EUROPE-IGM-ATLAS

*The European Atlas of Spatially Disaggregated Intergenerational Mobility*

# EUROPE-IGM-ATLAS Database Guide V1

Sarah McNamara, Guido Neidhöfer, and Patrick Lehnert

July 2, 2026

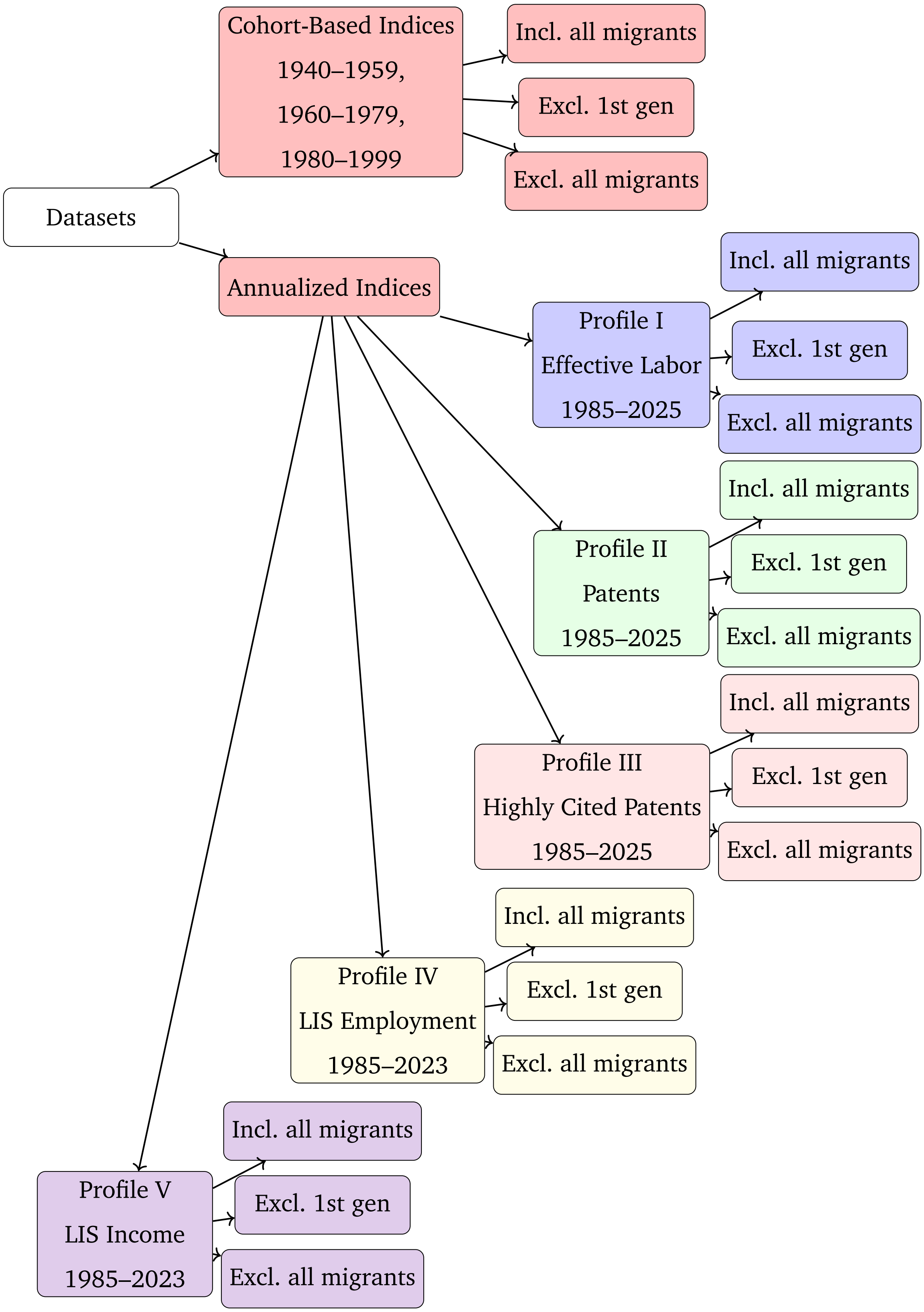
# Contents

# 1 Introduction

The European Atlas of Spatially Disaggregated Intergenerational Mobility (*EUROPE-IGM-ATLAS*), presented in McNamara et al. (2026), provides a database of indices of intergenerational educational mobility and human capital for 36 European countries, with spatially disaggregated measures available for 108 meso-regions (NUTS 1), and 225 micro-regions (NUTS 2). Indices are available for two measures of intergenerational educational mobility—*persistence* and *standardized persistence*, the average level of educational inequality for both the parents' and children's generations, and the average (harmonized) years of education for both the parents' and children's generations.

## 1.1 Database Structure

**Figure 1.1 The Structure of the EUROPE-IGM-ATLAS and Versions of Indices**



The structure of the *EUROPE-IGM-ATLAS* is summarized in Figure 1.1, and described in the following. The Atlas is currently available in the form of two distinct datasets:

- *Europe-IGM-Atlas-Cohorts* : Cohort-level measures
- *Europe-IGM-Atlas-Annualized* : An annualized panel of *effective* measures, based on weighted composite indices

The dataset *Europe-IGM-Atlas-Cohorts* is available for cohorts born 1940–59, 1960–79, and 1980–99, respectively. The dataset *Europe-IGM-Atlas-Annualized* is available for the years 1985–2025[1].

Both the cohort-level dataset and the annualized panel dataset contain three versions of the indices, based on the sample restrictions applied during the indices' construction. In both datasets, these are indexed using the variables *mig_subsample* (label) and *mig_subsample_id* (numeric code), such that the indices are available for sub-samples which i. exclude first-generation migrants, ii. exclude all individuals with a migration background[2], and iii. include all migrants. The choice of version is presumably conditional on the use case of those using our database.

