EUROPEAN COMMISSION
EUROSTAT

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

Luxembourg, 29 April 2021
ESTAT/F-2/GL*

TECHNICAL NOTE

Subject: **Methodology for the breakdown of the Eurostat Population Projections 2019-based (EUROPOP2019) by NUTS 3 region**

1. INTRODUCTION

Since almost two decades, Eurostat regularly produces population projections for the European Union (EU) and European Free Trade Association (EFTA) Member States. These projections are used as input to various policies, such as the assessment of the long-term sustainability of public finances[1]. The latest projections, labelled Eurostat Population Projections 2019-based (EUROPOP2019), were released in April 2020. However, since the exercise EUROPOP2013 there was not a disaggregation at regional level of the outcomes projected for the national level. The regional projections are meant to serve as input to policies addressing geographical areas smaller than a country, usually NUTS level 2 or level 3 regions. As several years had elapsed from the latest available regional projections, there was a pressing need for a fresh set. Eurostat has thus produced in 2020 a new set of sub-national population projections at NUTS level 3 with base year 2019. They expand the regional dataset available from the previous round in several aspects.

2. MAIN FEATURES

The methodology applied in the Eurostat projections has been constantly evolving, even though the theoretical framework of reference – the so-called 'convergence scenario' – has remained the same from the EUROPOP2008 exercise[2]. There are several features which make of these regional projections a very challenging exercise:

* Giampaolo Lanzieri (giampaolo.lanzieri@ec.europa.eu). This Note has not been edited and it is released to provide interested parties with preliminary information about the main features of the methodology.

[1] For instance, see *The Ageing Report 2018* by the European Commission, available at https://ec.europa.eu/info/sites/info/files/economy-finance/ip065_en.pdf.

[2] For information about the methodology of EUROPOP2019 at national level, please refer to the documentation available in the metadata of the Eurostat database (https://ec.europa.eu/eurostat/cache/metadata/en/proj_esms.htm) and in particular to the Eurostat

- the large number of geographical entities,
- the lack of data for most of its input components,
- the higher irregularity of demographic patterns,
- an additional component of demographic change (internal migration), and
- the perfect consistency with the projections results at national level.

The Eurostat regional projections are computed for the multi-regional countries, thus excluding Cyprus, Luxembourg, and Liechtenstein, and based on the NUTS-2016 classification. As the input data available by then were covering until the year 2018, the impact of the Covid-19 pandemic is not included.

These regional projections should be interpreted as a breakdown of the national projections consistent with the overall assumption of partial convergence, rather than as an independent set of projections. Like for the national level, they are not a forecast, but a what-if scenario aiming to show a plausible future demographic dynamic, would their assumptions hold.

3. PREPARATION OF DATA INPUT FOR POPULATION AND VITAL EVENTS

The data available to Eurostat at regional level do not have the level of disaggregation required by this exercise. In particular, as for population and vital events, are missing:

- population on 1 January by sex, single age and NUTS 3 region,
- deaths by sex, single age (reached during the year) and NUTS 3 region,
- live births by sex and NUTS 3 region, and
- live births by single age (reached during the year) of the mother and NUTS 3 region.

These data have been derived by applying an advanced version of the Iterative Proportional Fitting (IPF) method to data available for less granular geographic areas (usually NUTS level 2 regions) or available with less detail (e.g., by 5-year age group).

4. MORTALITY

Because the projections model works on a period-cohort observational plane, the probabilities must be computed with reference to the same plan. The probabilities of dying are estimated using the ratio of the annual single-age (reached during the year) deaths to the population that will be exposed to the risk of dying during the calendar year. Considering the period-cohort frame, for any given calendar year, this is implemented as follows:

Technical Note of 30 April 2020 on “*Methodology of the Eurostat population projections 2019-based (EUROPOP2019)*”.

$$
\begin{cases}
q_{t,s,0} = \dfrac{D_{t,s,0}}{B_{t,s}} & x = 0 \\
q_{t,s,x} = \dfrac{D_{t,s,x}}{P_{t,s,x-1}} & x \in I[1, \dots, 99] \\
q_{100+} = \dfrac{D_{100+}}{P_{99} + P_{100+}} & x = 100 +
\end{cases} \quad [1]
$$

where $D_{t,s,x}$ is the sex-specific number of deaths of age x reached at the end of the year occurred during the calendar year t, $P_{t,s,x-1}$ is the corresponding sex-specific population at the beginning of the year (on 1 January of the year t) of age reached $x - 1$, and $B_{t,s}$ are the births of sex s recorded during the calendar year t.

