Structural variables. Weighting the annual subsample
Summary of the process for calculating the weights

Abstract

This paper provides information about the weighting process in the Spanish LFS annual subsample. We describe the way to obtain the subsample, the calibration method employed and some comments about the results.

0.- The annual subsample

The annual subsample is selected from the Spanish LFS (EPA).

The complete LFS sample is composed of six waves (rotation groups), each of them also being a representative sample of the population.

The households of the sample are interviewed once per quarter, during six consecutive quarters.

All sample households of a specific rotation group are renewed each quarter, following the next scheme:

LFS Spain

		RG_1	RG_2	RG_3	RG_4	RG_5	RG_6	RG_1	RG_2	RG_3	RG_4	RG_5	RG_6	RG_1
					ROTATION GROUPS_Number of interview									
YEAR A	Q_1	6_Int	5_Int	4_Int	3_Int	2_Int	1_Int							
	Q_2		6_Int	5_Int	4_Int	3_Int	2_Int	1_Int						
	Q_3			6_Int	5_Int	4_Int	3_Int	2_Int	1_Int					
	Q_4				6_Int	5_Int	4_Int	3_Int	2_Int	1_Int				
YEAR A+1	Q_1					6_Int	5_Int	4_Int	3_Int	2_Int	1_Int			
	Q_2						6_Int	5_Int	4_Int	3_Int	2_Int	1_Int		
	Q_3							6_Int	5_Int	4_Int	3_Int	2_Int	1_Int	
	Q_4								6_Int	5_Int	4_Int	3_Int	2_Int	1_Int

The annual subsample is built with the four rotation groups, one per quarter, with households in their sixth and last participation in the survey. By construction, these four rotation groups have no overlap; therefore the theoretical subsample size is four-sixths of the LFS sample

1.- Weights

1.1 Before calibration

Weights entering the calibration process are calculated in the same way as in the quarterly LFS. Its general expression is:

$$\widehat{Y} = \sum_h \frac{P_h}{p_h} \sum_{i,j,k \in h} y_{ijk}$$

Where

P_h is the population aged 16 or over by stratum in the middle of the year. This information is provided by the INE Populations Unit

p_h is the population aged 16 or over by stratum in the subsample

h : stratum

ijk : PSU i, dwelling j, person k

Note. Strata are defined in the same way as in the LFS, i.e. inside NUTS3 (provinces) and by municipality size.

1.2.-Calibration

The calibration process is performed simultaneously at two geographical levels: National and NUTS2 .

Auxiliary variables at national level

- Employed persons by sex-and age groups (16-24; 25-34; 35-44; 45-54; 55 and over)

- Unemployed persons by sex and age groups (the same groups).

- Economically inactive persons by sex and age groups (the same groups).

Auxiliary variables at NUTS2 level

- Employed persons

- Unemployed persons

- Economically inactive persons

- Nationality (Two groups: Spanish, other)

Note. Spanish NUTS2 63 and 64, Ceuta and Melilla, are calibrated together, since they are too small to accept a separate calibration process.

Auxiliary variables vector

The sample records entering the calibration process are households.

Each record contains information on sample auxiliary variables regarding persons aged 16 and over residing in the sample household.

Since calibration variables are defined at two geographical levels, national and NUTS2, it is necessary to define a unique auxiliary vector including all the above variables in order to perform the calibration process in one step. By doing so, simultaneous calibration with all auxiliary variables is assured.

With this goal, auxiliary variables are defined by geographical level and population group. Each of the variables contain the number of people in the household who belong to this group.

Number of auxiliary variables

At national level there are 30 auxiliary variables.

At NUTS2 level there are 5 auxiliary variables, and 18 NUTS2 are used in the calibration process.

So that, in total each sample record has $30 + 5*18 = 120$ auxiliary variables.

Nevertheless some auxiliary variables must be removed in order to avoid linear combinations. (For example the sum of the number of employees per NUTS2 is the same as the total number of employees by age and sex at national level).

After removing the linear combinations, there are 99 auxiliary variables per record.

Population totals

In all these cases, population totals are obtained as the average of the LFS quarter estimations in the year.

Common weight in household members aged 16 and over

In order to ensure coherence between population and household data, within the calibration process, a common weight for all the persons aged 16 and over living in the

same dwelling is assigned. This common weight is also the weight for the household variables.

Procedure

The SAS macro CALMAR (Insee-Särndal, Deville, Sautory) is used in the calibration process, with parameter M=4, i.e. truncated linear method, with bounds LO=0,1 UP=10, in the same way as in LFS.

The main goal of these bounds is to ensure that the final weights are positive, and not too far from the starting weights.

Through this calibration process a household weight is obtained. This weight will also be the sample weight for people aged 16 and more residing in that dwelling

Variability of weights

As a consequence of the calibration process, an increase of the variability of the weights is observed. A summary of this variability rise is showed in the next table, where CV refers to the weights coefficient of variation

LFS. SUBSAMPLE 2013 WEIGHTS		
NUTS2	CV_PRECALMAR	CV_CALMAR
National	**74,6**	**90,6**
01	43,1	72,1
02	86,0	90,3
03	14,6	40,2
04	12,1	56,4
05	25,3	75,9
06	23,9	34,6
07	48,5	56,3
08	30,7	47,1
09	51,3	58,6
10	46,8	69,7
11	25,7	40,0
12	24,9	39,8
13	39,9	60,6
14	18,4	53,0
15	32,0	56,1
16	33,1	38,2
17	10,2	40,0
18	16,3	29,5

2.- Results

The annual sub-sample provides estimates that are consistent with the LFS results, in accordance with the requirements of the COMMISSION REGULATION (EC) No 377/2008.

As expected, consistency with the results of the LFS is complete in regard to the auxiliary variables used in the calibration process. Nevertheless, when increasing the disaggregation level of estimates, the differences are greater.

Results from annual subsample are disseminated at National and NUT2 levels.

Anonymized microdata files are provided by INE.