

Estimation of undercounting of Eurostat migration data using metadata and bilateral migration flows reported by Eurostat

Maciej J. Dańko

15 September, 2021

Contents

1	Estimation of undercount using metadata	2
2	Estimation of undercount using bilateral migration flow from Eurostat.	4
3	Combination of metadata and data estimated undercount	7
4	References	9
5	Supplementary figures	9

1 Estimation of undercount using metadata

The classification of undercounting cannot be efficiently done using metadata, because the metadata is very limited for both immigration (registration) and emigration (de-registration) data sources. The collected metadata is based on 2003 and 2015 Eurostat reports, Your Europe (2021), and government websites (see the **References 4** for detailed list and QuantMig Deliverable 6.2 for the summary of some of the metadata used)

The immigration metadata is shown in **Table 1** and contains only two variables:

- obligation of registration (“*obligation*”) - “Yes” / “No” / “Unknown” (Eurostat 2015 report)
- time limit for registration (“*time limit*”) (Eurostat 2015 report)

Most countries has obligatory registration (low undercounting), however there is no time limit for Spain and Italy (Medium undercounting). There was no metadata for Island, however, here we assume low undercounting, which characterizes all Nordic Countries.

Table 1: Immigration metadata (meaning of the columns is explained in the text)

iso2	country	obligation	time limit	comment	score
AT	Austria	Yes	3 days		low
BE	Belgium	Yes	90 days		low
BG	Bulgaria	Yes	at arrival		low
CH	Switzerland	Yes	14 days		low
CY	Cyprus	Yes	7 days		low
CZ	Czechia	Yes	90 days		low
DE	Germany	Yes	3 months		low
DK	Denmark	Yes	5 days	Nordic	low
EE	Estonia	Yes	1 month		low
EL	Greece	Yes	90 days		low
ES	Spain	Yes	No limit		medium
FI	Finland	Yes	7 days	Nordic	low
FR	France	No	-		high
HR	Croatia	Yes	2 days		low
HU	Hungary	Yes	90 days		low
IE	Ireland	No	-		high
IS	Iceland	Unknown	Unknown	Nordic	low
IT	Italy	Yes	No limit		medium
LI	Liechtenstein	Yes	not specified		low
LT	Lithuania	Yes	7 days		low
LU	Luxemburg	Yes	8 days		low
LV	Latvia	Yes	90 days		low
MT	Malta	Yes	1 month		low
NL	Netherlands	Yes	5 days		low
NO	Norway	Yes	8 days	Nordic	low
PL	Poland	Yes	4 days		low
PT	Portugal	No	-		high
RO	Romania	Yes	2 days		low
SE	Sweden	Yes	7 days	Nordic	low
SI	Slovenia	Yes	8 days		low
SK	Slovakia	Yes	5 days		low
UK	United Kingdom	No	-		low

The emigration metadata is shown in **Table 2** and contains the following variables:

- obligation to register (“*obligation*”) - “Yes” / “No” / “Unknown” (Eurostat 2015 report)
- obligation to register for third country nationals (“*obl. 3rdcn*”) - “Yes” / “No” / “Unknown” (Multiple sources)
- monitoring of third country nationals (“*monit. 3rdcn*”) - “Yes” / “No” / “Unknown” (Multiple sources)

- adjustment of under-/over- counting problems by the statistical office (“*SO corr.*”) - “yes” / “Unknown” (Eurostat 2003 report)

To calculate “*score num*” we use weighted sum of all “Yes” answers (or date for “*SO corr.*”) and divide them by weighted number of answers “Yes” or “No”. “*obligation*” and “*obl. 3rdcn*” have weights of 1, but “*monit. 3rdcn*” and “*SO corr.*” has a weight of 0.5, because we assume that these variables are both of lower quality and/or have lower impact on the undercounting. The resulting value is then subtracted from 1 so that the low values indicate low undercounting and high values indicate high undercounting. For example, while Slovakia has general obligation for de-registration ($1 \cdot 1$), there is no obligation for de-registration of nationals ($0 \cdot 1$) even if there is monitoring of nationals ($1 \cdot 0.5$). There is also unknown status off statistical adjustment of undercounting by Statistical office of Slovakian Republic ($0 \cdot 0.5$). The “*score num*” for Slovakia is thus calculated as $1 - (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0.5 + 0 \cdot 1) \div (1 \cdot 1 + 1 \cdot 1 + 1 \cdot 0.5 + 0 \cdot 0.5) = 1 - 1.5 \div 2.5 = 0.4$.

To categorize the calculated “*score num*” into the “*score*” we use simple thresholds:

score num threshold	score
0.00 - 0.29	low
0.30 - 0.59	medium
0.60 - 1.00	high

All Nordic Countries were classified as **low** independently of obtained score.

Table 2: Emigration metadata (meaning of the columns is explained in the text)

iso2	country	obligation	obl. 3rdcn	monit. 3rdcn	SO corr.	comment	score num	score
AT	Austria	Yes	Yes	No	Unknown		0.200	low
BE	Belgium	No	Yes	Yes	Yes		0.333	medium
BG	Bulgaria	No	No	Yes	Unknown		0.800	high
CH	Switzerland	Yes	Unknown	Unknown	Yes		0.000	low
CY	Cyprus	No	No	Yes	Unknown		0.800	high
CZ	Czechia	No	Yes	No	Unknown		0.600	high
DE	Germany	Yes	No	No	Yes		0.500	medium
DK	Denmark	Yes	Unknown	Unknown	Unknown	Nordic	0.000	low
EE	Estonia	Yes	Yes	No	Yes		0.167	low
EL	Greece	No	No	No	Unknown		1.000	high
ES	Spain	Yes	No	No	Unknown		0.600	high
FI	Finland	Yes	Yes	Yes	Yes	Nordic	0.000	low
FR	France	No	No	No	Unknown		1.000	high
HR	Croatia	Yes	No	No	Unknown		0.600	high
HU	Hungary	Yes	No	No	Unknown		0.600	high
IE	Ireland	No	Unknown	Unknown	Unknown		1.000	high
IS	Iceland	Unknown	Unknown	Unknown	Unknown	Nordic	0.000	low
IT	Italy	Yes	No	No	Yes		0.500	medium
LI	Liechtenstein	Yes	Unknown	Unknown	Unknown		0.000	low
LT	Lithuania	Yes	Yes	Yes	Unknown		0.000	low
LU	Luxemburg	Yes	Yes	Yes	Yes		0.000	low
LV	Latvia	Yes	Yes	Yes	Unknown		0.000	low
MT	Malta	No	No	No	Unknown		1.000	high
NL	Netherlands	Yes	Yes	No	Yes		0.167	low
NO	Norway	Yes	Unknown	Unknown	Unknown	Nordic	0.000	low
PL	Poland	Yes	No	No	Unknown		0.600	high
PT	Portugal	No	Yes	No	Unknown		0.600	high
RO	Romania	No	No	Yes	Unknown		0.800	high
SE	Sweden	Yes	No	No	Unknown	Nordic	0.600	low
SI	Slovenia	Yes	Yes	Yes	Unknown		0.000	low
SK	Slovakia	Yes	No	Yes	Unknown		0.400	medium
UK	United Kingdom	No	Unknown	Unknown	Unknown		1.000	high

2 Estimation of undercount using bilateral migration flow from Eurostat.

The alternative source of information about undercounting problem can be the data itself. The general idea behind the data-driven undercounting score is to take flows from a country X to a group of good data quality countries reported by country X and compare it with the same flow reported by the group of good data quality countries. Because the duration of stay may differs among countries the flows reported by each country should be somehow corrected. Here, we use Raymer's 2013 estimates for duration of stay (**Table 3**) as the duration of stay correction parameter.

