

Distr.: General
27 August 2014

English

Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics

Chisinau, Republic of Moldova

10-12 September 2014

Item 5 of the provisional agenda

Measurement of “hard-to-count” migrant groups

Migrant-specific use of the Labour Force Survey - Emigrants

Note by the Hungarian Central Statistical Office

Abstract

In the Hungarian Central Statistical Office (HCSO) considerable efforts were made recently for the better utilization of Labour Force Survey (LFS) data in migration statistics. The LFS is the most important regular survey in Hungary with the largest sample size, however till now data on migration were hardly utilized.

Concerning **emigrants** an attempt was made to capture the outflow of Hungarian citizens. In the framework of the **SEEMIG**¹ project on the migration of the South-east European region, LFS pilot surveys were executed in Hungary and Serbia. In the first phase questions were asked on emigrant household members with an extended migrant definition and contact details were compiled on Hungarian citizens living abroad. In the second phase these emigrants were surveyed with the use of the contact information. The target groups were the following: 1) any LFS household member who lived abroad at the time of the survey; 2) any person who have left abroad from this household since 1990; 3) emigrant siblings of any household member. Hereby the sample-size was increased to reach sufficient emigrants for the analysis and on the other side we could reach out to migrant persons who moved abroad together with all their previous household members.

¹ www.seemig.eu

Introduction

The SEEMIG project, funded by the European Union's South-East Europe Programme, was launched in June 2012 with the aim of better understanding of migration processes in the region. The project involves partners (statistical offices, research institutions, universities, local governments) from eight countries: Austria, Bulgaria, Hungary, Italy, Romania, Serbia, Slovakia and Slovenia. Within the project the long-term migratory processes were examined, comprehensive reports on available data sources were prepared as well as a public database on the region.

Shortages of migration statistics are well-documented in the relevant literature and they are also acknowledged and thoroughly analysed in earlier reports produced by the SEEMIG project (Gárdos - Gödri 2013). One of the key issues concerning the topic is the underestimation of emigration due to the lack of deregistration. SEEMIG tried to offer a solution through a pilot survey in two countries (Hungary and Serbia) based on the Labour Force Survey (LFS).²

The main problem of surveying emigrants is the lack of representativeness, since there are no appropriate sampling frames at the disposal and migrants are usually scattered in small ratios in the target countries. Therefore non-random sampling (e.g. snowball techniques) is widely used to reach the target group. Another important issue is that the topic of migration is often considered to be sensitive, hence low response rates can occur in surveys.

The idea of the methodology used in the SEEMIG pilot survey was based on a study of emigrants from Nepal (Ghimire, – Williams – Thornton – Young-DeMarco – Bhandari, 2013). During this study family-members of migrants were identified in a household survey and contact information were collected to reach the migrants abroad. The method proved to be a success, providing a large and representative sample of the emigrant population. However the authors of the study also acknowledged that non-response can be a significant obstacle. In the Nepal case the trust of respondents enabled researchers to reach migrants, nonetheless specific cultural and social context and circumstances can heavily influence the success of such surveys.

Based on this experience two pilot studies were carried out in Hungary and Serbia in the spring of 2013. Although the methodology was similar, the wider territorial scope and the European circumstances were important differences compared to the Nepalese study which heavily influenced the results as well. The main purpose of the pilot studies was to test this method for the countries in the region to provide a new, innovative technique to measure and estimate emigration.

The study also aimed to provide reliable estimates on the number of Hungarian emigrants abroad. Comparing the results with estimations from other data-sources was a proper tool to test the SEEMIG study. However it is hard to evaluate how accurate other estimates are, SEEMIG results seem to be rather underestimated concerning the number of emigrants from Hungary. The extent of this bias shall be evaluated later as well as the systematic analysis of potential reasons.

² SEEMIG project leader is **Attila Melegh** (Hungarian Central Statistical Office), while the leader of the work package responsible for the pilot surveys is **Zsuzsa Blaskó** (Hungarian Demographic Research Institute)

Data collection

The research design consisted of a two-stage methodology with the key idea to *derive a representative sample of emigrants from a representative national survey*. For the *first phase* of the study the LFS was used due to its large sample size and international comparability. In addition to the standard questionnaire a special SEEMIG battery was attached in order to collect some basic information on migrants from their family members. As a crucial part of the data collection contact information to migrants were also gathered from family-members. These information served as a basis for the *second phase*. During this second part in the autumn of 2013 migrants were interviewed via internet or telephone with the use of contact data.

