NORFACE MIGRATION

Institute for Employment Research
The Research Institute of the Federal Employment Agency

IAB

Education, gender and international migration: Insights from a panel dataset 1980-2010

August 2013

– Methodology Report –

Abstract

This note describes the content of the IAB brain-drain dataset and summarizes the methodological approach used to compute the emigration stocks and rates by country of origin, sex and educational attainment. Our methods mainly build on Docquier, Lowell and Marfouk (2007), - henceforth DLM07 - and Defoort (2008).

Preliminary Version

Acknowledgements

Financial support from NORFACE (New Opportunities for Research Funding Agency Cooperation in Europe), project TEMPO (Temporary Migration, Integration and the role of Policies) and the IAB (Institute for Employment Research, Nürnberg) is gratefully acknowledged. We thank the participants to the 4th NORFACE Migration Conference in London for useful comments and suggestions.

Herbert Brücker
IAB, Institute for Employment Research, Nürnberg and University of Bamberg

Stella Capuano
IAB, Institute for Employment Research, Nürnberg

Abdeslam Marfouk
ULB, Université Libre de Bruxelles

1 Emigration stocks

As already noticed in DLM07, reliable information on the number of immigrants by country or region of origin can be obtained only by harmonizing national censuses and population registers statistics from the receiving countries. We collected data from 20 OECD member states[1] on the immigrant population aged 25 years and older by gender, educational level and detailed country of birth from 1980 to 2010 (5 years intervals). For the same period we also compiled an auxiliary dataset featuring the number of immigrants (all ages) by gender and country of birth. Table 1 summarizes our data sources for each country and year. The computation of the immigration stocks implied a number of definitional and methodological assumptions, which we describe in further detail in what follows.

Migrant population. Following DLM07, migration is defined according to country of birth rather than foreign citizenship. We chose to rely on this definition for several reasons. First of all, the country of birth concept is time-invariant, whereas foreign citizenship changes with naturalization. OECD statistics report that, in the 20 destination countries that we cover, 17.8 million foreign-born people acquired citizenship between 2000 and 2010. The acquisition of citizenship is particularly high in the US, Canada, France, United Kingdom, Germany and Australia.

Moreover, the legislation regulating the acquisition of citizenship typically differs among countries and within the same country over time. Thus, especially in a dataset featuring a long time dimension, adopting the concept of citizenship instead of country of birth might compromise both the cross-country and the within-country analyses of the size and dynamics of immigration.

In some cases the difference between the foreign citizen and foreign-born groups can be substantial. As shown in Table 2 the former are in general (with the only exception of Luxembourg) less than the latter in our 20 destinations. As a consequence, counting the foreign citizens as migrants would underestimate the number of individuals who decide to migrate at any given point in time.

It is also worth mentioning that our figures may not correspond to the national official statistics on migration, because not every country adopts a similar concept as ours in defining the immigrant population. For example, while in countries such as Australia, New Zealand and the United Kingdom an immigrant corresponds to a person born-abroad (i.e. exactly the definition adopted in the present study), according to the US Census Bureau immigrants are individuals who are *born abroad*

[1] The number of receiving countries is kept constant over time, even though Chile became an OECD member state only in 2010. The data were obtained either directly from the National Statistical Offices and in a few cases through the IPUMS website (https://international.ipums.org/).

with foreign citizenship at birth. Individuals born in one of the US external territories or born abroad by at least one US citizen parent are counted as US citizens and will not be included in the US migration statistics. While the difference between citizens born abroad and immigrants is negligible in the US case, it is likely to be large in other countries which adopt a similar definition, especially former colonial powers.

France provides a clear example of the magnitude of this phenomenon. In 2008 25 percent of all foreign-born population had French citizenship at birth. Among them 56 percent were born in the former French colonies of Algeria, Morocco or Tunisia (Bouvier, 2008). For France and the US we were able to apply our definition in a detailed way. Not in all countries it is possible to make such a precise distinction, which calls for special attention when interpreting migration statistics.

Germany is the only destination country for which we used the concept of citizenship instead of country of birth. In fact, official statistics by country of birth are not available for this country before 2009. In the case of Greece, instead, we have data by country of birth and country of citizenship from 2000, but only by country of citizenship in the previous years. In this case we used the ratio between foreign citizens and foreign-born in 2000 in order to infer the number of foreign born in the previous years.

