13 covariates. These were in total $8\,191$ models. Each of these models include the ten covariates defined above plus one of the combinations of the other 13 covariates. We calculated these models using the statistical program R. Table 6 shows the list of additional covariates for the best 20 models according to AIC score. There is very little difference in these highest scores. If we want to include as few as possible covariates then we could choose the model with one additional covariate "Water Heater". On the other hand, if we include the covariate "Washing machine" as well, we will have the model with the best AIC score. In general defining "the best model" can be very subjective. In the next section the model with 12 covariates, that includes "Washing machine", "Water Heater" and the core model covariates, will be considered as the best model.

Figure 2 shows the QQ plots of the observed and the estimated income and log of observed and estimated income for the models with all 23 covariates (Model I) and with 12 covariates (Model II). Here we removed the smallest 4 and the largest 18 values from the observations and the estimates to avoid the effect of the outliers. On the plots for both models the estimates and the observed income values have big differences in the right tail. Modeling of households with high income is difficult since we have only few households in this group and the distribution of the income is very skew. On the other hand, if we look at the QQ plots of the log-incomes, the observed and the estimated values are much closer to each other. Therefore, considering the log of the income is the obvious decision. From these plots it is also difficult to decide which model to choose.

Figure 2 QQ plots for models I and II



6 Data Frame Stratification

We want to apply the models from the previous section to the IHS data frame. By using the beta coefficients of the models, we want to estimate the household income for households in the data frame. For this we need to have the information of the households in the data frame on the same variables as in the linear models in the previous section.

In February and March of 2013 interviewers were sent out in the Tbilisi area, with the improved asset ownership questionnaire, in order to conduct face to face household interviews. In total 8170 household addresses were chosen for these interviews. The response rate was reasonably high, 73.6%, see Table 7.

Table 7 Household response in data frame

Response	Frequency	Percentage
Completed interview	6012	73.59
Dwelling is closed (no one lives)	935	11.44
Is not used for living	48	0.59
Refused	679	8.31
Other reason	496	6.07

We want to compare the marginal distributions of the covariates in the two data sets, namely the data frame and the sample. We observe that in the data frame households have a higher

percentage for real estate property. For example, from the responded 6012 households, a total of 1085 (18.1%) had a "Home/apartment 1" in addition to the main dwelling. This is higher than 14.1% in our sample. Similar results hold for having a "Summer house" or a "Garage"; again the percentages are higher in the data frame than in the sample, see Tables 2 and 8. On the other hand, the sample households have a higher percentage for home equipment; see Appendix IV and Table 3. Overall the distributions are not very different.

Table 8 Household real estate property in data frame

Covariate	Frequency	Percentage
Main dwelling	6012	100.00
Home/apartment 1	1085	18.05
Home/apartment 2	87	1.45
Summer house 1	438	7.29
Summer house 2	39	0.65
Commercial property	87	1.45
Garage 1	819	13.62
Garage 2	28	0.47
Dwelling under construction	45	0.75
Nonagricultural land	162	2.69

In the previous section we calculated the estimates of the beta coefficients from the generalized linear models. We will use these estimates in order to estimate the income/expenditure of each household in the data frame. Again we consider two models: in Model I we included all covariates, see Appendix II; Model II is the model with the best AIC score in the previous section, Table 6. The second model includes 12 covariates, 10 from the core model and "Washing machine"and "Water Heater" as the additional two variables. To compare the estimated income score in the data frame with the observed income in the small sample we use QQ plots. In Figure 3 we present four QQ plots. These are:

- The estimated income based on Model I for data frame households against the observed income in the sample;
- The log of the estimated income based on Model I for data frame households against the log of the observed income in the sample;
- The estimated income based on Model II for data frame households against the observed income in the sample;
- The log of the estimated income based on Model II for data frame households against the log of the observed income in the sample;

In these plots we threshold the estimated and observed income by removing the households with income greater than 3000 lari. These values can be assumed to be outliers. Recall that in our models in Section 5 we fit the log of the income, $\log(Y)$, therefore the model did not suffer from outliers and we did not remove these before fitting the model. Also the QQ plots of log of estimated and observed incomes for both models show that these are much closer to each other than the income without log transformation. From these plots it is hard to choose the best model; both models seem to have similar estimated income. In Figure 4 we compare the estimated incomes from the two different models with each other. From the plots it is even more clear that the two models give almost the same household income estimates for the data frame. We recommend to base the data frame stratification on the second model, since it contains only 12 covariates.