In principle, q_x can only take values between 0 and 1; as for the extremes of this interval, whilst theoretically possible, they mean that, in the specific age class with 0 or 1 value for q_x, there is respectively no mortality at all or certainty of death. Given the limited number of events in smaller geographical areas, the q_x for NUTS 3 show higher variability by age and, due to data quality issues and/or peculiar outcomes of the previous estimation procedure and/or effect of migration, it might happen that the estimated annual q_x assume values out of or at the boundaries of the theoretical range [0,1]. In order to reduce such variability in the single ages (particularly at youngest ages), the q_x have been computed using the observations over the 5-year period 2014-2018:

$$
\begin{cases}
\bar{q}_{r,s,0} = \dfrac{\sum_{t=2014}^{2018} D_{r,s,0,t}}{\sum_{t=2014}^{2018} B_{r,s,t}} & x = 0 \\
\bar{q}_{r,s,x} = \dfrac{\sum_{t=2014}^{2018} D_{r,s,x,t}}{\sum_{t=2014}^{2018} P_{r,s,x-1,t}} & x \in \mathbb{Z}[1, \dots, 99] \\
\bar{q}_{r,s,100+} = \dfrac{\sum_{t=2014}^{2018} D_{r,s,100+,t}}{\sum_{t=2014}^{2018} P_{r,s,99+,t}} & x = 100 +
\end{cases} \quad [2]
$$

Further, if still taking values out of the range (0,1), the region-, sex- and age-specific averaged $\bar{q}_{r,s,x}$ for a NUTS 3 region have been replaced by the corresponding $\bar{q}_{r,s,x}$ at NUTS 2 level when available, otherwise with the corresponding $\bar{q}_{r,s,x}$ at NUTS 1 level, or even further at national level (NUTS 0).

The average $\bar{q}_x$ have then been modelled using the Heligman-Pollard mortality law and then further smoothed by applying a monotonic regression spline. Last step has been to shift the estimated age pattern of mortality such that the application of it to the observed data would reproduce the total number of death in the latest available year, which was by then the year 2018.

4.1. Mortality assumptions

In order to formulate assumptions for the NUTS 3 levels of a country consistently with the assumptions previously made for the country level, the regional $q_{r,s,x}$ are put in relation with the corresponding probabilities of dying for the entire country[3] $q_{c,s,x}$, so to obtain the following sex- and age-specific ratio of probabilities:

$$r^q_{r,s,x} = \frac{q_{r,s,x}}{q_{c,s,x}} \qquad [3]$$

The assumption of partial convergence of regional values towards the national level is then easily formulated by imposing a long-term convergence of the ratio to the unity, ending outside of the projections time horizon:

$$r^q_{r,s,x} \to 1 \qquad [4]$$

The final step is thus to multiply, in each year of the projections time horizon, the sex- and age-specific converging ratio by the corresponding assumed values of mortality at country level:

$$q_{r,s,x,t} = r^q_{r,s,x,t} \cdot q_{r,s,x,t} \quad ; \quad t \in \mathbb{Z}[2019, \dots, 2100] \qquad [5]$$

It should be noted that the regional probabilities of dying remains different from the national level all along the projections time horizon, even though their differential are projected to decrease over time. Partially converging towards the national patterns has also the effect of progressively removing peculiarities in the regional mortality profiles.

5. Fertility

Likewise for mortality, the aim here is to derive a period-cohort fertility age pattern to use as input to assumptions models. However, contrary to mortality where probabilities are used, the fertility indicator is an occurrence / exposure rate, that is computed as the ratio between the number of births from mothers of age reached *x* during the year *t* and the average number of women (subscript *f* in the formula) in that same age class[4]:

$$f_{r,x,t} = \frac{B_{r,x,t}}{0.5 \cdot \left(P_{r,f,x-1,t} + P_{r,f,x,t+1}\right)} = \frac{B_{r,x,t}}{E_{r,x,t}} \quad ; \quad x \in \mathbb{Z}[14, \dots, 50] \qquad [6]$$

In order to reduce the impact of random variability and thus to estimate smoother fertility age patterns, the data have been aggregated over the years 2014-2018:

$$\bar{f}_{r,x} = \frac{\sum_{t=2014}^{2018} B_{r,x,t}}{\sum_{t=2014}^{2018} E_{r,x,t}} \quad ; \quad x \in \mathbb{Z}[14, \dots, 50] \qquad [7]$$

In the case of fertility, it is not necessary to replace missing data with data from hierarchically higher NUTS regions. The $\bar{f}_{r,x}$ are then smoothed by applying a weighted regression B-splines with a concavity constraint. Last step has been to shift the fertility age pattern such to match the total number of births in the latest available year (by then, the year 2018) if multiplied by the estimated exposure to childbirth of the same year.

