Methodological Report

# New Models and Indices of Urban Network Sustainable Progress

## Application at Regional and Megaregional Levels

**Phase I - Methodology**


Department of Ecology and Territory

Barcelona Institute of Regional and Metropolitan Studies

Department of Mathematics

Autonomous University of Barcelona



Project CP 6.1.2

February 2018

# Contents

# Chapter 1

# Data

The data have been provided by the IERMB and the sources are Eurostar and NASA. We have used R software for data management and statistical analysis.

## 1.1 The original data sets

We deal with three original data files named `data`, `cdata` and `dt`. Each individual case in these files is a NUTS3 in a specific year and the columns are several attributes of the case, as you can see in the tables below.

The original data file is `data` (a $20704 \times 10$ matrix) and some of its rows are not complete, but have empty cells with missing values. The complete data file is obtained deleting cases with one or more missing values and we call it `cdata`; it is a $13237 \times 10$ matrix. The first two tables show the headers of these data files.

| | NUTS3 | Year | COUn | MGAn | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|---|---|
| AT111 1995 NMR | AT111 | 1995 | AT | NMR | | | 5.0 | 36.2 | 67.5 | 82.3 |
| AT111 1996 NMR | AT111 | 1996 | AT | NMR | 4.4 | | 5.1 | 36.3 | 78.4 | 70.7 |
| AT111 1997 NMR | AT111 | 1997 | AT | NMR | | | 5.4 | 39.0 | 82.5 | 66.9 |
| AT111 1998 NMR | AT111 | 1998 | AT | NMR | | | 5.2 | 36.6 | 75.0 | 73.1 |

**Table 1.1:** Header of the original data file: `data`.

| | NUTS3 | Year | COUn | MGAn | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|---|---|
| AT111 2003 NMR | AT111 | 2003 | AT | NMR | 75.1 | 17321.0 | 5.9 | 36.8 | 77.6 | 69.4 |
| AT111 2006 NMR | AT111 | 2006 | AT | NMR | 26.7 | 18422.5 | 6.5 | 37.2 | 83.5 | 64.0 |
| AT111 2007 NMR | AT111 | 2007 | AT | NMR | 4.5 | 18560.0 | 6.8 | 37.8 | 75.2 | 71.3 |
| AT111 2008 NMR | AT111 | 2008 | AT | NMR | 55.5 | 18293.3 | 6.1 | 37.6 | 82.6 | 64.9 |

**Table 1.2:** Header of the complete data file: `cdata`.

A method to impute a part of the missing values with certain requirements is supplied by the **mice** library in R (see the details in the Master Thesis of Alan Bernardo). The file of imputed data is named `dt` (a $17363 \times 10$ matrix).

*Remark:* Compare the number of raws of the original, the complete and the imputed data files: 20704, 13237 and 17306.

| | NUTS3 | Year | COUn | MGAn | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|---|---|
| AT111 2003 NMR | AT111 | 2003 | AT | NMR | 75.1 | 17321.0 | 5.9 | 36.8 | 77.6 | 69.4 |
| AT111 2006 NMR | AT111 | 2006 | AT | NMR | 26.7 | 18422.5 | 6.5 | 37.2 | 83.5 | 64.0 |
| AT111 2007 NMR | AT111 | 2007 | AT | NMR | 4.5 | 18560.0 | 6.8 | 37.8 | 75.2 | 71.3 |
| AT111 2008 NMR | AT111 | 2008 | AT | NMR | 55.5 | 18293.3 | 6.1 | 37.6 | 82.6 | 64.9 |

**Table 1.3:** Header of the complete-imputed data file: `dt`.

## 1.2 New variables and files

In order to estimate the values of the variables for larger regions (coutries or megaregions), we take weighted averages using convinient weights. The following variables in files *pop.csv* and *ilu.csv* are added to the data files: Identification number (ID), population (POP), Total Area (TArea), Lighted area (LArea) and Light per capita (*Lpc*). The enlarged file, `dM`, is a $20704 \times 12$ matrix.

| | ID | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor | POP | LArea | TArea | Lpc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AT111 1995 | | | 5.0 | 36.2 | 67.5 | 82.3 | 38725.2 | 576000000.0 | 700000000.0 | 82.3 |
| 2 | AT111 1996 | 4.4 | | 5.1 | 36.3 | 78.4 | 70.7 | 38586.4 | 495000000.0 | 700000000.0 | 70.7 |
| 3 | AT111 1997 | | | 5.4 | 39.0 | 82.5 | 66.9 | 38447.6 | 468000000.0 | 700000000.0 | 66.9 |
| 4 | AT111 1998 | | | 5.2 | 36.6 | 75.0 | 73.1 | 38239.4 | 512000000.0 | 700000000.0 | 73.1 |

**Table 1.4:** Header of the file `dM`, adding ID, population and lighting variables to the file `data`.

The averaged values for the variables ($y$ = PATth, GDPpc, PECpc and GREpc) in the regions are computed as follows: For each region $i$ and year $k$, the value of the variable, $y_{ik}$, is a weighted mean of the values $y_{jk}$ of the variable in the year $k$ in the subset of NUTS3$_j$ belonging to the region $i$, where weights $w_{jk}$ are the corresponding proportion of population:

$$y_{ik} \quad = \quad \sum_{\forall j \in i} y_{jk} \times w_{jk}, \qquad \text{where} \quad w_{jk} = \text{POPpr}_{jk} = \frac{\text{POP}_{jk}}{\sum\limits_{\forall m \in i} \text{POP}_{mk}}.$$

The variables URDpsk and URGpor in a region $i$ in year $k$ are computed as the following quotients of aggregated values, based on estimates derived from lighted areas and light intensities [see Marull, Galetto, Domene and Trullen, 2013].

$$\text{URGpor}_{ik} \quad = \quad \frac{\text{LArea}_{ik}}{\text{TArea}_{ik}} \times 10^2 = \frac{\sum\limits_{\forall j \in i} \text{LArea}_{jk}}{\sum\limits_{\forall j \in i} \text{TArea}_{jk}} \times 10^2$$

$$\text{URDpsk}_{ik} \quad = \quad \frac{\text{POP}_{ik}}{\text{LArea}_{ik}} \times 10^6 = \frac{\sum\limits_{\forall j \in i} \text{POP}_{jk}}{\sum\limits_{\forall j \in i} \text{LArea}_{jk}} \times 10^6$$

### 1.2.1 Description of the files

The following data files store the values for the variables by megaregions:

- In 4 specific years 1995, 2000, 2005, 2010, the file `dyMGA` contains the data in the megaregions.
- For each fixed year, the file `daMGA` contains the data in the megaregions.
- The file `dMGA` contains the data in the megaregions, all years averaged.

The following data files store the values for the variables by countries:

- In 4 specific years 1995, 2000, 2005, 2010, the file `dyCOU` contains the data in the countries.
- For each fixed year, the file `daCOU` contains the data in the countries.
- The file `dCOU` contains the data in the countries, all years averaged.

The files, *VarLinesyear.pdf*, *VarLines.pdf* and *VarVisualization.pdf* contain several lines and points graphical representations for megaregions and countries, all them are in folder `images`.

| | MGAn | Year | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|
| NMR 1995 | NMR | 1995 | 30.2 | 7806.6 | 3.5 | 29.8 | 90.0 | 45.7 |
| VIB 1995 | VIB | 1995 | 39.4 | 3461.9 | 2.1 | 47.2 | 105.7 | 91.8 |
| FRG 1995 | FRG | 1995 | 98.7 | 10722.2 | 3.7 | 43.2 | 51.1 | 86.9 |
| AMB 1995 | AMB | 1995 | 91.0 | 17576.9 | 4.3 | 20.0 | 192.1 | 97.3 |
| PRA 1995 | PRA | 1995 | 2.7 | 13896.0 | 3.0 | 55.3 | 76.5 | 98.4 |
| LIS 1995 | LIS | 1995 | 4.1 | 10614.4 | 2.5 | 37.9 | 106.6 | 89.9 |
| MAD 1995 | MAD | 1995 | 16.6 | 17478.7 | 1.2 | 38.6 | 681.7 | 92.4 |
| BAL 1995 | BAL | 1995 | 80.0 | 16322.9 | 3.3 | 37.3 | 249.6 | 74.3 |
| PAR 1995 | PAR | 1995 | 178.5 | 24241.9 | 1.8 | 44.1 | 560.6 | 95.8 |
| LON 1995 | LON | 1995 | 65.5 | 15009.1 | 3.6 | 44.1 | 265.4 | 97.1 |
| GLB 1995 | GLB | 1995 | 34.7 | 15982.4 | 4.1 | 44.5 | 354.2 | 58.7 |
| NMR 2000 | NMR | 2000 | 59.1 | 14657.6 | 3.8 | 39.9 | 112.9 | 40.5 |
| VIB 2000 | VIB | 2000 | 38.3 | 15729.3 | 2.7 | 34.7 | 136.9 | 84.6 |
| FRG 2000 | FRG | 2000 | 466.1 | 27126.9 | 3.5 | 52.8 | 314.2 | 88.0 |
| AMB 2000 | AMB | 2000 | 185.5 | 23512.0 | 3.9 | 40.1 | 374.2 | 96.7 |
| PRA 2000 | PRA | 2000 | 126.6 | 18067.7 | 3.6 | 51.8 | 155.5 | 94.4 |
| BER 2000 | BER | 2000 | 163.6 | 20977.8 | 1.6 | 47.5 | 1165.5 | 74.8 |
| LIS 2000 | LIS | 2000 | 2.6 | 13702.6 | 3.0 | 42.6 | 95.9 | 85.6 |
| MAD 2000 | MAD | 2000 | 25.0 | 25202.4 | 1.4 | 49.0 | 695.5 | 94.4 |
| BAL 2000 | BAL | 2000 | 112.4 | 21421.3 | 3.6 | 43.3 | 241.4 | 77.1 |
| PAR 2000 | PAR | 2000 | 237.7 | 32146.7 | 1.8 | 50.2 | 561.0 | 97.2 |
| RMT 2000 | RMT | 2000 | 82.0 | 23937.4 | 3.0 | 41.3 | 253.5 | 81.1 |
| LON 2000 | LON | 2000 | 107.0 | 20396.1 | 3.8 | 46.5 | 258.1 | 95.7 |
| GLB 2000 | GLB | 2000 | 54.7 | 20319.3 | 4.4 | 44.5 | 328.9 | 59.2 |
| NMR 2005 | NMR | 2005 | 61.3 | 17649.0 | 3.8 | 40.5 | 128.2 | 36.9 |
| VIB 2005 | VIB | 2005 | 40.7 | 20112.4 | 2.6 | 37.3 | 169.6 | 80.2 |
| FRG 2005 | FRG | 2005 | 494.0 | 31052.1 | 4.0 | 51.9 | 317.9 | 89.9 |
| AMB 2005 | AMB | 2005 | 188.7 | 26800.7 | 3.8 | 46.4 | 370.1 | 91.8 |
| PRA 2005 | PRA | 2005 | 136.6 | 21737.2 | 3.8 | 50.9 | 164.8 | 89.8 |
| BER 2005 | BER | 2005 | 194.0 | 21852.8 | 2.1 | 44.2 | 711.7 | 58.3 |
| LIS 2005 | LIS | 2005 | 12.2 | 17359.9 | 3.3 | 45.1 | 102.3 | 82.1 |
| MAD 2005 | MAD | 2005 | 39.5 | 28798.7 | 2.1 | 52.9 | 463.5 | 59.0 |
| BAL 2005 | BAL | 2005 | 140.5 | 24847.5 | 3.8 | 44.7 | 270.8 | 70.1 |
| PAR 2005 | PAR | 2005 | 229.6 | 34243.6 | 2.3 | 48.0 | 420.5 | 86.7 |
| RMT 2005 | RMT | 2005 | 98.1 | 25073.1 | 3.2 | 42.7 | 257.1 | 79.6 |
| LON 2005 | LON | 2005 | 82.7 | 21550.4 | 3.4 | 47.3 | 318.4 | 87.6 |
| GLB 2005 | GLB | 2005 | 57.4 | 25793.7 | 4.3 | 47.0 | 362.5 | 54.0 |
| NMR 2010 | NMR | 2010 | 33.0 | 19346.1 | 3.6 | 41.4 | 139.8 | 37.7 |
| VIB 2010 | VIB | 2010 | 18.5 | 21215.1 | 2.5 | 46.0 | 261.7 | 72.6 |
| FRG 2010 | FRG | 2010 | 245.2 | 34092.9 | 3.8 | 53.4 | 323.2 | 86.3 |
| AMB 2010 | AMB | 2010 | 92.4 | 28523.8 | 3.7 | 48.3 | 392.1 | 93.3 |
| PRA 2010 | PRA | 2010 | 74.8 | 23393.1 | 3.8 | 50.5 | 175.3 | 89.7 |
| BER 2010 | BER | 2010 | 108.4 | 25755.6 | 1.8 | 47.7 | 833.5 | 60.2 |
| LIS 2010 | LIS | 2010 | 5.7 | 19725.3 | 2.9 | 43.2 | 90.3 | 85.9 |
| MAD 2010 | MAD | 2010 | 20.9 | 30141.0 | 1.7 | 49.5 | 499.2 | 60.1 |
| BAL 2010 | BAL | 2010 | 58.1 | 25084.4 | 3.3 | 42.3 | 284.1 | 67.9 |
| PAR 2010 | PAR | 2010 | 97.9 | 39457.9 | 2.1 | 47.8 | 459.5 | 92.9 |
| RMT 2010 | RMT | 2010 | 46.9 | 23340.2 | 2.6 | 43.2 | 294.5 | 79.9 |
| LON 2010 | LON | 2010 | 31.9 | 19883.4 | 3.0 | 45.6 | 328.4 | 85.0 |
| GLB 2010 | GLB | 2010 | 21.9 | 23309.7 | 3.9 | 45.8 | 366.7 | 54.8 |

**Table 1.5:** Mega-regions aggregated data, in 4 specific years

| | MGAn | Year | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|
| NMR 1995 | NMR | 1995 | 30.2 | 7806.6 | 3.5 | 29.8 | 90.0 | 45.7 |
| VIB 1995 | VIB | 1995 | 39.4 | 3461.9 | 2.1 | 47.2 | 105.7 | 91.8 |
| FRG 1995 | FRG | 1995 | 98.7 | 10722.2 | 3.7 | 43.2 | 51.1 | 86.9 |
| AMB 1995 | AMB | 1995 | 91.0 | 17576.9 | 4.3 | 20.0 | 192.1 | 97.3 |