Formally, estimation of the undercounting ratio $U_{X,Y,y}^E$ of emigration data between country X and set of countries Y in year y , can be calculated as follows:

$$U_{X,Y,y}^E = \frac{\sum_c M(X_y \rightarrow Y_{c,y}, X_y) R_{X_y}}{\sum_c M(X_y \rightarrow Y_{c,y}, Y_{c,y}) R_{Y_{c,y}}}, \quad (1)$$

where $M(X_y \rightarrow Y_{c,y}, X_y)$ is the emigration flow from country X to country Y_c reported by country X in year y , $M(X_y \rightarrow Y_{c,y}, Y_{c,y})$ is the emigration flow from country X to country Y_c reported by country Y_c in year y , R_{X_y} is the Raymer's correction for duration of stay of country X in year y , and $R_{Y_{c,y}}$ is Raymer's correction for duration of stay of country Y_c reported in year y (**Table 3**). The group of good data quality countries (Y) includes all Nordic countries, Switzerland, Nederland, Belgium, Austria, Germany, and Lichtenstein.

Table 3: Raymer's parameters used to calculate correction of migration flows for duration of stay different than 12 months

duration of stay in months	Raymer's parameter
0	0.53
3	0.63
6	0.73
12	1.00
permanent	2.26

The undercounting ratio $U_{X,Y,y}^I$ of immigration data is calculated analogically:

$$U_{X,Y,y}^I = \frac{\sum_c M(X_y \leftarrow Y_{c,y}, X_y) R_{X_y}}{\sum_c M(X_y \leftarrow Y_{c,y}, Y_{c,y}) R_{Y_{c,y}}}, \quad (2)$$

where $M(X_y \leftarrow Y_{c,y}, X_y)$ is the immigration flow to country X from country Y_c reported by country X in year y , $M(X_y \leftarrow Y_{c,y}, Y_{c,y})$ is the migration flow to country X from country Y_c reported by country Y_c in year y ,

Because, bilateral migration for some years are missing a simple interpolation (spline) and extrapolation (weighted mean of last observations) for tails was used.

There are two main disadvantages of the constructed measure:

- It cannot exclude the effect of under-coverage, i.e., the index measures a combined effect of under-counting and coverage problems.
- It may not completely exclude the effect of the duration of stay that is different than 12 months, i.e., Raymer's estimates are taken from the model that does not include the most recent data. Furthermore, some duration seems to be badly reported in Eurostat metadata files.

The estimates of undercounting ratios for different countries can be found in supplementary figures (**Section 5**). The supplementary figures shows results for three different good register groups of countries as well as duration corrected and uncorrected measures. However, only results shown as blue solid lines (Bilaterally corrected Nordic + CH + NL + BE + AT + DE + LI) are used for the presented here classification.

The next step was calculation of 1998-2007 and 2008-2019 means of the undercounting ratios (**Figure 1**). These ratios were then classified according to the thresholds:

undercounting ratio	undercounting score
> 1.00	overcounting
0.750 - 1.000	very low
0.500 - 0.749	low
0.250 - 0.499	medium
0.125 - 0.249	high
< 0.125	very high

Values higher than 1 means overcounting, however they are likely a results of a lack of accuracy rather than true over-counting. The new data (starting 14.09.2021) assume that all overcounting are treated as **very low** records. This data will be used in next generations of models. Please notice that in supplementary figures (**Section 5**) **very low** and **low** classes are combined into a single class.

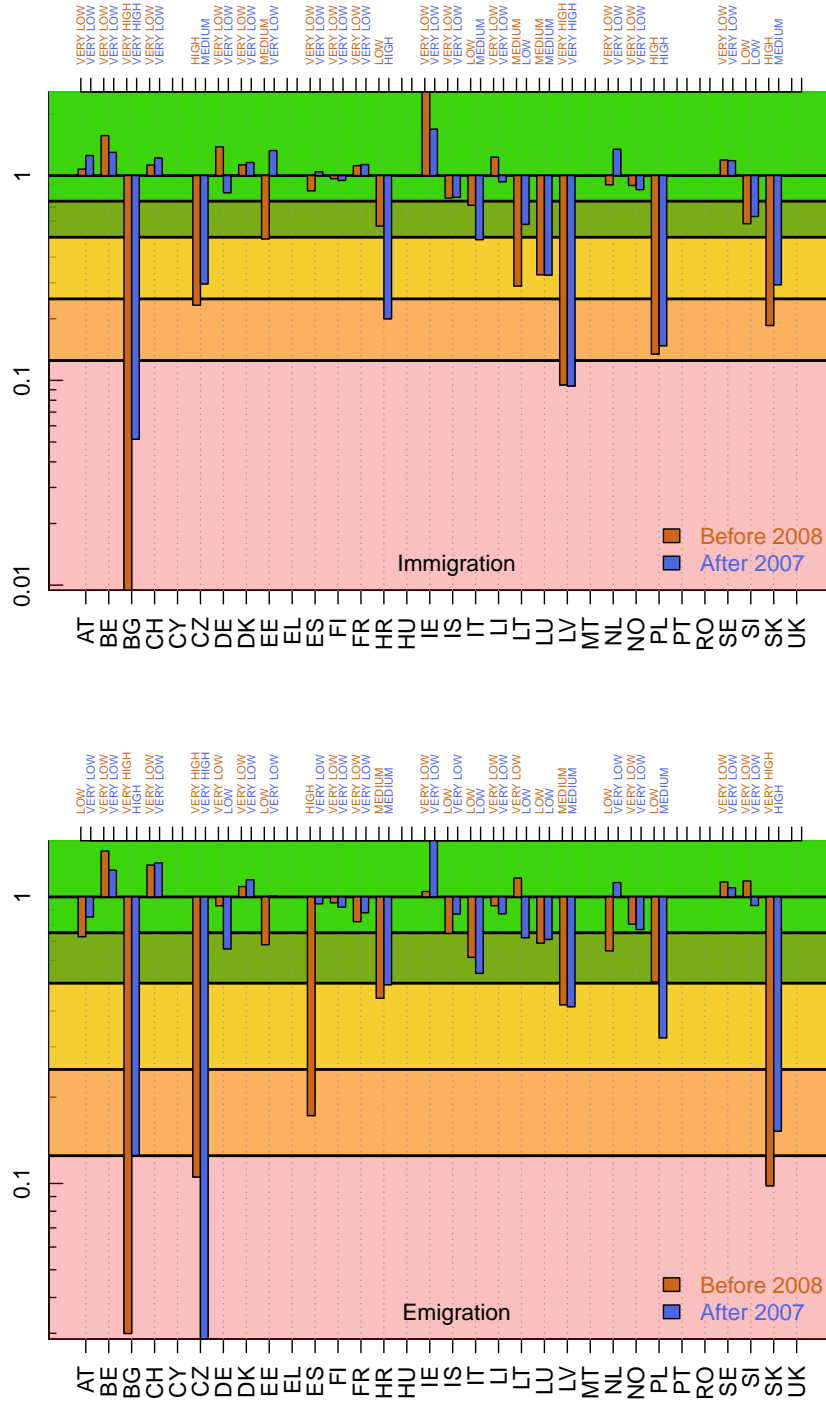


Figure 1: Classification of undercount. Upper panel - immigration, bottom panel - emigration

