

Checking the reliability and the comparability of international migration flows between Italy and Belgium

Submitted by the National Statistical Institutes of Italy and Belgium¹

Introduction

1. International migration is the only demographic statistic which is currently produced simultaneously by two National Statistical Institute, the one of the country of departure and the one of the country of arrival. Therefore it is a very interesting subject for studying on a general base the reliability of our statistical data in a case where, everybody will agree, the reliability problem is a major one. Moreover international migration is nowadays a very important matter on the policy point of view within Europe so that good data are urgently needed. In this context EUROSTAT and other international bodies try to improve the overall reliability and comparability of the data collection. This task is enormous and up to now most of our figures consist of two different numbers for an identical migration flow between a pair of countries.

A double matrice of intraeuropean international migration

2. On the base of the figures provided by each country on :

- * international immigrations by country of origin or country of last residence
- * international emigrations by country of destination or next residence

we can build a complete matrice where in each cell $M(i,j)$, for the migration flow between a specific pair of countries i and j , you can find two figures : the one on immigrations in the country of destination and the one on emigrations in the country of destination.

3. Such a matrice has been proposed by John KELLY since more than twenty years and the presentation and the discussion of this double matrice for the years 1994 and 1995 concerning 36 european countries will confirm that there is still a long way to achieve a minimal comparability level.

Harmonisation, comparability and reliability

4. A complete harmonisation of all criterium used to define an international migration is definitively an utopic target. A relatively high level of harmonisation may only be obtained when introducing a new data collection system like the Labor Force Survey, for example. When a data

¹ Prepared by Enrico Bisogno, ISTAT (Italy) and Michel Poulain, INS (Belgium).

collection system already exists in all countries, each country shows specificities in relation with facts from the national history and the development of national migration policies. Therefore, the achievement of a complete harmonisation will impose to change some of the national practices in order to introduce the internationally recommended definition. For a few countries, the gap between the national definition and the international recommended one is not so big so that an harmonisation may be expected. But in a large number of countries, the definition and the data source are so different that an harmonisation is quite impossible. Therefore a reachable objective should be to achieve a minimal comparability without forcing countries to an overall harmonisation.

5. In order to obtain this minimal comparability between data produced by all european countries, some approximation methods will be needed in order to take into consideration the fact that the national definition is different from the internationally recommended one but also considering the poor reliability of international migration data. In fact, even if two countries are using the same definition to measure international migration flows, the problem of non-reliability of the data collection system may involve very large differences between the two figures related to the same migration flow.

6. Therefore, it seems clear when analysing the dubble matrice that the problem on poor reliability of the international migration data collection systems in each countries is certainly as much important as the one of harmonisation of concept in order to improve the overall comparability of international migration statistics.

7. In the near future, some checks will be done, comparing different dubble matrices in time during the last 20 years, in order to look if such a overall improvement of the complete system is occurring

A case study : the flows between Italy and Belgium

8. One method to improve the actual situation consists of a bilateral cooperation between pairs of country in order to examine more in detail while the two data differ. Eurostat has given this possibility two years ago to Belgium, on one side, Denmark and Sweden, on the other side. Nowadays, we are trying to examine the situation between Belgium and Italy.

9. The main interest of this project is that for the first time individual data are linked between two data files produced by two completely different systems in two different countries. We are no longer comparing agregate data but individual data and the result may be quite different.

Data description

10. The data used are individual records of moves between Italy and Belgium during the years 1991-95, respectively recorded by ISTAT (Italy) and INS (Belgium).

11. In both countries data derive from administrative activity performed in order to update the Population Register. However, the Population Register is kept at the communal level in Italy (8100 Communes) while in Belgium it is central and computerized (Registre National).

12. In the former country individual forms (either on paper or on magnetic device) are sent to ISTAT by each Commune and then processed, while individual data are directly extracted from the Central Population Register in Belgium.

13. Each record of the two files provided by ISTAT and INS contains the following fields:

- * Direction of the move (Italy-Belgium; Belgium-Italy);
- * Year and month of registration;
- * Date of birth (yyymmdd);
- * Sex;
- * Citizenship (Italian, Belgian, Other)

- * Type of registration (for Italy: 1-Normal, 2-post-census regularization; for Belgium: 1-Normal, 2-Italian immigrants with unknown origin or Italian emigrants with unknown destination)

- * Commune code (referring to Commune of the country involved)

Checking for internal consistency of the records within each country file

14. Lacking an individual identity code, the basic rule followed in order to link records related to the same person within and between the two countries is that records related to persons having equal sex and date of birth are considered as referring to the same person. Doing so, there is a weak probability of misreporting and considering that two moves are related to the same person while it is two different person. This risk may be estimated by the probability to have two persons of the same sex and born on the same day when choosing two persons randomly in the total population. If the daily crude birth rate is about 4 per 100.000 inhabitants, then this risk, considering the sex is one per fifty thousand. As we consider about fifty thousand persons, we may expect to find one case of two persons of the same sex and date of birth and moving between Belgium and Italy. However, this doesn't take into consideration twins moving with their parents in the same family at the same time. We may suppose that twins aged less than twenty will have a very high propensity to move together and we will come back on this point when dealing with strictly double records.

