Working paper 2

Distr.: General
13 November 2024

English

Economic Commission for Europe
Conference of European Statisticians
Group of Experts on Measuring Poverty and Inequality
Geneva, Switzerland, 28-29 November 2024

Data sources for measuring poverty and inequality in Ukraine during the war

Note by V. Sarioglo, Institute for Demography and Life Quality Problems of the National Academy of Science of Ukraine

Abstract

The crisis triggered by Russia's invasion of Ukraine in February 2022 has led to large-scale displacement, both internally and across borders. This situation has necessitated the development of updated methods to generate relevant and timely statistics for stakeholders—particularly the Government, regional authorities, and international organizations—focused on social, economic, and community reconstruction efforts. The challenge intensified when the State Statistics Service of Ukraine (SSSU) suspended population and household surveys, along with the production of official statistics on demographics, living conditions, and the labor market, in accordance with martial law. Under these circumstances, it became crucial to estimate key indicators through alternative data sources, such as administrative registers and independent sample surveys conducted without the SSSU's direct involvement.

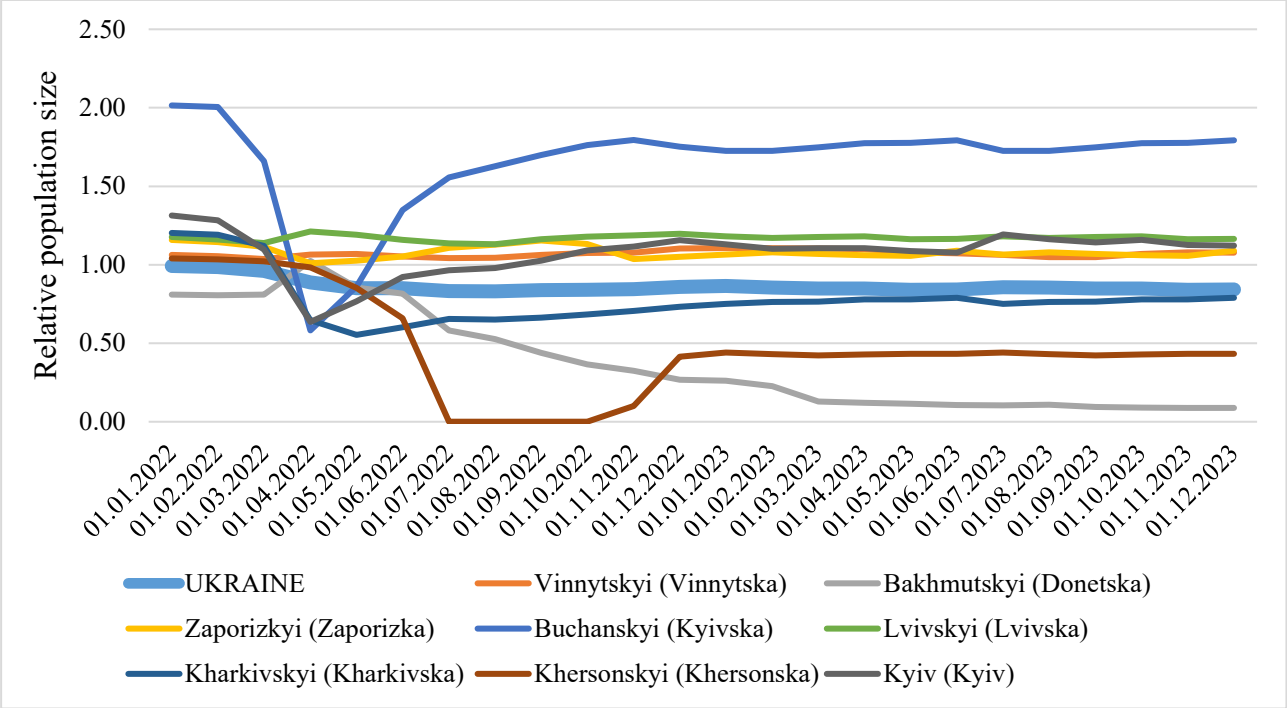
In 2023, the United Nations Population Fund (UNFPA), in collaboration with the Institute for Demography and Life Quality Problems of the National Academy of Sciences of Ukraine (IDLQP), the SSSU, and other organizations, developed an approach to estimate the population's size, structure, and distribution as of mid-2023 [1]. The estimates were based on the following main data sources:

- **Pre-war population statistics** (SSSU, UNFPA);
- **Aggregated mobile operators' data** on the number and location of subscribers;
- **Ministry of Justice of Ukraine** records on registered births and deaths;
- **Public Health Center of the Ministry of Health of Ukraine** data on newborns and children registered for medical care;
- **Ministry of Social Policy of Ukraine** data on the number and location of internally displaced persons (IDPs);
- **International Organization for Migration (IOM)** surveys providing humanitarian data within Ukraine;
- **United Nations Refugee Agency (UNHCR)** data on the number, demographic characteristics, and regions of origin of external migrants from Ukraine;
- **Pension Fund of Ukraine** data on the number and demographic characteristics of pensioners;
- **Ministry of Education and Science of Ukraine** data on the number, distribution, and gender-age structure of students across various levels of education.

Estimates for the population size and location within government-controlled areas (GCA) of Ukraine were primarily derived from mobile operators' data. This included information on the number and distribution of users from the three major Ukrainian mobile operators—Kyivstar, Vodafone Ukraine, and lifecell — as well as data from a specialized population survey on the use of mobile devices. The survey accounted for device numbers, mobile operator usage, and other relevant factors.

Figure 1 presents the population dynamics for selected regions of Ukraine during 2022–2023. The vertical axis (ordinate) shows relative values, calculated as the ratios of the estimates to the official population statistics as of January 1, 2022.

Figure 1. Population changes in Ukraine as a whole and in selected administrative units (regions) from January 1, 2022, to December 1, 2023 (with region names in parentheses).



Using this approach, it was determined that approximately 31.7 million people were living in the GCA of Ukraine as of July 2023. Additionally, estimates of the population's sex and age structure, broken down into five-year age groups, were produced at the national level, as well as for regions and administrative raions.

A critical data gap in 2022–2023 was the absence of information on household composition, living conditions, income and expenditures, economic activity, poverty, and inequality. To address this gap, a **Household Socio-Economic Status Survey (HSESS)** was conducted with organizational support from the Ministry of Social Policy of Ukraine and technical assistance from UNICEF. The survey took place at the end of 2023 and employed a probability sample of 8,023 households, following a methodology closely aligned with that of state sample surveys. This effort provided essential insights into household composition, poverty across various dimensions, the effectiveness of social programs, employment and unemployment, among other key indicators.

It should be noted that the population size and distribution estimates, obtained using the approach described above, were instrumental in designing the HSESS sample. Additionally, the sex and age structure estimates were used to calibrate the statistical weights, enhancing the representativeness of the survey results.

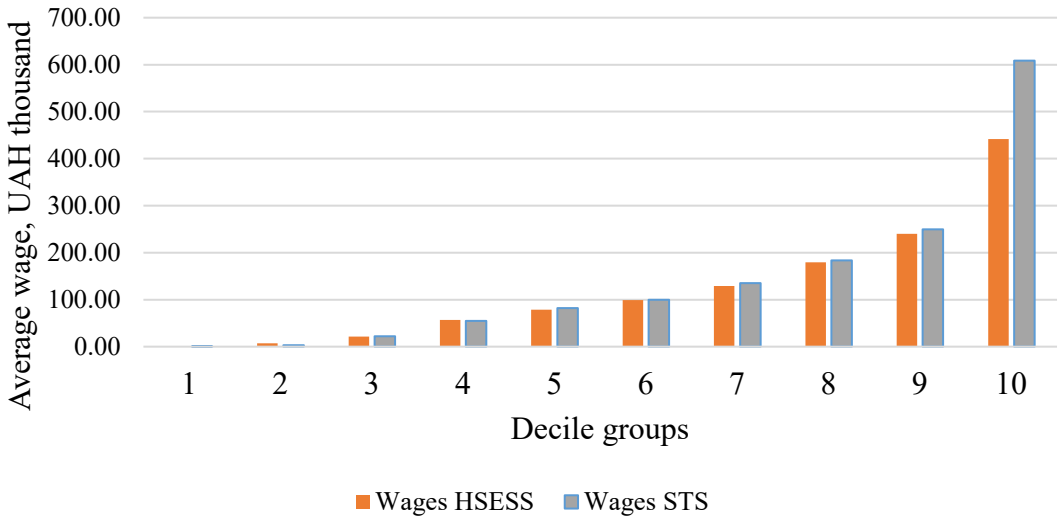
According to the population estimates and HSESS findings, the average household size in the GCA decreased from 2.57 to 2.29 persons compared to pre-war estimates. Furthermore, the number of households declined by approximately 975,000, from 14,549.2 thousand at the beginning of 2022 to 13,574.5 thousand at the beginning of 2024.