[3] See par.3.3 in the Eurostat Technical Note on "Methodology of Eurostat population projections 2019-based (EUROPOP2019)" of 30.04.2020, available at: https://ec.europa.eu/eurostat/cache/metadata/Annexes/proj_esms_an1.pdf.

[4] More correctly, the average is an estimate of the person-years of exposure *E* to childbirth of women aged *x* at the end of the year *t*.

5.1. Fertility assumptions

Likewise for the mortality assumptions, the regional $\hat{f}_{r,x}$ are put in relation with the corresponding fertility rates for the entire country $\hat{f}_x^C$[5], so to obtain the following age-specific ratio of fertility rates:

$$r_{r,x}^f = \frac{\hat{f}_{r,x}^R}{\hat{f}_{r,x}^C} \qquad [8]$$

The assumption of partial convergence of regional values towards the national level is then easily formulated by imposing a long-term convergence of the ratio to the unity, ending outside of the projections time horizon:

$$r_{r,x}^f \rightarrow 1 \qquad [9]$$

The final step is thus to multiply, in each year of the projections time horizon, the sex- and age-specific converging ratio by the corresponding assumed values of mortality at country level:

$$f_{t,x}^R = r_{t,x}^f \cdot f_{t,x}^C \quad ; \quad t \in \mathbb{Z}[2019, \dots ,2100] \qquad [10]$$

It should be noted that the regional fertility rates remain different from the national level all along the projections time horizon, even though their differentials are projected to decrease over time. Partially converging towards the national patterns has also the effect of progressively removing peculiarities in the regional fertility profiles.

6. INDICATORS OF MORTALITY AND FERTILITY

From the above-described set of assumptions it is possible to derive common indicators for fertility and mortality, such as the total fertility rate (TFR) as well as a life table and the associated life expectancy at birth (*e*^{0}). These measures are based on hypothetical underlying patterns of fertility and mortality and can be computed before ("*ex-ante*") the projections computations are actually carried out. Once the projections are computed, it is also possible to derive the same measures from the projected vital events and populations. The measures computed after the projections computations ("*ex-post*") may be different from those derived *ex-ante*. This may happen because the population sizes may be such that the assumed theoretical laws applied to the sex- and age-specific exposure to the risk, combined with the rounding and the constraint of matching the national values, do not generate the same number of vital events at some ages. This is what is also observed empirically, for instance for mortality at young ages when there are no recorded deaths: this does not mean that the risk of dying at those ages is nihil, it is rather the outcome of very small (but not zero) probabilities of dying in relatively small populations at risk. In fact, vital events are dichotomous in nature (i.e., either they occur or they do not) and thus their number are integers (i.e., there is no a "fraction of a death"). This feature is obviously potentially more visible in NUTS 3 regions than at national level or at lower geographic granularity.

---

5 See par.3.3 in the Eurostat Technical Note on "Methodology of Eurostat population projections 2019-based (EUROPOP2019)" of 30.04.2020, available at: https://ec.europa.eu/eurostat/cache/metadata/Annexes/proj_esms_an1.pdf.

Another feature of the fertility and mortality indicators computed for these regional projections is that they are based on a period-cohort observational plane (i.e., classifying events by age reached at the end of the year) and they are therefore not fully comparable with the indicators produced by Eurostat based on the age-period observational plane (i.e., classifying events by age in completed years). In particular, the life table for NUTS 3 regions is computed by applying an alternative methodology described in detail elsewhere[6].

7. International immigration

7.1. Data availability

In order to formulate quantitative assumptions on future regional migration flows, firstly it is necessary to know the distribution of the international migration from/to a country across its NUTS 3 regions. As there are no time series about international migration flows by NUTS 3 region available to Eurostat, in order to minimize the burden on data providers, regional distribution of international migration data have been sought in the websites and databases of the national statistical offices. For 20 out of 28 countries, it was possible to find data that could be used, directly or indirectly, as measure of the total international immigration and emigration flows to/from NUTS3 regions for the years 2014-2018. For the rest of countries (Germany, Ireland, Greece, France, Hungary, Malta, Poland, and Portugal), the only data source available to Eurostat was the Census Hub[7], from which data from the population censuses carried out in 2011 could be retrieved, and in particular the variable "*Residence one year before*" (R1Y). However, these data were based on a NUTS classification other than the one used in the projections exercise and they were not complete either for France and Germany. Therefore, the census data had to be converted in NUTS-2016 classification[8] and missing data for France and Germany had to be estimated[9]. For the former country, the estimated immigration for Mayotte (FRY50), the only region with missing data, has been derived by applying the same *per-capita* immigration value[10] of the Réunion (FRY40) to an estimate of its population in

[6] See the Eurostat Technical Note of 14 April 2020 on "*An alternative life table based on probabilities of dying within the calendar year (annual period-cohort life table)*". Available at https://ec.europa.eu/eurostat/cache/metadata/Annexes/proj_esms_an24.pdf.

