EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics
Unit F-2: Population and migration

Luxembourg, 5 July 2017
ESTAT/F-2/GL*

TECHNICAL NOTE

Subject: Methodology for the migration assumptions in the 2015-based population projections

1. DATA INPUT

In order to compute population projections, additionally to future births and deaths it is necessary to provide future migration flows. The required level of detail of the migration assumptions depends upon the level of disaggregation of the population projections. Given that the 2015-based population projections were to be provided by single age, sex and single year, the minimum required input on migration was the net[1] migration by single age and sex for each year from 2015 to 2080, for each country covered by this exercise. Whilst it may be easy to find speculations about future major migration trends, it is rather difficult to get those considerations being detailed. The task of identifying 367 640[2] (net) future migration figures is even more daunting considering the stretching 65 years ahead of their time horizon, the high uncertainty surrounding the future migration flows and the diversity of migration profiles of the countries.

Besides being notoriously the most volatile and the most difficult element of the population change to be forecasted, in recent years migration flows towards the EU have been characterized by a large number of asylum seekers, an even more volatile migration component. Additionally, current national practices on the inclusion/exclusion of asylum seekers and/or persons granted protection ('refugees') in the migration statistics are not fully harmonized across Member States[3].

* Giampaolo Lanzieri. Email: giampaolo.lanzieri@ec.europa.eu

[1] Whilst it is sounder to elaborate assumptions on immigration and on emigration separately, the computation formulas for population projections do not actually require distinct migration flows.

[2] 101 age classes by 2 sexes by 65 projections' years by 28 countries. In case of separated migration flows that number would become 735 280; any additional breakdown (e.g., by EU/non-EU citizenship, by country/region of birth, etc.) would inflate further the size of the data input.

[3] See Eurostat (2016): "*Classification of Asylum Seekers and Refugees in the EU Annual Demographic and Migration Statistics*". Paper for the Working Group on Population Statistics, Luxembourg, 2-3 June 2016.

Migration statistics have been partially harmonized by an EU Regulation in 2007[4], replacing former data collections where data were provided on voluntary basis only. The disaggregation of the data that countries are obliged to provide in accordance to this EU Regulation is however kept to a level that the Member States have by then considered feasible, mostly just the main aggregates. Still today, parts of the migration statistics which are disseminated by Eurostat are collected on voluntary basis.

Hence, the available data on immigration and emigration flows were very limited over time (in general covering only few years), with high variability and affected by the latest fluctuations of flows. An additional special data collection on past immigration and emigration data has been launched to address the former issue, but its outcome has led to consider net migration[5] as the only viable information on past trends in migration. In order to tackle the latter issues (variability and asylum seekers), the highest and the lowest values in the entire time series 1960-2016[6] of net migration were replaced respectively by the second highest and the second lowest value. By doing so, extreme peaks and troughs, usually due to special events (including the above-mentioned latest high inflows), would not affect the identification of migration trends. The input dataset consisted then of 'winsorized'[7] time series of total net migration starting from 1960, but for Portugal (1976) and Malta (1996), for which data referring to earlier years were not adequate.

2. THE MIGRATION MODEL

The model used to produce migration assumptions for the 2015-based population projections is built upon four components:

a) nowcast;

b) trend model;

c) convergence model;

d) working-age population 'feedback mechanism'.

The weight of the first three components in the overall migration assumptions varies depending on the year of reference (see below). This model tries to take into account past migration trends, very latest evidences, driving demographic factors as well as a vision about future developments in migration flows. On purpose, it does not require non-demographic data input.

[4] European Parliament and Council Regulation (EC) No 862/2007 on Community statistics on migration and international protection.

[5] By definition, net migration is the difference between immigration and emigration numbers. However, it may be computed also as residual component from the demographic balance (i.e., as population change minus the natural change). Using this method – the only feasible in lack of data on separated flows – net migration incorporates also any sort of statistical adjustments made to fit the demographic balance.

[6] These time series included the migration nowcasts 2016 provided by the countries.

[7] 'Winsorizing' or 'winsorization' is the transformation of data by limiting extreme values to reduce the effect of possibly spurious outliers.

2.1. Nowcast

Being the jump-off time of reference of the population projections the 1 January 2015 ('2015-based' projections), the migration events that should have been considered were those until the year 2014. However, because of the timing of the exercise, provisional data for the year 2015 and for part of the year 2016 were available at the time of the projections computations. In order to incorporate the latest empirical evidence, the net migration observed in 2015 has been directly taken as 'assumption' for the year 2015.