3 Combination of metadata and data estimated undercount

The calculation method of the metadata-data combined scores are presented in **Table 4** and **5**. The following thresholds were used to classify “*cbefore*” and “*cafter*” scores into “*cbs*” and “*cas*”:

score num threshold	score
0.00 - 0.24	low
0.25 - 0.59	medium
0.60 - 1.00	high

Table 4: Combined results of metadata and undercount index for immigration data. **IMEM** is undercounting assumed in the IMEM model (0 = low, 1 = high); **meta** is a score calculated using immigration metadata (0 = low, 0.5 = medium, 1=high) obtained in **Section 1** (**Table 1**); **before** and **after** are scores for 1998-2007 and 2008-2019 respectively obtained in **Section 2** (0 = very low, 0.25 = low, 0.5 = medium, 0.75 = high, 1 = very high); **cbefore** and **cafter** are metadata - model combined scores, $\text{cbefore} = (\text{meta} + 2 \cdot \text{before}) \div 3$, **cafter** is calculated analogically; **cbs** and **cas** are classified **cbefore** and **cafter** scores respectively. Some countries (Cyprus, Greece, Hungary, Malta, Portugal, Romania, and UK) have no bilateral data thus **before** and **after** cannot be calculated. In such cases **before** and **after** are replaced with **IMEM** score).

iso2	country	IMEM	meta	before	after	cbefore	cafter	cbs	cas
AT	Austria	0	0.0	0.00	0.00	0.00	0.00	low	low
BE	Belgium	0	0.0	0.00	0.00	0.00	0.00	low	low
BG	Bulgaria	1	0.0	1.00	1.00	0.67	0.67	high	high
CH	Switzerland	0	0.0	0.00	0.00	0.00	0.00	low	low
CY	Cyprus	0	0.0	NA	NA	0.00	0.00	low	low
CZ	Czechia	1	0.0	0.75	0.50	0.50	0.33	medium	medium
DE	Germany	0	0.0	0.00	0.00	0.00	0.00	low	low
DK	Denmark	0	0.0	0.00	0.00	0.00	0.00	low	low
EE	Estonia	1	0.0	0.50	0.00	0.33	0.00	medium	low
EL	Greece	1	0.0	NA	NA	0.67	0.67	high	high
ES	Spain	0	0.5	0.00	0.00	0.17	0.17	low	low
FI	Finland	0	0.0	0.00	0.00	0.00	0.00	low	low
FR	France	0	1.0	0.00	0.00	0.33	0.33	medium	medium
HR	Croatia	1	0.0	0.25	0.75	0.17	0.50	low	medium
HU	Hungary	1	0.0	NA	NA	0.67	0.67	high	high
IE	Ireland	0	1.0	0.00	0.00	0.33	0.33	medium	medium
IS	Iceland	0	0.0	0.00	0.00	0.00	0.00	low	low
IT	Italy	0	0.5	0.25	0.50	0.33	0.50	medium	medium
LI	Liechtenstein	1	0.0	0.00	0.00	0.00	0.00	low	low
LT	Lithuania	1	0.0	0.50	0.25	0.33	0.17	medium	low
LU	Luxemburg	0	0.0	0.50	0.50	0.33	0.33	medium	medium
LV	Latvia	1	0.0	1.00	1.00	0.67	0.67	high	high
MT	Malta	1	0.0	NA	NA	0.67	0.67	high	high
NL	Netherlands	0	0.0	0.00	0.00	0.00	0.00	low	low
NO	Norway	0	0.0	0.00	0.00	0.00	0.00	low	low
PL	Poland	1	0.0	0.75	0.75	0.50	0.50	medium	medium
PT	Portugal	1	1.0	NA	NA	1.00	1.00	high	high
RO	Romania	1	0.0	NA	NA	0.67	0.67	high	high
SE	Sweden	0	0.0	0.00	0.00	0.00	0.00	low	low
SI	Slovenia	1	0.0	0.25	0.25	0.17	0.17	low	low
SK	Slovakia	1	0.0	0.75	0.50	0.50	0.33	medium	medium
UK	United Kingdom	0	0.0	NA	NA	0.00	0.00	low	low