[7] https://ec.europa.eu/CensusHub2

[8] Sub-national data from the Eurostat Census Hub are available by the NUTS classification in force by then, which was the NUTS-2010. In the meanwhile, the NUTS classification has changed twice, and the current projections exercise uses data with NUTS-2016 classification. The conversion of the census data to the NUTS-2016 classification was made using the NUTS converter developed by the Joint Research Centre of the European Commission (https://urban.jrc.ec.europa.eu/nutsconverter/#/). Regional data resulting from this conversion might not have the same level of precision of data produced directly with the NUTS-2016 classification.

[9] In the analysis to identify the best predictor(s) for the missing migration inflows have been considered various variables as well as various model specifications. In particular, the population size, the network of migrants (as measured by the stock of foreign-born population) and the economic differentials (as measured by the GDP *per capita*) have been tested.

[10] The resulting estimated inflow of 512 immigrants would refer to the legal inflow only. Mayotte is however subject to important irregular migration – see e.g. Torre, H. (2008): "*Rapport d'information n° 461 (2007-2008) fait au nom de la commission des finances, sur l'immigration clandestine à*

2011[11]. As for Germany, data on regional immigration flows from the census 2011 were missing for five regions (DE913, DEA24, DEA32, DEA41, and DEA55) and they have been estimated by means of a partial least squares model using measures of population size and migrants network[12].

At this point it has been possible to compute the regional shares of the international immigration $w_{r,t}^{IMM}$ (only totals) for all the countries:

$$w_{r,t}^{IMM,c} = \begin{cases} \dfrac{I_{r,t}}{\sum_{r=1}^{R} I_{r,t}} & c \neq DE, IE, EL, FR, HU, MT, PL, PT \\ \dfrac{I_{r,2011}^{R1Y}}{\sum_{r=1}^{R} I_{r,2011}^{R1Y}} & c = DE, IE, EL, FR, HU, MT, PL, PT \end{cases} \qquad [11]$$

where $I_{r,t}$ is the total international immigration arrived in the region r of country c in the year t as from annual statistics, and $I_{r,2011}^{R1Y}$ is the number of persons – resident in the region r – who were residing abroad one year before the census date.

7.2. Past international immigration to NUTS 3 regions

The annual regional weight of international immigration, computed for each year from 2014 to 2018, is then applied to the corresponding international immigration flows, broken down by sex and single age (see below), for those same years.

$$\hat{I}_{r,s,x,t} = w_{r,t}^{IMM,c} \cdot I_{c,s,x,t} \qquad [12]$$

By doing so, it is assumed that there are no regional differentials in the age and sex patterns of international immigration, and they are therefore mimicking the national patterns.

The resulting outcomes are re-proportionated and rounded to get perfect consistency between the sex- and age-specific values at national level and the corresponding sum of the annual estimates of international immigration to NUTS 3 regions:

$$\hat{\hat{I}}_{r,s,x,t} = k_{r,s,x,t}^{IMM} \cdot \hat{I}_{r,s,x,t} \qquad [13]$$

with:

$$k_{r,s,x,t}^{IMM} \ni \left[\left(\sum_{r=1}^{R} \left(k_{r,s,x,t}^{IMM} \cdot \hat{I}_{r,s,x,t} \right) = I_{c,s,x,t} \right) \cap \left(\hat{\hat{I}}_{r,s,x,t} \in \mathbb{Z}_{+} \right) \right] \qquad [14]$$

Mayotte", which estimates the annual irregular migration by then to about 16,000 migrants, broadly corresponding to the number of enforcements of immigration law.

[11] The French region of Mayotte (FRY50) is not covered by the census data available at Eurostat. For its population, data are retrieved from a former Eurostat estimate (see Eurostat Technical Note of 20 September 2012 on "Estimation of the current population size of Mayotte (FR)").

[12] The former is a general indicator of capacity to attract and receive immigration flows, the latter aims to capture the propensity of migrants to settle in the proximity of an existing community of immigrants.

At the end of this process, all regional immigration values are positive integers (as they should be) and consistent with the totals.