Using the estimated income from Model II we now can define the strata for the data frame as

Figure 3 QQ plots of the estimated and observed income

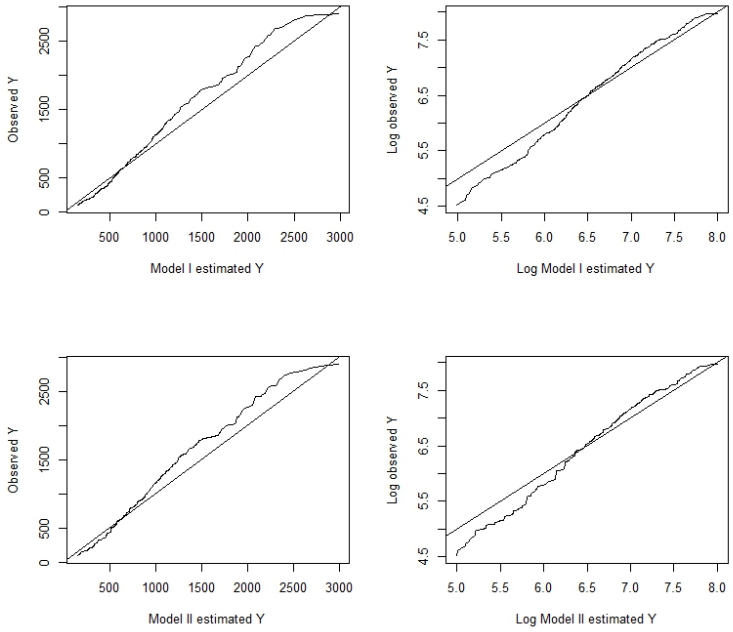
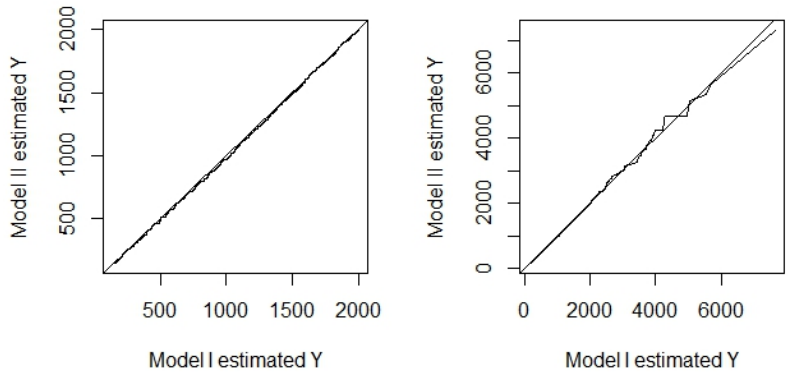


Figure 4 Estimated income, Models I vs II



follows: Let us suppose that we want to define 10 strata's within the data frame. We first order the estimated income and then calculate the 10-th, 20-th, etc. percentiles for the cumulative empirical distribution. In this way we obtain 10 groups of households with equal inclusion probabilities. We can then sample the same number of households from each group. Any other number of strata can be defined similarly.

7 Conclusions

The goal of the pilot study was the improvement of the stratification and the sampling design for the IHS; one of the issues is the stratification of the IHS data frame according to the household income. However, the household income is one of the most difficult and unreliable variables in the IHS questionnaire and there is no other source available for GeoStat for household income.