**Table 1.6:** Header of aggregated data in mega-regions every year.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| NMR | 54.1 | 15125.5 | 3.7 | 37.8 | 122.3 | 38.2 |
| VIB | 33.8 | 16864.7 | 2.6 | 41.2 | 164.6 | 77.4 |
| FRG | 437.8 | 29237.7 | 3.8 | 51.9 | 291.8 | 85.9 |
| AMB | 168.1 | 24821.9 | 3.9 | 41.5 | 354.4 | 93.2 |
| PRA | 117.0 | 19829.0 | 3.8 | 50.0 | 157.6 | 89.5 |
| BER | 164.0 | 21464.9 | 1.8 | 44.4 | 814.8 | 60.9 |
| LIS | 7.7 | 15697.7 | 3.0 | 43.1 | 98.8 | 83.3 |
| MAD | 31.8 | 26856.8 | 1.6 | 49.0 | 595.1 | 71.2 |
| BAL | 111.8 | 22803.3 | 3.5 | 42.5 | 263.0 | 70.8 |
| PAR | 217.9 | 32759.9 | 2.0 | 47.5 | 486.3 | 91.7 |
| RMT | 77.1 | 17207.4 | 2.9 | 35.6 | 268.4 | 78.6 |
| LON | 82.5 | 19789.4 | 3.5 | 46.5 | 293.4 | 90.2 |
| GLB | 50.2 | 22285.8 | 4.1 | 46.0 | 361.5 | 54.9 |

**Table 1.7:** Mega-regions aggregated data

| | COUn | Year | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|
| 1995 BE | BE | 1995 | 80.5 | 18424.0 | 4.7 | 37.2 | 344.5 | 97.2 |
| 1995 BG | BG | 1995 | 1.0 | 4662.6 | 2.5 | 43.0 | 175.8 | 41.9 |
| 1995 CZ | CZ | 1995 | 1.6 | 11396.0 | 3.8 | 49.9 | 118.1 | 93.8 |
| 1995 EE | EE | 1995 | 2.1 | 5298.3 | 3.7 | 44.2 | 177.5 | 18.0 |
| 1995 ES | ES | 1995 | 10.3 | 13344.8 | 2.4 | 34.4 | 180.9 | 40.2 |
| 1995 FI | FI | 1995 | 140.4 | 3739.0 | 5.5 | 10.1 | 53.4 | 28.0 |
| 1995 FR | FR | 1995 | 88.7 | 17149.4 | 3.8 | 38.1 | 165.4 | 64.1 |
| 1995 IE | IE | 1995 | 27.1 | 15187.1 | 2.9 | 35.7 | 114.1 | 45.7 |
| 1995 LT | LT | 1995 | 0.8 | 5200.2 | 2.2 | 40.8 | 440.1 | 12.9 |
| 1995 LU | LU | 1995 | 79.2 | 32618.7 | 8.1 | 52.9 | 159.0 | 99.7 |
| 1995 PT | PT | 1995 | 1.6 | 9961.3 | 2.5 | 40.4 | 51.5 | 47.5 |
| 1995 SE | SE | 1995 | 172.5 | 18323.3 | 5.6 | 46.9 | 64.4 | 30.3 |
| 1995 UK | UK | 1995 | 57.0 | 14843.9 | 4.1 | 44.0 | 222.7 | 67.4 |
| 2000 AT | AT | 2000 | 147.1 | 25069.5 | 3.4 | 49.3 | 159.0 | 60.6 |
| 2000 BG | BG | 2000 | 0.9 | 5388.9 | 2.1 | 40.8 | 282.2 | 25.4 |
| 2000 CY | CY | 2000 | 8.9 | 16652.7 | 3.3 | 45.5 | 123.1 | 58.7 |
| 2000 CZ | CZ | 2000 | 6.3 | 13495.2 | 3.8 | 47.4 | 147.7 | 89.1 |
| 2000 DE | DE | 2000 | 283.9 | 23142.9 | 3.7 | 49.3 | 252.8 | 84.2 |
| 2000 EE | EE | 2000 | 4.1 | 8569.1 | 3.5 | 41.9 | 164.6 | 18.9 |
| 2000 EL | EL | 2000 | 5.9 | 16361.2 | 2.8 | 42.3 | 22.5 | 36.1 |
| 2000 ES | ES | 2000 | 20.7 | 18453.0 | 2.9 | 41.3 | 186.1 | 39.8 |
| 2000 FI | FI | 2000 | 277.9 | 22251.0 | 6.1 | 44.3 | 73.5 | 20.6 |
| 2000 FR | FR | 2000 | 123.4 | 22159.1 | 4.0 | 42.5 | 168.1 | 64.4 |
| 2000 IE | IE | 2000 | 53.7 | 25045.6 | 3.6 | 44.6 | 93.0 | 59.3 |
| 2000 IT | IT | 2000 | 65.3 | 21823.5 | 3.0 | 38.9 | 227.1 | 75.0 |
| 2000 LT | LT | 2000 | 1.3 | 7500.8 | 1.8 | 40.2 | 280.4 | 19.3 |
| 2000 LU | LU | 2000 | 182.7 | 46555.1 | 8.3 | 60.5 | 169.3 | 100.0 |
| 2000 LV | LV | 2000 | 3.8 | 6868.2 | 1.6 | 39.4 | 330.4 | 11.1 |
| 2000 NL | NL | 2000 | 234.2 | 25483.9 | 4.2 | 52.0 | 387.7 | 99.3 |
| 2000 PT | PT | 2000 | 1.4 | 13798.8 | 3.1 | 48.9 | 52.3 | 47.9 |
| 2000 RO | RO | 2000 | 0.3 | 4943.4 | 1.6 | 48.0 | 321.4 | 29.5 |
| 2000 SE | SE | 2000 | 259.2 | 24266.7 | 5.3 | 48.7 | 78.8 | 24.8 |
| 2000 SK | SK | 2000 | 2.1 | 9546.9 | 3.1 | 37.6 | 150.0 | 74.0 |
| 2000 UK | UK | 2000 | 92.5 | 19916.1 | 4.4 | 45.9 | 227.4 | 67.7 |
| 2005 AT | AT | 2005 | 185.1 | 28089.6 | 4.0 | 49.4 | 167.9 | 58.9 |
| 2005 BE | BE | 2005 | 143.8 | 26890.9 | 4.9 | 40.6 | 358.0 | 96.7 |
| 2005 BG | BG | 2005 | 3.1 | 8201.6 | 2.4 | 47.0 | 310.3 | 21.6 |
| 2005 CY | CY | 2005 | 22.2 | 20307.5 | 3.3 | 48.6 | 134.8 | 58.5 |
| 2005 CZ | CZ | 2005 | 10.5 | 17757.7 | 4.1 | 48.2 | 150.4 | 87.0 |
| 2005 DE | DE | 2005 | 304.2 | 26811.3 | 3.7 | 48.4 | 270.8 | 79.1 |
| 2005 DK | DK | 2005 | 107.0 | 23665.5 | 4.3 | 47.5 | 23.6 | 56.9 |
| 2005 EE | EE | 2005 | 4.8 | 13981.1 | 4.0 | 45.2 | 164.5 | 18.3 |
| 2005 EL | EL | 2005 | 7.1 | 20240.1 | 2.9 | 43.5 | 23.7 | 35.7 |
| 2005 ES | ES | 2005 | 33.1 | 22936.2 | 3.2 | 45.6 | 219.8 | 36.1 |
| 2005 FI | FI | 2005 | 252.7 | 25660.5 | 6.4 | 45.9 | 61.6 | 25.0 |
| 2005 FR | FR | 2005 | 136.5 | 24961.8 | 4.2 | 42.2 | 189.4 | 59.2 |
| 2005 IE | IE | 2005 | 65.4 | 32392.9 | 3.5 | 47.2 | 117.4 | 51.3 |
| 2005 IT | IT | 2005 | 79.8 | 23054.6 | 3.2 | 40.6 | 234.6 | 73.9 |
| 2005 LT | LT | 2005 | 2.6 | 11944.3 | 2.3 | 42.8 | 308.4 | 16.8 |
| 2005 LU | LU | 2005 | 216.0 | 57067.9 | 10.3 | 66.1 | 180.7 | 99.9 |
| 2005 LV | LV | 2005 | 8.1 | 10732.1 | 2.0 | 43.8 | 349.8 | 9.9 |
| 2005 MT | MT | 2005 | 27.9 | 17971.0 | 2.4 | 37.4 | 1276.9 | 100.0 |
| 2005 NL | NL | 2005 | 229.4 | 29055.9 | 4.4 | 51.6 | 402.1 | 98.0 |
| 2005 PL | PL | 2005 | 4.9 | 15268.9 | 1.6 | 41.5 | 57.4 | 64.9 |
| 2005 PT | PT | 2005 | 4.6 | 16566.2 | 3.3 | 48.6 | 50.9 | 50.7 |
| 2005 RO | RO | 2005 | 1.3 | 7838.3 | 1.7 | 42.8 | 401.2 | 22.8 |
| 2005 SE | SE | 2005 | 266.6 | 27337.7 | 5.4 | 48.3 | 88.2 | 22.6 |
| 2005 SK | SK | 2005 | 5.9 | 13481.6 | 3.3 | 38.9 | 192.9 | 57.4 |
| 2005 UK | UK | 2005 | 74.8 | 21986.3 | 3.8 | 47.0 | 287.1 | 61.2 |
| 2010 AT | AT | 2010 | 117.1 | 30771.7 | 3.9 | 51.5 | 176.6 | 57.0 |
| 2010 BE | BE | 2010 | 70.7 | 29386.4 | 4.9 | 41.2 | 370.7 | 96.7 |
| 2010 BG | BG | 2010 | 1.1 | 10895.3 | 2.3 | 49.1 | 271.0 | 24.4 |
| 2010 CY | CY | 2010 | 7.2 | 23626.7 | 3.1 | 48.4 | 133.2 | 65.1 |
| 2010 CZ | CZ | 2010 | 10.9 | 19653.3 | 4.0 | 48.3 | 163.8 | 82.0 |
| 2010 DE | DE | 2010 | 163.3 | 29762.6 | 3.7 | 50.5 | 273.8 | 79.6 |
| 2010 DK | DK | 2010 | 101.2 | 30198.3 | 3.6 | 49.7 | 210.3 | 61.8 |
| 2010 EE | EE | 2010 | 8.0 | 15965.0 | 4.5 | 41.2 | 145.0 | 20.4 |
| 2010 EL | EL | 2010 | 0.2 | 20990.2 | 2.5 | 42.8 | 24.4 | 35.7 |
| 2010 ES | ES | 2010 | 17.8 | 24113.6 | 2.7 | 42.1 | 236.5 | 38.6 |
| 2010 FI | FI | 2010 | 94.5 | 27880.4 | 6.6 | 46.3 | 73.5 | 21.4 |
| 2010 FR | FR | 2010 | 65.2 | 26873.4 | 4.0 | 41.7 | 189.4 | 60.9 |
| 2010 HU | HU | 2010 | 6.8 | 16112.8 | 2.4 | 40.0 | 288.8 | 37.5 |
| 2010 IE | IE | 2010 | 26.4 | 32167.3 | 3.3 | 41.3 | 120.7 | 54.7 |
| 2010 IT | IT | 2010 | 37.3 | 21247.7 | 2.5 | 40.8 | 270.9 | 74.3 |
| 2010 LT | LT | 2010 | 2.5 | 14330.8 | 1.9 | 40.3 | 272.2 | 17.8 |
| 2010 LU | LU | 2010 | 98.6 | 65226.8 | 9.1 | 70.9 | 196.7 | 100.0 |
| 2010 LV | LV | 2010 | 4.6 | 12658.0 | 2.1 | 40.1 | 324.2 | 10.2 |
| 2010 MT | MT | 2010 | 4.8 | 20945.8 | 2.1 | 39.6 | 1318.2 | 99.1 |
| 2010 NL | NL | 2010 | 70.8 | 26834.4 | 3.7 | 52.9 | 483.1 | 97.9 |
| 2010 PL | PL | 2010 | 6.0 | 16374.2 | 2.4 | 42.1 | 144.2 | 69.2 |
| 2010 PT | PT | 2010 | 1.9 | 18227.9 | 2.9 | 45.4 | 46.9 | 56.4 |
| 2010 RO | RO | 2010 | 0.7 | 11689.0 | 1.6 | 42.7 | 291.8 | 31.0 |
| 2010 SE | SE | 2010 | 106.0 | 30171.0 | 5.2 | 48.2 | 92.6 | 22.3 |
| 2010 SK | SK | 2010 | 4.4 | 18000.4 | 3.1 | 40.3 | 227.8 | 48.7 |
| 2010 UK | UK | 2010 | 29.6 | 20310.9 | 3.4 | 45.6 | 300.2 | 60.5 |

**Table 1.8:** Countries aggregated data, in 4 specific years

| | COUn | Year | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|---|---|
| 1995 BE | BE | 1995 | 80.5 | 18424.0 | 4.7 | 37.2 | 344.5 | 97.2 |
| 1995 BG | BG | 1995 | 1.0 | 4662.6 | 2.5 | 43.0 | 175.8 | 41.9 |
| 1995 CZ | CZ | 1995 | 1.6 | 11396.0 | 3.8 | 49.9 | 118.1 | 93.8 |
| 1995 EE | EE | 1995 | 2.1 | 5298.3 | 3.7 | 44.2 | 177.5 | 18.0 |

**Table 1.9:** Header of aggregated data in countries every year.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| AT | 154.8 | 20012.2 | 3.7 | 49.4 | 140.2 | 56.8 |
| BE | 121.9 | 24714.2 | 4.9 | 32.3 | 354.6 | 97.0 |
| BG | 1.7 | 7005.7 | 2.4 | 45.4 | 290.8 | 23.8 |
| CY | 10.2 | 18811.4 | 3.3 | 47.2 | 129.4 | 58.7 |
| CZ | 10.0 | 16055.4 | 4.0 | 48.6 | 147.8 | 84.7 |
| DE | 272.9 | 25309.9 | 3.7 | 48.2 | 234.9 | 80.2 |
| DK | 192.0 | 28986.3 | 3.6 | 51.6 | 62.2 | 66.0 |
| EE | 9.8 | 11224.6 | 3.9 | 44.4 | 169.2 | 18.0 |
| EL | 5.1 | 17858.3 | 2.8 | 30.2 | 23.6 | 35.1 |
| ES | 23.7 | 20588.7 | 2.8 | 41.9 | 212.4 | 37.3 |
| FI | 226.9 | 19142.1 | 6.4 | 35.1 | 69.2 | 22.1 |
| FR | 120.1 | 23069.6 | 4.1 | 41.2 | 180.2 | 61.3 |
| HU | 11.8 | 9520.3 | 2.4 | 18.8 | 236.9 | 46.3 |
| IE | 55.1 | 27530.1 | 3.5 | 43.8 | 109.7 | 53.2 |
| IT | 61.6 | 15652.8 | 2.9 | 34.5 | 243.2 | 73.3 |
| LT | 1.9 | 9811.8 | 2.3 | 41.7 | 317.3 | 16.7 |
| LU | 161.7 | 51293.1 | 8.9 | 63.5 | 176.2 | 99.7 |
| LV | 4.1 | 8564.2 | 1.9 | 42.4 | 365.0 | 9.8 |
| MT | 12.4 | 16842.2 | 2.2 | 34.9 | 871.9 | 99.9 |
| NL | 193.6 | 25404.6 | 4.1 | 37.4 | 414.0 | 98.1 |
| PL | 6.6 | 15566.5 | 2.0 | 42.8 | 45.5 | 62.9 |
| PT | 3.2 | 14850.0 | 3.0 | 46.6 | 52.1 | 48.8 |
| RO | 0.7 | 5961.0 | 1.7 | 30.0 | 345.1 | 26.9 |
| SE | 239.2 | 25322.8 | 5.4 | 48.2 | 83.3 | 23.8 |
| SK | 4.4 | 12783.0 | 3.2 | 39.2 | 157.2 | 61.7 |
| UK | 73.7 | 19909.7 | 3.9 | 46.2 | 260.1 | 63.8 |