7.2.1. Breakdown of the international immigration to countries by age and sex

International immigration data are available from the Eurostat database[13], but to a different extent: the breakdown by sex and single *age reached* is available for Belgium, Bulgaria, Czechia, Denmark, Germany[14], Estonia, Spain, France, Croatia, Italy, Latvia, Lithuania, Hungary, the Netherlands, Poland, Portugal, Slovakia, Finland, Sweden, Iceland, Norway, and Switzerland (22 countries); for Ireland, Greece, Malta, Austria, Romania, and Slovenia (6 countries) it is available the breakdown by sex and single *age completed* and therefore the distribution by *age reached* is obtained by conversion. The conversion from immigration values by age completed to values by age reached in done as follows:

$$I_x^R = \begin{cases} \dfrac{I_0^C}{2} & x = 0 \\ \dfrac{I_{x-1}^C + I_x^C}{2} & 1 \leq x \leq 99 \\ \dfrac{I_{99}^C}{2} + I_{100+}^C & x \geq 100 \end{cases} \qquad [15]$$

All the converted values are rounded to integers and such that their sum is consistent with the total immigration.

7.3. Assumptions on the evolution of the international immigration pull factors

In the framework of the overarching assumption of partial convergence, it can be assumed that the international immigration pull factors may evolve towards reducing their differential impact across regions. For instance, for the migrants' network this would imply assuming that the migrants' communities will tend to be more equally distributed across regions and therefore shares of foreign-born people will be closer to equality; for the economic attractiveness, this would mean as well a long-term tendency towards equal level of attractiveness across regions. Unlike the former pull factors, it is much less likely that the population shares of the regions evolve towards common levels, as this would mean that each region in a country would host the same number of people; on the contrary, these shares change year after year and therefore, the converging tendency related to the regional population shares must be expressed differently. If the other two pull factors tend to take common values across regions, their relative influence on the international immigration share should decrease as compared to the regional population share. Eventually, the only variability in the regional immigration shares would come from the variability in the regional population shares, i.e.:

$$\frac{I_r}{I_c} = f\left(\frac{P_r}{P_c}\right) \qquad [16]$$

[13] Table "Immigration by age and sex" (*migr_imm8*).

[14] Data for Germany stop at age 95. They are then uniformly redistributed in the age class 95-99.

where $I_c = \sum_{r=1}^{R} I_r$ is the total immigration to the country c, which is the sum of the international immigration I_r to each of its regions, and P are the corresponding population sizes. Assuming a direct relation between the regional population shares and the regional immigration shares, the [16] becomes:

$$\frac{I_r}{I_c} \equiv \frac{P_r}{P_c} \quad \Rightarrow \quad i_r = \frac{I_r}{P_r} = \frac{I_c}{P_c} \qquad [17]$$

where i_r is the *per-capita* immigration[15] to the region r. Consequently, the converging process related to population sizes as determinant of the international immigration can be expressed as:

$$i_{r,t} = \frac{I_{r,t}}{P_{r,t}} \longrightarrow \frac{I_c}{P_c} = i_c \qquad [18]$$

This partial convergence process would apply for each sex and single age, thus $i_{r,s,x,t} \longrightarrow i_{c,s,x}$. In order to obtain a more robust estimate of the $i_{r,s,x}$ at the beginning of the projections period, the regional immigration estimates are aggregated over the period 2014-2018 and related to the corresponding population aggregates:

$$i_{r,s,x} = \frac{\sum_{t=2014}^{2018} \hat{\hat{I}}_{r,s,x,t}}{\sum_{t=2014}^{2018} P_{r,s,x,t}} \qquad [19]$$

With high data granularity it may happen that $i_{r,s,x,t}$ takes indefinite (0/0) or infinite (∞) values, both results caused by a null population size in the region-, year-, sex-, and age-specific category. In the former case, it can be considered that $i_{r,s,x,t} = 0$. On the contrary, in the latter case (infinite value) the estimated immigration $\hat{\hat{I}}_{r,s,x}$ is actually positive but, if there is no resident population, then there cannot be a *per-capita* measure. For the sake of simplicity and considering the marginal cases where such situation might occur[16], this issue is solved by imposing value zero as well to the $i_{r,s,x}$. Further, $i_{r,s,x}$ for ages above 80 are imposed to follow a non-monotonic decrease, whose upper values are those from the previous age:

$$i_{r,s,x} = \begin{cases} \dfrac{I_{r,s,x}}{P_{r,s,x}}, & (x \leq 80) \cup \left[(x > 80) \cap \left(\dfrac{I_{r,s,x}}{P_{r,s,x}} \leq \dfrac{I_{r,s,x-1}}{P_{r,s,x-1}}\right)\right] \\ \dfrac{I_{r,s,x-1}}{P_{r,s,x-1}}, & (x > 80) \cap \left(\dfrac{I_{r,s,x}}{P_{r,s,x}} > \dfrac{I_{r,s,x-1}}{P_{r,s,x-1}}\right) \\ 0, & P_{r,s,x,t} = 0 \end{cases} \qquad [20]$$