We designed a pilot study in order to estimate household income and stratify the IHS data frame. The pilot study was carried out only in the Tbilisi area. The main idea of the study was to find covariates that would explain the household income best. We chose household assets as possible covariates. The GeoStat team developed the asset ownership questionnaire. The questionnaire was first applied on a small group (790 households) from the Tbilisi area. For these households the income was known beforehand. Using a linear regression model we identified the significant asset ownership variables for household income and changed the questionnaire by removing the least significant variables. Next, the improved questionnaire was applied to a bigger group in the Tbilisi area, in total 8170 households. This group is used as the data frame for the IHS in the Tbilisi area. We recommend GeoStat to use the response and our linear regression model coefficients to estimate the household income in the data frame. Based on the estimated income we defined a stratification method for the IHS frame in the Tbilisi area according to household income.

Availability of household income information for sampling/weighting purposes can potentially bring very significant benefits to the IHS by making the income distribution of households more representative. Among other statistics the IHS data are also used for determining the poverty line, and also for this purpose it is important to have a well stratified sample frame for households income and wealth.

Note that the asset ownership questionnaire is designed and tested for the Tbilisi area. The questionnaire therefore will not yet be suitable for the rural area. The relation between the asset ownership and income is very different between urban and rural areas. We recommend to design a new questionnaire for the rural areas. A similar approach as described in this paper can be applied for rural areas.

References

P.M. Hauser and M.H. Hansen. On sampling in market surveys. *Journal of Marketing*, 9(1), 1944.

C. Kruiskamp. Report on integrated household survey questionnaire design. Internal report, Statistics Netherlands, 2013.

N. Mushkudiani. Bilateral collaboration between statistics netherlands and the national statistics office of georgia, project scope. Internal report, Statistics Netherlands, 2013.

D. Woensdrecht and P. van Teeffelen. Bezoekrapport geostat 2011. Internal report, Statistics Netherlands, 2011.

Appendices

I Sample frequency table

Covariate	Frequency	Percentage
Second Home/apartment	5	0.81
Third Home/apartment	0	0.00
Second Summer house	2	0.32
Third Summer house	0	0.00
Commercial property	5	0.81
Second Garage	1	0.16
Third Garage	0	0.00
Dwelling under construction	1	0.16
Refrigerator 2	5	0.81
Washing machine 2	1	0.16
Gas stove 2	0	0.00
Dish washing machine	4	0.65
Water heater 2	4	0.65
Personal computer 2	1	0.16
Laptop 2	10	1.62
Tablet computer/Kindly 1	6	0.97
Tablet computer/Kindly 2	1	0.16
Home theater	0	0.00
Digital camera 2	1	0.16
Video camera 1	21	3.40
Video camera 2	0	0.00
Power generator	2	0.32
Air-conditioner 2	1	0.16
LCD or plasma TV-set 2	8	1.29
LCD or plasma TV-set 3	0	0.00
Individual heating system 2	3	0.49
Central heating system	0	0.00

II Full model

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.544	0.267	24.539	0.000
Bedrooms1	0.021	0.073	0.288	0.773
Bedrooms2	0.105	0.074	1.418	0.157
Bedrooms3	0.069	0.092	0.749	0.454
Bedrooms4	0.085	0.142	0.598	0.550
Kitchen1	0.064	0.137	0.469	0.639
Kitchen2	0.075	0.343	0.219	0.827
Bath1	0.057	0.125	0.454	0.650
Bath2	0.168	0.246	0.681	0.496
Cellar1	0.034	0.048	0.713	0.476
Cellar2	−0.048	0.176	−0.271	0.786
HolidayHome1	0.041	0.147	0.282	0.778
HolidayHome2	−0.242	0.081	−2.994	0.003
HolidayAbroad1	0.489	0.252	1.940	0.053
HolidayAbroad2	−0.175	0.195	−0.894	0.372
BuiltYear1	0.040	0.161	0.248	0.804
BuiltYear2	0.086	0.056	1.535	0.125
LastRepairYear1	0.073	0.059	1.236	0.217
LastRepairYear2	0.028	0.066	0.426	0.670
LastRepairYear3	−0.052	0.079	−0.660	0.510
Ownership1	−0.055	0.114	−0.480	0.631
Ownership2	0.545	0.154	3.532	0.000
Ownership3	0.178	0.196	0.912	0.362
Score1	0.136	0.082	1.670	0.095
Score2	0.002	0.076	0.030	0.976
Score3	−0.142	0.071	−2.001	0.046
Score4	−0.083	0.070	−1.186	0.236
HomeApartment1	0.052	0.067	0.782	0.434
SummerHouse1	−0.059	0.109	−0.543	0.587
Garage1	0.079	0.076	1.029	0.304
Land1	0.049	0.118	0.417	0.677
FamilySize1	−1.053	0.094	−11.186	0.000
FamilySize2	−0.474	0.071	−6.711	0.000
FamilySize3	−0.177	0.068	−2.597	0.010
ScoreCar1	0.163	0.114	1.428	0.154
ScoreCar2	0.186	0.071	2.615	0.009
ScoreCar3	0.125	0.082	1.531	0.126
Frige1	0.282	0.082	3.436	0.001
Washing machine1	0.096	0.062	1.533	0.126
Kitchenmachine1	0.040	0.080	0.492	0.623
Foreignvacuumcleaner1	0.218	0.054	4.040	0.000
Water Heater1	0.091	0.052	1.742	0.082
LCDorPlasmaTVset1	0.179	0.056	3.195	0.001
Laptop1	0.171	0.056	3.038	0.002