**Table 1.10:** Countries aggregated data

# Chapter 2

# Confirmatory factorial model

The *pairwise-complete* correlation matrix is performed by pairwise correlations using the subset of complete data for each pair of variables, this subset and its size usually changes with the pair. The *complete* correlation matrix is performed by pairwise correlations using the subset of complete data having no missing values for any variable, so the subset is the same for any pair. These two matrices are shown below. Notice that the highest differences between these two correlation matrices are in the correlations involving PECpc.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| PATth | 1.00 | 0.37 | -0.01 | 0.29 | 0.07 | 0.29 |
| GDPpc | 0.37 | 1.00 | -0.05 | 0.70 | 0.41 | 0.42 |
| PECpc | -0.01 | -0.05 | 1.00 | -0.21 | -0.42 | -0.04 |
| GREpc | 0.29 | 0.70 | -0.21 | 1.00 | 0.31 | 0.25 |
| URDpsk | 0.07 | 0.41 | -0.42 | 0.31 | 1.00 | 0.29 |
| URGpor | 0.29 | 0.42 | -0.04 | 0.25 | 0.29 | 1.00 |

**Table 2.1:** Correlation matrix of pairwise complete data.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| PATth | 1.00 | 0.38 | -0.05 | 0.28 | 0.08 | 0.31 |
| GDPpc | 0.38 | 1.00 | -0.22 | 0.76 | 0.38 | 0.32 |
| PECpc | -0.05 | -0.22 | 1.00 | -0.28 | -0.47 | -0.25 |
| GREpc | 0.28 | 0.76 | -0.28 | 1.00 | 0.33 | 0.25 |
| URDpsk | 0.08 | 0.38 | -0.47 | 0.33 | 1.00 | 0.29 |
| URGpor | 0.31 | 0.32 | -0.25 | 0.25 | 0.29 | 1.00 |

**Table 2.2:** Correlation matrix of complete data.

## 2.1 Factorial models

We present here a CFA, not an explorarory analysis, because of the expertise criterion postulating that three interrelated latent factors can help in explaining the correlations between the observed measures of GDPpc, PECpc, PATth, GREpc, URDpsk and URGpor in the urban networks of regional and mega-regional level in UE. One latent factor relates to the economic impact (PATth, URGpor and GDPpc), a second factor involves urban ecology (PECpc and URDpsk) and a third one focuses in social cohesion (GREpc and GDPpc). The three factors are necessarely correlated in this setting, so orthogonal factors are unexpected.

The 3-factors model will be estimated using only complete data (`cdata` file), nevertheless, the scores in the model will be computed for more cases after the controlled-imputation of a subset of missing values (`dt` file). `Lavaan` package in R (see Rosseel Y., 2012)) is used to perform factorial models.

The linear equation in factorial models in matrix form and applied to scaled data is

$$Z = QF + U \tag{2.1}$$

where $Z$ are the scaled observed variables, $F$ the latent or common factors, $U$ the unique factors (or residuals) and $Q$ are the factors loadings. A set of assumptions on the model is described in the sequel.

### 2.1.1 The model proposed

The model is defined by the following equations expressing variables in terms of factors, that is, the *pattern model*:

$$\begin{aligned}
\text{zPATpc} &= q_{11}\,F_1 + 0\,F_2 + 0\,F_3 + U_1\\
\text{zGDPpc} &= q_{21}\,F_1 + 0\,F_2 + q_{23}\,F_3 + U_2\\
\text{zPECpc} &= 0\,F_1 + q_{32}\,F_2 + 0\,F_3 + U_3\\
\text{zGREpc} &= 0\,F_1 + 0\,F_2 + q_{43}\,F_3 + U_4\\
\text{zURDpsk} &= q_{51}\,F_1 + q_{52}\,F_2 + 0\,F_3 + U_5\\
\text{zURGpor} &= q_{61}\,F_1 + 0\,F_2 + 0\,F_3 + U_6
\end{aligned}$$

The notation with a "z" preceeding the variable name (zGDPpc, and so on) indicates that the observed variables are normalized before the model fitting. The coefficients ($q_{ij}$) perform the matrix of loadings $Q$. Three latent factors $F_1, F_2$ and $F_3$, and six unique, residual or specific factors $U_1, \ldots, U_6$ are taken. We assume that there exist non-null correlation between each pair of latent factors which have unit variance, that is, a non-diagonal correlations matrix $R_F$ with diagonal elements equal to one. It is assumed that the specific factors are uncorrelated having covariance matrix $\Psi$, and that common and specific factors are uncorrelated one to each other. It is easy to prove that this set of assumptions together with equation 2.1 imply the following decomposition of the correlation matrix $R = (r_{ij})$ (the complete data correlation matrix in Table 2.2):

$$R = Q\,R_F\,Q^t + \Psi \tag{2.2}$$

where

$$Q = \begin{pmatrix} q_{11} & 0 & 0\\ q_{21} & 0 & q_{23}\\ 0 & q_{32} & 0\\ 0 & 0 & q_{43}\\ 0 & q_{52} & 0\\ q_{61} & 0 & 0 \end{pmatrix} \quad \Sigma_F = \begin{pmatrix} 1 & \alpha & \beta\\ \alpha & 1 & \gamma\\ \beta & \gamma & 1 \end{pmatrix} \quad \Psi = \begin{pmatrix} \psi_{11} & 0 & 0 & 0 & 0 & 0\\ 0 & \psi_{22} & 0 & 0 & 0 & 0\\ 0 & 0 & \psi_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & \psi_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & \psi_{55} & 0\\ 0 & 0 & 0 & 0 & 0 & \psi_{66} \end{pmatrix} \tag{2.3}$$

## 2.2 Parameters estimates

The parameters in $Q$, $R_F$ and $\Psi$ must be estimated from data. Taking the factorial model applied to the complete data correlation matrix, we use funtion `cfa()` in `lavaan` library to fit the model by the method of unweighted least squares (ULS). The algorithm converges for this model and the output gives the implied or reproduced loadings matrix, also called *pattern matrix*, $Q$, shown in Table 2.3, the estimates of the unique variances diagonal matrix $\Psi$, shown in Table 2.5, and the estimates of the within common factors correlations $R_F$, shown in Table 2.4.

In Figure 2.1 Confirmatory Factor Analysis *pattern matrix* is represented. One direction arrows represent loadings (regression coefficients expressing variables in terms of latent factors) in Table 2.3 and double arrows represent *implied* correlations between the common factors in Table 2.4, with red colored lines corresponding to negative coefficients and lines width being proportional to the absolute value.

| | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| zPATth | 0.50 | 0.00 | 0.00 |
| zGDPpc | 0.27 | 0.00 | 0.68 |
| zPECpc | 0.00 | -0.58 | 0.00 |
| zGREpc | 0.00 | 0.00 | 0.92 |
| zURSpsk | 0.00 | 0.80 | 0.00 |
| zURGpor | 0.60 | 0.00 | 0.00 |

**Table 2.3:** Estimates of the loadings matrix or *pattern matrix* $Q$.

In Figure 2.2, the Confirmatory Factor Analysis *structure matrix* $R_{XF}$ is represented. The formula (2.4) establishes that the matrix $R_{XF}$ can be computed as the product of the loadings matrix $Q$ and

| | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| Factor 1 | 1 | 0.45 | 0.50 |
| Factor 2 | 0.45 | 1 | 0.47 |
| Factor 3 | 0.50 | 0.47 | 1 |

**Table 2.4:** Estimates of the within comon factor correlations matrix $R_F$.

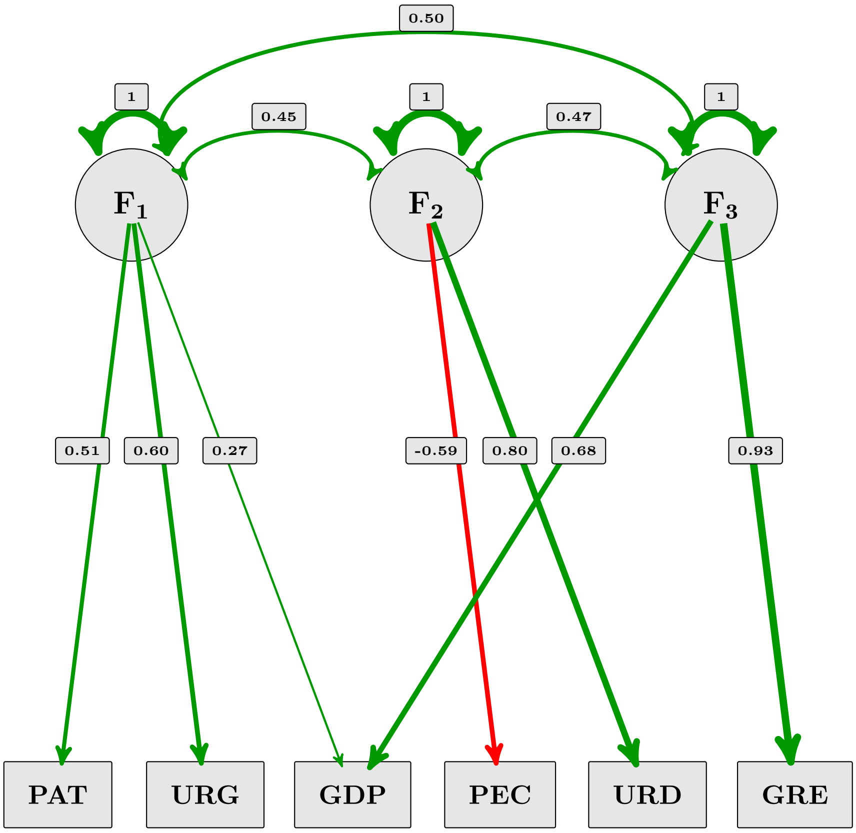


**Figure 2.1:** Graphical representation of the *pattern* matrix $Q$ of the CFA model. Numerical values on arrows correspond to Table 2.3 and Table 2.4, the red colored lines indicate negative coefficients.

the within factors correlation matrix $R_F$. This can be deduced from the equation of the model and the assumptions describe below. In particular, formula (2.4) shows that the correlations between variables and factors differs from the loadings matrix $Q$ when there are non-null correlations between common factors. In this picture, *implied* correlations are represented by double arrows, with red colored lines corresponding to negative coefficients and lines width being proportional to the absolute value. The numerical values recorded in Table 2.6 and Table 2.4 are not printed in the picture.

$$R_{XF} = Q\,R_F \tag{2.4}$$

### 2.2.1 The implied correlation matrix

The implied correlation matrix $\tilde{R}$, also called reproduced correlation matrix and shown in Table 2.7, is computed using the decomposition on the right hand side of (2.2), that is, $\tilde{R} = Q\,R_F\,Q^t + \Psi$, taking the estimates of $Q$, $R_f$ and $\Psi$ given in Table 2.3, Table 2.4 and Table 2.5.

In order to compare $\tilde{R}$ with the actual sample correlation matrix in Table 2.2, we compute the elementwise differences shown in Table 2.8. The most important differences correspond to some variable variances (diagonal elements), but does'nt affect the correlations which are the terms of interest in the CFA; out of the diagonal, the largest difference in absolute value is 0.10.

| | zPATth | zGDPpc | zPECpc | zGREpc | zURSpsk | zURGpor |
|---|---|---|---|---|---|---|
| zPATth | 0.74 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| zGDPpc | 0.00 | 0.28 | 0.00 | 0.00 | 0.00 | 0.00 |
| zPECpc | 0.00 | 0.00 | 0.64 | 0.00 | 0.00 | 0.00 |
| zGREpc | 0.00 | 0.00 | 0.00 | 0.66 | 0.00 | 0.00 |
| zURSpsk | 0.00 | 0.00 | 0.00 | 0.00 | 0.36 | 0.00 |
| zURGpor | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 |

**Table 2.5:** Estimates of the diagonal matrix $\Psi$ containing uniquenesses.

| | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| zPATth | 0.50 | 0.23 | 0.25 |
| zGDPpc | 0.62 | 0.44 | 0.82 |
| zPECpc | -0.26 | -0.58 | -0.28 |
| zGREpc | 0.47 | 0.44 | 0.92 |
| zURSpsk | 0.36 | 0.80 | 0.38 |
| zURGpor | 0.60 | 0.27 | 0.30 |

**Table 2.6:** Estimates of the *structure matrix* $R_{XF}$, containing the correlations between variables and common factors.

The fact that the algorithm has converged, along with a not to large amount residuals out off the diagonal in the correlation matrix, says that the model has an acceptable goodness of fit to the EU regional data. By construction, factors $F_1$, $F_2$ and $F_3$ can be labeled as *Economic growth?*, *Social cohesion?* and *Urban Ecology?*, because they load preferently on the desired variables.

## 2.3 Scores

According to the CFA model, the Thompson type regression scores (see Thomson, 1951) can be computed. Using equation (2.5), we get the whole scores data matrix $F$ from the scaled data file $Z$, the complete data sample correlation matrix $R$ and the estimates of $Q$ and $R_F$,

$$F = Z\, R^{-1}\, Q\, R_F \tag{2.5}$$

Notice that scores are computed from scaled data, but they are not scaled anymore.