A similar approach is applied for the estimation of the *per-capita* immigration at national level $i_{c,s,x}$. Sex- and age-specific immigration and population counts at national level are cumulated for the years from 2014 to 2018, from which:

---

[15] Because immigration is not an event occurring to the resident population, it is here preferred to refer to i_r as a relative measure and not as a probability. Further, in principle the immigration could also be larger than the resident population: this would lead to values of i_r higher than one, which is incompatible with the measure of probability.

[16] This is more likely to happen in small regions at very high ages for men.

$$i_{c,s,x} = \frac{\sum_{t=2014}^{2018} I_{c,s,x,t}}{\sum_{t=2014}^{2018} P_{c,s,x,t}} \qquad [21]$$

Also the $i_{c,s,x}$ are corrected for non-regular values and imposed a non-monotonic decrease from age 80, like in [20].

7.4. Assumptions on the regional international immigration

Coherently with the assumed evolution in the determinants of immigration, the assumptions are formulated in terms of partial convergence of each regional $i_{r,s,x}$ to the correspondent sex- and age-specific ratios $i_{c,s,x}$ in the total population. To take into account the evolutions in these latter patterns, this is implemented by linking the regional and national patterns as follows:

$$r^{i}_{r,s,x,t} = \frac{i_{r,s,x,t}}{i_{c,s,x,t}} \to 1 \qquad [22]$$

Eventually, once the ratio takes values equal to one (which it does not happen in the time horizon covered by the projections), all regions would have an equal share of *per capita* immigration, which however is translated in different immigration levels depending on the regional population size. The use of the ratios $r^{i}_{r,t,s,x}$ ensures as well that, in this process of partial convergence, the regional values follow the general trends whilst reducing the regional differentials.

The ratios are assumed to decrease linearly towards the unitary value. There are two cases in which the ratio can take non-valid values:

a) when both $i_{r,s,x,t}$ and $i_{c,s,x,t}$ are equal to zero, which returns an indefinite value (0/0): in this case, it is $r^{i}_{r,t,s,x} \equiv 1$ for the entire time horizon of the projections and consequently the regional value will be always equal to the national one:

b) when $i_{r,s,x,t} > 0$ but $i_{c,s,x,t} = 0$, which returns the infinite value (∞): in this case, in principle the estimate of the regional value should not be positive, given that there are no occurrences at national level; it is then imposed $i_{r,s,x,t} = 0$ and the case becomes equal to the previous one.

The assumptions on the regional shares of *per capita* immigration are thus obtained as:

$$i_{r,s,x,t} = r^{i}_{r,s,x,t} \cdot i_{c,s,x,t} \quad ; \quad t \in [2019, \ldots, 2100] \qquad [23]$$

8. INTERNATIONAL EMIGRATION

8.1. Data availability

For the international emigration, the procedure is very similar to the international immigration assumptions. The first step is the computation of the regional weights, but because there are no data available from the EU censuses 2011 about emigration, for the countries for which Eurostat did not have annual data about the regional distribution of international emigration, the regional shares of emigration have been assumed equal to the regional shares of immigration[17], the assumption behind being that a region that attracts a proportionally higher number of international immigrants compared to other regions is also likely to have an higher number of international emigrants, due to return migration as well. In formula:

$$w_{r,t}^{EMI,c} = \begin{cases} \dfrac{E_{r,t}}{\sum_{r=1}^{R} E_{r,t}} & c \neq DE, IE, EL, FR, HU, MT, PL, PT \\ \dfrac{I_{r,2011}^{R1Y}}{\sum_{r=1}^{R} I_{r,2011}^{R1Y}} & c = DE, IE, EL, FR, HU, MT, PL, PT \end{cases} \quad [24]$$

8.2. Past international emigration from NUTS 3 regions

Next phase is to distribute the observed international emigration during the years 2014-2018 across NUTS 3 regions. International emigration data are available from the Eurostat database[18], but to a different extent: the breakdown by sex and single *age reached* is available for Belgium, Bulgaria, Czechia, Denmark, Germany[19], Estonia, Spain, Croatia, Italy, Latvia, Lithuania, Hungary, the Netherlands, Poland, Slovakia, Finland, Sweden, Iceland, Norway, and Switzerland (20 countries); for Ireland, Greece, Malta, Austria, Romania, and Slovenia (6 countries) it is available the breakdown by sex and single *age completed* and therefore the distribution by *age reached* is obtained by conversion; last, for France and Portugal (2 countries), the sex-specific age distribution is taken from the EUROPOP2019 exercise at national level and applied to the available breakdown by sex. The objective here is to obtain international emigration flows broken down by sex and single age reached for all the countries for the period 2014-2018.