To enhance the reliability of living standards estimates, some HSESS results were cross-checked with data from other sources. **Figure 2** compares the distribution of average yearly wages by decile groups for working individuals, based on HSESS data and data from the State Tax Service of

Ukraine (hereinafter, STS). The deciles were determined using wage distribution from STS microdata.

It is important to note that the data were not adjusted for full employment, which accounts for the low values observed in the lower decile groups. As shown in the figure, significant wage differences exist in the tenth decile group. As expected, wage data from the STS are substantially higher than those from the HSESS.

Figure 2. Comparison of average wages based on STS and HSESS data by decile groups, with deciles constructed using STS data



It is well known that such effects are primarily associated with the undercoverage of wealthy households in sample surveys, due to their refusal to participate, inaccessibility, or reluctance to disclose sensitive information about income and expenditures. As our studies demonstrate, this undercoverage can introduce significant biases in estimates of average household income and expenditures, as well as in poverty and inequality measures. Consequently, the reliability characteristics of these estimates may be significantly worse than those calculated without taking bias into account. In such cases, the use of external data sources to enhance the reliability of survey results is particularly appealing—especially when access to a comprehensive microdata set, rather than isolated variables, is available. In practice, improvements to sample survey outcomes at the estimation stage are achievable primarily through modifications to estimation procedures, particularly by refining the system of sample weights.

One effective method for this is the **weight calibration procedure** [2]. The calibration process minimizes the distance between the calibrated weights and the original design weights. The calibrated weights must meet specific conditions, defined based on the selected auxiliary data, to ensure alignment between the survey results and external data sources. In microdata sets, users typically see and use only a single variable representing sample weights, making any reduction in bias achieved through calibration procedures difficult to detect. However, for experts in statistical agencies, understanding the magnitude of bias can reveal the extent to which certain population strata are underrepresented, if such issues exist.

In Ukraine, the weights for the state **Household Living Conditions Survey (HLCS)** are calibrated using auxiliary information, including population counts by strata (region and area type) and the population's age-sex structure [3]. Over the past five years, the calibration process has been refined to better reflect the living conditions of higher-income groups. One promising approach involves calibrating the sample weights using wage distribution data from the STS [4]. This results in a more intricate calibration process, significantly adjusting the original design weights to improve the representativeness and reliability of the survey findings.

For example, our estimates for the 2019 HLCS show that the bias in the average cash income estimate can be as high as 10.9% for urban households and 3.3% for rural households. This discrepancy arises because urban areas have a higher concentration of high-wage earners compared to rural areas. Notably, this level of bias significantly impacts the national-level reliability of income estimates. Specifically, the actual reliability corresponds to a **relative root mean square error (RRMSE)** of 10.98%, compared to a **coefficient of variation** of 1.00%, which was calculated without accounting for the bias.

Incorporating STS data into the analysis results in a more than 10% increase in the **Gini coefficient**, indicating higher income inequality, with other measures of income distribution showing similar increases. Additionally, the poverty rate, measured against the actual subsistence minimum, would decrease by 8%.

Based on these findings, several key conclusions can be drawn. In the absence of official statistics during wartime, Ukraine possesses a wide range of alternative data sources that can provide approximate estimates for many key indicators. Moreover, it is feasible to conduct sample surveys of household income, expenditures, and economic activity using methodologies closely aligned with state surveys. While such surveys cannot fully replace official statistics, they can still provide critical information to support the development of socio-economic policies and the implementation of social and humanitarian programs.

An effective way to enhance the reliability of living standards, poverty, and inequality estimates is to account for the incomes of high-earning households through the **calibration of statistical weights** using external data sources. A prime example of such external data is wage information from the STS. Although the use of external data may increase the variability of statistical weights, it significantly reduces bias in income and expenditure estimates. This also refines poverty indicators and improves the measurement of income inequality.

References

[1] Silva, R., Snyder, M., Han, M. D., Sarioglo, V., and Libanova, E. (2023). Subnational population projections for humanitarian response in Ukraine: integration and cross-validation of traditional and non-traditional sources: Quetelet 2023 Seminar, (November 9–10, 2023). Leuven. Retrieved from https://uclouvain.odoo.com/en_US/event/quetelet-2023-seminar-403/page/introduction-quetelet-2023-seminar
[2] Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87:376–382.
[3] 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).
[4] 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).