For the annualized panel dataset, *Europe-IGM-Atlas-Annualized*, for each of the three migrant sample types, five versions of the indices are provided based on the choice of weighting profile used in their construction. The weighting profiles are indexed using the variable *weight_type*, summarized as follows:

- *Profile I* : Based on per-capita labor profiles over the life cycle, from Mason et al. (2022).
- *Profile II* : Based on innovation life-cycle profiles for all patenting activity, from Bell et al. (2016).

[1]Specifically, the annualized panel is available 1985–2025 for Profile I-III weighting types. For the versions of the indices constructed using Profile IV and V weights, extracted from the Luxembourg Income Study (LIS), the panel is only available until 2023 in this version of the database. Future releases will extend the date range, given any changes in availability of the LIS data used to compute said weights.

[2]Where in addition to first generation migrants, a migration background includes both second generation migrants, and later generation migrants who are not citizens of the country in which they reside.

- *Profile III* : Based on innovation life-cycle profiles for highly cited patenting activity, from Bell et al. (2016).
- *Profile IV* : Based on employment contribution profiles from the Luxembourg Income Study (LIS) database, accessed via LISSY.
- *Profile V* : Based on total individual income contribution profiles from the Luxembourg Income Study (LIS) database, accessed via LISSY.

See Section 3.6.1 for a more in-depth description. The goal of the annualized panel is to provide *effective* measures of intergenerational educational mobility and human capital. To this end, the indicator value of each region-cohort pair is weighted according to the expected economic contribution of said cohort in a given year, and the composite *effective* measure for that year is then a weighted sum across cohorts within said region. See Section 3.6 for further details. Similar to choosing the migrant sample type, the choice of weighting profile is presumably conditional on the use case of those using our database.