III Core Model

Covariate	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.801	0.226	30.063	0.000
HolidayHome1	0.033	0.144	0.229	0.819
HolidayHome2	−0.290	0.077	−3.746	0.000
HolidayAbroad1	0.485	0.244	1.984	0.048
HolidayAbroad2	−0.160	0.186	−0.859	0.391
Ownership1	0.051	0.100	0.506	0.613
Ownership2	0.604	0.146	4.147	0.000
Ownership3	0.229	0.187	1.221	0.222
Score1	0.111	0.073	1.517	0.130
Score2	−0.042	0.068	−0.618	0.536
Score3	−0.217	0.062	−3.486	0.001
Score4	−0.100	0.062	−1.600	0.110
FamilySize1	−1.179	0.088	−13.422	0.000
FamilySize2	−0.537	0.068	−7.957	0.000
FamilySize3	−0.205	0.066	−3.086	0.002
ScoreCar1	0.192	0.113	1.704	0.089
ScoreCar2	0.209	0.070	3.002	0.003
ScoreCar3	0.160	0.080	1.988	0.047
Frige1	0.412	0.076	5.401	0.000
Foreignvacuumcleaner1	0.260	0.051	5.074	0.000
LCDorPlasmaTVset1	0.221	0.054	4.125	0.000
Laptop1	0.180	0.055	3.273	0.001

IV Home equipments in the data frame

Covariate	Frequencies	Percentages
Refrigerator	5031	83.68
Washing machine	4267	70.97
Kitchen machine	177	2.94
Dish washing machine	80	1.33
Foreign vacuum cleaner	1432	23.82
Water heater	3333	55.44
Personal computer	2670	44.41
Laptop	1331	22.14
Tablet computer	88	1.46
Mini system	378	6.29
Home theater	39	0.65
Professional digital camera	100	1.66
Video camera	173	2.88
Satellite TV antenna	348	5.79
Air-conditioner	453	7.53
LCD or plasma TV-set 1	1581	26.30
LCD or plasma TV-set 2	121	2.01
LCD or plasma TV-set 3	17	0.28
Central heating system	656	10.91

V Asset Ownership Questionnaire

National Statistics Office of Georgia
Geostat

Monitoring of Households

Economic condition of the household

The day of the interview			
	Date	**Month**	**Year**
The area to be described	**Household**		

Name of the respondent				
Respondent's phone number	**Fixed:**		**Mobile:**	
	(code)	(number)	(code)	(number)
Interviewer				
Supervisor				

According to the Georgian law **"Regarding the Official Statistics" 25th Article, point first,** Geostat is authorized to ask and get all statistical and other kind (including confidential) information from administrative authorities and other physical and judicial person, for implementing its duties.

The information provided by you is confidential and protected by the "Regarding the Official Statistics" law of Georgia 28th Article, point first. It is only used for calculating the general statistical indexes.

Thank you for your cooperation!