At this point it is possible to estimate the international emigration from each NUTS 3 region. This is done by applying the regional weights to the reported international emigration flows broken down by sex and age, which returns the same emigration flows further broken down by NUTS 3 region for each of the years in the period 2014-2018.

$$\hat{E}_{r,s,x,t} = w_{r,t}^{EMI,c} \cdot E_{c,s,x,t} \quad [25]$$

that it is rounded and matched with the national level:

$$\hat{\hat{E}}_{r,s,x,t} = k_{r,s,x,t}^{EMI} \cdot \hat{E}_{r,s,x,t} \quad [26]$$

with:

$$k_{r,s,x,t}^{EMI} \ni \left[\left(\sum_{r=1}^{R} \left(k_{r,s,x,t}^{EMI} \cdot \hat{E}_{r,s,x,t} \right) = E_{c,s,x,t} \right) \cap \left(\hat{E}_{r,s,x,t} \in \mathbb{Z}_{+} \right) \right] \quad [27]$$

[17] It should be noted that this assumption does not imply an equal number of international immigrants and emigrants for a region.

[18] Table "Emigration by age and sex" (*migr_emi2*).

[19] Data for Germany stop at age 95. They are then uniformly redistributed in the age class 95-99.

At the end of this process, all regional emigration values are positive integers (as they should be) and consistent with the totals.

8.3. Probability of emigrating abroad

The last step is to compute the sex- and age-specific probabilities of emigrating abroad from each region. The estimated regional annual emigration flows and populations are aggregated over the 5 years from 2014 to 2018 and put in a ratio such to give the estimated probabilities of emigrating abroad (indicated by *e*) for each region *r*, sex *s* and age *x*:

$$e_{r,s,x} = \frac{\sum_{t=2014}^{2018} \hat{E}_{r,s,x,t}}{\sum_{t=2014}^{2018} P_{r,s,x,t}} \qquad [28]$$

This set of probabilities is adjusted for non-regular values and for non-monotonic decrease above age 80 like in immigration, and it is then used as starting distribution for the assumption on regional international emigration

8.4. Assumptions on the regional international emigration

Like for international immigration to regions, the assumptions on international emigration are formulated in terms of partial convergence to the sex- and age-specific national emigration probabilities:

$$r^e_{r,s,x,t} = \frac{e_{r,s,x,t}}{e_{c,s,x,t}} \rightarrow 1 \qquad [29]$$

and, consequently, the regional sex- and age-specific probabilities of emigrating abroad are given by the following formula:

$$e_{r,s,x,t} = r^e_{r,s,x,t} \cdot e_{c,s,x,t} \quad ; \quad t \in [2019, \ldots, 2100] \qquad [30]$$

The assumptions on the regional international immigration can therefore be computed only once completed the projections (*ex-post*). The consistency with the national total is controlled for each sex and single age.

9. INTERNAL MIGRATION

Data on internal migration during the period 2014-2018 were kindly provided by Belgium, Bulgaria, Czechia, Denmark, Germany (only totals by sex 2014-2017), Estonia (2014-2016), Ireland (2016), Greece (2011), Spain, Croatia, Italy, Latvia, Lithuania, Hungary (only breakdown by age), the Netherlands, Austria, Poland, Slovenia, Slovakia, Finland, Sweden, Iceland, Norway, and Switzerland. Data for the countries that had not provided data, or they were incomplete, or for which data were considered not adequate (Greece, Germany[20], France, Malta, Portugal, and Romania), have been estimated building upon a migration matrix derived using a gravity model. In this latter, a measure of economic differential has been used as “distance” between regions is, and the share of

[20] The data provided by Germany have been merged with the estimates from the gravity model.

persons who did not change their NUTS 3 region of residence (as from the censuses 2011) has also been taken into account. These estimated total in- and out-flows have been broken down by sex and age using the patterns from international migration, and therefore the estimated flows may not fully capture the age and sex specificities of the internal migration such as the inter-regional post-retirement migration.