# 2 Codebook

## 2.1 *Europe-IGM-Atlas-Cohorts* – Cohort-Based Measures

- *mig_subsample* : Sub-sample used in indices construction (descriptive).[3]

| id | Value | Desc. |
|---|---|---|
| 1 | excl. all migrants | excluding all migrants |
| 2 | excl. first-gen migrants | excluding first-gen migrants |
| 3 | incl. migrants | including all migrants |

- *mig_subsample_id* : Sub-sample used in indices construction (numeric).[4]

| id | Value | Desc. |
|---|---|---|
| 1 | 1 | excluding first-gen migrants |
| 2 | 2 | excluding all migrants |
| 3 | 3 | including all migrants |

- *country* : Name of top-level country (descriptive).

| id | Value | Desc. |
|---|---|---|
| *n* | *country name* | *country name* |

- *nuts_id* : Regional identifier (NUTS code) (descriptive).

| id | Value | Desc. |
|---|---|---|
| *n* | *NUTS code* | *NUTS code* |

- *nuts_level* : Level of regional identifier (NUTS code) (numeric).

[3] "All migrants" refers to those individuals with a migration background, which, in addition to first generation migrants, includes both second generation migrants, and later generation migrants who are not citizens of the country in which they reside.

[4] See previous footnote.

| id | Value | Desc. |
|---|---|---|
| 1 | 0 | NUTS 0 |
| 2 | 1 | NUTS 1 |
| 3 | 2 | NUTS 2 |

- *N* : Unweighted number of observations used to construct estimates (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *N* |

- *wN* : Weighted number of observations used to construct estimates (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *N* |

- *cohort* : Birth cohort (numeric).

| id | Value | Desc. |
|---|---|---|
| 1 | 1 | [1940,1959] |
| 2 | 2 | [1960,1979] |
| 3 | 3 | [1980,1999] |

- *b* : Intergenerational persistence (slope coefficient) for members of the cohort and their parents' (harmonized) years of education (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *persistence* |

- *beta* : Standardized intergenerational persistence (slope coefficient multiplied by the ratio of standard deviations of parents' and children's years of education) for members of the cohort and their parents' (harmonized) years of education (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *standardized persistence* |

- *educ* : Average (harmonized) years of education for the cohort (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *average education* |

- *educ_parents* : Average (harmonized) years of education for the cohort's parents (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *parents' average education* |

- *coefvar_educ* : Educational inequality for the cohort (coefficient of variation in harmonized years of education) (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *educational inequality* |

- *coefvar_pareduc* : Educational inequality for the cohort's parents (coefficient of variation in parents' harmonized years of education) (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *parents' educational inequality* |

## 2.2 *Europe-IGM-Atlas-Annualized* – Annualized Measures

- *mig_subsample* : Sub-sample used in indices construction (descriptive).[5]

| id | Value | Desc. |
|---|---|---|
| 1 | excl. all migrants | excluding all migrants |
| 2 | excl. first-gen migrants | excluding first-generation migrants |
| 3 | incl. migrants | including all migrants |

- *mig_subsample_id* : Sub-sample used in indices construction (numeric).[6]

| id | Value | Desc. |
|---|---|---|
| 1 | 1 | excluding first-gen migrants |
| 2 | 2 | excluding all migrants |
| 3 | 3 | including all migrants |

- *country* : Name of top-level country (descriptive).

| id | Value | Desc. |
|---|---|---|
| *n* | *country name* | *country name* |

- *nuts_id* : Regional identifier (NUTS code) (descriptive).

| id | Value | Desc. |
|---|---|---|
| *n* | *NUTS code* | *NUTS code* |

- *nuts_level* : Level of regional identifier (NUTS code) (numeric).

[5]"All migrants" refers to those individuals with a migration background, which, in addition to first generation migrants, includes both second generation migrants, and later generation migrants who are not citizens of the country in which they reside.

[6]See previous footnote.

| id | Value | Desc. |
|---|---|---|
| 1 | 0 | NUTS 0 |
| 2 | 1 | NUTS 1 |
| 3 | 2 | NUTS 2 |

- *weight_type* : Identifies the version of the indices based on weighting profile used in their construction (numeric).

| id | Value | Desc. |
|---|---|---|
| 1 | 1 | Profile I |
| 2 | 2 | Profile II |
| 3 | 3 | Profile III |
| 4 | 4 | Profile IV |
| 5 | 5 | Profile V |

- *year* : Year (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *year* |

- *b* : Effective intergenerational persistence (slope coefficient), a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective persistence* |