To get the equations defining $F_1, F_2$ and $F_3$ as functions of the scaled initial random vectors $Z$, we use the transposed form. Indeed, define $B = R^{-1}QR_F$ and write it with a 3-digit rounding:

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = B^t \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \end{pmatrix}, \text{ with } B^t = \begin{pmatrix} 0.242 & 0.380 & -0.042 & -0.018 & 0.073 & 0.383 \\ 0.141 & 0.024 & -0.257 & 0.109 & 0.639 & -0.053 \\ -0.064 & 0.275 & 0.009 & 0.714 & 0.029 & 0.053 \end{pmatrix} \tag{2.6}$$

where $Z_1 =$ zPATth, $Z_2 =$ zGDPpc, $Z_3 =$ zPECpc, $Z_4 =$ zGREpc, $Z_5 =$ zURDpsk and $Z_6 =$ zURGpor. This formula can be applied to compute the scores both for the scaled sample observations and for new observations previously scaled with respect to centers and standard deviations of the sample observations:

$$\begin{aligned} F_1 &= 0.242Z_1 + 0.38Z_2 - 0.042Z_3 - 0.018Z_4 + 0.073Z_5 + 0.383Z_6 \\ F_2 &= 0.141Z_1 + 0.024Z_2 - 0.257Z_3 + 0.109Z_4 + 0.639Z_5 - 0.053Z_6 \\ F_3 &= -0.064Z_1 + 0.275Z_2 + 0.009Z_3 + 0.714Z_4 + 0.029Z_5 + 0.053Z_6 \end{aligned} \tag{2.7}$$

Moreover, taking into account that $Z = \frac{X-\overline{x}}{s}$ and renaming $A^t = B^t S^{-1}$, where $S^{-1}$ is the diagonal matrix of the reciprocoa of the standard deviations $\frac{1}{s}$ of the original variables, the scores can be written in terms of the original unscaled variables:

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = A^t \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} - A^t \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \\ \overline{x}_4 \\ \overline{x}_5 \\ \overline{x}_6 \end{pmatrix} \quad \text{where } A^t = B^t S^{-1} \text{ and } S^{-1} = \text{diag}(\tfrac{1}{s_i}) \tag{2.8}$$

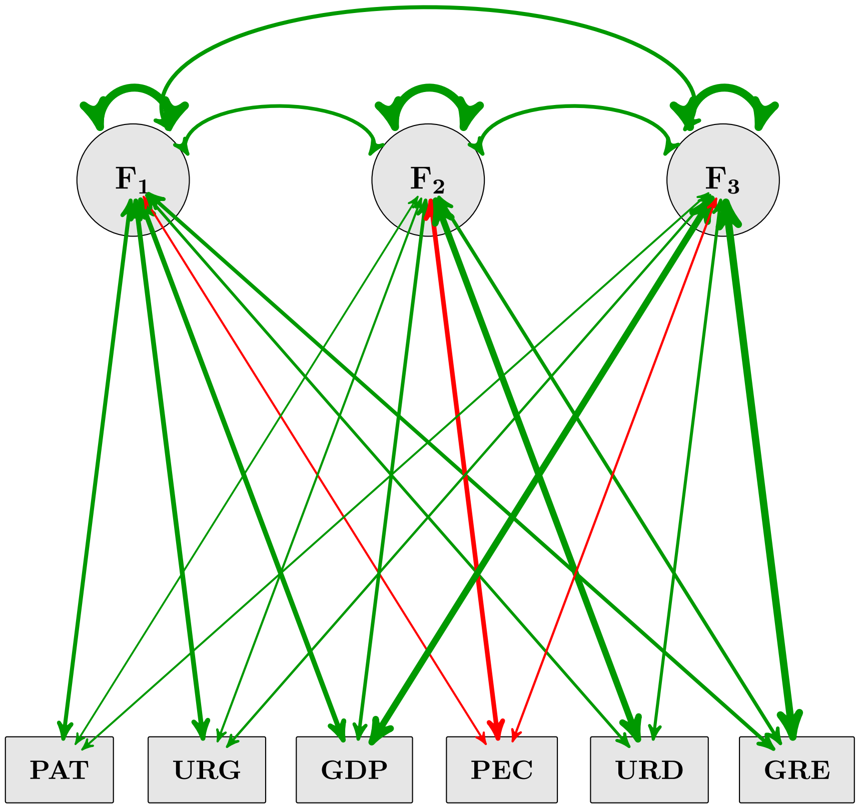


**Figure 2.2:** The thicknesses of the lines are proportional to the absolute values of the estimates in the *structure matrix* $R_{XF}$, containing the correlations between variables and common factors (Table 2.6), and the within common factors correlations (Table 2.4), with red colored lines indicating negative coefficients.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| zPATth | 1.00 | 0.31 | -0.13 | 0.24 | 0.18 | 0.31 |
| zGDPpc | 0.31 | 1.00 | -0.26 | 0.76 | 0.35 | 0.37 |
| zPECpc | -0.13 | -0.26 | 0.98 | -0.26 | -0.47 | -0.16 |
| zGREpc | 0.24 | 0.76 | -0.26 | 1.51 | 0.35 | 0.28 |
| zURDpsk | 0.18 | 0.35 | -0.47 | 0.35 | 1.00 | 0.22 |
| zURGpor | 0.31 | 0.37 | -0.16 | 0.28 | 0.22 | 0.51 |

**Table 2.7:** Reproduced correlation matrix in the cFA, $\tilde{R}$, estimated by the right hand side of equation (2.2).

The matrix $A^t$ estimated by our CFA model is

$$A^t = \begin{pmatrix} 0.001292 & 0.000044 & -0.019810 & -0.001635 & 0.000064 & 0.014809 \\ 0.000752 & 0.000003 & -0.120969 & 0.009838 & 0.000559 & -0.002065 \\ -0.000341 & 0.000032 & 0.004020 & 0.064529 & 0.000026 & 0.002067 \end{pmatrix}$$

where the elements of matrix $A$ are rounded to six digits and the mean vector is $\overline{X}$ =(147.96, 2.18724×$10^4$, 4.45, 44.9, 554.96, 78.15). Then, the equations are:

$$\begin{aligned} F_1 &= 0.001292X_1 + 4.4\times10^{-5}X_2 - 0.01981X_3 - 0.001635X_4 + \\ &\quad 6.4\times10^{-5}X_5 + 0.014809X_6 - 2.178924 \\ F_2 &= 7.52\times10^{-4}X_1 + 3\times10^{-6}X_2 - 0.120969X_3 + 0.009838X_4 + \\ &\quad 5.59\times10^{-4}X_5 - 0.002065X_6 - 0.224602 \\ F_3 &= -3.41\times10^{-4}X_1 + 3.2\times10^{-5}X_2 + 0.00402X_3 + 0.064529X_4 + \\ &\quad 2.6\times10^{-5}X_5 + 0.002067X_6 - 3.732656 \end{aligned} \tag{2.9}$$

Recall that $X_1$ = PATth, $X_2$ = GDPpc, $X_3$ = PECpc, $X_4$ = GREpc, $X_5$ = URDpsk and $X_6$ = URGpor. Equations in (2.7) express the scores in terms of scaled original variables and equations in (2.9) are the scores expressed in terms of their scaled versions.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| PATth | -0.00 | -0.07 | -0.08 | -0.04 | 0.10 | -0.00 |
| GDPpc | -0.07 | 0.00 | -0.04 | -0.00 | -0.03 | 0.05 |
| PECpc | -0.08 | -0.04 | -0.02 | 0.02 | -0.00 | 0.09 |
| GREpc | -0.04 | -0.00 | 0.02 | 0.51 | 0.02 | 0.03 |
| URDpsk | 0.10 | -0.03 | -0.00 | 0.02 | 0.00 | -0.07 |
| URGpor | -0.00 | 0.05 | 0.09 | 0.03 | -0.07 | -0.49 |

**Table 2.8:** Differences between the correlations matrix $\tilde{R}$ reproduced from CFA model and the observed correlations matrix in the complete data file $R$.

### 2.3.1 Transforming scores. The Laplace distribution

It is well known that the Box-Cox $\{\varphi_\lambda, \lambda\}$ family of functions (Box and Cox, 1964) can be helpful to reduce the skewness in data and, moreover, correct some of the effects due to the lack of normality. The function `powerTransform()` in R-library `car` (see Fox and Weisberg, 2011) provides tools to estimate the optimal value of the parameter $\lambda$. For that, it is necessary to avoid that the variables take negative values. In our setting, a (+10) translation assuring positive scores is applied to each factor scores $F_j$, followed by a Box-Cox transformation with an optimal value $\lambda_j$ giving rise to the transformed scores $tF_j$, for $j = 1, 2, 3$:

$$tF_j = \frac{(F_j + 10)^{\lambda_j} - 1}{\lambda_j}, \qquad \lambda_1 \approx -0.964,\ \lambda_2 \approx -4.505,\ \lambda_3 \approx -3.933. \tag{2.10}$$

The marginal distributions of the transformed scores seem to adjust better to the Laplace or double exponential distribution than to the Gaussian, as it can be seen in Figure 2.3, and the skewness is clearly reduced. Recall that Laplace density and distribution fuctions are, rescpectively:

$$f(x) = \frac{1}{2\beta}\exp\left(-\frac{|x-m|}{\beta}\right) \qquad \Phi(x) = \frac{1}{2} + \frac{1}{2}\text{sgn}(x-m)\left(1 - \exp\left(-\frac{|x-m|}{\beta}\right)\right) \tag{2.11}$$

Given a sample $y_1, \ldots, y_n$, the Maximum Likelihood Estimators (MLE) of the parameters $m$ and $\beta$ are the sample median (50th-percentile) and the mean absolute deviations from the sample, respectively:

$$\widehat{m} = C_{50} \qquad \widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{m}|$$

Using these estimates for each transformed factor $tF_j$, the Laplace density is represented over the histograms of $tF_j$ in the pictures on the right hand side of Figure 2.3. Compared to the adjust of scores $F_j$ to the Gaussian (left left side) and the adjust of transformed scores $tF_j$ to the Gaussian, the Laplace low gives clearly better fits to $tF_j$, for each $j$. Elsewhere, the corresponding Laplace distribution function $\Phi_j$ is compared to the empirical distribution function (ecdf) of the same transformed scores (see Figure 2.4).

### 2.3.2 Scores of imputed data

Once the model has been identified and its parameters estimated, the formula (2.8) and/or the formula (2.9) can be used to compute the score values for any complete data row vector, whatever these data were complete in the original data file or completed after imputation. Notice that the imputed cases are not used to define nor to estimate the model, onli to obtain predictions for them. In this bloc, we proceed to obtain the scores for the imputed-complete cases in dataframe `dt`.

| | Factor1 | Factor2 | Factor3 |
|---|---|---|---|
| AT111 2003 NMR | -0.5 | -0.6 | -0.7 |
| AT111 2006 NMR | -0.6 | -0.7 | -0.6 |
| AT111 2007 NMR | -0.5 | -0.7 | -0.5 |
| AT111 2008 NMR | -0.5 | -0.6 | -0.6 |

**Table 2.9:** Header of scores in the three factors CFA model, for the imputed-complete data file.

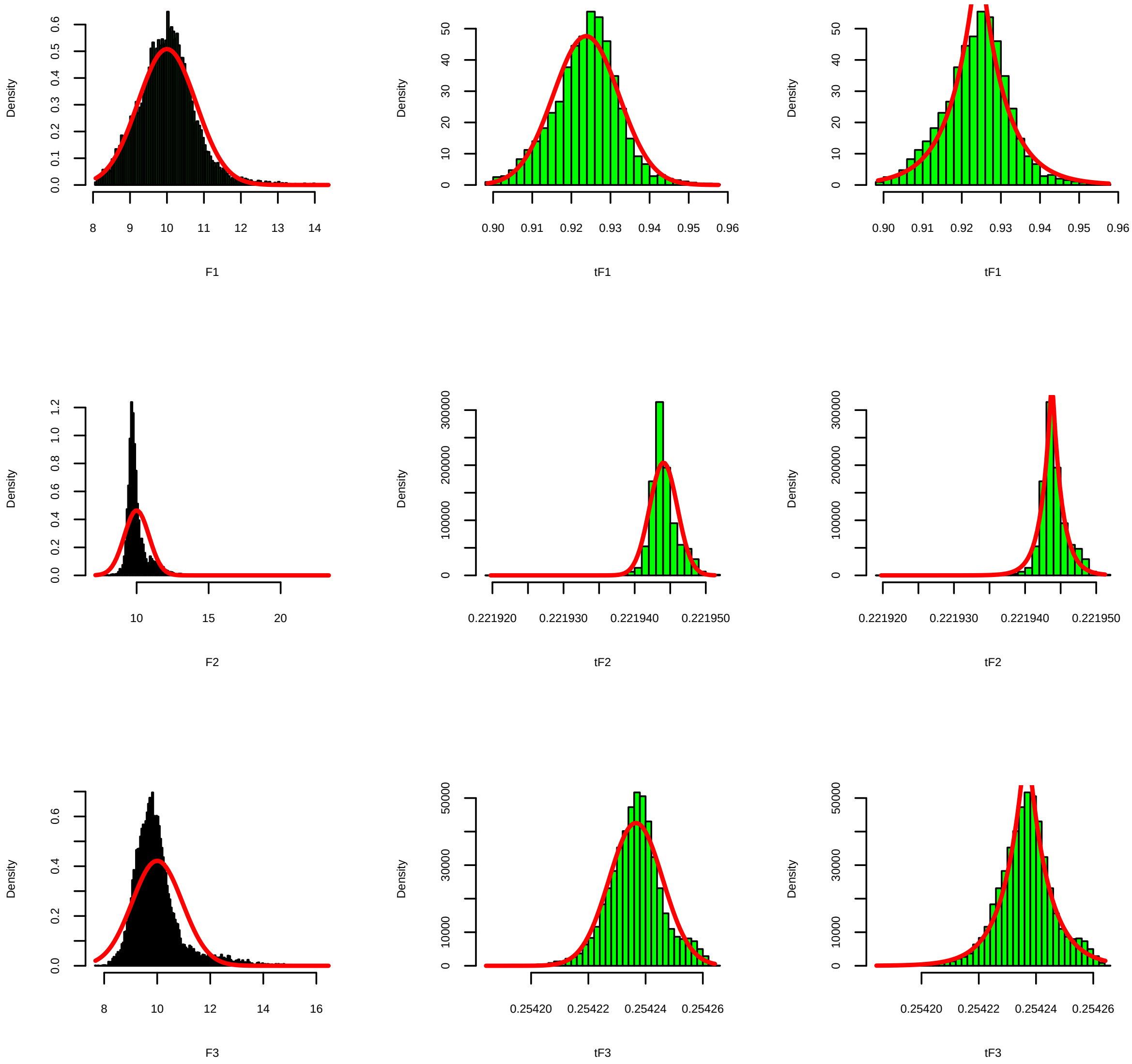


**Figure 2.3:** From top to bottom, for $j = 1, 2, 3$, in each row of the picture we show the histograms of $F_j$ (left) and $tF_j$ (center) jointly with Gaussian density. On the right, the best fit correspond to the Laplace density adjusted to the histogram of $tF_j$.

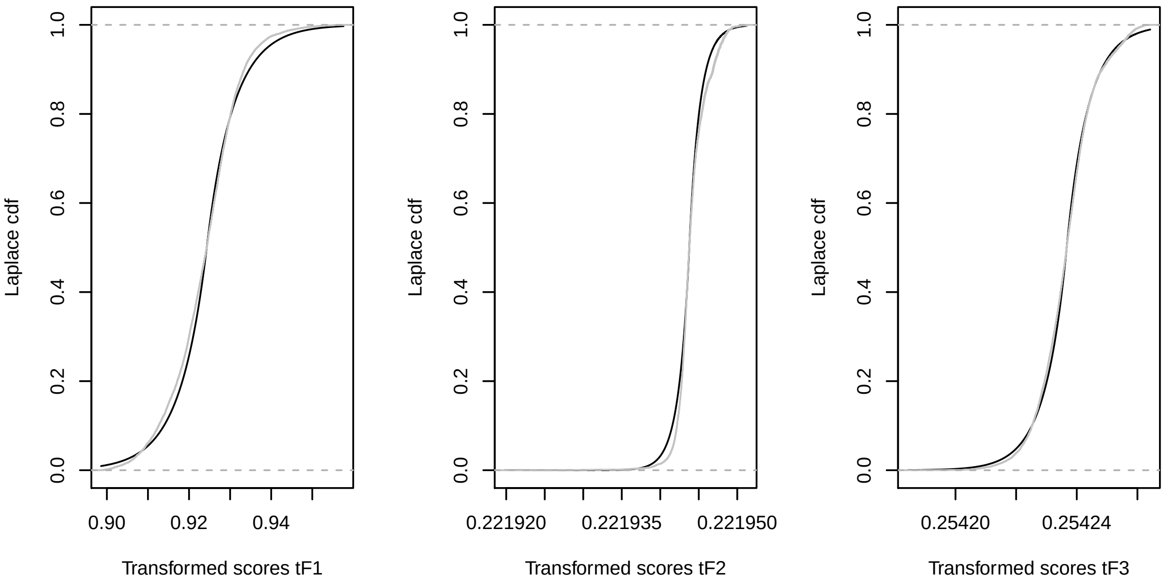


**Figure 2.4:** Cummulative Laplace distribution function (black line) and empirical distribution function (gray line) for the transformed factors $tF_j$, $j = 1, 2, 3$.