For the year 2016, the Member States have been invited to provide a statistically sound forecast of net migration, using all the latest (usually monthly or quarterly) available data. Such forecast being carried out in the same year to which it was referring to (the 2016), following a widespread terminology it is named 'nowcast'. The nowcast for the net migration in 2016 has been provided by all countries except Belgium, Estonia, France, Hungary, Romania and Slovakia (see Table 1). For these latter countries, assumptions for the year 2016 where then produced using the other components of the migration model (see below).

Table 1: observed (2015) and 'nowcasted' (2016) net migration by country as provided by the national statistical authorities

Country	*2015*	*2016*	*Country*	*2015*	*2016*
BE	67 123	n.a.	LU	11 159	10 800
BG	-4 247	-4 321	HU	15 119	n.a.
CZ	15 977	18 600	MT	4 176	3 480
DK	42 532	36 700	NL	56 368	85 513
DE	1 139 000	750 000	AT	113 067	73 813
EE	2 410	n.a.	PL	-12 792	4 900
IE	-571	14 779	PT	-10 481	-10 481
EL	-44 905	-23 900	RO	-57 932	n.a.
ES	-8 390	12 939	SI	507	155
FR	45 820	n.a.	SK	3 127	n.a.
HR	-17 945	-21 500	FI	12 441	15 861
IT	133 123	134 493	SE	78 410	103 500
CY	-2 000	1 000	UK	333 000	244 000
LV	-10 640	-9 400	NO	29 802	27 450
LT	-22 403	-28 182			

2.2. Trend models

In order to take into account past migration in the formulation of assumptions on future flows, net migration trends were identified and extrapolated by applying Auto-Regressive Integrated Moving Average (ARIMA) models selected by an automated model specification procedure[8]. Because the 'winsorized' time series used as input to the ARIMA modelling was including net migration data until 2016, the extrapolations have covered the time horizon 2017-2080[9]. The extrapolated trends can point to any direction, i.e. indicating increase, decrease or stability of the future flows, depending on the past migration trends. They are mainly an attempt to incorporate any past regularity in

[8] See Hyndman R. and Y. Khandakar (2008): "Automatic Time Series Forecasting: The forecast Package for R". *Journal of Statistical Software*, 27(3):1-22.

[9] From 2016 for the six countries which did not provided the nowcast for 2016.

migration flows into the assumption for the future; in several cases, however, the best possible model was a 'random walk'[10] (see the Table 2).

Table 2: ARIMA models selected for the trend extrapolation

Country	*Model*	*Country*	*Model*
BE	ARIMA(0,1,1)	LU	ARIMA(1,1,0)
BG	ARIMA(0,0,0) with non-zero mean	HU	ARIMA(0,1,1)
CZ	ARIMA(0,1,0)	MT	ARIMA(0,1,0)
DK	ARIMA(0,1,0)	NL	ARIMA(0,1,0)
DE	ARIMA(1,0,0) with non-zero mean	AT	ARIMA(0,1,0)
EE	ARIMA(0,1,0)	PL	ARIMA(2,0,2) with non-zero mean
IE	ARIMA(1,0,1) with zero mean	PT	ARIMA(1,0,2) with zero mean
EL	ARIMA(1,1,0)	RO	ARIMA(0,1,1)
ES	ARIMA(0,1,0)	SI	ARIMA(1,0,0) with non-zero mean
FR	ARIMA(2,0,0) with non-zero mean	SK	ARIMA(0,1,0)
HR	ARIMA(1,0,0) with zero mean	FI	ARIMA(0,1,0)
IT	ARIMA(0,1,0)	SE	ARIMA(0,1,2)
CY	ARIMA(0,1,0)	UK	ARIMA(0,1,1)
LV	ARIMA(0,1,0)	NO	ARIMA(0,1,0)
LT	ARIMA(0,1,0)		

The values extrapolated from these country-specific ARIMA models have been used jointly with those from the convergence model to produce the net migration assumptions.

2.3. The convergence model

Considering that the prolongation of latest migration trend very far in the future may return implausible assumptions, an additional component of the migration model was dealing with a longer term view on migration.