Countries of birth. We report data on the stock of immigrants coming from 195 countries (193 UN member states plus Holy See and Taiwan).[2] To harmonize the group of source countries during 1980-2010 we aggregated East and West Germany as well as Yemen.

We also provide statistics from a few origin countries that did not exist before a given year, i.e. all the States belonging to the former Soviet Union (existing only after 1991), the ones belonging to the Socialist Federal Republic of Yugoslavia and former Czechoslovakia (existing only after 1992), and all former colonies which gained independence during the considered period.[3] In these cases, we estimated the immigrant stock from each post-secession origin by multiplying the total migration stock of the pre-secession State by the gender and skill-specific share of the post-secession country population over the total pre-secession country migration stock.

Age groups. Migration figures by education, gender and countries of birth refer to individuals aged 25 years and older. We chose the 25+ age group in order to maximize the comparability between our data and the educational attainment figures in the source countries and in a number of international migration databases.

[2] Due to data limitations and underreporting, we aggregate South and North Korea, Serbia, Kosovo and Montenegro, North and South Sudan; we also consider two Chinese independent territories (Hong Kong and Macao) and Palestinian Territories.

[3] Antigua and Barbuda (1981), Belize (1981), Saint Kitts and Nevis (1983), Marshall Islands (1986), Federal States of Micronesia (1986), Namibia (1990), Eritrea (1993), Palau (1994).

The 25+ age group is also less likely to contain a large number of students who temporarily emigrate for educational purposes.

Educational categories. We distinguish three levels of education: primary (low skilled: includes lower secondary, primary and no schooling); secondary (medium-skilled: high-school leaving certificate or equivalent) and tertiary education (high-skilled: higher than high-school leaving certificate or equivalent).

Imputation of missing information. Most destination countries included in our dataset take the Census every 10 years (with the exception of Canada, Australia and New Zealand, which take the Census every 5 years). In order to address this issue, we applied the same imputation procedure as Defoort (2008). This procedure can be described as follows:

- Suppose we have information on the number of immigrants by country of birth, gender and educational level for the years 1980, 1990, 2000, and 2010 and we have to impute the information for 1985, 1995 and 2005. First of all, we need to know the total number of male and female individuals aged 25 years and over for the years 1985, 1995 and 2005. We obtained this piece of information either directly from the statistical offices from which we collected the data, or we estimated it from our auxiliary dataset on foreign born individuals by sex and country of birth for all ages, in the following way. Knowing the share of male and female immigrants aged 25+ in 1980, 1990, 2000 and 2010, we can apply the average share of any two extreme years to the total number of immigrants (all ages) in 1985, 1995 and 2005 in order to recover the total number of migrants (males and females) aged 25+ in those years. In this way we are able to use for the imputations the same data sources as the figures on the stock of migrants, which represents a departure from the procedure described in Defoort (2008).[4]

- At this point, we are able to fully apply Defoort (2008). First, we compute the proportion of low -, medium - and high-skilled over the total immigrant population aged 25+ in 1985, 1995 and 2005 by applying the formula reported in Defoort (2008), i.e.

$$h_{e,m,t} = \frac{\left(\frac{h_{e,m}}{h_{e,r}}\right)_{t-1} + \left(\frac{h_{e,m}}{h_{e,r}}\right)_{t+1}}{2} h_{e,r,t}$$

  Where, for each destination country, $h_{e,m,t}$ is the proportion of persons with skill level e among the immigrants at time t (known for 1980, 1990 and 2000 and 2010), and $h_{e,r}$ is the proportion of individuals with the same skill level in the resident population. The latter figures were obtained for all countries of

[4] The author used the OECD data for the total number of immigrants in the destination country.

origin by applying the population shares of high, medium and low skilled male and female individuals provided in Barro and Lee (2013) to United Nations (2011) data on population by sex.[5] These proportions are applied to the total number of individuals aged 25+ in order to compute the total number of low, medium and highly skilled immigrants.

- Finally, the total number of low, medium and highly skilled immigrants is distributed across the different countries of birth by applying the share of each origin country over total migration (by gender and skill level) to the total immigration stock by skill level computed at the previous point.

Whenever the value for 2010 had to be imputed, the information on the last available year was used to carry forward the information in 2010.[6] Similarly, the first available information was used to perform retropolation for the previous years (e.g., when our time series started from 1990).