### 2.3.3 Scores of megaregions and countries

The scores can also be computed in megaregions and countries (considered as the weighted means of their NUTS3) in an specific year or aggregating years, using the data files described in Chapter 1. Taking these values in the original variables, we project megaregions and countries into the factorial space in two steps: Firstly, scaling these data with respect to the center and the standard deviation of the complete data set and, secondly, applying formula (2.8) or (2.9) to obtain de scores $F_j$ on the common factors.

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| NMR | -0.5 | -0.8 | -0.3 | -0.6 | -0.4 | -1.5 |
| VIB | -0.6 | -0.6 | -0.9 | -0.3 | -0.3 | -0.0 |
| FRG | 1.5 | 0.8 | -0.3 | 0.6 | -0.2 | 0.3 |
| AMB | 0.1 | 0.3 | -0.3 | -0.3 | -0.2 | 0.6 |
| PRA | -0.2 | -0.2 | -0.3 | 0.5 | -0.3 | 0.4 |
| BER | 0.1 | -0.0 | -1.2 | -0.0 | 0.2 | -0.7 |
| LIS | -0.7 | -0.7 | -0.7 | -0.2 | -0.4 | 0.2 |
| MAD | -0.6 | 0.6 | -1.3 | 0.4 | 0.0 | -0.3 |
| BAL | -0.2 | 0.1 | -0.4 | -0.2 | -0.3 | -0.3 |
| PAR | 0.4 | 1.3 | -1.1 | 0.2 | -0.1 | 0.5 |
| RMT | -0.4 | -0.5 | -0.7 | -0.8 | -0.3 | 0.0 |
| LON | -0.3 | -0.2 | -0.4 | 0.1 | -0.2 | 0.5 |
| GLB | -0.5 | 0.0 | -0.2 | 0.1 | -0.2 | -0.9 |

**Table 2.10:** Scores in megaregions (years aggregated).

| | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor |
|---|---|---|---|---|---|---|
| AT | 0.0 | -0.2 | -0.4 | 0.4 | -0.4 | -0.8 |
| BE | -0.1 | 0.3 | 0.2 | -1.1 | -0.2 | 0.7 |
| BG | -0.8 | -1.7 | -1.0 | 0.0 | -0.2 | -2.1 |
| CY | -0.7 | -0.4 | -0.5 | 0.2 | -0.4 | -0.8 |

**Table 2.11:** Header of scores in countries (years aggregated).

# Chapter 3

# Probability-based undices

In this chapter, we construct indices in basis of the probability distribution function of one or many random variables jointly evaluated on the same cases, provided that the distribution belongs to a parametric family of densities (Gaussian, Laplace or another continous family). The parameters of the distribution will be estimated from a set of observations $y_1, \ldots, y_n$ where these variables are jointly measured. Indices will be standard (having no units) and will take values in (0,1). Every index will define a total order on the cases and satisfy other expected properties like monotonicity.

Consider any random variable, say $Y$. Assume that the law of $Y$ belongs to a parametric continous family of laws and that the parameters are estimated from the sample, if necessary. Say $\Phi$ the estimated distribution function of $Y$. In Chapter 4, indices apply to the transformed factors $Y = tF_j, j = 1, 2, 3$, defined in Chapter 2. The corresponding transformed scores $\{tf_{j1}, \ldots, f_{jn}\}$ will be considered the sample of $tF_j$.

The natural order in $\mathbf{R}$, defines an obvious total ordering between any pair of cases, in this way:

$$i \prec j \text{ if and only if } y_i \leq y_j. \tag{3.1}$$

Assume that scores indicate best performance of the case as greater their value is. The downside is that the $y_j$ are usually unbounded and isolated values can not be interpreted. To circumvent this problem, any bounded monotonic function should be applied. We propose using the transformation given by the distribution function $\Phi$ of the random variable $Y$. As the distribution function is increasing, we have:

$$i \prec j \text{ if and only if } y_i \leq y_j \text{ if and only if } \Phi(y_i) \leq (y_j). \tag{3.2}$$

This function is intrinsic to the observations and the nature of the variable. The value $\Phi(y_i)$ is the cummulative percentage of cases that are smaller or equal to $y_i$. In this way, it is the *percentile position* of the case acording to the sample distribution.

## 3.1 Basic and weighted indices: Definition and properties

*Definition 1.* Given any random variable $Y$ taking real values, assume that its law belongs to a known family of laws. We define index $I$ by means of the distribution function $\Phi$ of $Y$:

$$I \colon = \Phi(F) \tag{3.3}$$

The index $I$ takes values between 0 and 1 and preserves the order relation. Moreover, as close to 1 as better the position is and, conversely, values near 0 indicate bad positionings. Notice that the index $I$ is a random variable following a uniform distribution in the unit interval. If $\Phi$ is not known but it belongs to some parametric family, parameters can be estimated from a sample.

In factorial analysis, often factorial models involve more than one factor. Consider two factors $F_1, F_2$ jointly evaluated on the same cases and denote $I_1$ and $I_2$ the corresponding indices in Definition 1. The goal could be indexing the cases by some kind of ordering taking into account both $I_1$ and $I_2$, in a way that priorizes the first index $I_1$, but penalizes negatively cases showing low values in $I_2$ and positively cases taking high values on the second index. In this way, we suggest a weighted-averaged index depending on a weight parameter $w \in [0, \frac{1}{2}]$ which expresses the *penalty degree.*

*Definition 2.* Given any pair of random variables $Y_1$ and $Y_2$ taking real values and laws belonging to some known family of distributions, we define a weighted index $I_{12}^w$ depending on their distributions functions $\Phi_1$, $\Phi_2$ and $w \in [0, \frac{1}{2}]$:

$$I_{12}^w : = \Phi_1(Y_1) - w(\Phi_1(Y_1) - \Phi_2(Y_2)) = (1-w)\Phi_1(Y_1) + w\Phi_2(Y_2) = (1-w)I_1 + wI_2. \tag{3.4}$$

*Remark 1.* Notice that, for any $w \in [0, 1]$, the right hand side in (3.4) is a convex combination of $I_1 = \Phi(Y_1)$ and $I_2 = \Phi(Y_2)$, thus giving intermediate values. The constraint $w \le \frac{1}{2}$ implies $(1-w) \ge w$ and expresses the idea that $Y_1$ is the primer factor to define the ordering and $Y_2$ plays a secondary role, the role being paritary if $w = \frac{1}{2}$.

In a more general way, more than two factors can be combined to define a weighted bounded index, taking values in $[0, 1]$.

*Definition 3.* Given random variables $Y_1, \ldots, Y_k$ taking real values, consider the index $I_{1\ldots k}^w$ defined by the distribution functions $\Phi_1, \ldots, \Phi_k$ and scalars $w = (w_1, \ldots, w_k)$, $w_j \in [0, 1]$ and satisfying $\sum_{j=1}^k w_j = 1$:

$$I_{1\ldots k}^w : = \sum_{j=1}^k w_j \Phi_j(Y_j) = \sum_{j=1}^k w_j I_j. \tag{3.5}$$

*Remark 2.* Notice that the index in (3.4) is a particular case of this general index, with $k = 2$, $w_1 = 1-w$ and $w_2 = w$. Another example, with $k = 2$, $w_1 = w_2 = w_3 = \frac{1}{3}$, is

$$I_{123}^w : = \frac{1}{3}\left(\Phi_1(Y_1) + \Phi_2(Y_2) + \Phi_3(Y_3)\right) = \frac{1}{3}\left(I_1 + I_2 + I_3\right). \tag{3.6}$$

A new variant of index can be derived from a given set of indices $I_1, \ldots, I_m$: Take a finite set of descriptives of position (median, mode, median, among others) of them and define weighted averages of these descriptives (see the next definition).

*Definition 4.* Assume that indices $I_1, \ldots, I_m$ are given and that we take $k$ descriptives of the indices $(k \le m)$, say $\varphi_j = \varphi_j(I_1, \ldots, I_m)$, $j = 1, \ldots, k$. Take scalars $w = (w_1, \ldots, w_k)$, $w_j \in [0, 1]$ and satisfying $\sum_{j=1}^k w_j = 1$. Then, we define:

$$I_{1\ldots k}^{\varphi, w} : = \sum_{j=1}^k w_j \varphi_j. \tag{3.7}$$

An example of this last type of indices $I_{1\ldots k}^{\varphi, w}$ is

$$\mathbf{I} = \left(\frac{1}{3} + 2\beta\right) \mathrm{Min}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\} + \frac{1}{3}\mathrm{Med}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\} + \left(\frac{1}{3} - 2\beta\right) \mathrm{Max}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\} \tag{3.8}$$

with $\beta \le \frac{1}{6}$. This indicator penalizes unbalanced values of the three basic indicators.

Some immediate relations can be easily shown.

*Properties.*

1. For any $w \in [0, 1]$, $I_{12}^w \in [0, 1]$.
2. For any $w \in [0, 1/2]$, for the cases satisfying $Y_2 \le Y_1$, then we have $I_{12}^w \le I_1$.
3. For any $w \in [0, 1/2]$, for the cases satisfying $Y_1 \le Y_2$, then we have $I_{12}^w \ge I_1$.
4. $I_{12}^w(y_1, y_2) \le I_{12}^w(y_1', y_2')$ if and only if $\frac{I_1(y_1) - I_1(y_1')}{I_2(y_2) - I_2(y_2')} \le \frac{1-w}{w}$.
5. $\forall \{w_j\}$, such that $\sum_{j=1}^k w_j = 1$, $I_{1\ldots k}^w \in [0, 1]$.
6. $\forall \{w_j\}$, such that $\sum_{j=1}^k w_j = 1$, $I_{1\ldots k}^{\varphi, w} \in [0, 1]$.

### 3.1.1 Comparison of basic and weighted indices

To understand better the use of indices defined in (3.3) and (3.4), a bivariate Gaussian sample of size 100 with scaled marginals and some amount of correlation ($\rho = 0.3$) is generated (see the head of the data in Table 3.1). We divide the sample into two subsamples, the first one is characterized by cases satisfying $Y_1 \leq Y_2$ and the second one is composed by cases satisfying $Y_1 > Y_2$. Some values of both subsamples, sorted in increasing order are shown in Table 3.2.

| | Y1 | Y2 |
|---|---|---|
| 1 | 0.90 | 0.97 |
| 2 | -0.31 | -1.00 |
| 3 | -0.36 | 1.15 |
| 4 | -0.12 | 0.45 |
| ⋮ | ⋮ | ⋮ |

**Table 3.1:** Header of a bvariate standard Gaussian sample of size 100 and correlation 0.3.

| | $Y_1$ | $Y_2$ |
|---|---|---|
| 87 | -1.82 | -0.84 |
| 97 | -1.81 | -1.73 |
| 41 | -1.81 | -0.39 |
| 91 | -1.53 | -1.16 |
| ⋮ | ⋮ | ⋮ |

| | $Y_1$ | $Y_2$ |
|---|---|---|
| 39 | -0.89 | -1.87 |
| 85 | -0.65 | -1.82 |
| 71 | -0.63 | -0.67 |
| 68 | -0.62 | -2.75 |
| ⋮ | ⋮ | ⋮ |

**Table 3.2:** Header of subsample $Y_1 \leq Y_2$ on the left hand side and subsample $Y_1 > Y_2$ on the right. Cases are sorted by $Y_1$ in increasing order.

At the top left corner of Figure 3.1, we represent index $I_1 = \Phi(Y_1)$, where $\Phi$ is the standard Gaussian cummulative distribution function and, from left to right and top to bottom, we show the effect of penalizing cases using index $I_{12}^w$, with weights $w = 0.1, 0.2, 0.3, 0.4, 0.5$, respectively. In Figure 3.2, the same indices ar shown together with the two subsamples behavior: the red points corresponding to cases with $Y_2 < Y_1$ are negatively penalized, and the blues correspond to $Y_2 \geq Y_1$ being positively penalized. The penalizing effects can be seen in $I_{12}^w$ compared to $I_1$ that projects in $(0, 1)$ without penality. In Figure 3.3 we appreciate more clearly the differences between a low penalizing effect $w = 0.1$, on the left, and a high penalizing effect $w = 0.5$, on the right.

