Creating a synthetic database for research in migration and subjective well-being

Statistical Matching techniques for combining the basic and complementary questionnaires of the Hungarian Microcensus 2016

**Keywords:** Migration and new techniques of data collection, Statistical Matching, Migration aspirations, Subjective well-being, Hungarian Microcensus

# Introduction

In 2016, with the aim of tracking social trends between full-scope censuses, the Hungarian Central Statistical Office (HCSO) carried out Microcensus, a population survey based on an unusually large sample covering 10 percent of the Hungarian households. Apart from the basic questionnaires on dwellings and personal information, selected households were asked to fill in one of the following complementary surveys on a) international migration, b) subjective well-being, c) social stratification, d) occupational prestige, and e) health problems.

From a methodological point of view, the Microcensus dataset with the above described structure of basic and complementary questionnaires invites for performing a statistical – or *synthetic* – matching exercise. This method, in accordance with the UNECE Data Integration Guide [1], “involves the integration of data sources with usually distinct samples from the same target population, in order to study and provide information on the relationship of variables not jointly observed in the data sets”. That is, the statistical matching exercise resembles an “imputation problem of the target variables from a donor to a recipient survey” on the basis of common variables [2].

The HCSO methodologists and experts of population and migration statistics embarked on creating such a synthetic Microcensus database of the complementary modules on the basis of the variables from the basic questionnaires. The resulting dataset – unique in terms of sample size – will contain, apart from the information obtained in the basic questionnaires, the estimated/imputed data from each of the complementary sets of questions. Thus, it will serve to develop analytic richness through making possible the study of the diverse relationships between a wide range of variables never observed together in Hungary.

This presentation – instead of entering in the details of the analytical results – focuses on the methodological questions of creating such a synthetic database and the evaluation of the usability and potential of the output dataset. It also gives an overview of the first phase of and the lessons already learnt from the ongoing experiment and insight into the next steps.

# Methods

# In their joint effort, as a first step, the multidisciplinary project team is currently working on identifying the best solutions to combine the basic and two of the complementary questionnaires: the one on migration and another on the aspects of subjective well-being. In the next phases, the methodological solutions identified in this step will be expanded in order to include other complementary questionnaires in the exercise. The reason for choosing the migration and the subjective well-being questionnaires to be combined in the first step of the matching exercise is the growing interest of social researchers for understanding the interconnections between migration behaviour and the subjective evaluations of the different aspects of the contexts in which migration decisions take place. Innumerable studies have been carried out in recent years regarding the relations of well-being and migration behaviour (for a summary, see e.g. [3], [4], [5]).

The HCSO team sees statistical matching – proved to be a valid and useful method in diverse fields of socio-economic research (e.g. [6]) – as a means of creating an appropriate database for studying such relationships in details on the basis of the Microcensus dataset.

The Microcensus dataset fulfils the conditions or prerequisites recommended for carrying out the statistical matching exercise [7]:

* detailed methodological information available on the sampling design and the variables of data collection,
* common methodology for data collection across the basic and complementary questionnaires,
* the sub-samples for complementary questionnaires belong to the same target population,
* data collection was carried out in the same period,
* similar distributions of common key variables,
* lack of data (missing data) negligible in both datasets.

The following approaches are foreseen to be used in order to find out which method gives the results desired in terms of imputation quality evaluation (methodological perspective) and in terms of comparability with national and international empirical data regarding the key variables (analytical perspective):

* random hot deck matching techniques (nonparametric micro approach),
* nearest neighbour distance hot deck techniques (nonparametric micro approach),
* rank hot deck distance method (nonparametric micro approach),
* regression based methods (parametric approaches)
* multiple imputation methods (parametric approaches)
* mixed methods.

# Expected Results

Since well-being has become one of the key concepts of migration research today, more and more empirical data is available on this issue.[[1]](#footnote-1) Existing analyses focusing on the key variables of such relationships will be compared with the results we obtain once the synthetic Microcensus database is created.