Results of the interview

1. Completed interview
2. Dwelling is closed (no one lives)
3. Is not used for living
4. Refused
5. Other reason (please specify) ----------

I. Real Estate module

1. Which of the listed below items does your household owns or uses?

(Please define the name of block and street in case of Tbilisi)

		Location				Type of ownership					Date of built	Whole Area (sq.m)	Among area of stores (basement and etc.) (sq.m)	Year of the last major repair (whole or part)
		District	**Name of the settlement** (city, village)	**Block** (fill only in case of Tbilisi)	**Street**	Belongs to the household only	Rented	Mortgaged	Shared	Is in free use				
	1	*2*	*3*	*4*	*5*	*6*					*7*	*8*	*9*	*10*
1.0	Main dwelling					1	2	3	4	5				
1.1	Home/apartment 1					1	2	3	4	5				
1.2	Home/apartment 2					1	2	3	4	5				
2.1	Summer house 1					1	2	3	4	5				
2.2	Summer house 2					1	2	3	4	5				
3.0	Commercial property					1	2	3	4	5				
4.1	Garage 1					1	2	3	4	5				
4.2	Garage 2					1	2	3	4	5				
5.0	Dwelling under construction					1			4					
6.0	Nonagricultural land					1	2	3	4	5				

2. How many rooms are there in your main dwelling?

(Except the kitchen, corridor, bathroom, toilet, waiting room and other premises)

1. Total: _____
2. Of which bedrooms: _____

II. Durable goods module

3. Do you owe below listed durable goods, which are in working condition? (Includes only durable goods owned by the household)

(Encircle the appropriate code)

		Manufacturer and model		Release year	The year when it was bought	Cost	
						Currency	Cost
	1	*2*	*3*	*4*	*5*	*6*	*7*
		Manufacturer	Model				
1.1	Car 1						
1.2	Car 2						
1.3	Car 3						
2.1	Motorcycle						
3.1	Other *(please specify) 1* _____						
3.2	Other *(please specify)* 2 _____						

4. Does your household owe below listed durable goods, which are in working condition? (Includes only durable goods owned by the household)

(Encircle the appropriate code)

		Type	The year when it was bought	Cost		Condition at the time of purchase	
				Currency	Cost	New	Used
	1	*2*	*3*	*4*	*5*	*6*	
1.1	Refrigerator	1. One compartment 2. Other				1	2
2.1	Washing machine					1	2
3.1	Kitchen machine					1	2
4.1	Dish washing machine					1	2
5.1	Foreign vacuum cleaner					1	2
6.1	Water heater	1. Electric 2. Gas				1	2

		Type	The year when it was bought	Cost		Condition at the time of purchase	
				Currency	Cost	New	Used
	1	*2*	*3*	*4*	*5*	*6*	
7.1	Personal computer					1	2
8.1	Laptop					1	2
9.1	Tablet computer					1	2
10.1	Mini system					1	2
11.1	Home theater					1	2
12.1	Professional digital camera					1	2
13.1	Video camera					1	2
14.1	Satellite TV antenna					1	2
15.1	Air-conditioner					1	2
16.1	LCD or plasma TV-set 1					1	2
16.2	LCD or plasma TV-set 2					1	2
16.3	LCD or plasma TV-set 3					1	2
17.1	Central heating system					1	2

III. Accessibility to various services

5. Does your household have financial ability to go on vacation at least one time an year (at least one household member) inside Georgia or abroad?

(The question implies only vacation by the household's own expenses. Does not include vacation on free e.g. at the village or summer house of relatives/friends)

1. **In Georgia:**
   - 1.1. Our household has financial ability and go on vacation ---- times an year (write down how many times)
   - 1.2. Our household has the financial ability to go on vacation (at least one household member) --- times an year (write down how many times) but we do not go
   - 1.3. Our household does not have financial ability to go on vacation
2. **Abroad:**
   - 1.4. Our household has financial ability and go on vacation ---- times an year *(write down how many times)*
   - 1.5. *Our household has the financial ability to go on vacation (at least one household member) --- times an year (write* down *how many times) but we do not go*
   - 1.6. Our household does not have financial ability to go on vacation

IV. Interviewer's notes and comments