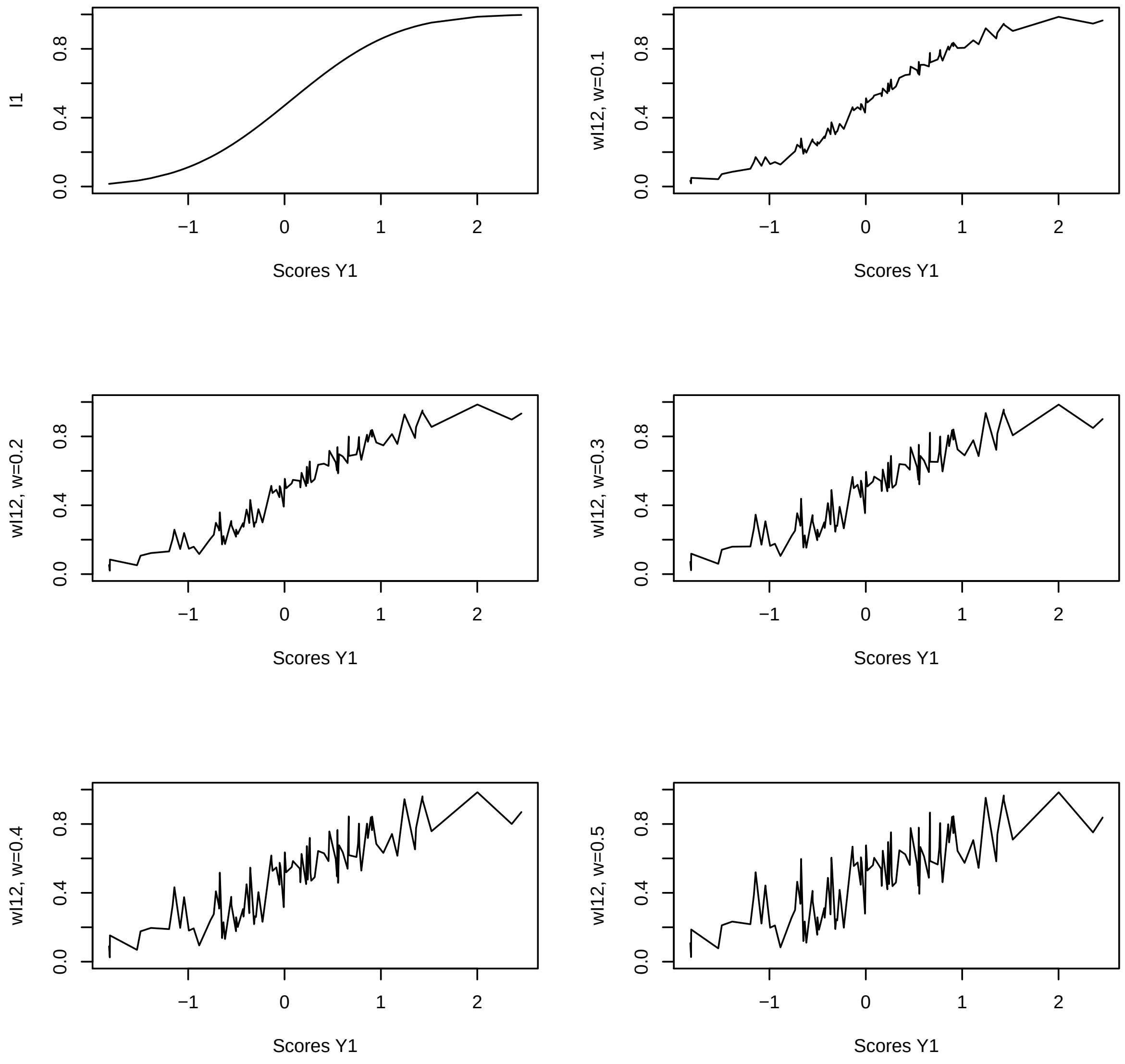


**Figure 3.1:** Index $I_1$ is represented at the top left corner. From top to bottom and left to right, $I_{12}^{w}$ modifies the ordering defined by $I_1$ (quantiles of $Y_1$) by the value of $I_2$ (quantiles of $Y_2$), with an increasing effect as $w$ increases.

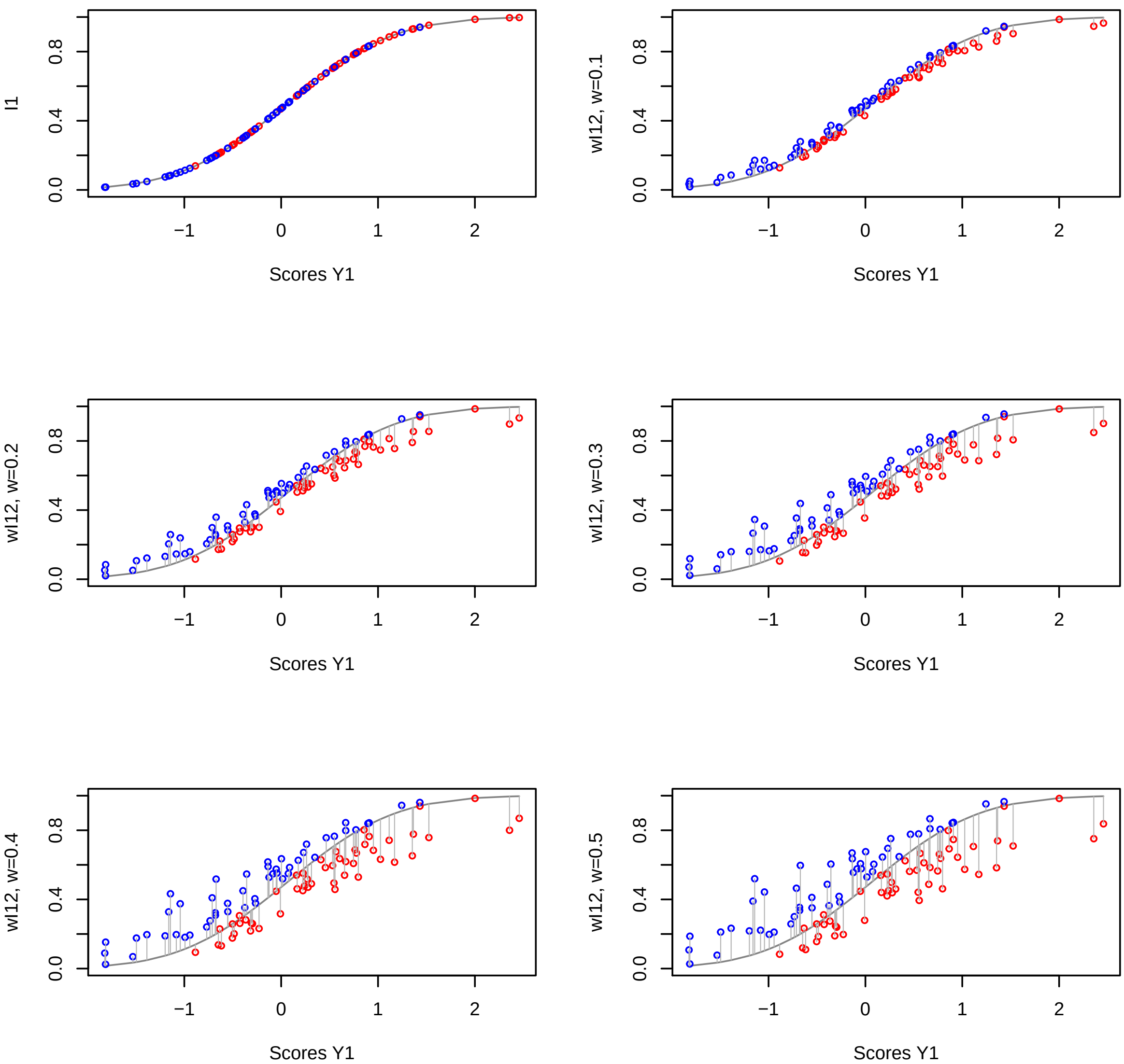


**Figure 3.2:** The same representation as in Figure 3.1. Here, the red points corresponding to $Y_2 < Y_1$ are negatively penalized and the blue points corresponding to $Y_2 \geq Y_1$ are positively penalized in $I_{12}^{w}$.

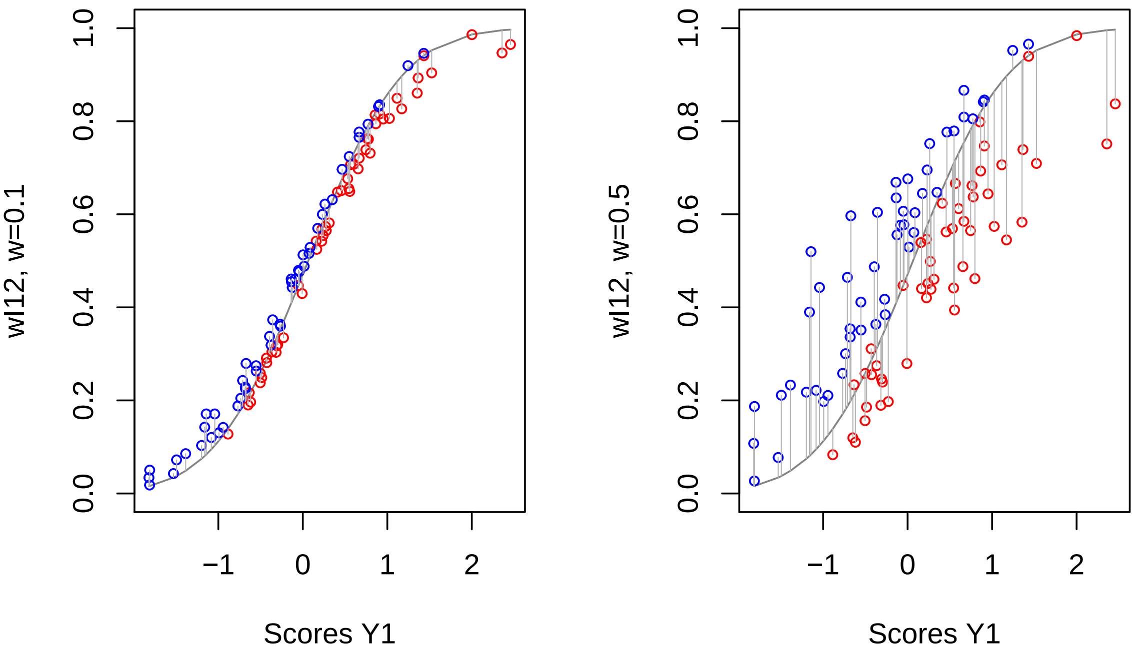


**Figure 3.3:** The effect of index $I_{12}^{w}$ with $w = 0.1$ on the left is compared wiht the one caused by the same index with $w = 0.5$.

# Chapter 4

# Indices to classify regions and megaregions

In Chapter 2 we compute the scores of the three factor CFA model, $F_1, F_2$ and $F_3$, for the individual NUTS3, for megaregions and for countries, every year or aggregated. We use the variables $Y_j = tF_j$, $j = 1, 2, 3$ defined in (2.10) and the Laplace distributions adjusted to the whole complete data set in that subsection to define several indices.

Following the development in Chapter 3, basic indices $\mathbf{I_1}$, $\mathbf{I_2}$ and $\mathbf{I_3}$ can be defined using the Laplace distribution function $\Phi$ adjusted to the transformed factors $tF_1, tF_2$ and $tF_3$. Moreover, we derive weighted convex combinations of these single indices that we call *scenarios.*

The scheme to obtain indices $\mathbf{I_j}$ can be summarized as follows:

$$\begin{array}{ccccccccc} X & \longrightarrow & F_j & \longrightarrow & F_j + c & \longrightarrow & tF_j & \longrightarrow & \mathbf{I_j} = \Phi(tF_j) \\ & \uparrow & & \uparrow & & \uparrow & & \uparrow & \\ \text{data} & \text{CFA} & \text{scores} & \text{translation} & & \text{Box-Cox} & \text{trsf.sco.} & \text{cdf } \Phi & \text{index} \end{array} \tag{4.1}$$

In our case,$j = 1, 2, 3$, the constant in the translation is $c = 10$ and $\Phi$ is the cummulative distribution function (cdf) of the Laplace distribution. Table 4.1 contains the head and the tail of the data file that includes initial variables, the scores in facto $F_1$, the transformed scores $tF_1$ and the values of index $\mathbf{I_1} = \Phi(tF_1)$. Cases are sorted in decreasing order with respect to $F_1$, which is equivalent to ordering with respect to $tF_1$, beacause the Box-Cox transformation is not decreasing, and equivalent to ordering with respect to $I_1$ too, being $\Phi$ an increasing function. Recall that $\mathbf{I_1}$ has the advantage of taking standard values between 0 and 1.

| NUTS3 | Year | COUn | MGAn | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor | $F_1$ | $tF_1$ | $\mathbf{I_1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DE252 | 2008 | DE | PRA | 1943.42 | 60543.89 | 2.17 | 89.65 | 1518.84 | 100 | 14.37 | 0.95761 | 0.99709 |
| NL414 | 2001 | NL | AMB | 3108.16 | 26425.28 | 4.61 | 54.76 | 497.36 | 100 | 14.32 | 0.95737 | 0.99698 |
| FR101 | 2008 | FR | PAR | 353.18 | 72335.45 | 0.14 | 85.11 | 22227.00 | 100 | 14.20 | 0.95669 | 0.99665 |
| DE252 | 2009 | DE | PRA | 1750.33 | 60550.81 | 1.91 | 91.23 | 1526.09 | 100 | 14.12 | 0.95627 | 0.99643 |
| FR101 | 2007 | FR | PAR | 351.38 | 69863.77 | 0.14 | 86.18 | 22022.00 | 100 | 14.07 | 0.95601 | 0.99628 |
| FR101 | 2009 | FR | PAR | 338.32 | 68541.31 | 0.12 | 83.80 | 22390.00 | 100 | 14.03 | 0.95575 | 0.99612 |
| DE252 | 2007 | DE | PRA | 1827.64 | 55930.90 | 2.18 | 87.70 | 1510.14 | 100 | 14.02 | 0.95570 | 0.99610 |
| FR101 | 2006 | FR | PAR | 350.50 | 68366.86 | 0.15 | 85.74 | 21872.00 | 100 | 14.00 | 0.95558 | 0.99602 |
| FR101 | 2010 | FR | PAR | 126.39 | 73515.07 | 0.13 | 83.53 | 22469.00 | 100 | 13.98 | 0.95546 | 0.99595 |
| FR101 | 2005 | FR | PAR | 351.01 | 67924.94 | 0.15 | 85.81 | 21768.00 | 100 | 13.97 | 0.95544 | 0.99594 |

| NUTS3 | Year | COUn | MGAn | PATth | GDPpc | PECpc | GREpc | URDpsk | URGpor | $F_1$ | $tF_1$ | $\mathbf{I_1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LV008 | 2005 | LV | NMR | 0.57 | 6449.63 | 2.71 | 42.54 | 211.06 | 8 | 8.11 | 0.89917 | 0.01028 |
| RO216 | 2006 | RO | NMR | 2.89 | 4048.29 | 1.07 | 45.29 | 932.45 | 9 | 8.11 | 0.89914 | 0.01022 |
| BG341 | 1999 | BG | NMR | 0.47 | 5515.63 | 2.06 | 62.69 | 542.53 | 10 | 8.11 | 0.89912 | 0.01020 |
| BG311 | 2001 | BG | NMR | 0.83 | 4360.33 | 1.54 | 29.10 | 494.76 | 9 | 8.10 | 0.89899 | 0.00999 |
| BG311 | 2006 | BG | NMR | 8.60 | 5193.47 | 1.97 | 33.53 | 642.54 | 6 | 8.10 | 0.89895 | 0.00994 |
| LT007 | 2002 | LT | NMR | 7.47 | 4977.58 | 1.56 | 41.33 | 390.09 | 8 | 8.09 | 0.89891 | 0.00987 |
| EE008 | 2002 | EE | NMR | 0.28 | 6726.44 | 3.80 | 36.29 | 371.35 | 6 | 8.09 | 0.89886 | 0.00979 |
| LV005 | 2006 | LV | NMR | 2.80 | 5688.69 | 1.57 | 39.31 | 453.30 | 5 | 8.09 | 0.89884 | 0.00977 |
| LV005 | 2005 | LV | NMR | 2.76 | 5144.99 | 1.61 | 35.46 | 386.45 | 7 | 8.08 | 0.89872 | 0.00959 |
| EE008 | 1995 | EE | NMR | 5.41 | 3657.02 | 3.97 | 40.03 | 150.77 | 15 | 8.07 | 0.89858 | 0.00938 |

**Table 4.1:** Header and tail of the complete data file adding the scores $F_1$, the transformed scores $tF_1$ and the index $\mathbf{I_1}$ . Cases are ordered by $\mathbf{I_1}$ , (equivalently by either $F_1$ or $tF_1$) in decreasing order.