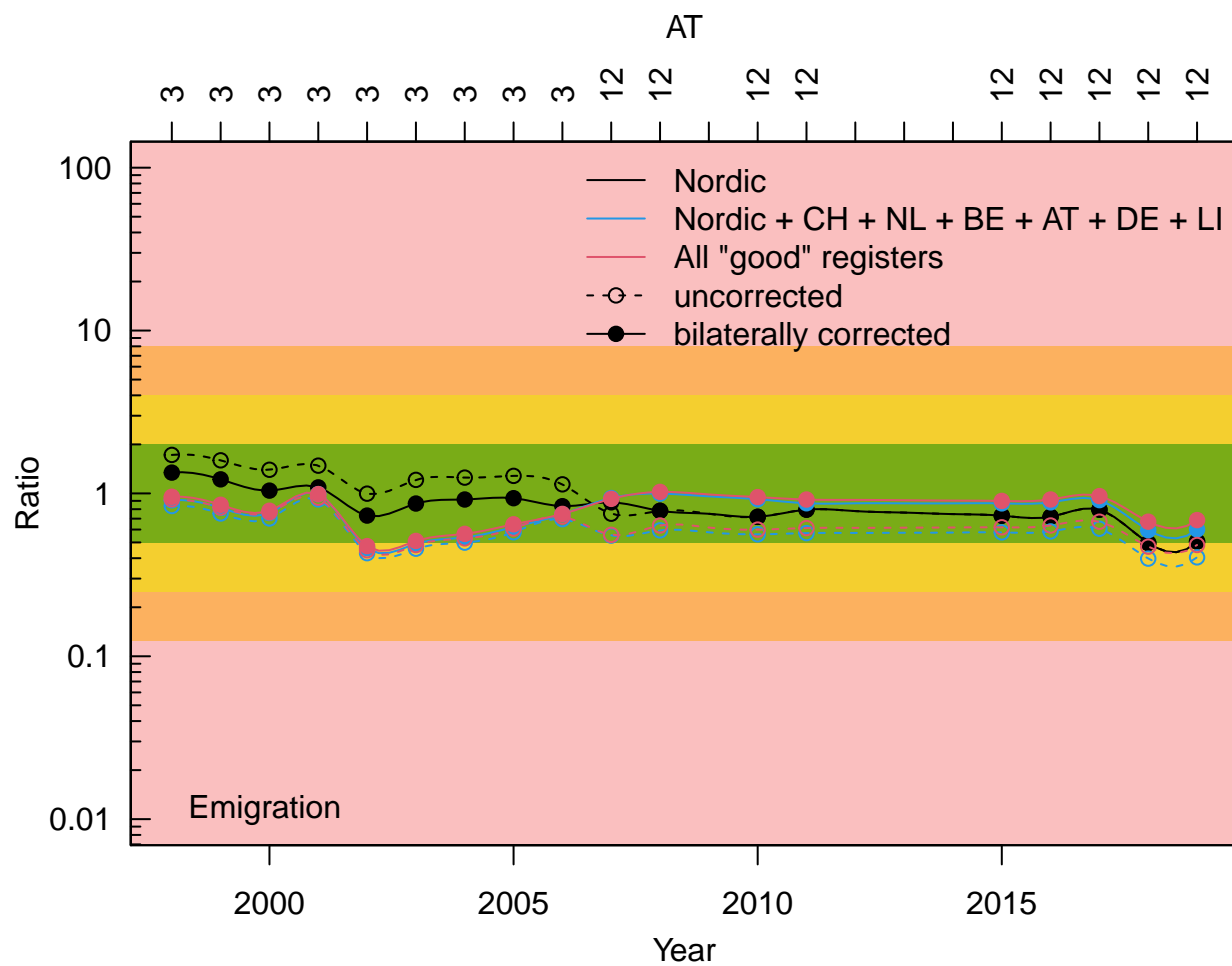
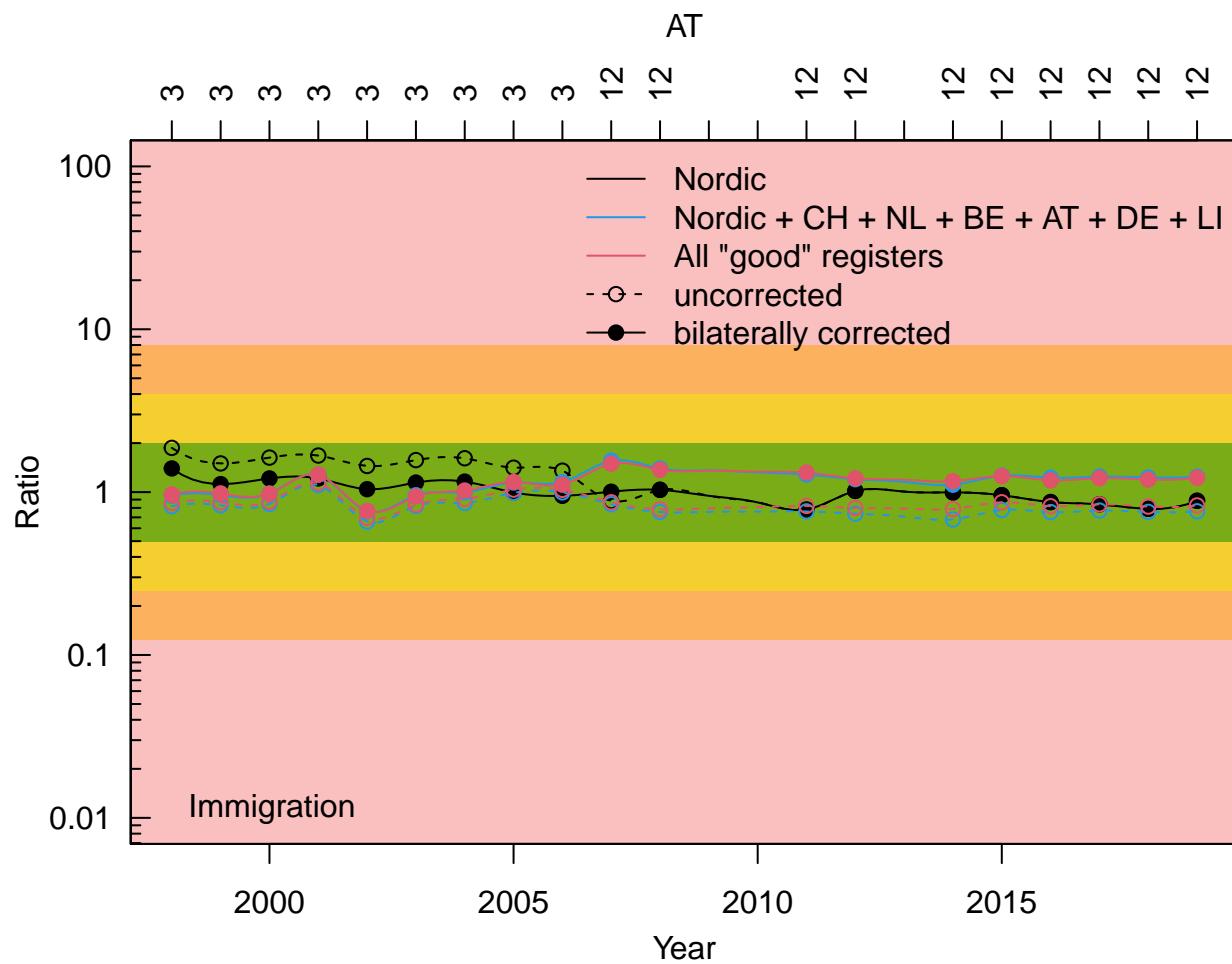
Table 5: Combined results of metadata and undercount index for emigration data. **IMEM** is undercounting assumed in the IMEM model (0 = low, 1 = high); **meta** is the score calculated using emigration metadata (**score num** of **Table 2**) obtained in **Section 1**; **before** and **after** are scores for 1998-2007 and 2008-2019 respectively obtained in **Section 2** (0 = very low, 0.25 = low, 0.5 = medium, 0.75 = high, 1 = very high); **cbefore** and **cafter** are metadata - model combined scores, **cbefore** = (**meta** + 2 · **before**) ÷ 3, **cafter** is calculated analogically; **cbs** and **cas** are classified **cbefore** and **cafter** scores respectively. Some countries (Cyprus, Greece, Hungary, Malta, Portugal, Romania, and UK) have no bilateral data thus **before** and **after** cannot be calculated. In such cases **before** and **after** are replaced with **IMEM** score).