- *beta* : Effective standardized intergenerational persistence (slope coefficient multiplied by the ratio of standard deviations of parents' and children's years of education), a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective standardized persistence* |

- *educ* : Effective average (harmonized) years of education, a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective average education* |

- *educ_parents* : Effective average (harmonized) years of parents' education, a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective parents' average education* |

- *coefvar_educ* : Effective educational inequality (coefficient of variation in harmonized years of education), a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective educational inequality* |

- *coefvar_pareduc* : Effective parents' educational inequality (coefficient of variation in parents' harmonized years of education), a composite measure constructed according to the weighting profile indicated in *weight_type* (numeric).

| id | Value | Desc. |
|---|---|---|
| *n* | *N* | *effective parents' educational inequality* |

# 3 Constructing the Atlas

The information in the following is summarized from the *Methods* section of McNamara et al. (2026), with expose for clarity. Please see the original article for further discussion.

## 3.1 Data Sources

As the basis of our measures of intergenerational educational mobility, we rely on 11 full waves of the European Social Survey (ESS) 2002–2023, a representative cross-national survey. In addition to socio-demographic characteristics of the respondent (i.e., gender, region of residence, citizenship, country of birth, etc.), it records information about the respondent's level of education, and country-specific qualifications, as well as information about their parents. The ESS also includes survey weights to ensure representativeness of the weighted sample. We normalize these so they are consistent across waves, and perform weighted regression when constructing our mobility measures.

As the basis of the weighting profiles used in the computation of the annualized panel of intergenerational educational mobility indices, we rely on three additional sources of data. First, per-capita labor profiles over the life cycle come from Mason et al. (2022), which we use to construct cohort-year *labor contribution profiles* (Profile I). Second, innovation life-cycle profiles for all patenting activity and highly cited patenting activity come from Bell et al. (2016), which we use to construct cohort-year *innovation contribution profiles* (Profiles II and III, respectively). Third, we rely on the Luxembourg Income Study (LIS) database, accessed via LISSY, to construct cohort-year-country *labor contribution profiles* and *total individual income contribution profiles* (Profiles IV and V, respectively).

## 3.2 Sample Selection

When constructing our measures of intergenerational educational mobility, we restrict the ESS sample to respondents who were at least 22 years of age when the survey was conducted, and for whom information about gender, education, and parental education, etc. is not missing. We then define three sub-samples, based on the operationalization of our definition of migrants. The first (and least restrictive) sub-sample includes all respondents, independent of their migration background. The second sub-sample excludes first-generation migrants, in particular, which we define as those individuals who were not born in their country of residence. Finally, the third sub-sample is the most restrictive, and excludes all individuals with a "migration background" where, in addition to first generation migrants, a migration background includes both second generation migrants, and later generation migrants who are not citizens of their country of residence. We estimate the cohort-level measures of intergenerational educational mobility, average education, and educational inequality separately for each of these three sub-samples, and report both the weighted and unweighted number of observations used to estimate each index value in *Europe-IGM-Atlas-Cohorts* (variables *N* and *wN*, respectively).

## 3.3 Identifying Regions

The goal of the EUROPE-IGM-ATLAS is to provide disaggregated measures of intergenerational educational mobility across European regions. Thus, in addition to national level estimates, we compute estimates at the sub-national level using the region of residence of the respondent.

In the ESS, the decision of how to report regional information is left to the individual countries participating in the survey. First, while most report this information using the NUTS (Nomenclature of Territorial Units for Statistics) classification system, some countries report this information using country-specific administrative codes. Our definition of regions is thus augmented by the addition of countries not covered by the NUTS classification system, such as Albania, Kosovo, Montenegro, North Macedonia, Serbia, Türkiye, and Ukraine. Second, even for those using the NUTS classification system, the hierarchical level at which this information is recorded varies. Some countries report region of residence at the NUTS 1 level, some at the NUTS 2 level, and others at the NUTS 3 level. Third, there have been temporal changes in regional administrative boundaries over the last two decades in which the survey has been conducted, though this affects smaller spatial scales in particular, and is less of an issue when it comes to larger macro-regions.