In some cases, official Statistics in the destination countries by skill level were not available, but it was still possible to obtain the total number of male and female immigrants aged 25+ for a given year. In order to distribute the total stock of migrants by country of birth into the low, medium and high educational categories we applied the same procedure as above, but we did not need to carry out the first step of the imputation procedure. When 1990 and 2000 had to be imputed we used the shares by skill level implied by DLM07.

2 Emigration rates

Emigration rates by gender and educational level are indicators that compare the total number of immigrants living in all the considered OECD countries with the total number of individuals coming from a given source country (i.e. residents and migrants) and belonging to the same educational category.

The total stock of immigrants coming from country *i* living in the 20 destination countries as a whole (denoted as *OECD20*), by gender *s*, level of education *e* at time *t* is given by:

$$M^{i}_{OECD20,s,e,t} = \sum_{j=1}^{20} M^{i}_{j,s,e,t}$$

[5] In some cases population data are missing, and we used the World fact book (https://www.cia.gov/library/publications/the-world-factbook/) and national statistical sources to complement these data.

[6] In most cases the imputation of year 2010 was needed because the last year of the Census was not available at the time we started our data collection. We are in the process of carrying out a further wave of data collection where we will replace the imputed data with actual Census data.

Let $R^i_{s,e,t}$ denote the total number of residents in source country *i*, with gender *s* and educational level *e* at time *t* (computed from Barro and Lee, 2013 as explained above). The emigration rate for any level of education is then defined as:

$$m_{i,t} = \frac{M^i_{OECD20,s,e,t}}{R^i_{s,e,t} + M^i_{OECD20,s,e,t}}$$

The emigration rates computed in this way weight the total number of migrants from a given country of origin by the total workforce in that country, and hence provides a measure of the loss of labor market potential experienced by a given source country when part of its population migrates.

Table 1. Data sources

	1980	1985	1990	1995	2000	2005	2010
Australia	1981 Census data from Australian Bureau of Statistics	1986 Census data from Australian Bureau of Statistics	1991 Census data from Australian Bureau of Statistics	1996 Census data from Australian Bureau of Statistics	2001 Census data from Australian Bureau of Statistics	2006 Census data from Australian Bureau of Statistics	-
Austria	1981 Census data from Statistics Austria	-	1991 Census data from Statistics Austria	-	2001 Census data from Statistics Austria	-	-
Canada	1981 Census data from Statistics Canada (20% sample)	1986 Census data from Statistics Canada (20% sample)	1991 Census data from Statistics Canada (20% sample)	1996 Census data from Statistics Canada (20% sample)	2001 Census data from Statistics Canada (20% sample)	2006 Census data from Statistics Canada (20% sample)	-
Chile	1982 IPUMS (10% sample)	-	1992 IPUMS (10% sample)	-	2002 IPUMS (10% sample)	-	-
Denmark	1981 Statistics Denmark population and education registers	1985 Statistics Denmark population and education registers	1990 Statistics Denmark population and education	1995 Statistics Denmark population and education registers	2000 Statistics Denmark population and education	2005 Statistics Denmark population and education	2010 Statistics Denmark population and education

			registers		registers	registers	registers
France	1981 Census data from National Institute of Statistics and Economic studies	-	1989 Census data from National Institute of Statistics and Economic studies	-	2001 Census data from National Institute of Statistics and Economic studies	-	2009 Census data from National Institute of Statistics and Economic studies
Finland	Census data from Statistics Finland	Census data from Statistics Finland	Census data from Statistics Finland	Census data from Statistics Finland	Census data from Statistics Finland	Census data from Statistics Finland	Census data from Statistics Finland
Germany	Mikrozensus (data by citizenship) from the German Federal Statistical Office	Mikrozensus (data by citizenship) from the German Federal Statistical Office	1991 Mikrozensus (data by citizenship) from the German Federal Statistical Office	Mikrozensus (data by citizenship) from the German Federal Statistical Office	Mikrozensus (data by citizenship) from the German Federal Statistical Office	Mikrozensus (data by citizenship) from the German Federal Statistical Office	2009 Mikrozensus (data by citizenship) from the German Federal Statistical Office
Greece	1981 IPUMS (data by citizenship,	-	1991 IPUMS (data by citizenship,	-	2001 IPUMS (10% sample)	-	-