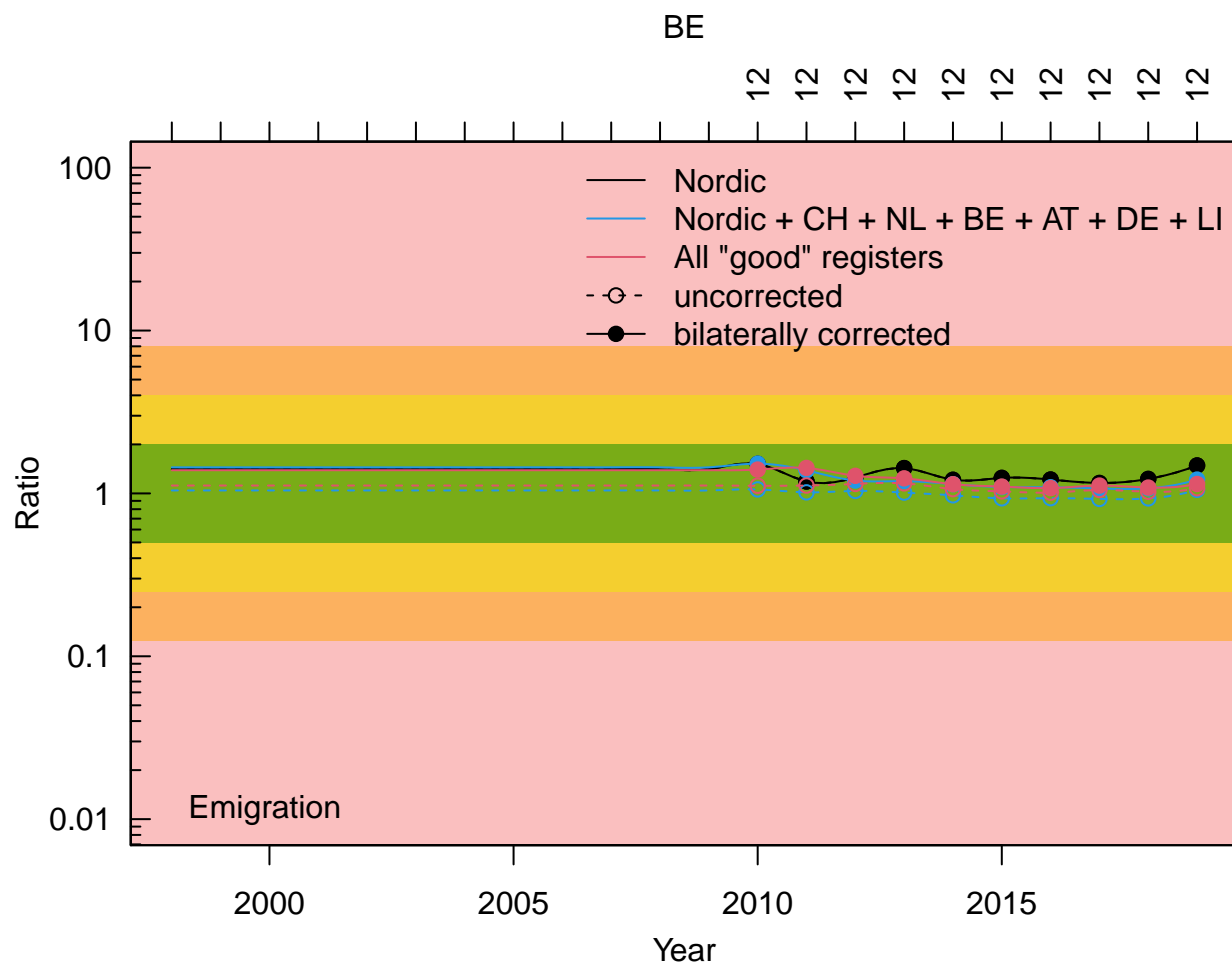
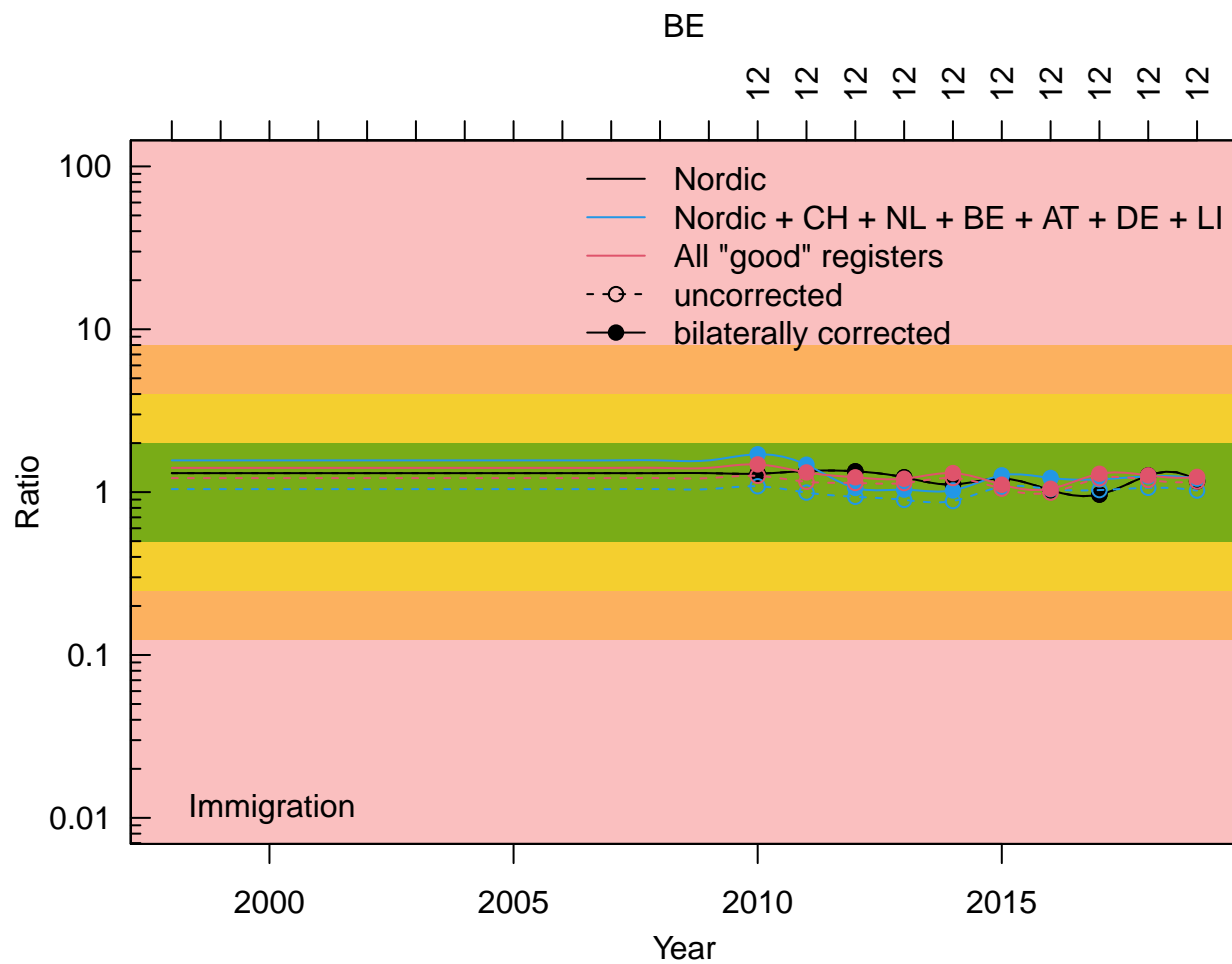
iso2	country	IMEM	meta	before	after	cbefore	cafter	cbs	cas
AT	Austria	0	0.200	0.25	0.00	0.23	0.07	low	low
BE	Belgium	0	0.333	0.00	0.00	0.11	0.11	low	low
BG	Bulgaria	1	0.800	1.00	0.75	0.93	0.77	high	high
CH	Switzerland	0	0.000	0.00	0.00	0.00	0.00	low	low
CY	Cyprus	0	0.800	NA	NA	0.27	0.27	medium	medium
CZ	Czechia	1	0.600	1.00	1.00	0.87	0.87	high	high
DE	Germany	0	0.500	0.00	0.25	0.17	0.33	low	medium
DK	Denmark	0	0.000	0.00	0.00	0.00	0.00	low	low
EE	Estonia	1	0.167	0.25	0.00	0.22	0.06	low	low
EL	Greece	1	1.000	NA	NA	1.00	1.00	high	high
ES	Spain	1	0.600	0.75	0.00	0.70	0.20	high	low
FI	Finland	0	0.000	0.00	0.00	0.00	0.00	low	low
FR	France	0	1.000	0.00	0.00	0.33	0.33	medium	medium
HR	Croatia	1	0.600	0.50	0.50	0.53	0.53	medium	medium
HU	Hungary	1	0.600	NA	NA	0.87	0.87	high	high
IE	Ireland	0	1.000	0.00	0.00	0.33	0.33	medium	medium
IS	Iceland	0	0.000	0.25	0.00	0.17	0.00	low	low
IT	Italy	0	0.500	0.25	0.25	0.33	0.33	medium	medium
LI	Liechtenstein	1	0.000	0.00	0.00	0.00	0.00	low	low
LT	Lithuania	1	0.000	0.00	0.25	0.00	0.17	low	low
LU	Luxemburg	0	0.000	0.25	0.25	0.17	0.17	low	low
LV	Latvia	1	0.000	0.50	0.50	0.33	0.33	medium	medium
MT	Malta	1	1.000	NA	NA	1.00	1.00	high	high
NL	Netherlands	0	0.167	0.25	0.00	0.22	0.06	low	low
NO	Norway	0	0.000	0.00	0.00	0.00	0.00	low	low
PL	Poland	1	0.600	0.25	0.50	0.37	0.53	medium	medium
PT	Portugal	1	0.600	NA	NA	0.87	0.87	high	high
RO	Romania	1	0.800	NA	NA	0.93	0.93	high	high
SE	Sweden	0	0.600	0.00	0.00	0.20	0.20	low	low
SI	Slovenia	1	0.000	0.00	0.00	0.00	0.00	low	low
SK	Slovakia	1	0.400	1.00	0.75	0.80	0.63	high	high
UK	United Kingdom	0	1.000	NA	NA	0.33	0.33	medium	medium