This long-term vision is built on the assumption of converging attractiveness[11] of the hosting countries for the would-be migrants, except for the different population structures projected in the future, factor which was differently tackled (see below). Such tendency of the national pull forces to converge leads to assume a trend towards equal immigration and emigration rates across countries in the very far future. Although in principle these rates may take different values, the extreme uncertainty of the far future and the large number of factors which may influence migration[12] have suggested considering a long-term tendency towards common equal rates of immigration and emigration. While this translates in the 'easy-to-implement' assumption of zero net migration in the far future, it would be misleading to refer to these migration assumptions as 'migration going to zero' (as shown in Table 4 as well). A few points are worth noting in this respect:

[10] In a 'random walk', identified as ARIMA(0,1,0), in each time period the value of the variable is composed of the past value plus a random term.

[11] There could be several reasons for that, such as sharing of best integration practices, convergence of the economic performance of the countries or similar relative size of migrants' networks.

[12] For instance, the impact of high migration flows on the population composition of low-fertility countries, emerging aspects of economic production which may affect migration (e.g., delocalisation of production, increasing use of robotics, etc.), increasing practice of temporary/circular migration, etc.

1. The long-term assumption of zero net migration applies only to the part of the migration model based on the assumption of convergence: different components of the model deal with the short- and medium-term and with the population structure factor.

2. Zero net migration is never reached within the time horizon of the projections: this is only a hypothetical value of reference necessary to model convergence in the national pull forces.

3. Zero net migration does not denote 'no migration': immigration and emigration flows may well be much different from zero, implying a relevant migration turnover.

4. The constraint only applies on the total size, and not to the age- and sex-specific flows.

In practice, the values of net migration based on the convergence assumption for the long term are derived by a piecewise linear interpolation between the last observed value (2015) and the common reference value in the far future. In order to reduce the influence of the last observation, the linear interpolation has been applied first between the net migration value in the year 2015 and an intermediate point value estimated for the year 2020 (see the Table 3), obtained as the average of the net migration observed in the last 20 years (1996-2015). Afterwards, a second linear interpolation was done between the intermediate value in 2020 and the reference value of convergence (here equal to zero). By doing so, the potential impact of an extreme starting value in 2015 is smoothed by forcing it towards a more 'stable' value derived from a much longer time period.

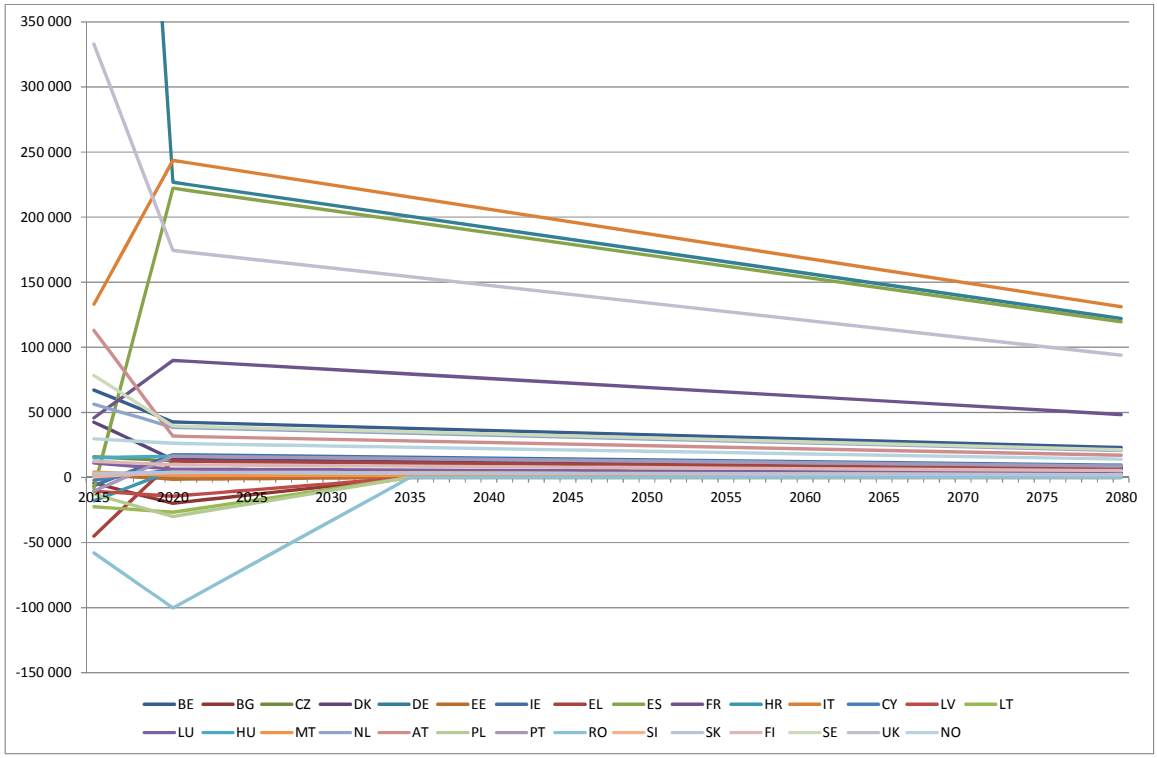
The reference value of convergence was attributed to different years depending upon the sign of the net migration in 2020 as identified by the 20-year average. For the countries where this net migration was positive, the convergence was assumed to occur in the very long term, far ahead of the time horizon of the projections. For the six countries[13] for which the intermediate value was negative, thus with negative (convergence) net migration projected all over the following period, the convergence was assumed to occur by 2035 and values for the years after 2035 were also set to zero, avoiding the prolongation over the entire time horizon of the projections of negative net migration assumption (for the convergence part of the migration model). The assumptions for net migration as from the convergence model are shown in the Figure 1.

