Active population flow statistics

Introduction

Under the denomination of "Active Population Flow Statistics" we have been considering all information regarding the monitoring of the population or specific groups of population in time with respect to their relationship with the labour market. This information links the situations of interviewee in two moments in time based on the Economically Active Population Survey, with the object of studying the interviewee's evolution.

Survey design is reviewed in sixths every quarter, keeping 5/6ths fixed between two consecutive quarters (4/6ths between one quarter and the subsequent one, etc.) and which includes among its questions the situation of the interviewee at a time in the past, which allows obtaining flow statistics in two different manners: that based on retrospective questions, with information regarding the two time references of interest compiled into a single quarter and exploited together with the remaining quarterly variables and that which makes use of the common part of the sample between two issues of the same survey. The latter requires compiling special files of flows that allow linking common records in different quarters.

Each one of these procedures has its advantages and its weaknesses. Nevertheless, the limitation of retrospective questions that would excessively overload the questionnaire, which, on the other hand, would be negatively influenced by the lack of memory, makes the analysis of the common part of the sample by means of special flow files the most interesting part of the survey.

The present publication focuses precisely on this last analysis. The reason for this, on the one hand, is to offer the users some reference tables that will allow comparing the data from the special flow files, the exploiting of which is technically more complex. On the other hand, it seems reasonable to prioritise the use of longitudinal information from the EAPS common sample that is exclusive of a continuous survey with partial rotation of the sample (retrospective questions can be included in any type of survey).

Coinciding with the renovation of the Economically Active Population Survey during the second quarter of 1987, the INE started series and studies on population flows with respect to economic activity based on the longitudinal analysis of the common sample in successive periods, which implied an important novelty with respect to traditional methods used until that date.

Detailed analyses of flow data and the methodology used can be found in the following publications:

Active Population Survey. Flow Statistics. 2nd quarter 1987 - 2nd quarter 1988. INE. Madrid 1989.
Active Population Survey. Flow Statistics. 1988-1990. INE. Madrid 1992
Active Population Survey. Flow Statistics. 1990-1992. INE. Madrid 1995
Active Population Survey. Flow Statistics. 1992-1994. INE. Madrid 1997
Active Population Survey. Flow Statistics. Methodology, INE. Madrid, 1996
Active Population Survey. Flow Statistics. 1984-1996. INE. Madrid, 1999
Labour market flows analysis from data from the of Economically Active Population Survey. Social indicators of Spain 2003. INE. Madrid, 2003

The publication of the flows of the economically active population, up until the fourth quarter o2012 in INEbase intends to continue with the dissemination of flow statistics, offering for the first time the results from the EAPS methodology of 2005 (that is, with the base 2001 populations introduced in March 2005, which led to the revision of the EAPS series for the period 1996-2004).

The series from 1996 can be consulted, separated into three different blocks: 1996-2000, 2001-2004 and since 2005 into two-year periods. The first division (2000 and 2001) was due to the implementation of the operative harmonised definition on the consideration of unemployed of Regulation 1897/2000 of the EC and its consequences on the group of inactive persons. From 2001 the data on unemployed and inactive persons in the EAPS is cannot be directly compared with previous periods. The second block (2004 and 2005) follows the variations in the series introduced by the change in collection methods of 2005. Since then, despite the results being presented as two-year periods, the data is homogeneous for all purposes.

In addition to the biannual dissemination of results, each quarter, together with the publication of the EAPS, a table regarding 'mobility as related to economic activity between the current quarter and the previous quarter' is included in the corresponding press release, this being calculated using the same methodology described below.

Analysis of flows from the common sample of the EAPS.

EAPS flow statistics based on the use of the common part of the sample comprising 5/6 parts of households between one quarter and the following one, 4/6 parts between one quarter and the subsequent one, etc., are based on a *panel* survey in which each individual 16 years of age and over has a record with data from all the interviews.

Chart 1. Shift rotation scheme of the Economically Active Population Survey

	Quarter					
Rotation shift	**T+1**	**T+2**	**T+3**	**T+4**	**T+5**	**T+6**
1	1st interv.	2nd interv.	3rd interv.	4th interv.	5th interv.	6th interv.
2	6th interv.	1st interv.	2nd interv.	3rd interv.	4th interv.	5th interv.
3	5th interv.	6th interv.	1st interv.	2nd interv.	3rd interv.	4th interv.
4	4th interv.	5th interv.	6th interv.	1st interv.	2nd interv.	3rd interv.
5	3rd interv.	4th interv.	5th interv.	6th interv.	1st interv.	2nd interv.
6	2nd interv.	3rd interv.	4th interv.	5th interv.	6th interv.	1st interv.

Colour shading corresponds to rotation shifts forming the common sample with respect to quarter T. After six quarters there has been a complete renovation of the sample

LIMITATIONS OF THE PROCEDURE.

