Sample Weights Calibration with Aim to Reduce the Estimation Bias Due to Under Coverage of the Well-Off Population

Volodymyr Sarioglo[1] and Nataliia Romanchuk[2]

[1] Ptoukha Institute for Demography and Social Studies, e-mail: sarioglo@idss.org.ua
[2] Ptoukha Institute for Demography and Social Studies, e-mail: romanchuk_nataliia@ukr.net

Abstract

In modern household sample surveys many indicators are estimated with significant bias due to the unwillingness of households to answer some questions, and undercoverage of the well-to-do population strata. When there is no access to personalized register data, an effective approach to mitigating these problems can be calibration of survey design weights with the use of relevant external information.

In the State Household Living Conditions Survey (HLCS) provided by the State Statistics Service of Ukraine (SSSU) a complex calibration procedure is tested in 2015 – 2017 in order to reduce the bias of income, income differentiation and income related estimates. Main data sources for this procedure are data on household disposable income from the National Accounting System (NAS) and data from the Tax Administration (TA) on wages and salary distribution. Received results testify to potential efficiency of such approach for increase of HLCS basic indicators reliability.

Keywords: calibration, estimation bias, reliability

1 Introduction

One of the serious problems that official statistics have to deal is reliable estimation of real income of households and their members. This problem is especially relevant for countries with a high level of shadow and informal economy, which is, in particular, Ukraine. In such conditions assessment of household ability-to-pay for utilities and services, level of social support programs targeting, household tax burden, and other issues become very problematic. Accordingly, the efficiency and effectiveness of socio-economic and fiscal policies are reduced.

Data on household incomes derived by modern household sample surveys are characterized by such disadvantages as underestimation of income due to the unwillingness of households to answer questions about the level and sources of income, and inadequate coverage of the well-to-do population strata due to their refusal to participate in surveys. Over the past decade these problems have become much worse

which negatively affects the reliability of the direct estimation of many important indicators by the results of sample surveys.

One of the main approaches to overcome these problems is the use of additional information (auxiliary in relation to the survey data). The use of such information is possible at different survey stages and depends on the research objectives, available auxiliary information, its quality and compatibility with the main source of data, etc.

Calibration of sample design weights is one of such approaches used at the indicator estimation stage (Deville, J.-C. and Särndal, C.-E. 1992). Calibration allows you to take into account available reliable auxiliary information in the indicator estimates and to provide analysis using full survey data set. In the HLCS during 2015 – 2017 a complex calibration procedure with a view to reduce possible biases in estimates of indicators due to households refuses to participate in the survey was tested. This procedure use demographic data, data from the NAS and the TA.

2 Methodological approach and results

The calibration procedure is carried out in 3 main stages: preparatory, and two calibration stages. At the preparatory stage data from additional sources are prepared for use in the calibration procedure. Numbers of population by sex-age groups and regions, numbers of households by regions and type of area are calculated. This information is used for calibration in the HLCS more than ten years. The new auxiliary information is the percentile distribution of the TA data on wages and salary by regions and data from NAS on disposable income by regions. For this information at the preparatory stage some data harmonization procedures are implemented: for instance, in the NAS data the amount of imputed rent is excluded from the amount of disposable income in every region; in the TA data the amounts of social contribution, income tax and military tax are excluded from the total amount of wages and salary in percentiles.

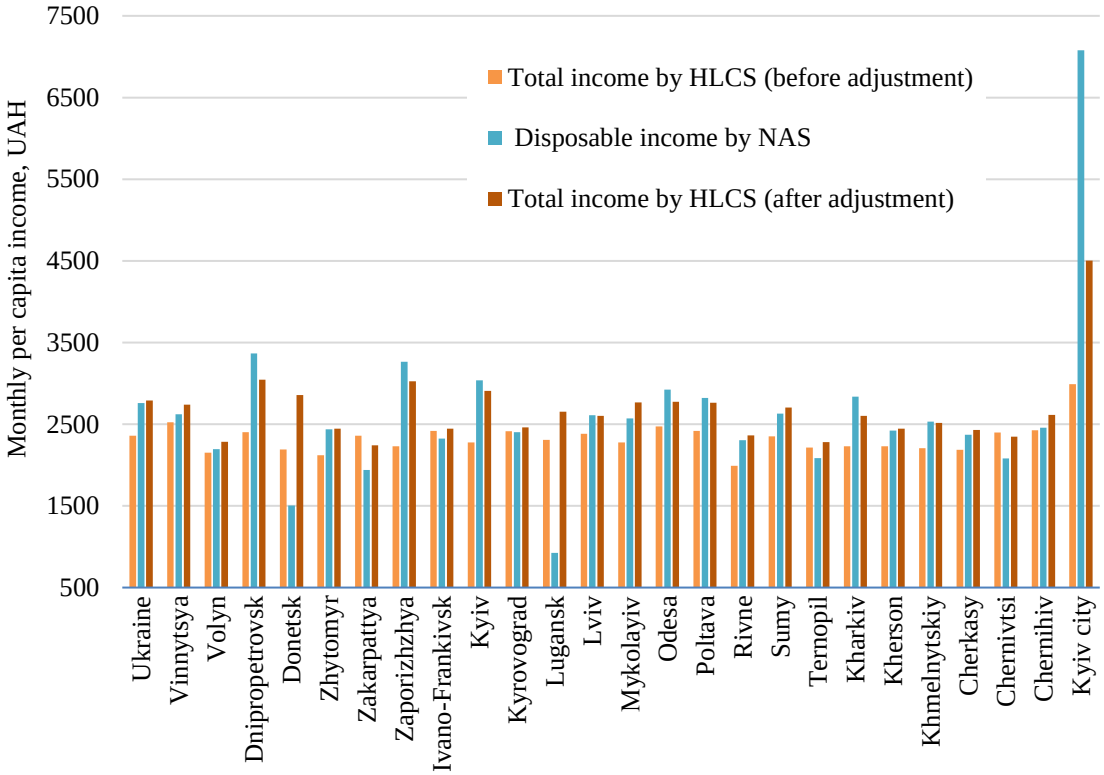
At the first stage of calibration of the HLCS sample weights the prepared auxiliary information on population by strata (regions and area type), household numbers by strata, and age - sex structure of population is used (SSSU 2011).