[13] They were: Bulgaria, Estonia, Latvia, Lithuania, Poland, and Romania.

Table 3: net migration in 2020 in the convergence model

Country	*Net migration*	*Country*	*Model*
BE	42 662	LU	6 434
BG	-19 776	HU	16 422
CZ	12 724	MT	1 652
DK	13 876	NL	38 428
DE	226 800	AT	31 776
EE	-1 310	PL	-29 947
IE	17 585	PT	16 355
EL	12 640	RO	-100 203
ES	222 219	SI	4 085
FR	89 901	SK	3 762
HR	5 653	FI	9 801
IT	243 567	SE	39 628
CY	5 916	UK	174 430
LV	-14 336	NO	26 207
LT	-26 615		

Figure 1: projected net migration as from the convergence model



Note: for the sake of better readability of the graph, DE out of the frame for the first years

2.4. Preliminary net migration assumptions from 2017 to 2080

Once projected values of the total net migration are available from both the trends and the convergence models, they are pooled giving progressively more weight to the convergence model. This was done by means of a simple weighted average, where the weight attributed to the trend component goes from one in 2015 to zero in 2050, year by which the transition from the trends to the convergence is completed.

2.5. Breakdown by age and sex

Given that a 'net migrant' does not exist, it is necessary to decompose the net migration assumptions to reach the level of detail required for the projections computations. The breakdown by age and sex of the projected total net migration has been implemented by temporarily decomposing the net migration in separated flows for immigration and emigration. First, emigration levels are taken from the average of last observed 5 years (2010-2014). This level is kept constant all over the projections period and the corresponding immigration level is then derived as the sum of net migration and emigration. In the few cases where this operation was returning negative immigration levels, immigration was taken to be equal to the last observed value (2014) and the emigration level adjusted accordingly to match the projected total net migration. The two time series so derived (immigration and emigration flows 2015-2080) have then been divided by two to obtain the sex-specific flows.

The age pattern for each migration flow (male immigrants, female immigrants, male emigrants, female emigrants) has been obtained by linear interpolation between the initial values and age profiles common for all countries, derived in a previous study[14]. The initial values were computed as the average of the latest observed 3 years (2012-2014) of age-specific migration levels and normalized to give a sum equal to one over the entire age range. The common age profiles at the end of the interpolation period were also standardized in the same way. The linear interpolation between these two sets was then returning standardized migration age profiles for each projections year, sex and flow. The entire set of migration assumptions was obtained by multiplying these age-, sex- and flow-specific profiles by the projected corresponding sex- and flow-specific migration, and then re-aggregating age- and sex-specific immigration and emigration to get the projected net migration by age and sex.

2.6. The working-age population 'feedback mechanism'

As mentioned above, countries are characterized by different demographic profiles, including their population structure and the way this latter is projected to change over time. In countries where the size of the population in working ages (conventionally 15-64 years old) is projected to shrink, a 'feedback' correction factor for immigration is applied. This additional immigration is limited to 10% of the projected shrinkage of the working-age population between two consecutive years. This quantity is estimated as overall volume, added in one round to the corresponding annual assumptions for each year of the projections period and distributed by age and sex in accordance with the country- and year-specific immigration patterns.

Such procedure requires to project a first time the population over the entire projections period using the preliminary net migration assumptions, then to compute year by year the projected shrinkage, to estimate the annual additional immigration and finally to re-run the projections computations again for the entire period. This demographic feedback mechanism is only applied for the shrinkage of the working-age population (thus increasing immigration) and not on its growth (thus increasing emigration).

