The rent imputation method impact on the accuracy of the estimates of poverty and income inequality indicators

Inga Masiulaitytė

Statistics Lithuania, Vilnius University, Lithuania
e-mail: inga.masiulaityte@stat.gov.lt

Abstract

The aim of the study is to analyze the impact of imputation methods for rent on the accuracy of the estimates of poverty indicators. The analysis of the imputation methods for rent has been done using Lithuanian Household Budget Survey in 2006.

1 Introduction

Analysis of poverty and income inequality indicators is very important part in analyzing of persons' standard of living. The most useful poverty and income inequality indicators are headcount indicator, poverty gap index and Gini coefficient. To calculate these indexes we use total disposable income variable.

Disposable income is the total income from work in cash and in kind (income and social insurance tax excluded), income from farming, business, independent professional work, social benefits, property income, rent, regular support from other persons.

The imputation of rental expenditures is an important step in the estimation of a household's standard of living. An income comparison of two households having the same income but with one household renting and the other being an owner-occupier would, in absence of imputation, conclude that their position is the same; in reality the owner household is better-off because it enjoys housing services for free.

The aim of this research is to calculate all defined indicators, also analyze the rent imputation impact on these indicators.

2 Poverty and income inequality indicators

Let us consider a population U of N earners and let y_i be the total disposable income of the individual i.

2.1 Headcount indicator

Headcount indicator (Khandker, 2004) is a proportion of individuals who are below the poverty line:

$$P_0 = \frac{1}{N}\sum_{i=1}^{N} I_{(y_i<z)} = \frac{N_q}{N}, \qquad I_{y_i<z} = \begin{cases} 1, & if\ y_i < z; \\ 0, & if\ y_i \geq z. \end{cases} \tag{1}$$

here z is the poverty line, N is the total population size.

The headcount indicator can be estimated by:

$$\hat{P}_0 = \frac{1}{\hat{N}}\sum_{i=1}^{n} w_i I_{(y_i<\hat{z})} = \frac{\hat{N}_q}{\hat{N}}, \qquad I_{y_i<z} = \begin{cases} 1, & if\ y_i < \hat{z}; \\ 0, & if\ y_i \geq \hat{z}. \end{cases} \tag{2}$$

here w_i is a sample design weight, $\hat{z}$ is the estimate of poverty line, $\hat{N} = \sum_{i=1}^{n} w_i$ is the estimate of population size.

2.2 Poverty gap index

The poverty gap index is defined as follows:

$$P_1 = \frac{1}{N}\sum_{i=1}^{N} \frac{(z - y_i) I_{(y_i<z)}}{z}, \tag{3}$$

We estimate the poverty gap index by the estimator:

$$\hat{P}_1 = \frac{1}{\hat{N}}\sum_{i=1}^{n} \frac{(\hat{z} - y_i) w_i I_{(y_i<\hat{z})}}{\hat{z}}. \tag{4}$$

2.3 Gini coefficient

Assume that all total disposable income are sorted in ascending order:

$$y_1 \leq y_2 \leq \ldots \leq y_N.$$

Gini coefficient is defined as follows:

$$G(y)=\frac{\sum_{i=1}^{N}(2r(i)-1)y_i}{N\sum_{i=1}^{N}y_i}-1, \quad r(i)=\sum_{l=1}^{N}I_{y_l\le y_i}, \qquad (5)$$

where $r(i)$ is the rank of the individual i.

The Gini coefficient G(y) can be estimated by:

$$\hat{G}(y)=\frac{\sum_{i=1}^{n}(2\hat{r}(i)-1)w_i y_i}{\hat{N}\sum_{i=1}^{n}w_i y_i}-1, \qquad \hat{r}(i)=\sum_{l=1}^{n}w_l I_{y_l\le y_i}, \quad (6)$$

where w_l is the weight of the element l.

3 Imputation methods for rent

In our study we analyse following imputation methods:

a) Homogeneity groups method;

b) Heckman method;

c) Log-linear regression model;

d) Multiple imputations.

We analyze each of these methods under different percent of tenants in the population.

4 Monte Carlo simulation design

An artificial population has been formed using the Lithuanian HBS 2006 data. Population size N is 19130 individuals. The population of individuals is divided into seven strata: cities – Vilnius, Kaunas, Klaipėda, Šiauliai, Panevėžys, other small towns, and rural areas. A large number J (1000) of possible samples (of size n=2000) is drawn from the given population according to the stratified sampling design (see Table 1). An allocation scheme of the sample in each stratum is proportional to the stratum size.

The estimates $\hat{P}_0$ and the variance estimates $\hat{V}(\hat{P}_0)$ are calculated to each sample. Let us denote by $\hat{P}_0^{(j)}$ and $\hat{V}(\hat{P}_0^{(j)})$ the estimation results obtained for the jth sample. Then we calculate mean of the estimates:

$$\bar{\hat{P}}_0 = \frac{1}{1000}\sum_{j=1}^{1000}\hat{P}_0^{(j)}$$

and empirical variance

$$V(\hat{P}_0) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{P}_0^{(j)} - \bar{\hat{P}}_0\right)^2.$$

They are considered to be estimates of indicators P_0 and $V(\hat{P}_0)$. The number of samples 1000 is large enough, the distribution of 1000 estimates that can be considered as an empirical sampling distribution, will closely approximate the exact sampling distribution that we cannot obtain easily.

Table 1. Size of the artificial population and sample

Number	Strata	Artificial population N_h	Sample size n_h
1	Vilnius	1997	206
2	Kaunas	1854	191
3	Klaipėda	764	79
4	Šiauliai	625	64
5	Panevėžys	628	65
6	Other small towns	6262	645
7	Rural areas	7000	750
	Total	19130	2000

5 Conclusions

Simulation is still in progress. The results will be presented at the workshop.

References

Bethlehem J., Cobben F., Schouten B., May 2006, *Nonreponse in Household Surveys*. Version 1, Statistics Netherlands, Methods and Informatics Department, p. 141-142.

DOC HBS/161/2006/EN, DOC EU-SILC/162/06/EN. *Meeting of the working group on living conditions (HBS, EU-SILC). HBS and EU-SILC Imputed rent.* Eurostat-Luxembourg, May 2006.

Eurostat Phare multicountry Programme, Organise by ICON-Institute gmbh/Germany of behalf of Eurostat. *Report on the Methodological Study: "Possible data sourses and methods to calculate the target variable "imputed rent" for Lithuania".* Action code: 1138-ILC-sercon-Lithuania-1

EU–SILC DOC TFMC-12/06. *Third Meeting of the EU-SOLC task force on methodological issues. Imputed rent.*. Eurostat-Luxembourg, April 2006.

Jeroen Smits, (2003) *Estimating the Heckman two-step procedure to control for selection bias with SPSS*. Http://home.planet.nl/~smits.jeroen/selbias/Heckman-SPSS.doc .

Khandker S., (2004) *Poverty Manual,* World Bank.