To gain in volume of the flows and thus in robustness of the results, the internal migration data have been compiled over the available years, producing for each country an internal migration matrix containing the flows during the period 2014-2018. From this matrix have been computed the totals by row and by column. These marginal row and column represent respectively the total internal in-migration to the specific region and the total internal out-migration from the specific region. Such computations have been done for each sex and each single age of migrants, therefore for 202 migration matrices for each country. These values of the marginal column have then been related to the corresponding sex- and age-specific population, obtaining the sex- and age-specific internal out-migration probabilities. These rates have been smoothed using Rogers-Castro models[21]. The application of these out-migration rates to the population generates the sex- and age-specific internal migration outflows from each region. For each sex and single age, these region-specific outflows are then pooled and re-distributed across regions. The sex- and age-specific shares of internal migrants attributed to each region are derived from the marginal row of the sex- and age-specific migration matrix. For instance, all 25-year old male internal migrants from every region are grouped together (that would give the overall total as the sum of the marginal column of the sex=M and age=25 migration matrix) and redistributed to the regions according to the defined proportion (that would give the marginal row of the same sex-age-specific migration matrix). The number resulting from the redistribution according to the shares are rounded such that their sum matches exactly the total outflows. This methods ensures consistency between out-migration and in-migration flows, the latter being generated from the former ones.

For the projections time horizon, it is assumed that the regional differentials that triggers internal migration are likely to decrease over time, which is translated in shrinking out-migration rates. For the same reason, also the regional shares are assumed to reduce the inter-regional differences and therefore to evolve towards a situation of equal in-migration attractiveness. The ultimate level for this latter is estimated as the population share at the beginning of the projections period; in other words, each region is attributed a share of in-migrants that over time becomes closer to its population weight in the country. This ultimate situation corresponds in fact to the case in which the regional *per-capita* in-migration are equal for all regions, and equal to the national value. In formulas:

$$o_{r,s,x,t} \rightarrow 0 \qquad [31]$$

where $o_{r,s,x,t}$ is the region-, sex- and age-specific out-migration probability in the year *t*, and

[21] The variant here applied is the one with 13 parameters, thus including in the age pattern the so-called "education peak" – if present. See Wilson, T. (2010): "Model migration schedules incorporating student migration peaks", *Demographic Research*, 23(8):191-222.

$$s_{r,s,x,t} = \frac{I^{INT}_{r,s,x,t}}{I^{INT}_{c,s,x,t}} \rightarrow \frac{P_{r,s,x}}{P_{c,s,x}} \qquad [32]$$

where $s_{r,s,x,t}$ is the region-, sex- and age-specific in-migration share in the year *t*. It may be noted that this latter equation can also be formulated as:

$$\frac{I^{INT}_{r,s,x,t}}{P_{r,s,x}} \rightarrow \frac{I^{INT}_{c,s,x,t}}{P_{c,s,x}} \qquad [33]$$

that is an approximate version of the regional *per-capita* in-migration converging towards the national value.

10. PROJECTIONS COMPUTATION

The computations for each year of the projections time horizon and for each country are run as follows:

1. Prepare the population by sex and age at the beginning of the year, which is the base population at the beginning of the projections time horizon, otherwise the population at the end of the previous year.

2. Compute the number of deaths by sex and age (except age 0).

3. Compute the number of international emigrants by sex and age (except age 0).

4. Compute the number of international immigrants by sex and age (except age 0).

5. Compute the population without internal migration at the end of the year by sex and age (except age 0).

6. Compute the average population in the year by sex and age (except age 0).

7. Compute the number of live births by age of the mother and break them down by sex.

8. Compute the number of deaths by sex at age 0.

9. Compute the number of international emigrants by sex at age 0.

10. Compute the number of international immigrants by sex at age 0.

11. Compute the population without internal migration at the end of the year by sex at age 0.

12. For each sex and age, use the relative distribution across regions of births, deaths, international emigration and international immigration to re-proportionate if necessary ensuring the match between national values and sum of regional values. All the resulting figures are integer.

13. Compute the number of internal emigrants (out-migration) for each sex, age and region. The figures are rounded to the nearest integer.

14. For each sex and age, pool the number of internal emigrants from all regions and redistribute them across the same regions based on regional in-migration shares. All the resulting figures on internal immigrants (in-migration) are integers and their sum consistent with the total number of internal emigrants.

15. Compute the population at the end of the year by sex and age, including internal migration. All figures of the demographic balance are integers.

Age for events is intended as age reached during the year. It may be noted that the re-proportioning to the national values may cause a discrepancy between the (*ex-ante*) events generated by the theoretical patterns and the number of projected (*ex-post*) events, particularly in the first year of the projections where the national values are taken from the nowcast.