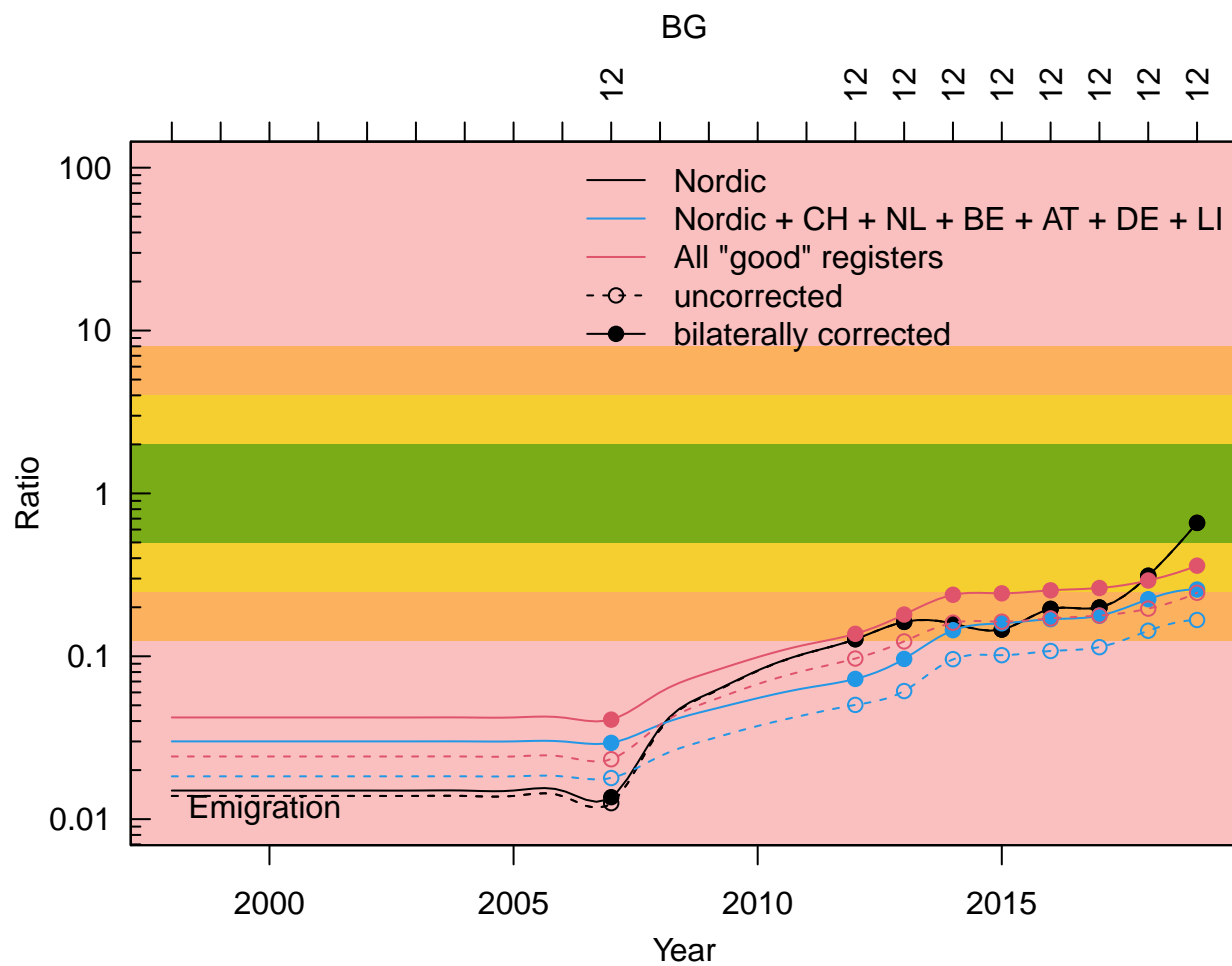
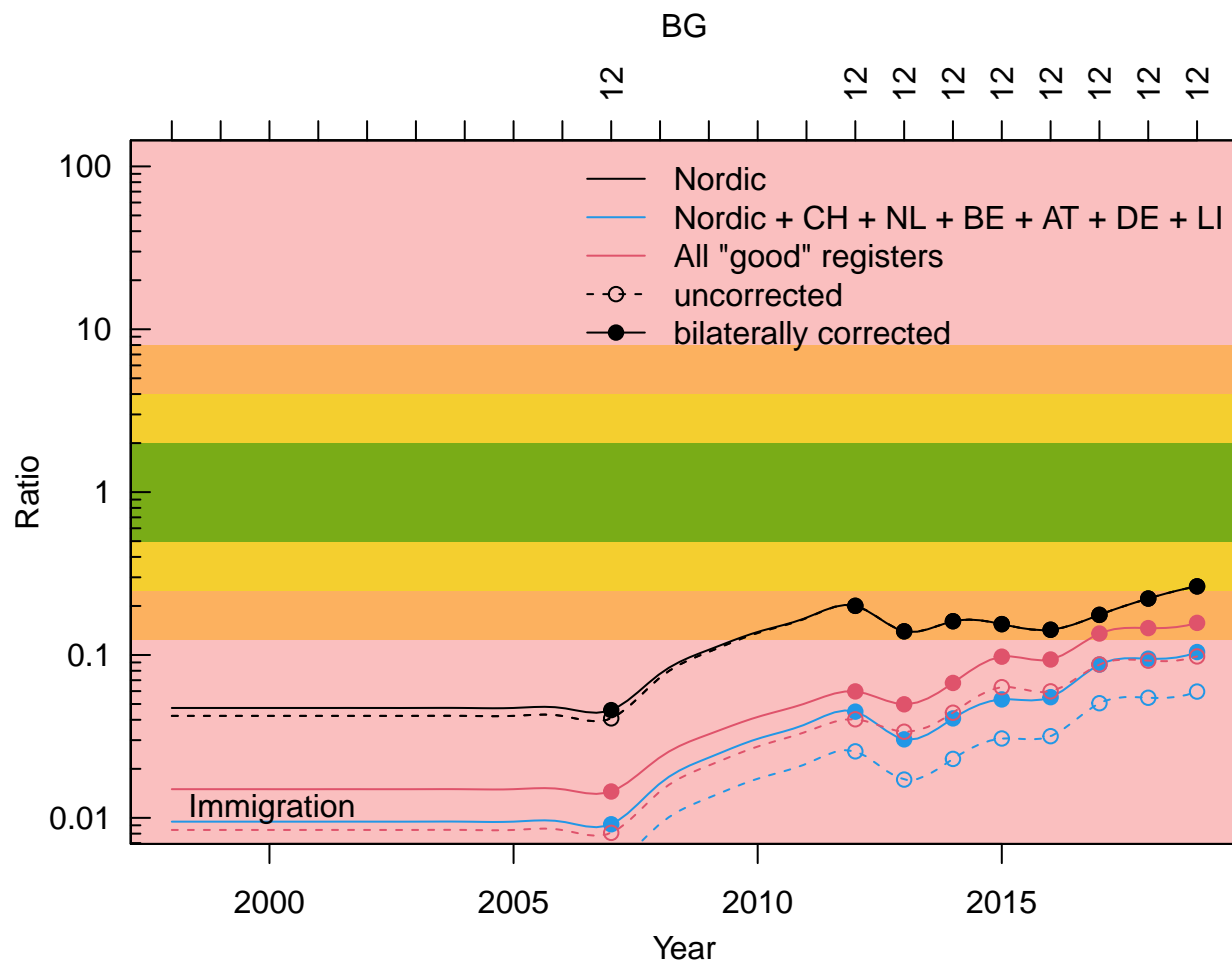
4 References