## 4.1 Probability-based indices on NUTS3

Considering now the imputed data file, **dt**, we obtain the values of the three indices $\mathbf{I_1}$, $\mathbf{I_2}$ and $\mathbf{I_1}$ for the whole set of imputed data using the scheme above. Several tabular and graphic summaries can be performed, for instance, Table 4.2 relating the indices of Barcelona in different years to minimum, medium and maximum values of these indices.

| | Position | identif1 | $\mathbf{I_1}$ | identif2 | $\mathbf{I_2}$ | identif3 | $\mathbf{I_3}$ |
|---|---|---|---|---|---|---|---|
| 1 | Barcelona | ES511 1995 BAL | 0.3296 | ES511 1995 BAL | 0.8451 | ES511 1995 BAL | 0.2211 |
| 2 | Min | EE004 1995 NMR | 0.0076 | BE342 1995 AMB | 0.0002 | ITI45 1995 RMT | 0.0002 |
| 3 | Med | DE127 1995 NMR | 0.3542 | BE251 1995 AMB | 0.4052 | DE276 1995 FRG | 0.2422 |
| 4 | Max | UKI11 1995 LON | 0.9915 | FR101 1995 PAR | 0.9978 | UKI11 1995 LON | 0.9855 |
| 5 | Barcelona | ES511 2000 BAL | 0.5624 | ES511 2000 BAL | 0.8711 | ES511 2000 BAL | 0.7329 |
| 6 | Min | LT007 2000 NMR | 0.0099 | BE342 2000 AMB | 0.0000 | PL631 2000 NMR | 0.0003 |
| 7 | Med | DE936 2000 AMB | 0.4351 | DEA38 2000 AMB | 0.4810 | BE324 2000 AMB | 0.4404 |
| 8 | Max | UKI11 2000 LON | 0.9985 | FR101 2000 PAR | 0.9979 | UKI11 2000 LON | 0.9920 |
| 9 | Barcelona | ES511 2005 BAL | 0.6518 | ES511 2005 BAL | 0.8979 | ES511 2005 BAL | 0.8310 |
| 10 | Min | LV005 2005 NMR | 0.0096 | BE342 2005 AMB | 0.0000 | UKM63 2005 NMR | 0.0015 |
| 11 | Med | DEA28 2005 AMB | 0.4761 | BE242 2005 AMB | 0.5261 | AT124 2005 NMR | 0.4956 |
| 12 | Max | UKI11 2005 LON | 0.9992 | FR101 2005 PAR | 0.9979 | UKI11 2005 LON | 0.9928 |
| 13 | Barcelona | ES511 2010 BAL | 0.6781 | ES511 2010 BAL | 0.8938 | ES511 2010 BAL | 0.7974 |
| 14 | Min | LV009 2010 NMR | 0.0148 | BE342 2010 AMB | 0.0004 | HU313 2010 NMR | 0.0027 |
| 15 | Med | DE939 2010 AMB | 0.4966 | DE923 2010 AMB | 0.5051 | DEG0F 2010 PRA | 0.5637 |
| 16 | Max | UKI11 2010 LON | 0.9992 | FR101 2010 PAR | 0.9979 | UKI11 2010 LON | 0.9929 |

**Table 4.2:** Unweighted basic indices $\mathbf{I_1}$, $\mathbf{I_2}$ i $\mathbf{I_3}$: Barcelona compared to the nuts that reach the maximum, the median and the minimum of the whole imputed data set.

## 4.2 Weighted probability-based indices on NUTS3

Weighted general indices are defined in (3.5) and (3.8). As we are dealing with three factors, our interest is in indices **I** of the following type:

$$\mathbf{I} = \omega_1 \mathbf{I_1} + \omega_2 \mathbf{I_2} + \omega_3 \mathbf{I_3}$$

where $w_1 + w_2 + w_3 = 1$. In the sequel, we have chosen specific values of $w_i$ to illustrate the results (see the maps in the next chapter), but other selections could be done if interested the researchers.

We call *scenarios* and denote they $\mathbf{S_k}$ the weighted indices that we define in our theoretical approach to sustainability:

1. Taking $\omega_1 = 0.8$, $\omega_2 = 0.1$ and $\omega_3 = 0.1$, define

$$\mathbf{S_1} = 0.8\mathbf{I_1} + 0.1\mathbf{I_2} + 0.1\mathbf{I_3}$$

   $\mathbf{S_1}$ priorizes the economic component (80%) and gives a low (but unvanishing) weight (10%) to both the environmental and the social components.

2. Taking $\omega_1 = 0$, $\omega_2 = 0.2$ and $\omega_3 = 0.8$, define:

$$\mathbf{S_2} = 0.2\mathbf{I_2} + 0.8\mathbf{I_3}$$

   It priorizes the environmental component (80%) and gives a low weight or penalitation (20%) to the social component.

3. Taking $\omega_1 = 0$, $\omega_2 = 0.2$ and $\omega_3 = 0.8$, define:

$$\mathbf{S_3} = 0.8\mathbf{I_2} + 0.2\mathbf{I_3}$$

   It is similar to the previous one, priorizing the social component (80%) and giving a low weight (20%) to the environmental component.

| | Position | identif1 | $\mathbf{S_1}$ | identif2 | $\mathbf{S_2}$ | identif3 | $\mathbf{S_3}$ |
|---|---|---|---|---|---|---|---|
| 1 | Barcelona | ES511 1995 BAL | 0.3703 | ES511 1995 BAL | 0.3459 | ES511 1995 BAL | 0.7203 |
| 2 | Min | FI1D7 1995 NMR | 0.0104 | BE334 1995 AMB | 0.0090 | BE353 1995 AMB | 0.0126 |
| 3 | Med | NL230 1995 AMB | 0.3450 | SE311 1995 NMR | 0.2766 | SI012 1995 NMR | 0.3828 |
| 4 | Max | UKI11 1995 LON | 0.9914 | UKI11 1995 LON | 0.9876 | UKI11 1995 LON | 0.9940 |
| 5 | Barcelona | ES511 2000 BAL | 0.6104 | ES511 2000 BAL | 0.7605 | ES511 2000 BAL | 0.8435 |
| 6 | Min | UKM63 2000 NMR | 0.0268 | BE342 2000 AMB | 0.0066 | BE342 2000 AMB | 0.0017 |
| 7 | Med | DE401 2000 BER | 0.4305 | ES611 2000 NMR | 0.4448 | AT311 2000 NMR | 0.4721 |
| 8 | Max | UKI11 2000 LON | 0.9977 | UKI11 2000 LON | 0.9930 | UKI11 2000 LON | 0.9958 |
| 9 | Barcelona | ES511 2005 BAL | 0.6944 | ES511 2005 BAL | 0.8444 | ES511 2005 BAL | 0.8846 |
| 10 | Min | UKM63 2005 NMR | 0.0400 | BE334 2005 AMB | 0.0148 | BE342 2005 AMB | 0.0167 |
| 11 | Med | AT314 2005 NMR | 0.4636 | DEA1B 2005 AMB | 0.4970 | UKL13 2005 LON | 0.5047 |
| 12 | Max | UKI11 2005 LON | 0.9983 | UKI11 2005 LON | 0.9936 | UKI11 2005 LON | 0.9962 |
| 13 | Barcelona | ES511 2010 BAL | 0.7116 | ES511 2010 BAL | 0.8167 | ES511 2010 BAL | 0.8745 |
| 14 | Min | EE006 2010 NMR | 0.0360 | BE345 2010 AMB | 0.0169 | BE353 2010 AMB | 0.0182 |
| 15 | Med | UKF15 2010 LON | 0.4802 | DE931 2010 NMR | 0.5396 | ITI18 2010 RMT | 0.4999 |
| 16 | Max | UKI11 2010 LON | 0.9984 | UKI11 2010 LON | 0.9938 | UKI11 2010 LON | 0.9963 |

**Table 4.3:** Weighted indices (scenarios) $\mathbf{S_1}$, $\mathbf{S_2}$ i $\mathbf{S_3}$: Barcelona compared to the nuts that reach the maximum, the median and the minimum of the whole imputed data set.

4. $\mathbf{S_{4.1}}$: is a weighted balanced index, with $\omega_1 = \frac{1}{3}$, $\omega_2 = \frac{1}{3}$ i $\omega_3 = \frac{1}{3}$:

$$\mathbf{S_{4.1}} = \frac{1}{3}\mathbf{I_1} + \frac{1}{3}\mathbf{I_2} + \frac{1}{3}\mathbf{I_3}$$

   It gives the same weight (33.3%) to any component.

5. $\mathbf{S_{4.2}}$ is a composed index of type (3.8), with $\beta = \frac{1}{12}$

$$\mathbf{S_{4.2}} = \frac{1}{2}\text{Min}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\} + \frac{1}{3}\text{Med}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\} + \frac{1}{6}\text{Max}\left\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}\right\}$$

   It penalizes cases showing unbalanced equilibrium between economic, social and environmental indices.

| | Position | identif1 | $\mathbf{S_{4.1}}$ | identif2 | $\mathbf{S_{4.2}}$ |
|---|---|---|---|---|---|
| 1 | Barcelona | ES511 1995 BAL | 0.4653 | ES511 1995 BAL | 0.3613 |
| 2 | Min | FI1D7 1995 NMR | 0.0166 | FI1D7 1995 NMR | 0.0126 |
| 3 | Med | UKF16 1995 LON | 0.3330 | EL134 1995 NMR | 0.2725 |
| 4 | Max | UKI11 1995 LON | 0.9910 | UKI11 1995 LON | 0.9893 |
| 5 | Barcelona | ES511 2000 BAL | 0.7222 | ES511 2000 BAL | 0.6707 |
| 6 | Min | UKM63 2000 NMR | 0.0256 | UKM63 2000 NMR | 0.0175 |
| 7 | Med | DEB3E 2000 FRG | 0.4439 | DE12A 2000 FRG | 0.3766 |
| 8 | Max | UKI11 2000 LON | 0.9958 | UKI11 2000 LON | 0.9947 |
| 9 | Barcelona | ES511 2005 BAL | 0.7936 | ES511 2005 BAL | 0.7526 |
| 10 | Min | UKM63 2005 NMR | 0.0541 | UKM63 2005 NMR | 0.0332 |
| 11 | Med | UKL15 2005 LON | 0.4679 | NL132 2005 AMB | 0.3922 |
| 12 | Max | UKI11 2005 LON | 0.9963 | UKI11 2005 LON | 0.9953 |
| 13 | Barcelona | ES511 2010 BAL | 0.7898 | ES511 2010 BAL | 0.7539 |
| 14 | Min | EE006 2010 NMR | 0.0659 | EE006 2010 NMR | 0.0493 |
| 15 | Med | DE258 2010 PRA | 0.4732 | ITF14 2010 RMT | 0.4090 |
| 16 | Max | UKI11 2010 LON | 0.9964 | UKI11 2010 LON | 0.9954 |

**Table 4.4:** Weighted indices (scenarios) $\mathbf{S_{4.1}}$ and $\mathbf{S_{4.2}}$: Barcelona compared to the nuts that reach the maximum, the median and the minimum of the whole imputed data set.

## 4.3 Indices and *scenarios* on megaregions

Averaged weighted values of variables in megaregions, computed as described in Section 1.2, can be used to project the scores and the indices, in specific years or aggregated over time. The results are summarized in several tables.

| | MGA | $\mathbf{I_1}$ | identif2 | $\mathbf{I_2}$ | identif3 | $\mathbf{I_3}$ |
|---|---|---|---|---|---|---|
| 1 | NMR 1995 | 0.0480 | NMR 1995 | 0.3392 | NMR 1995 | 0.0186 |
| 2 | VIB 1995 | 0.1359 | VIB 1995 | 0.6180 | VIB 1995 | 0.3317 |
| 3 | FRG 1995 | 0.2368 | FRG 1995 | 0.4063 | FRG 1995 | 0.2969 |
| 4 | AMB 1995 | 0.5416 | AMB 1995 | 0.2068 | AMB 1995 | 0.0075 |
| 5 | PRA 1995 | 0.3267 | PRA 1995 | 0.5539 | PRA 1995 | 0.8172 |
| 6 | LIS 1995 | 0.2168 | LIS 1995 | 0.4530 | LIS 1995 | 0.1485 |
| 7 | MAD 1995 | 0.4525 | MAD 1995 | 0.8269 | MAD 1995 | 0.2733 |
| 8 | BAL 1995 | 0.2661 | BAL 1995 | 0.5573 | BAL 1995 | 0.1843 |
| 9 | PAR 1995 | 0.7643 | PAR 1995 | 0.8448 | PAR 1995 | 0.6494 |
| 10 | LON 1995 | 0.4091 | LON 1995 | 0.5249 | LON 1995 | 0.4611 |
| 11 | GLB 1995 | 0.1447 | GLB 1995 | 0.5859 | GLB 1995 | 0.4552 |
| 12 | NMR 2000 | 0.0798 | NMR 2000 | 0.4798 | NMR 2000 | 0.2050 |
| 13 | VIB 2000 | 0.3076 | VIB 2000 | 0.4675 | VIB 2000 | 0.1272 |
| 14 | FRG 2000 | 0.8535 | FRG 2000 | 0.8387 | FRG 2000 | 0.8360 |
| 15 | AMB 2000 | 0.7430 | AMB 2000 | 0.6368 | AMB 2000 | 0.4413 |
| 16 | PRA 2000 | 0.5293 | PRA 2000 | 0.6076 | PRA 2000 | 0.7796 |
| 17 | BER 2000 | 0.5253 | BER 2000 | 0.9246 | BER 2000 | 0.6960 |
| 18 | LIS 2000 | 0.2381 | LIS 2000 | 0.4471 | LIS 2000 | 0.3506 |
| 19 | MAD 2000 | 0.6926 | MAD 2000 | 0.8592 | MAD 2000 | 0.8025 |
| 20 | BAL 2000 | 0.4419 | BAL 2000 | 0.6067 | BAL 2000 | 0.5445 |
| 21 | PAR 2000 | 0.8756 | PAR 2000 | 0.8760 | PAR 2000 | 0.8528 |
| 22 | RMT 2000 | 0.5566 | RMT 2000 | 0.6385 | RMT 2000 | 0.5140 |
| 23 | LON 2000 | 0.6042 | LON 2000 | 0.5724 | LON 2000 | 0.6791 |
| 24 | GLB 2000 | 0.2168 | GLB 2000 | 0.5573 | GLB 2000 | 0.5652 |
| 25 | NMR 2005 | 0.0940 | NMR 2005 | 0.5176 | NMR 2005 | 0.2705 |
| 26 | VIB 2005 | 0.3871 | VIB 2005 | 0.5671 | VIB 2005 | 0.2501 |
| 27 | FRG 2005 | 0.8937 | FRG 2005 | 0.8238 | FRG 2005 | 0.8476 |
| 28 | AMB 2005 | 0.7679 | AMB 2005 | 0.7015 | AMB 2005 | 0.7491 |
| 29 | PRA 2005 | 0.6029 | PRA 2005 | 0.6038 | PRA 2005 | 0.7933 |
| 30 | BER 2005 | 0.3752 | BER 2005 | 0.8784 | BER 2005 | 0.5515 |
| 31 | LIS 2005 | 0.2911 | LIS 2005 | 0.4687 | LIS 2005 | 0.5622 |
| 32 | MAD 2005 | 0.4337 | MAD 2005 | 0.8368 | MAD 2005 | 0.8617 |
| 33 | BAL 2005 | 0.5042 | BAL 2005 | 0.6545 | BAL 2005 | 0.6599 |
| 34 | PAR 2005 | 0.8585 | PAR 2005 | 0.8381 | PAR 2005 | 0.8344 |
| 35 | RMT 2005 | 0.5870 | RMT 2005 | 0.6487 | RMT 2005 | 0.6050 |
| 36 | LON 2005 | 0.5409 | LON 2005 | 0.6593 | LON 2005 | 0.7170 |
| 37 | GLB 2005 | 0.2937 | GLB 2005 | 0.6453 | GLB 2005 | 0.7387 |
| 38 | NMR 2010 | 0.1045 | NMR 2010 | 0.5449 | NMR 2010 | 0.3478 |
| 39 | VIB 2010 | 0.3244 | VIB 2010 | 0.6960 | VIB 2010 | 0.6626 |
| 40 | FRG 2010 | 0.8507 | FRG 2010 | 0.7729 | FRG 2010 | 0.8882 |
| 41 | AMB 2010 | 0.7584 | AMB 2010 | 0.6838 | AMB 2010 | 0.8124 |
| 42 | PRA 2010 | 0.5983 | PRA 2010 | 0.5714 | PRA 2010 | 0.8056 |
| 43 | BER 2010 | 0.4419 | BER 2010 | 0.8936 | BER 2010 | 0.7534 |
| 44 | LIS 2010 | 0.3864 | LIS 2010 | 0.4848 | LIS 2010 | 0.5324 |
| 45 | MAD 2010 | 0.4859 | MAD 2010 | 0.8445 | MAD 2010 | 0.8304 |
| 46 | BAL 2010 | 0.4162 | BAL 2010 | 0.6470 | BAL 2010 | 0.5783 |
| 47 | PAR 2010 | 0.8868 | PAR 2010 | 0.8206 | PAR 2010 | 0.8710 |
| 48 | RMT 2010 | 0.4933 | RMT 2010 | 0.6932 | RMT 2010 | 0.5977 |
| 49 | LON 2010 | 0.4119 | LON 2010 | 0.6585 | LON 2010 | 0.6397 |
| 50 | GLB 2010 | 0.2312 | GLB 2010 | 0.6462 | GLB 2010 | 0.6761 |