[14] Eurostat (2007): "*EUROPOP2007 Convergence Scenario: Summary Note*". Paper for the joint meeting of the Working Group on Population Projections and Working Group on Ageing Populations and Sustainability, Luxembourg, 29-30 November 2007.

2.7. Final net migration assumptions from 2017 to 2080

Putting all parts together, the assumptions on total net migration are derived from observed data for 2015, from national nowcasting for the year 2016[15], from a mix of trends extrapolation and long-term convergence between 2017 and 2050, almost exclusively from the trends component at the beginning and progressively more from the 'convergence' values until entering the long-term period (2050 onwards) in which the convergence assumption defines the migration values. All over the projections horizon, net migration flows may be increased due to the additional feedback mechanism depending on the working-age population change. The results are shown in the Table 4 for selected years.

Table 4: assumptions on net migration at selected years (in persons)

	2020	2030	2040	2050	2060	2070	2080
BE	53 228	48 304	41 508	32 819	29 538	26 247	22 977
BG	-11 934	-9 082	499	3 858	666	1 258	1 690
CZ	21 499	17 499	20 522	14 037	8 808	8 540	8 985
DK	33 431	26 828	18 872	10 678	11 366	9 295	7 477
DE	327 319	268 069	206 033	199 012	175 047	143 466	135 589
EE	2 334	1 441	1 190	673	91	259	301
IE	9 877	7 504	11 408	13 671	12 179	10 820	9 512
EL	-16 765	-4 149	7 943	13 297	10 464	10 953	9 204
ES	51 185	119 385	163 414	170 939	153 842	136 754	119 658
FR	77 044	85 898	77 328	69 155	62 243	55 322	48 411
HR	-1 732	4 194	5 024	6 016	5 217	4 558	3 858
IT	161 150	209 659	217 702	197 397	176 737	163 777	138 746
CY	1 711	2 911	3 855	4 894	4 396	3 742	3 340
LV	-8 034	-6 138	-1 451	1 192	-2	121	586
LT	-23 791	-16 965	-6 289	1 337	220	-2	818
LU	10 210	8 743	6 992	4 954	4 452	3 963	3 469
HU	19 884	16 221	20 831	15 341	13 836	11 182	9 877
MT	3 214	2 641	1 992	1 402	1 258	1 015	920
NL	66 913	59 482	43 700	29 561	28 609	24 537	20 696
AT	67 802	55 401	40 288	26 342	24 751	20 598	18 524
PL	-33	-2 359	16 243	29 737	11 648	7 300	9 979
PT	2 427	12 804	18 247	15 796	14 601	14 150	11 122
RO	-65 128	-51 088	-8 942	7 746	1 601	2 631	953
SI	4 206	4 126	4 284	3 765	2 831	2 513	2 445
SK	5 919	4 988	6 769	6 466	3 789	3 225	3 222
FI	15 808	13 681	10 723	8 517	7 815	6 765	5 846
SE	67 873	57 155	44 691	30 491	27 439	24 385	21 346
UK	251 507	220 071	180 958	134 178	121 112	107 338	93 924
NO	27 268	26 047	23 684	20 164	18 144	16 128	14 106

[15] Belgium, Estonia, France, Hungary, Romania, and Slovakia have not provided the net migration nowcast for 2016. For these countries the migration assumption for 2016 is mostly taken from the trend extrapolation.

3. COMPARISON WITH THE METHODOLOGY APPLIED IN EUROPOP2013

The methodology applied for the 2015-based population projections is the same applied in the previous round of projections (Eurostat Population Projections 2013-based – EUROPOP2013), except for the following changes:

1. The intermediate point for net migration used in the double linear interpolation of the convergence model is computed over the latest available 20 years instead than over the latest available 10 years as in the EUROPOP2013 model.

2. The transition from trend to convergence starts at the beginning of the projections period (i.e., in 2015), while in EUROPOP2013 the transition was starting in 2020.

3. In EUROPOP2013, the transition for countries with negative net migration at the intermediate point above described was shortened to be completed by 2035; in the 2015-based projections, the final year of the transition remains the same (2050) for all countries.

4. Emigration levels used to break down the net migration by flow are estimated as average over the latest 5 years (2010 – 2014) rather than over the latest 3 years (2010 – 2012) as in EUROPOP2013.