Georgios Aristotelous

Database: Migration estimates

Deliverable 6.4

History of changes

Version	Date	Changes
1.0	23 June 2022	Issued for Consortium Review
1.1	29 June 2022	First version submitted as official deliverable to the EC
1.2	7 August 2023	Public version submitted to the open data repository (Zenodo)

Suggested citation

Aristotelous G (2022) Database: Migration estimates. QuantMig Project Deliverable D6.4. Southampton: University of Southampton.

Dissemination level

PU Public

Acknowledgments

Cover photo: iStockphoto.com/Guenter Guni

Database:
Migration estimates[*]

QuantMig Deliverable 6.4

Georgios Aristotelous [†]

This document is a 'read me' type of cover note for the files that form the database of migration estimates. Note that deliverable 6.4 (D6.4) is the database of migration estimates, restricted to the consortium and the European Commission, and not the present document per se. That is to say, that the present document only serves as a general manual for the database and not as a public document.

The database consists of three sets of migration estimates. The first set of estimates are disaggregated by origin, destination and time, a breakdown which we denote as ODT. The second and third sets of estimates are again disaggregated by origin, destination and time, but they are additionally disaggregated by other factors. The second set is further disaggregated by age and sex while the third set by birth region. We respectively denote these breakdowns as ODAST and ODBT. The model that produces the ODT estimates is described in detail in Aristotelous et al. (2022) while the methodology followed to further disaggregate these estimates and produce the ODAST and ODBT estimates is described in Wiśniowski et al. (2016). The sourcing and cleaning of the migration flow data used in the modelling process are described in Aristotelous et al. (2020).

The three sets of estimates, ODT, ODAST and ODBT, are stored in three `csv` files respectively named `flows_ODT_long.csv`, `flows_ODAST_long.csv` and `flows_ODBT_long.csv`. These three files form the migration flow database. The files are all in long format and

[†]University of Southampton, UK. Email: G.Aristotelous@soton.ac.uk

have as variables (i.e. columns) the disagreggation factors and the migration estimates, so that each row provides the estimates for each combination of the factor levels. For example, in the `flows_ODT_long.csv` file, the second row provides the migration estimates for the flow where the origin is Belgium, the destination is Austria and the year is 2009. Similarly, in the `flows_ODAST_long.csv` file, the second row provides the migration estimates for the flow where the origin is Belgium, the destination is Austria, the age group is less than 5 years old, the sex group is female and the year is 2009.

In all files we provide three migration estimates, namely the median estimate along with the 2.5% lower and the 97.5% upper percentile estimates, respectively named `flow_50%`, `flow_2.5%` and `flow_97.5%`. These three values are providing a central estimate (`flow_50%`), a lower bound estimate (`flow_2.5%`) and an upper bound estimate (`flow_97.5%`) of a given flow. Furthermore, the interval (`flow_2.5%`, `flow_97.5%`) is a 95% probability interval meaning that, according to the model, the true flow lies in (`flow_2.5%`, `flow_97.5%`) with 95% probability.

We now provide a description of the levels of the factors for each of the three files. For completeness and clarity we also summarize this information in Tables 1, 2 and 3. In all files time is in years (named in the files as `year`) and its levels are the years 2009 to 2019, accordingly named 2009, 2010, ..., 2019. In all files the destination factor is named `dest`.

For the ODT estimates (file `flows_ODT_long.csv`), the levels for origin and destination are the 32 EU+ countries, North Macedonia and 8 rest of the world (RW) regions, in total 41 levels. By EU+ we mean the 28 EU-member countries[1], and the 4 EFTA countries, namely Iceland, Liechtenstein, Norway and Switzerland. The 8 RW regions are Other Europe, North Africa, Sub-Saharan Africa, West Asia, East Asia, South-Southeast Asia, North America and Oceania and Latin America. The countries which comprise each of the 8 RW regions are provided in the file `RW_regions.xlsx` for completeness.

---

[1] For our considered time period, 2009-2019, the United Kingdom was an EU-member country and is therefore one of the 28 EU-member countries.

For the ODAST estimates (file `flows_ODAST_long.csv`), the levels for origin and destination are the 32 EU+ countries, North Macedonia and a single RW region, in total 34 levels. The difference compared to the ODT estimates is that in the ODAST estimates the 8 RW regions are combined into a single RW region (named `Rest World` in the file). This is because of data availability and more specifically because migration data disaggregated by age and sex are not available for specific RW regions but rather only for RW in general. Regarding age, the levels are 18 non-overlapping 5-year age groups, namely less than five years old, 5-9, 10-14, and so on up until 80-84 and finally the greater than or equal to 85 years old group. In the file these levels are respectively named `Y_LT5`, `Y5-9`, `Y10-14`, . . . ,`Y80-84`, and `Y_GE85`. As far as sex, the levels are female and male, respectively named `F` and `M` in the file.

Finally, for the ODBT estimates (file `flows_ODBT_long.csv`), the levels for origin and destination are the 32 EU+ countries and a single RW region, a total of 33 levels. Same as for the ODAST estimates, and for the same reasons, the 8 RW regions are combined into a single RW region (named `Rest World` in the file). Note also that North Macedonia is included into the RW, unlike the ODT and the ODAST files. Regarding birth region (named `birth_region` in the file), the levels are EU-born and non-EU-born, and are respectively named `EU` and `non-EU` in the file.

We note that in all files flow estimates of same-origin-destination entries are recorded as NA (Not Applicable) since we do not consider within-country migration. We also note that the estimates are rounded to one demimal point for clarity.

Table 1: Disaggregation factors and their levels for the ODT estimates (file `flows_ODT_long.csv`).

factor	levels
origin and destination	32 EU+ countries, North Macedonia and 8 RW regions
time	years 2009 to 2019

Table 2: Disaggregation factors and their levels for the ODAST estimates (file `flows_ODAST_long.csv`).

factor	levels
origin and destination	32 EU+ countries, North Macedonia and 1 RW region
age	18 non-overlapping 5-year age groups
sex	female and male
time	years 2009 to 2019

Table 3: Disaggregation factors and their levels for the ODBT estimates (file `flows_ODBT_long.csv`).

factor	levels
origin and destination	32 EU+ countries and 1 RW region
birth region	EU-born and non-EU-born
time	years 2009 to 2019

References

Aristotelous, G., Smith, P. W. F., and Bijak, J. (2020). Database: Flows, stocks and quality for modelling. QuantMig Deliverable D5.3, University of Southampton, Southampton.

Aristotelous, G., Smith, P. W. F., and Bijak, J. (2022). Technical report: Estimation methodology. QuantMig Deliverable D6.3, University of Southampton, Southampton.

Wiśniowski, A., Forster, J. J., Smith, P. W. F., Bijak, J., and Raymer, J. (2016). Integrated modelling of age and sex patterns of European migration. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):1007–1024.