To resolve the potential problem of cross-national and cross-temporal inconsistency in how we measure an individual's region of residence, we harmonize region prior to computing our estimates. In the EUROPE-IGM-ATLAS, we provide our indices at three hierarchically nested levels: NUTS 0, NUTS 1, and NUTS 2 (or equivalent). The first step for countries with time-inconsistent boundaries is to simply aggregate upward if responses are coded at the NUTS 3 level, given the higher level boundaries are more stable. Similarly, for regions in which the NUTS 2 level boundaries also changed substantially between waves, we aggregate further to NUTS 1 to preserve temporal consistency. In general, the 2016 version of NUTS serves as the basis of our harmonized definition of region, though with a notable exception for Poland. The sub-national measures for Poland are included based on the 2008 version of NUTS, as top-level regional divisions occurred in later versions.

## 3.4 Harmonizing Education

Between countries, the number of years of schooling needed to achieve a given qualification (e.g., the general high school diploma) varies. Further, some countries consider pre-school, kindergarten, etc., to be part of the formal education system, while others do not. To account for this variation between otherwise similar qualifications in terms of level, we use a modified ISCED scale to generate a harmonized measure of years of schooling. We use the ESS-ISCED measure as the basis of this measure, and use country-specific information about own and parents' education recorded in the ESS to harmonize qualifications from earlier waves of the survey, as well as for the non-EU countries. See the *Methods* section of McNamara et al. (2026) for further discussion, as well as cross-walk tables.

## 3.5 Estimating Intergenerational Mobility

To measure intergenerational educational mobility, we estimate *persistence* in the relationship between the levels of education of children (who are themselves adults when surveyed) and their parents. We do so by regressing the (harmonized) years of education of individual survey respondents on those of their highest educated parent[7], while separately controlling for both self-reported gender and survey year fixed effects. We divide the sub-samples into birth cohorts, and repeat this procedure separately for each region-cohort pair.

[7]If information for one parent is missing, we defer to the (harmonized) years of education reported for the available parent.

The slope coefficient from this procedure is our first measure of intergenerational educational mobility, which we refer to as *persistence*. The higher this measure, the stronger the relationship between the years of education of parents and their children for those born to that cohort-region pair, suggesting that parents' education more strongly determined the amount of education their child accrued.

An important consideration, however, is that there may be distributional differences between the generations. That is, if average levels of education rise over time, policy changes see changes to universal secondary education, or there are expansions to the tertiary education sector, etc., the same "level" of education for a parent and child may represent very different positions in a distributional sense. To account for this, in a second measure we multiply the slope coefficient *persistence* by the ratio of standard deviations for parents' and children's (harmonized) years of education. We refer to this second measure as *standardized persistence*. Importantly, though these two measures determine the *strength* of the relationship in absolute terms, they are not directional and are origin-independent; capturing both upward and downward movements. For example, a low value for a given region-cohort pair suggests that for that cohort, on average, their parents' level of education only weakly determined the level of education they in turn accrued, and that intergenerational mobility was thus high. But in relative terms, individuals could have accrued more or less than their highest educated parent.

For this reason, in addition to the measures of *persistence* and *standardized persistence*, for each region-cohort pair we additionally report measures of average (harmonized) years of education for both the survey respondents and their parents, as well as educational inequality (i.e., the coefficient of variation) for both the survey respondents and their parents. These measures provide contextual information about the relative shape of the educational distribution applicable to said region-cohort pair.

## 3.6 Building a Panel

The basis of the *EUROPE-IGM-ATLAS* is the *Europe-IGM-Atlas-Cohorts* dataset, which provides the cohort-level measures described thus far. To create the annualized panel of *effective* measures reported in the *Europe-IGM-Atlas-Annualized* dataset, we construct weighted composites of the cohort-level measures. We do this by weighting the indicator value of a given cohort-region pair by the expected contribution of cohort members to the economy in a given year. The composite *effective* measure for that year is then a weighted sum across cohorts within said region.