15. Before trying to link records of the two countries, a consistency check has been carried out, for each file separately, in order to identify registrations related to the same person and showing an illogical migratory path (e.g.: two successive emigrations of the same person). This step has proven to be necessary in order to avoid the matching of inconsistent records, an anomaly which would have created some heavy difficulties when reconstructing the migration history of all migrants, the second purpose of the study. Moreover, the consistency check can give a first impression about the of general quality of the data.

16. The consistency procedure has then developed as follows:

* In the first step of consistency check double registrations - e.g. records being completely alike - were identified. A first check in the Belgian data shows that most of the 42 double records are related to twins moving together. This check cannot be processed with the Italian data. Moreover, for Belgium, the age distribution of these "double" records confirm that most of them are less than 25 years old. A similar check will be done with the Italian data. The number of forty double records in Belgium seems to be a reasonable number taken into consideration multiple births with the same sex.

* Afterwards, all records referring to the same person were put in chronological order in order to check their consistency in terms of migratory path; in particular, not admissible records were identified according to the following rules:

- two successive moves in the same direction;
- two moves with opposite direction but in the same date.

17. As a result of these two first operations, consistent and inconsistent records were put in two distinct files, separately for each country. Doing so we suppose there is a very probability that the inconsistent records concern two different persons with the same sex and the same date of birth. But as we have seen this probability is very weak. There is also a possibility that a person leaves for example Italy to Belgium, then move to Germany and returns to Italy before to move again to Belgium. In the particular case, the two following records between Italy and Belgium are migration from Italy to Belgium. But we shall suppose this migratory situation to be exceptional. Finally, there is also a weak possibility to find a very short migration and return migration between the two countries during the same month and this situation is consistent. We will also suppose that this situation is quite exceptional.

Merging the two database

18. Even if movements of the whole period 1991-95 were taken into consideration, only the three years 1992, 1993 and 1994 were actually considered as object of the study. In fact, it often occurs that the same move is recorded with a considerable time lag in one of the two countries involved, therefore it seems very much probable that many movements registered in one country in 1991 or 1995 have been recorded, in the other country, during years not covered by observation, for example in 1990 or 1996. Records of 1991 and 1995 have then been considered only if matching with registrations of the other country relative to years 1992-94.

19. Inconsistent records of each country were then confronted with exact records of the other country, in order to identify good records matching with wrong registrations of the other country: (this has been necessary since linkage criteria are slightly different if records are logically ordered or not).

20. Three separated files were then identified for each country:

- * file A of consistent migratory sequences (immi, emi, immi-emi, emi-immi, immi-emi-immi,...)
- * file B of consistent migratory sequences matching, even only partially, with inconsistent paths of the other country
- * file C of inconsistent migratory sequences

21. At this point, files of exact records (files A) of the countries were confronted in order to link records referring to the same move, according to the following rules:

- * moves of the same person were temporally ordered in each file separately;
- * records of the two files were then linked if having equal sex, date of birth, direction and place of order referring to migration sequence of the same person.

22. Afterwards, file B of Italy was confronted with file C of Belgium, linking records with the same sex, date of birth, direction and the least difference in period of migration (the same was obviously done between file B of Belgium and file C of Italy).

23. Records not yet linked of the two files C were then confronted, in order to link registrations having the same sex, date of birth, direction and the least difference in period of migration; in this case an ex-post check was necessary in order to eliminate not consistent migratory paths.

24. Linkage procedure can now be considered as terminated, even if a last passage was necessary in order to exclude records of the years 1991 and 1995 which were not linked to registrations relative to the period 1992-94.

Conclusion

25. This research project is still in course but a first indication shows that we will link approximatively 40% of all moves. After considering the link between same moves, we will analyse individual migratory behaviour as recorded by each of the two country and finally we will estimate the impact of missing records on the stock of italians living or supposed to live in Belgium and belgians living or supposed to live in Italy.