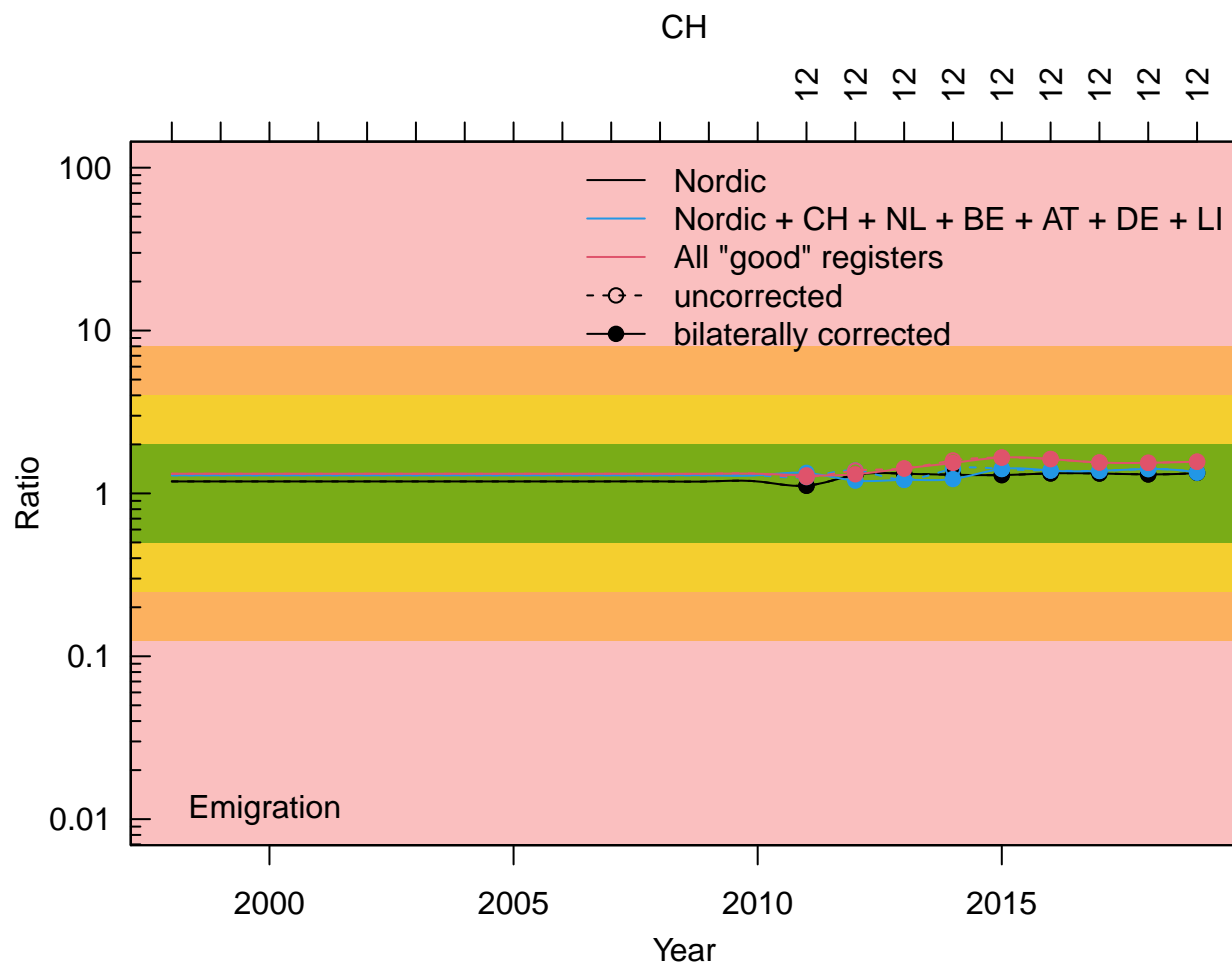
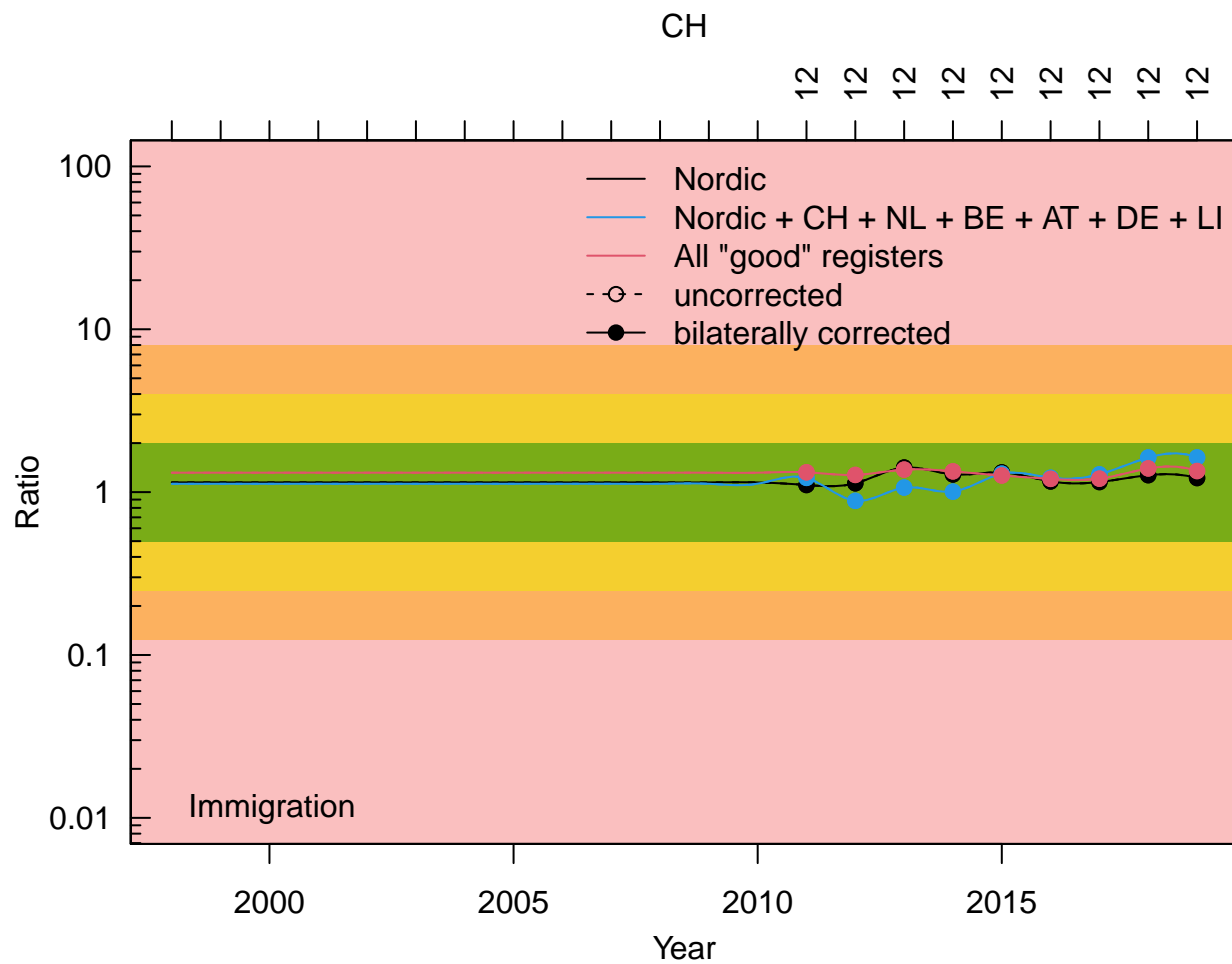
- Eurostat (2001) Documentation of Eurostat’s database on international migration: Central European Countries, Cyprus and Malta. Working Paper. <https://www.um.edu.mt/library/oar/bitstream/123456789/4860/1/Documentation%20of%20Eurostat%E2%80%99s%20database%20on%20international%20migration.pdf>
- Eurostat (2003) Demographic statistics: Definitions and methods of collection in 31 European Countries. ISSN 1725-065X ISBN 92-894-6051-2. <https://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-CC-03-005-EN.pdf>
- Eurostat (2015) Demographic statistics: a review of definitions and methods of collection in 44 European countries. Luxembourg: Eurostat. doi:10.2785/717072.
- Eurostat (2021a) https://ec.europa.eu/eurostat/cache/metadata/en/migr_immi_esms.htm. Accessed June 2021 27 Quality assessment of European migration data
- Eurostat (2021b) https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm. Accessed June 2021
- European Commission (2019) Legal Migration Fitness Check. https://ec.europa.eu/home-affairs/what-we-do/policies/legal-migration/fitness-check_en. Accessed June 2021
- Jarl Mooyaart, Maciej J. Dańko, Rafael Costa, and Michaël Boissonneault (2021) Quality assessment of European migration data. Deliverable 6.2 http://quantmig.eu/res/files/QuantMig_Deliverable%206.2%20vf.pdf
- Your Europe (2021) residence formalities. https://europa.eu/youreurope/citizens/residence/documents-formalities/reporting-presence/index_en.htm. Accessed June 2021