That is, the resulting dataset will be evaluated in two steps:

1. evaluating imputation quality,
2. *(if imputation quality appears to be satisfactory*) comparing the relationships of main variables (e.g. migration aspirations and satisfaction with life as a whole) with national and international empirical data.

That is, on the one hand, our results are expected to contribute to the methodological literature on statistical matching with the technical details of the experiences and lessons learnt (including the methodological challenges of carrying out statistical matching and evaluating its quality) anticipating the completion of the synthetic database with the other complementary questionnaires of Microcensus.

On the other hand, the newly created dataset will expectedly bear a great analytical potential as regards the relations between diverse subjective well-being indicators (as overall satisfaction, satisfaction with different aspects of life, mental states, trust or others) and migration characteristics (such as motivations for migration behaviour, sending remittances or over-qualification) involved in the statistical matching exercise. Thus, it will make possible to get newer and more detailed insights into the migration and well-being interconnections in cross-sectional analyses and international comparisons.

# Conclusions

Despite statistical matching appears to be a promising tool for the creation of databases appropriate for the analysis of variables not observed together, this method is still underexploited for the aims of developing official statistics as well as for research purposes. The joint effort of HCSO methodologists and experts in population and migration statistics aims at making the first step in changing this tradition in Hungary: it contributes to deepen our knowledge regarding statistical matching techniques by applying this method to new data sources, and as such, also to open new possibilities for future work and research both on the domains of migration and well-being, and on using statistical matching techniques to support such research.

# References

[1] UNECE HLG-MOS, A Guide to Data Integration for Official Statistics,Version 2.0, (2018)

[2] A. Leulescu and M. Agafitei, Statistical matching: a model based approach for data integration,Eurostat Methodologies and Working Papers (2013)

[3] N. Simpson, Happiness and Migration. In: K. F. Zimmerman and A. F. Constant, (eds.) International Handbook on the Economics of Migration, Edward Elgar Publishing Limited (2013) 393-407.

[4] IOM, World Migration Report 2013: Migrant well-being and Development.

 Geneva, Switzerland: IOM. (2013)

[5] A. Ivlevs. Happy Moves? Assessing the Link Between Life Satisfaction and Emigration Intentions. IZA Discussion Paper Series No. 9017. (2015)

[6] P. Serafino and R. Tonkin, Statistical matching of European Union statistics on income and living conditions (EU-SILC) and the household budget survey, Eurostat Statistical Working Papers (2017)

[7] M. D'Orazio, M. Di Zio, and Mauro Scanu, Statistical Matching: Theory and Practice, Wiley, Chichester (2006)

[8] V. Otrachshenko and O. Popova, Life (Dis)Satisfaction and the Intention to Migrate: Evidence from Central and Eastern Europe. Journal of Socio-Economics 48, (2014) 40–49.

[9] S. Migali and M. Scipioni, A global analysis of intentions to migrate, European Comission, JRC Technical Reports (2018)

[10] I. Gödri, Elvándorlási szándékok – álmok és konkrét tervek között: A migrációs potenciál jellemzői és meghatározó tényezői a 18–40 évesek körében Magyarországon (Migration intentions - dreams and concrete plans: Characteristics and determinants of Migration Potential among the 18-40 years old population in Hungary), Hungarian Demographic Research Institute, Research Reports No.98 (2016)

[11] D. Bartram, Migration, Return and Happiness: On seeking one ́s fortune in a wealthier country. SSRN Electronic Journal 10.2139/ssrn.1735367 (2012)

1. In this respect, we should highlight the studies focusing on the relations of migration aspirations and satisfaction in the cases of both potential migrants and returners. As regards the former, most studies concluded that subjective well-being and the willingness to migrate are negatively associated (see [8], [9], for a Hungarian case study see [10]), while others found a U-shaped relationship: usually the most and least satisfied people [5]. Regarding the latter, [11] found a negative relationship between return migration and subjective well-being in Eastern-European countries. [↑](#footnote-ref-1)