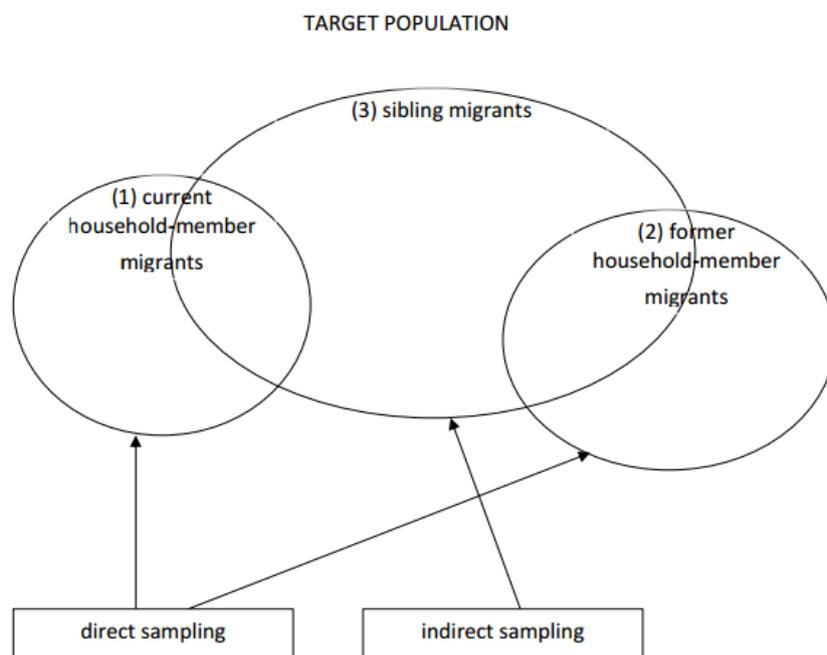
As for the definitions concern those were considered to be migrants who at the time of the survey were identified as „currently living abroad” by his/her household member and who were not born in the country of their current residence. In line with LFS age-definitions those were included in the sample who were between 15 and 74 years of age. Concerning the time criteria, the study followed the 862/2007 European regulation on migration statistics e.g. persons on temporary absence were excluded.

Besides LFS household members who lived abroad at the time of the survey³ the scope was widened to 'any person who left abroad from this household' with a time limit in 1990. This means that those were included who left the country in 1990 or later⁴. As a further widening of the target group, information of migrant siblings of any household member were collected⁵. This was an important step since on one hand the sample size was larger, on the other hand with this method those households could have been reached as well, where all members left the country (therefore not available in the regular LFS). The method has a great impact on both the questionnaire and the weighting scheme. Each of (1)-(3) defines a migrant sub-population, the first two of which are disjoint. As for sampling, reaching sub-population 1 and 2 means direct sampling, while reaching the third one is an indirect sampling method.

³ (1) current LFS-household members who are migrants;

⁴ (2) former LFS-household members of an existing LFS-household who are migrants;

⁵ (3) migrants who are brothers or sisters of a current LFS-household member living in Hungary;



Source: Blaskó–Jamalia (2014): *Surveying emigration I*.

Since during the data collection not only data on respondents but also on third party persons were collected, a detailed data protection protocol was elaborated and applied during the study. This was also to improve the trust towards the whole study, since the most challenging part with the highest expected non-response was the collection of contact information. Respondents were also invited to get in touch with their migrant household members and ask their permission to provide contact data. Other methods like the 'SEEMIG Research Participant Card' with contact data so that migrants can reach the researchers via the internet proved to be less successful. Interviewers also had a special training with the emphasis of the handling of the situation when contact data are asked, and extra bonus was granted for every contact information.

The reason for all these mentioned measures was to reach the highest possible sample size in order to ensure representativeness in the migrant sample. Non-response is also a problem because it is likely to be unevenly distributed across the various segments of the target population.

During the first phase 1,090 emigrant persons were identified in Serbia and 2,401 in Hungary. However when trying to collect more information on migrants the attrition rate was 25% in Serbia and 30% in Hungary (these figures are not fully comparable due to differences in the techniques applied). Basic data (sex, age, date of emigration, destination country etc.) were collected on migrants from household members in 819 cases in Serbia and in 1,659 cases in Hungary.

The largest attrition occurred when contact information to the migrants were requested. In the case of emigrants on whom detailed data were provided by family members, only 23 percent in Hungary, and 27 percent in Serbia were given contact details. This meant a very low sample-size which did not seem to be sufficient to gain representative results. However it was important to collect all

information on testing the method under the circumstances of the region, therefore it was worth to continue and collect all experiences.

Contacting migrants in LFS

| LFS | Serbia | | Hungary | |
|--|--------------|-------------|---------------|-------------|
| | Counts | % | Counts | % |
| <i>Households successfully interviewed</i> | <i>7 986</i> | <i>100%</i> | <i>23 393</i> | <i>100%</i> |
| Number of households reporting migrants | 816 | 10% | 1 785 | 8% |
| <i>Migrants – identified</i> | <i>1 090</i> | <i>100%</i> | <i>2 401</i> | <i>100%</i> |
| Migrants – statistical details provided | 819 | 75% | 1 659 | 69% |
| Migrants – contact information provided | 298 | 27% | 546 | 23% |

Source: Blaskó–Jamalia (2014): *Surveying emigration II*.