### 3.6.1 Weighting Profiles

The five weighting profiles, indexed using the variable *weight_type*, are as follows:

- *Profile I* : Based on per-capita labor profiles over the life cycle, from Mason et al. (2022), we construct cohort-year *labor contribution profiles*. The weight ascribed to a birth cohort in a given year consists of their expected relative labor supply (given their age) over the total labor supply in said year. Importantly, we assume no variation in cohort-year contribution profiles between countries.

- *Profile II* : Based on innovation life-cycle profiles for all patenting activity, from Bell et al. (2016), we construct cohort-year *innovation contribution profiles*. The weight ascribed to a birth cohort in a given year consists of their respective expected relative contribution to total patenting activity in said year. Importantly, we assume no variation in cohort-year contribution profiles between countries.

- *Profile III* : Based on innovation life-cycle profiles for highly cited patenting activity, from Bell et al. (2016), we construct cohort-year *innovation contribution profiles*. The weight ascribed to a birth cohort in a given year consists of their respective expected relative contribution to highly cited patenting activity in said year. Importantly, we assume no variation in cohort-year contribution profiles between countries.

- *Profile IV* : Based on the Luxembourg Income Study (LIS) database, accessed via LISSY, we construct cohort-year-country *labor contribution profiles*. The weight ascribed to a birth cohort in a given year consists of their actual relative employment over total employment in said year. Importantly, though contribution profiles are computed using actual data for each cohort-year-country triplet, we assume no variation across regions.

- *Profile V* : Based on the Luxembourg Income Study (LIS) database, accessed via LISSY, we construct cohort-year-country *total individual income contribution profiles*. The weight ascribed to a birth cohort in a given year consists of their actual relative total individual income over actual collective total individual income in said year. Importantly, though contribution profiles are computed using actual data for each cohort-year-country triplet, we assume no variation across regions.

See the *Methods* section of McNamara et al. (2026) for further details. The weighting profiles themselves are also presented graphically in the paper's Supplementary Information.

### 3.6.2 Creating Weighted Composites

As described in Section 3.6.1, our weighting profiles can be grouped into two categories: (1) for Profiles I-III, we assume no variation in cohort-year contribution profiles between countries, (2) for Profiles IV and V, we compute participation using actual data for each cohort-year-country triplet, though assume no variation across regions.

For the first group, the structure of our annualized *effective* measures can be described as follows:

$$\textit{effective measure}_{rt} = \sum_{c=1}^{C} \textit{weight}_{ct} \times \textit{cohort-level measure}_{cr} \tag{1}$$

where the index value of the *effective* measure in a given region $r$ in year $t$ is the weighted sum of each cohort-level measure multiplied by the cohort-participation weight applicable to the cohort $c$ for said year $t$.

For the second group, the structure of our annualized *effective* measures additionally takes into account this country-level variation in the weighting component, i.e.:

$$\textit{effective measure}_{rt} = \sum_{c=1}^{C} \textit{weight}_{nct} \times \textit{cohort-level measure}_{cr} \tag{2}$$

where the index value of the *effective* measure in a given region $r$ in year $t$ is the weighted sum of each cohort-level measure multiplied by the cohort-participation weight applicable to the cohort $c$ in country $n$ for said year $t$. See the *Methods* section of McNamara et al. (2026) for further details.

# Bibliography

Bell, A., Chetty, R., Jaravel, X., Petkova, N. and Van Reenen, J. (2016). The lifecycle of inventors. *SSRN Electronic Journal*. Link: `https://www.atarikafa.com/wp-content/uploads/2016/07/2016_06_14_patents.pdf`.

Mason, A, Lee, R. and members of the NTA Network (2022). Six ways population change will affect the global economy. *Population and Development Review*, **48**(1), pp.51–73. DOI: `https://doi.org/10.1111/padr.12469`.

McNamara, S., Neidhöfer, G., Lehnert, P. (2026). Intergenerational Mobility Fosters Innovation in Europe. *Nature*. DOI: `https://www.nature.com/articles/s41586-026-10736-9`.