Discussion on the report: "The added informative value of combining different data sources"

Big Data training program: final webinar, December, 6, 2022

Volodymyr Sarioglo

Ptoukha Institute for Demography and Social Studies

The problem of "combining" official statistics/administrative data with web data

Web surveys become more and more used mode of dada collection in the official statistics (NSO). In the OS there is a different problem on combining of data (in one survey) collected using different mode: face-to-face ---- phone ----- web

So, we should speak about combining of Big Data and official statistics

Danny Pfeffermann, one of the leading contemporary statisticians in 2015 said:

Big Data $\longrightarrow$ *Big Problems* $\longrightarrow$ *Big Headache*

- ✓ Coverage/selection bias
- ✓ Data accessibility
- ✓ New legislation
- ✓ Privacy (data protection)
- ✓ Disclosure control
- ✓ Computer storage
- ✓ Computation and Analysis
- ✓ Linkage of different files
- ✓ Risk of data manipulation

But some of these problems have already been solved or are being successfully solved at the present time.

Data collection and analysis must change, as well as statistical production

In our opinion, if we talk about official statistics and Big Data, it is advisable to distinguish between the stages of the of the formation of information support, in particular for LMI:

✓ data collection stage

✓ data processing stage

✓ indicator estimation stage

✓ data analysis stage (extraction of information from the data)
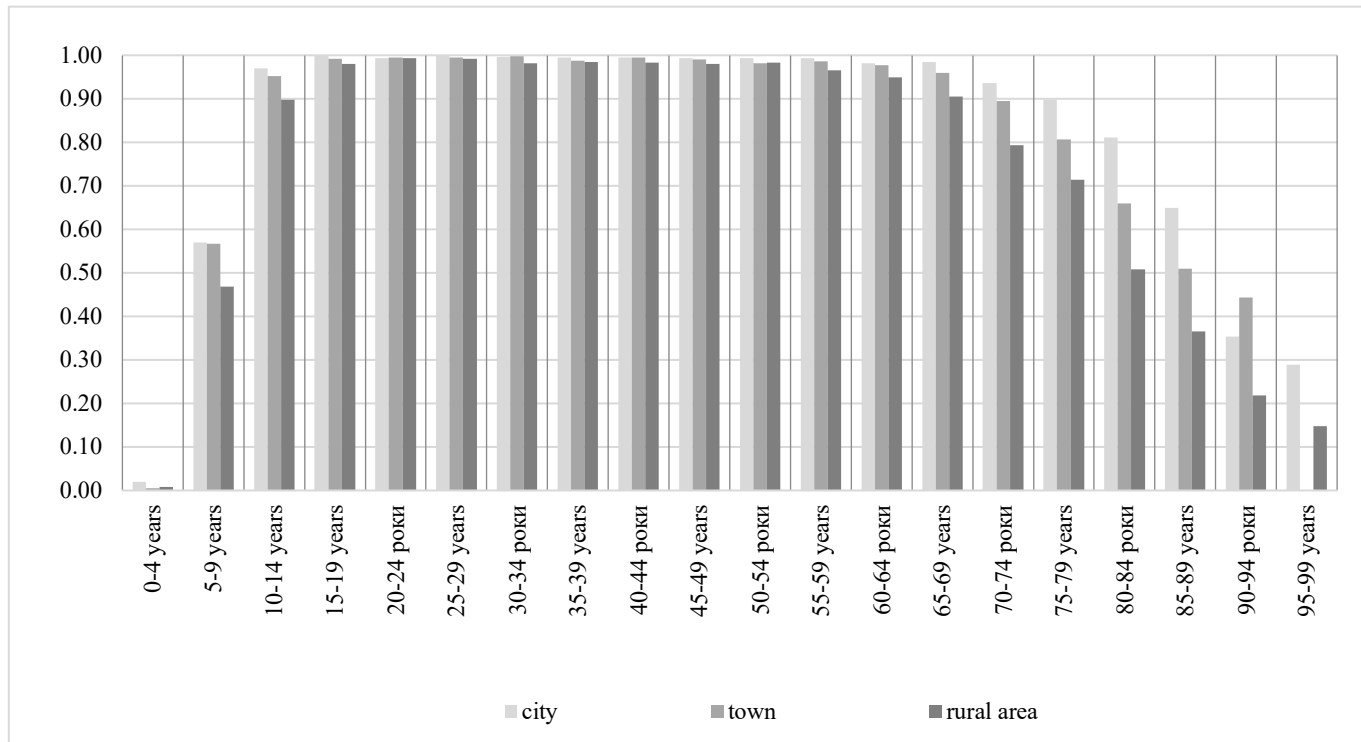
Big data can be used both at each of these stages separately, and at once at several stages or at all stages at once.