The low sample size made additional sample-boosting necessary, for this purpose the method of Respondent Driven Sampling (RDS) was chosen. In this method the first respondent's network is utilized to increase sample size. In order to meet RDS requirements a special set of questions were added to the emigrant-questionnaire, which referred to the number and socio-demographic characteristics of their acquaintances. Two contact information of the migrants' acquaintances were also asked. Regarding the content of the second (emigrant) questionnaire it focused primarily on in-depth migration-history information, in order to benefit from the opportunity of contacting emigrants directly. Since contact information were either e-mail addresses or telephone numbers, a mixed method survey was executed: both CAWI (Computer Assisted Web Interviewing) and CATI (Computer Assisted Telephone Interviewing) methods were applied.

Results and conclusions

Out of the total contact of 546 in Hungary, altogether 125 successful interviews were made: 66 on the web, and 59 via telephone. Corresponding figures in Serbia are: out of 298 persons with a contact information 98 were successfully interviewed – the majority of them (88) via telephone and only 10 had filled out the electronic questionnaire. These add up to a success rate of 22 per cent in Hungary and 33 per cent in Serbia.

Within the RDS phase 33 emigrants provided 54 contact detail to further emigrant acquaintances in Hungary and the situation was quite similar in Serbia. These results suggest that using the RDS method does not provide a proper solution to boost the sample.

From the results we can draw the conclusion that low response rate in the first phase with the small number of collected contact information and an especially high attrition rate with CAWI method in the second phase led to an **insufficiently large sample size**. The data collection could not fulfil the original goals i.e. to gain a large, unbiased sample of emigrants on the basis of an internationally standardized national representative survey (LFS). The results can be mainly traced back to the low

level of confidence towards interviewers when contact information were asked. During the RDS phase it seems that low response rate can be attributed to the lack of personal contact with interviewers, since in a previous study the method worked successfully (Hárs, 2009).

However there are some **useful results** from the pilot study as well. On the basis of the collected detailed data from emigrants **further qualitative analyses** can be carried out to find common patterns in migration histories. Apart from geographical biases a **plausible distribution** of migrants was gained from the study, which with the utilization of this detailed dataset some „common knowledge” e.g. on the educational level of migrants could have been tested. The results can also serve as a **starting point for new studies** e.g. information on remittances were also collected.

Another achievement is that **the reliability of the results can be tested with a one-by-one comparison** of the questionnaires of the two phases referring to the same emigrant (i.e. data collected from household-members in the first phase, and emigrant questionnaires in the second phase). This offers the possibility to carry out further analyses on emigration from a dual point-of-view, which is exceptional in this field regarding the two countries.

It was also proved that the method can be utilized to obtain reliable **estimates on the number of emigrants** and their distribution by basic variables like age, sex or country of residence. However further analysis is needed to find potential reasons for the possible underestimation. The low response rate was likely to play a crucial role in this issue.

Concerning the further application of the method it seems that it is difficult to reach an unbiased sample with sufficiently enough number of cases with the use of large, highly formalized, national surveys. Smaller, local-level surveys are rather more appropriate for emigration studies, building on the local knowledge of the interviewers.

References

Blaskó, Zsuzsa – Jamalia, Natalie (2014): Surveying emigration I. Report on the first stage of the SEEMIG pilot study in Hungary and Serbia. Research report developed within the project ‘SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies’.

<http://www.seemig.eu/downloads/outputs/SEEMIGPilotReport1.pdf>

Blaskó, Zsuzsa – Jamalia, Natalie (2014): *Surveying Emigration II. Report on the second stage of the SEEMIG pilot study in Hungary and Serbia*. Research report developed within the project ‘SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies’. <http://www.seemig.eu/downloads/outputs/SEEMIGPilotReport2.pdf>

Gárdos, É. – Gödri, I. (2013): Analysis of existing migratory data production systems and major data sources in eight South-East European countries. Research paper developed within the project ‘SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies’.

Ghimire, Dirgha J. – Williams, Nathalie - Thornton, Arland – Young-DeMarco, Linda – Bhandari Prem B. (2013): Innovation in the Study of International Migrants. Presented at the Population Association of America Annual Meeting, April 11-13, New Orleans, LA.

Hárs, Ágnes (2009): Magyarok az osztrák munkaerőpiacon. Ingázók, bevándorlók, munkaerő-migránsok? Kutatási Zárójelentés. (Hungarians in the Austrian Labour Market. Research Report.)
<https://onedrive.live.com/view.aspx?resid=20226F10B70B2C25!311&cid=20226f10b70b2c25&app=WordPdf&wdo=2&authkey=!AEZMjxUfjQcQMng>