	10% sample)		10% sample)				
Ireland	1981 Census data from the Central Statistical Office Ireland	-	1991 Census data from the Central Statistical Office Ireland	1996 Census data from the Central Statistical Office Ireland	2001 Census data from the Central Statistical Office Ireland	2006 Census data from the Central Statistical Office Ireland	-
Luxembourg	1981 Statistics Luxembourg	-	1991 Statistics Luxembourg	-	2001 Statistics Luxembourg	-	-
Netherlands	-	-	-	DLM07 and population registers data from Central Bureau of Statistics Netherlands	DLM07	Population registers data from Central Bureau of Statistics Netherlands and authors' imputations.	Population registers data from Central Bureau of Statistics Netherlands and authors' imputations.
New Zealand	1981 Census data from Statistics New Zealand	1986 Census data from Statistics New Zealand	1991 Census data from Statistics New Zealand	1996 Census data from Statistics New Zealand	2001 Census data from Statistics New Zealand	2006 Census data from Statistics New Zealand	-

Norway	Population registers data from Statistics Norway and authors' imputations.	Population registers data from Statistics Norway and authors' imputations.	DLM07 and population register data from Statistics Norway	Population registers data from Statistics Norway and authors' imputations.	DLM07 and population register data from Statistics Norway	Population registers data from Statistics Norway and authors' imputations.	Population registers data from Statistics Norway and authors' imputations.
Portugal	1981 Census data from Statistics Portugal	-	1991 Census data from Statistics Portugal	-	2001 Census data from Statistics Portugal	-	-
Spain	1981 Census data from National Statistics Institute	-	1991 Census data from National Statistics Institute	-	2001 Census data from National Statistics Institute	-	-
Sweden	Population registers data from Statistics Sweden and authors' imputations	Population registers data from Statistics Sweden and authors' imputations	Population registers data from Statistics Sweden	Population registers data from Statistics Sweden	Population registers data from Statistics Sweden	Population registers data from Statistics Sweden	Population registers data from Statistics Sweden
Switzerland	Census data from Swiss Statistics	-	Census data from Swiss	-	Census data from Swiss Statistics	-	-

			Statistics				
United Kingdom	Census data from the Office for National Statistics UK and authors' imputations	-	DLM07	-	Census data from the Office for National Statistics UK	-	-
United States	IPUMS	-	IPUMS	-	IPUMS	ACS (American Community Survey)	ACS (American Community Survey)

"-" means that no data were available for that year. In these cases the figures have been imputed as explained in the main text.

Table 2. Foreign citizens and foreign born individuals in the 20 OECD countries

	Foreign citizens	Foreign born	Ratio citizens/f. born
Austria	928	1316	0.71
Belgium	1058	1504	0.70
Canada	1759	6187	0.28
Denmark	346	429	0.81
Finland	168	248	0.68
France	3731	5342	0.70
Germany	6754	10591	0.64
Ireland	413	602	0.69
Italy	4235	4799	0.88
Luxembourg	221	189	11.7
Netherlands	760	1869	0.41
Norway	369	569	0.65
Portugal	448	669	0.67
Spain	5731	6660	0.86
Sweden	633	1385	0.46
Switzerland	1720	2075	0.83
United Kingdom	4524	7056	0.64
United States	21581	39917	0.54

Note: Data refer to 2006 for Canada, 2008 for France, and 2009 for Belgium and Italy. Source: OECD (2010) and authors' calculation.

References

Barro, R. and Lee J: (2013): A New Data Set of Educational Attainment in the World, 1950-2010. Forthcoming, Journal of Development Economics.

Bouvier, G. (2008): Les descendants d'immigrés plus nombreux que les immigrés: une position française originale en Europe. Vue d'ensemble. INSEE.

Defoort, C. (2008): Long-term trends in international migration: an analysis of the six main receiving countries. Population, 63, 285-317.

Docquier, F., Lowell, B. L. and Marfouk, A. (2009): A Gendered Assessment of Highly Skilled Emigration. Population and Development Review, 35, 297–321.

OECD (2010): International migration database, *OECD International Migration Statistics* (database).

United Nations, Department of Economic and Social Affairs, Population Division (2011): World Population Prospects: The 2010 Revision. CD-ROM Edition.