3C steps to Compare, Correct, and Combine OJA-based analyses with official statistics to improve trend believability

The general problem of the 3C steps from the OS point of view is harmonization of data from different sources and/or results of the analysis. Harmonization in different aspects:
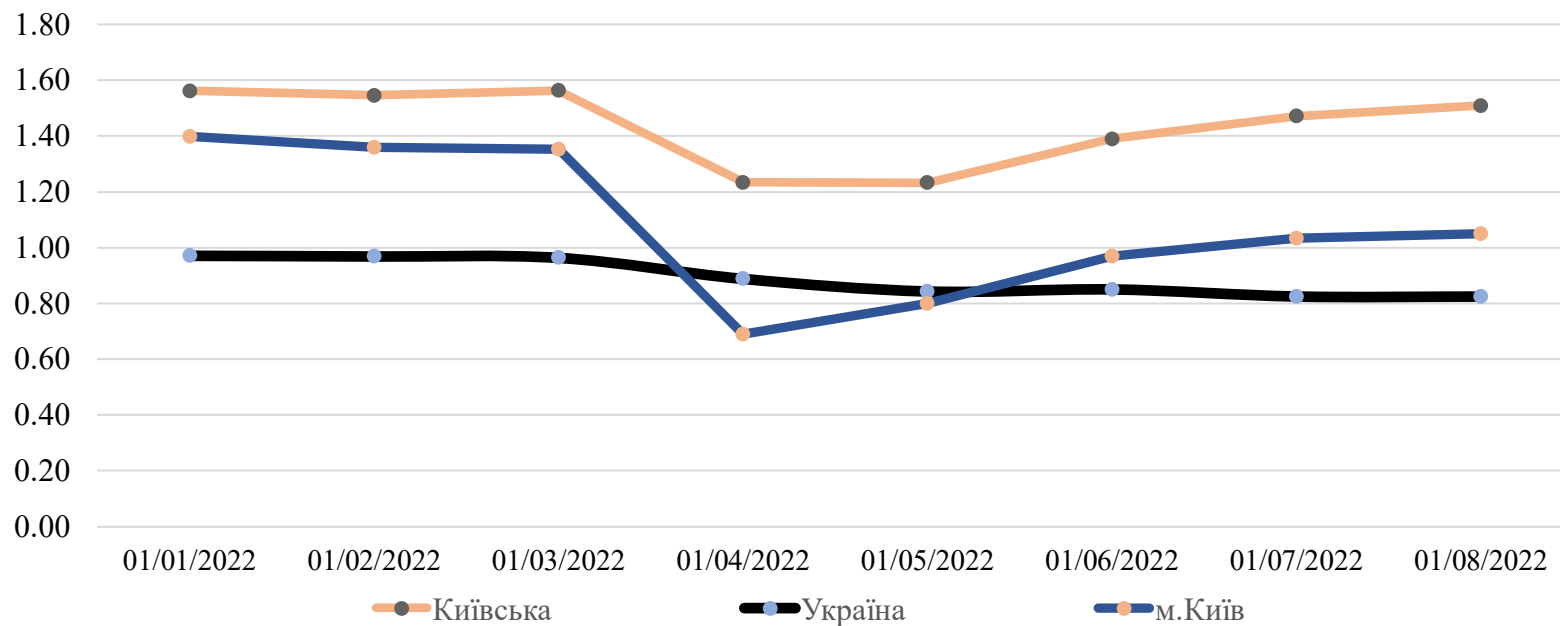
- ✓ by units of observation
- ✓ by time (by period of time)
- ✓ by population coverage
- ✓ by characteristics of units of observation and indicators
- ✓ by used classifications
- ✓ etc.

Proportion of mobile users by type of area and age group

Source: built by authors using data of the State Statistics Survey

Dynamics of the population of Ukraine in 2022

Source: State Statistics Service of Ukraine, experimental estimates on base of MO data

Conclusions

The presented draft report "The added informative value of combining different data sources" contains very useful information for all those involved in the problems of LMI, highlights modern approaches to the problem of using Big Data in official statistics.

In fact, there is no other way for official statistics to increase their potential and demand than modernization, including through the use of Big Data. This has become especially clear in the last decade with increased mobility of the population and labor force, the impact of the Covid 19 pandemic, military actions in different countries.

References

The draft report "The added informative value of combining different data sources", ETF, 2022. Authors: Emilio Colombo, Pietro Giorgio Lovaglio, Fabio Mercorio, Mario Mezzanzanica, and Francesco Trentini

Pfeffermann, D. Official Statistics for the Next Decade - Methodological Issues and Challenges / Presentation / Conference on New Techniques and Technologies for Statistics (NTTS), March 2915

Pfeffermann, D. (2015). Methodological Issues and Challenges in the Production of Official Statistics. Journal of Survey Statistics and Methodology, Vol. 3, 4 425-483 [in English].

Mode and web panel experiments in the European Social Survey – Lessons for EU-SILC / Improving the measurement of poverty and social exclusion in Europe: reducing non-sampling errors. Edited by Peter Lynn and Lars Lyberg, 2022 edition

Thanks for your attention!