1. Loss of the sample.

The information link may not be established for all the records of the common initial sample. There are always records of the common sample for which there is no information available in some of the periods to be analysed (refusals or absence of a family group, dwellings not accessible, registrations and cancellations in collaborating households, etc.)
Furthermore, the link is performed by disregarding the records imputed in each period. That is, those records that were "no response" one quarter and who had answered in

the previous quarter. The lack of response in the current quarter is corrected by copying the data from the preceding quarter, thus partly solving the lack of response in the quarterly EAPS. In the case of the flow analysis its inclusion would artificially increase the number of persons who do not change their situation from one period to the next.

2. Inconsistencies in the response

Flow statistics as a product derived from the Economically Active Population Survey present inconsistencies between information compiled in the different periods analysed. Thus, for example, it is possible to find persons who have worked in one quarter and who subsequently declare that they are looking for their first job, or they have a level of studies that is less than that previously declared. We must take into account that he EAPS admits information provided by another person from the household different to the reference person; these changes of informant can give rise to inconsistencies of the type described.

3. Significance of the results

Since the EAPS is a sample survey, the results it provides are associated with a smaller sampling error that is usually greater the smaller the group and the more uniformly distributed the variable or variables it refers to.

In the case of flow data the problem of sampling errors associated to small groups increases, since when linking two time periods, analysis categories are in fact being multiplied, with the corresponding risk that some crossings are left with a small sample size and therefore the estimates may suffer from strong sampling errors.

VALIDITY OF FLOW STATISTICS

An essential requirement for estimate validity is that part of the sample considered is representative of total population. All rotation shifts have the same distribution by strata, such that the lost sample would have the same features as the linked one.
In second place, the group of lost records due to lack of response and registrations and cancellations of persons in one of the quarters do not involve a high figure with respect to potential common records.

Therefore, neither the sample loss involved in the rotation of 1/6 thereof every quarter or that due to the loss of records for persons 16 years of age and over or to possible inconsistencies between quarters imply any risk regarding representation, more so taking into account the low levels of category breakdown reflected in the flow tables, in which there is no data from subjective response variables. When considering relatively important groups, that is, with a high sample representation, sampling errors remain at tolerable limits.

THE PROBLEM OF THE ELEVATION FACTORS.

The elevation factors associated to each individual of the sample are equivalent to the number of individuals of the total population they represent, therefore, the total figures for the flow statistics represent only the common part of the sample with respect to the number of consecutive quarters under study. Thus, for example, between two

consecutive quarters, the total figure corresponds to somewhat less than 5/6 parts of the population (taking into account the deficiencies quoted by refusals, absences, etc.), between three consecutive quarters it will be reduced to somewhat less than 4/6 parts, and so on. For this reason they cannot be interpreted as gross input and output flows from one quarter to the next, nor can they provide a coherent estimation of transfer flows with respect to the contingents of the two quarters. Their representation is coherent when observed in terms of percentage.

For the gross flows from one quarter to the next to be representative of the entire population a new elevation factor would be necessary starting from the common sample, also calculating the quotient between the population value and the sample value for each stratum. Which would further require obtaining new population projections (for the numerator of the elevation factor), which would imply the problem of choosing a reference date for them and would further provide different estimations for the different population categories and for those of the EAPS for each quarter. Neither would estimations for a given quarter and the next coincide with those obtained for that same quarter linked to the two following quarters, since although the starting point is the same in both cases, both the common samples and the projections are different in each case. Only percentage data would be considered representative, as in the previous procedure.

GENERAL SCHEME OF FLOW TABLES FROM THE COMMON SAMPLE.

The advantages implied by studying flows from the common part of the sample in two consecutive quarters with elevation factors from each quarterly sample have been decisive in choosing said method for the analysis of the flow tables of the present publication.

Monitoring of each person 16 years of age and over is thus performed so that the information that said person has provided in the two linked periods is grouped as to combine both. Crossings between the categories of a variable for one period with categories of the same or other variables in another period are calculated and presented in relative figures that we could consider in terms of probability. Thus, for example, in the chart's table the 'probability of transition between the 'states' Ei - E' j would be pij.

General scheme of flow tables from the common sample

	Tb			
Ta	E1'	E2'	……	Se'
E1	p11	p12	……	p1m
E2	p21	p22	……	p2m
….	….	….	….	….
En	pn1	pn2	….	pnm

pij: probability of 'passing/origin' from Ei to Ej'
Ta, Tb: Reference periods a and b
Ei/Ej': Analysed status

In the case that the elevation factor used is that of the first period, we would have an **evolution analysis**, that is, the future situation in terms of probability of persons observed in the starting quarter (for example, "a" being the starting period marking the elevation factors).

If the factor used is that of the final or arrival period, we would have a **source analysis**, or similarly the original situation, in terms of probability, in the arrival quarter (for example, with elevation factors for period "b").

TABLES

The results that the flow publication of the economically active population up until the fourth quarter of 2012 show the **source analysis**, in other words, the distribution of the original situation into which converges each category studied in the arrival quarter the data refers to.

There are three types of tables. The so-called "**Mobility**" tables include the percentage distribution of the main categories (employed, unemployed and inactive persons) with respect to those of the previous quarter. The "**National**" tables include the probabilities of origin between two categories in a longitudinal analysis, by age groups and sex. And finally, the tables by "**Autonomous Communities**" show those same probabilities in the autonomic territorial scope.