At the second stage of calibration information on regional distribution of disposable income and decile distribution of population by TA data is used (SSSU 2014). Wherein only TA data for decile groups in which number of people by TA data is higher than by HLCS estimates is considered. For all regions these are 8 -10 or 9, 10 decile groups only. Also numbers of population and households by regions and type of area are taken into account.

As it can be seen from the data presented on the Fig. 1, results of adjustment (calibration) are more significant for regions with higher household disposable income – Dnipropetrovsk, Zaporzhzhya, Kyiv city. It should be noted that in these regions the amount of wages and salary for highest decile groups by TA data is also higher. For some western regions – Zakarpattya, Ivano-Frankivsk, Ternopil, Chernivtsi – the

disposable income is lower than income, estimated by HLCS, but TA data nevertheless lead to a correction of income upwards.

Figure 1: Monthly per capita income by region of Ukraine, 2016



In the Table 1 some differences in estimates of household expenditures before and after weights calibration are presented. As it can be seen for some groups of expenditures the differences are quite significant.

It should be noted that in practice full compliance of HLCS adjusted estimates and auxiliary data is generally not achieved. This is due to restrictions on the minimal quality level of calibrated weights – maximum and minimum values, correlations with the design weights etc.

As it follows from the obtained results, some direct estimates of the HLCS can be substantially biased. Accordingly, their reliability in reality can be much lower than estimated without taking this fact into account. In our opinion the proposed calibration scheme can significantly improve the reliability of the HLCS results.

Table 1: Differences in estimates of household expenditures
before and after weights calibration

	Per household (UAH)		***% of total consumption expenditures***	
	before adjustment	*after adjustment*	*before adjustment*	*after adjustment*
Consumption expenditures				
food and non-alcoholic beverages	2852.69	3055.22	53.6	51.5
alcoholic beverages and tobacco	168.14	185.31	3.10	3.20
clothing and footwear	314.65	367.47	5.9	6.2
housing, water, gas, electricity and other fuels	917.53	999.77	17.2	16.8
furnishings, household equipment and routine maintenance of the house	97.16	116.10	1.8	2.0
transport	205.33	270.94	3.9	4.6
recreation and culture	80.63	106.40	1.5	1.8
restaurants and hotels	125.04	169.04	2.3	2.9
miscellaneous goods and services	142.06	165.86	2.7	2.8
Non consumption expenditures				
help relatives, other people	186.74	224.53	3.2	3.4
purchase of real estate, bank deposits, construction, overhaul	141.73	271.11	2.5	4.2
GINI index (by the total income)			0.220	0.234
Number of persons with equivalent per capita total income lower than the minimum subsistence level, %			51.1	35.2

References

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87:376–382.

Methodology of calculating the sampling weights for grossing up of the results of the State Household Living Conditions Sample Survey on the population. – Kyiv, State Statistics Service of Ukraine (SSSU), 2011 (in Ukrainian). Available at: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2006/521/metod.htm (Accessed: 18 May 2018).

Methodology of adjustment of household's living condition indicators with the purpose of taking into account income and expenditures of the well-to-do population. – Kyiv, State Statistics Service of Ukraine (SSSU), 2014 (in Ukrainian). Available at: http://www.ukrstat.gov.ua/metod_polog/metod_doc/2014/415/metod_pol_koryg_zhrn.zip (Accessed: 18 May 2018).