5 Supplementary figures

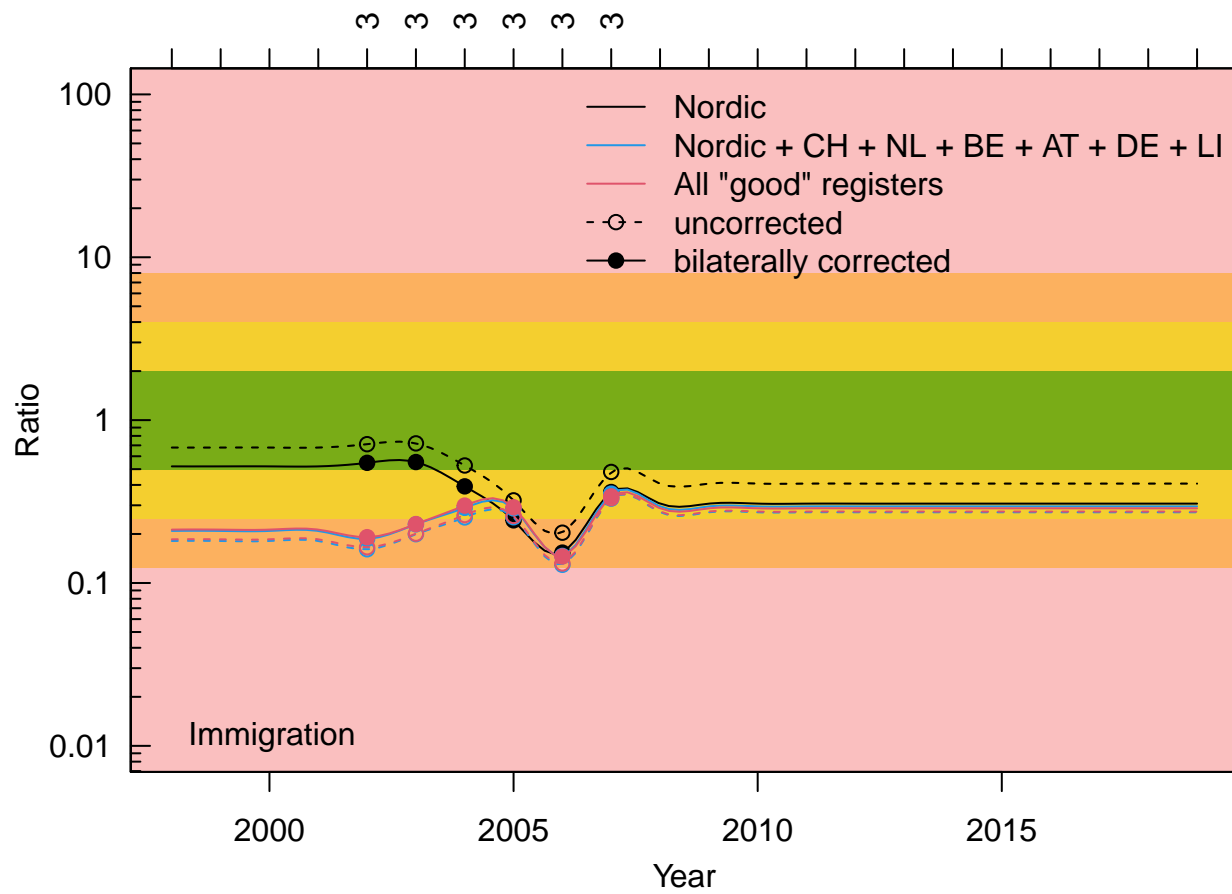








CZ



CZ