**Table 4.5:** Indices $\mathbf{I_1}$, $\mathbf{I_2}$ and $\mathbf{I_3}$, positioning every megaregion, in years 1995, 2000, 2005 and 2010.

Maps are included as added pages at the end, illustrating several 5-groups classification of regions. Each classification is given by an specific index. The five ordered groups correspond to the 5-quantile partition, one color for every group.

# References

Thomson, G. H. (1951), The Factorial Analysis of Human Ability. London University Press)

Box G. E. P. Box and Cox D. R., An analysis of transformations. Journal of the Royal Statistical Society Series B 26(2). 1964. S. 211–252.

lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. `http://www.jstatsoft.org/v48/i02/`.

John Fox and Sanford Weisberg (2011). An R Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`

Joan Marull, Vittorio Galletto, Elena Domene and Joan Trullén. Emerging megaregions: A new spatial scale to explore urban sustainability. Land Use Policy 34 (2013) 353-366.

Alan Bernardo. Development of Sustainable Macroeconomic Models and Indicators. Master Thesis. Master in Modelling for Science and Engineering, Dept. Math, UAB. 2017

| | MGA | $\mathbf{S_1}$ | identif2 | $\mathbf{S_2}$ | identif3 | $\mathbf{S_3}$ |
|---|---|---|---|---|---|---|
| 1 | NMR 1995 | 0.0742 | NMR 1995 | 0.0827 | NMR 1995 | 0.2751 |
| 2 | VIB 1995 | 0.2037 | VIB 1995 | 0.3890 | VIB 1995 | 0.5607 |
| 3 | FRG 1995 | 0.2598 | FRG 1995 | 0.3188 | FRG 1995 | 0.3844 |
| 4 | AMB 1995 | 0.4547 | AMB 1995 | 0.0474 | AMB 1995 | 0.1670 |
| 5 | PRA 1995 | 0.3985 | PRA 1995 | 0.7645 | PRA 1995 | 0.6066 |
| 6 | LIS 1995 | 0.2336 | LIS 1995 | 0.2094 | LIS 1995 | 0.3921 |
| 7 | MAD 1995 | 0.4721 | MAD 1995 | 0.3841 | MAD 1995 | 0.7162 |
| 8 | BAL 1995 | 0.2870 | BAL 1995 | 0.2589 | BAL 1995 | 0.4827 |
| 9 | PAR 1995 | 0.7609 | PAR 1995 | 0.6885 | PAR 1995 | 0.8058 |
| 10 | LON 1995 | 0.4259 | LON 1995 | 0.4738 | LON 1995 | 0.5121 |
| 11 | GLB 1995 | 0.2198 | GLB 1995 | 0.4814 | GLB 1995 | 0.5598 |
| 12 | NMR 2000 | 0.1323 | NMR 2000 | 0.2600 | NMR 2000 | 0.4248 |
| 13 | VIB 2000 | 0.3056 | VIB 2000 | 0.1953 | VIB 2000 | 0.3994 |
| 14 | FRG 2000 | 0.8503 | FRG 2000 | 0.8366 | FRG 2000 | 0.8382 |
| 15 | AMB 2000 | 0.7022 | AMB 2000 | 0.4804 | AMB 2000 | 0.5977 |
| 16 | PRA 2000 | 0.5621 | PRA 2000 | 0.7452 | PRA 2000 | 0.6420 |
| 17 | BER 2000 | 0.5823 | BER 2000 | 0.7417 | BER 2000 | 0.8789 |
| 18 | LIS 2000 | 0.2703 | LIS 2000 | 0.3699 | LIS 2000 | 0.4278 |
| 19 | MAD 2000 | 0.7203 | MAD 2000 | 0.8138 | MAD 2000 | 0.8478 |
| 20 | BAL 2000 | 0.4686 | BAL 2000 | 0.5569 | BAL 2000 | 0.5942 |
| 21 | PAR 2000 | 0.8734 | PAR 2000 | 0.8575 | PAR 2000 | 0.8713 |
| 22 | RMT 2000 | 0.5605 | RMT 2000 | 0.5389 | RMT 2000 | 0.6136 |
| 23 | LON 2000 | 0.6085 | LON 2000 | 0.6578 | LON 2000 | 0.5938 |
| 24 | GLB 2000 | 0.2857 | GLB 2000 | 0.5636 | GLB 2000 | 0.5588 |
| 25 | NMR 2005 | 0.1540 | NMR 2005 | 0.3199 | NMR 2005 | 0.4682 |
| 26 | VIB 2005 | 0.3914 | VIB 2005 | 0.3135 | VIB 2005 | 0.5037 |
| 27 | FRG 2005 | 0.8821 | FRG 2005 | 0.8429 | FRG 2005 | 0.8285 |
| 28 | AMB 2005 | 0.7594 | AMB 2005 | 0.7396 | AMB 2005 | 0.7110 |
| 29 | PRA 2005 | 0.6220 | PRA 2005 | 0.7554 | PRA 2005 | 0.6417 |
| 30 | BER 2005 | 0.4432 | BER 2005 | 0.6169 | BER 2005 | 0.8130 |
| 31 | LIS 2005 | 0.3360 | LIS 2005 | 0.5435 | LIS 2005 | 0.4874 |
| 32 | MAD 2005 | 0.5168 | MAD 2005 | 0.8567 | MAD 2005 | 0.8418 |
| 33 | BAL 2005 | 0.5348 | BAL 2005 | 0.6588 | BAL 2005 | 0.6555 |
| 34 | PAR 2005 | 0.8541 | PAR 2005 | 0.8351 | PAR 2005 | 0.8374 |
| 35 | RMT 2005 | 0.5950 | RMT 2005 | 0.6137 | RMT 2005 | 0.6399 |
| 36 | LON 2005 | 0.5703 | LON 2005 | 0.7054 | LON 2005 | 0.6708 |
| 37 | GLB 2005 | 0.3734 | GLB 2005 | 0.7200 | GLB 2005 | 0.6640 |
| 38 | NMR 2010 | 0.1729 | NMR 2010 | 0.3872 | NMR 2010 | 0.5055 |
| 39 | VIB 2010 | 0.3954 | VIB 2010 | 0.6692 | VIB 2010 | 0.6893 |
| 40 | FRG 2010 | 0.8467 | FRG 2010 | 0.8651 | FRG 2010 | 0.7959 |
| 41 | AMB 2010 | 0.7564 | AMB 2010 | 0.7867 | AMB 2010 | 0.7095 |
| 42 | PRA 2010 | 0.6164 | PRA 2010 | 0.7588 | PRA 2010 | 0.6182 |
| 43 | BER 2010 | 0.5182 | BER 2010 | 0.7814 | BER 2010 | 0.8656 |
| 44 | LIS 2010 | 0.4109 | LIS 2010 | 0.5229 | LIS 2010 | 0.4943 |
| 45 | MAD 2010 | 0.5562 | MAD 2010 | 0.8332 | MAD 2010 | 0.8417 |
| 46 | BAL 2010 | 0.4555 | BAL 2010 | 0.5920 | BAL 2010 | 0.6333 |
| 47 | PAR 2010 | 0.8786 | PAR 2010 | 0.8609 | PAR 2010 | 0.8307 |
| 48 | RMT 2010 | 0.5237 | RMT 2010 | 0.6168 | RMT 2010 | 0.6741 |
| 49 | LON 2010 | 0.4594 | LON 2010 | 0.6434 | LON 2010 | 0.6547 |
| 50 | GLB 2010 | 0.3172 | GLB 2010 | 0.6701 | GLB 2010 | 0.6522 |

**Table 4.6:** Indices $\mathbf{S_1}$, $\mathbf{S_2}$ and $\mathbf{S_3}$, positioning every megaregion, in years 1995, 2000, 2005 and 2010.

| | identif1 | $\mathbf{S_{4.1}}$ | identif2 | $\mathbf{S_{4.2}}$ |
|---|---|---|---|---|
| 1 | NMR 1995 | 0.1353 | NMR 1995 | 0.0818 |
| 2 | VIB 1995 | 0.3619 | VIB 1995 | 0.2815 |
| 3 | FRG 1995 | 0.3133 | FRG 1995 | 0.2851 |
| 4 | AMB 1995 | 0.2520 | AMB 1995 | 0.1630 |
| 5 | PRA 1995 | 0.5659 | PRA 1995 | 0.4842 |
| 6 | LIS 1995 | 0.2728 | LIS 1995 | 0.2220 |
| 7 | MAD 1995 | 0.5176 | MAD 1995 | 0.4253 |
| 8 | BAL 1995 | 0.3359 | BAL 1995 | 0.2737 |
| 9 | PAR 1995 | 0.7528 | PAR 1995 | 0.7203 |
| 10 | LON 1995 | 0.4650 | LON 1995 | 0.4457 |
| 11 | GLB 1995 | 0.3953 | GLB 1995 | 0.3217 |
| 12 | NMR 2000 | 0.2549 | NMR 2000 | 0.1882 |
| 13 | VIB 2000 | 0.3008 | VIB 2000 | 0.2441 |
| 14 | FRG 2000 | 0.8428 | FRG 2000 | 0.8398 |
| 15 | AMB 2000 | 0.6070 | AMB 2000 | 0.5567 |
| 16 | PRA 2000 | 0.6388 | PRA 2000 | 0.5971 |
| 17 | BER 2000 | 0.7153 | BER 2000 | 0.6488 |
| 18 | LIS 2000 | 0.3453 | LIS 2000 | 0.3104 |
| 19 | MAD 2000 | 0.7848 | MAD 2000 | 0.7570 |
| 20 | BAL 2000 | 0.5310 | BAL 2000 | 0.5035 |
| 21 | PAR 2000 | 0.8681 | PAR 2000 | 0.8643 |
| 22 | RMT 2000 | 0.5697 | RMT 2000 | 0.5489 |
| 23 | LON 2000 | 0.6186 | LON 2000 | 0.6008 |
| 24 | GLB 2000 | 0.4464 | GLB 2000 | 0.3883 |
| 25 | NMR 2005 | 0.2940 | NMR 2005 | 0.2234 |
| 26 | VIB 2005 | 0.4014 | VIB 2005 | 0.3486 |
| 27 | FRG 2005 | 0.8550 | FRG 2005 | 0.8434 |
| 28 | AMB 2005 | 0.7395 | AMB 2005 | 0.7284 |
| 29 | PRA 2005 | 0.6667 | PRA 2005 | 0.6349 |
| 30 | BER 2005 | 0.6017 | BER 2005 | 0.5178 |
| 31 | LIS 2005 | 0.4406 | LIS 2005 | 0.3955 |
| 32 | MAD 2005 | 0.7107 | MAD 2005 | 0.6394 |
| 33 | BAL 2005 | 0.6062 | BAL 2005 | 0.5802 |
| 34 | PAR 2005 | 0.8437 | PAR 2005 | 0.8396 |
| 35 | RMT 2005 | 0.6136 | RMT 2005 | 0.6033 |
| 36 | LON 2005 | 0.6390 | LON 2005 | 0.6097 |
| 37 | GLB 2005 | 0.5592 | GLB 2005 | 0.4851 |
| 38 | NMR 2010 | 0.3324 | NMR 2010 | 0.2590 |
| 39 | VIB 2010 | 0.5610 | VIB 2010 | 0.4991 |
| 40 | FRG 2010 | 0.8373 | FRG 2010 | 0.8181 |
| 41 | AMB 2010 | 0.7515 | AMB 2010 | 0.7301 |
| 42 | PRA 2010 | 0.6584 | PRA 2010 | 0.6194 |
| 43 | BER 2010 | 0.6963 | BER 2010 | 0.6210 |
| 44 | LIS 2010 | 0.4679 | LIS 2010 | 0.4435 |
| 45 | MAD 2010 | 0.7203 | MAD 2010 | 0.6605 |
| 46 | BAL 2010 | 0.5472 | BAL 2010 | 0.5087 |
| 47 | PAR 2010 | 0.8595 | PAR 2010 | 0.8484 |
| 48 | RMT 2010 | 0.5947 | RMT 2010 | 0.5614 |
| 49 | LON 2010 | 0.5700 | LON 2010 | 0.5289 |
| 50 | GLB 2010 | 0.5178 | GLB 2010 | 0.4437 |

**Table 4.7:** Indices $\mathbf{S_{4.1}}$ and $\mathbf{S_{4.2}}$, positioning every megaregion, in years 1995, 2000, 2005 and 2010.