**Coherence between surveys and register-based data in labour market statistics**

Sara Carrascosa García, INE, Sara.carrascosa.garcia@ine.es

Laura Grande Freire, INE, Laura.grande.freire@ine.es

**Abstract**

The study of the labour market has always had a special interest in society, both economically and socially. In Spain, there are different sources of information that allow us to approach and measure this reality. Statistics based on surveys are available, such as the Labour Force Survey (LFS) - carried out by the National Statistics Institute of Spain (INE) -; and others using administrative data, such as those based on the social security affiliation register and the unemployed persons registered in public employment services, provided by the Ministry of Employment and Social Security.

Therefore it is interesting to study the coherence among these sources of information as an aspect of improving the quality of statistics.

Under the framework of the High Council on Statistics the working group of Short-Term Labour Market Statistics draws up periodically a report with the aim of analyzing and comparing the data provided by the LFS and those provided by the Ministry of Employment, using for that purpose, a harmonized methodology.

In addition, a series of short-term analyses are also carried out to study the coherence between them from the point of view of the final published data.

This project allows, on the one hand, to compare, study and review methodological differences, and on the other to analyze the information that users receive from different sources of information.

**Keywords:** coherence, administrative data, labour market statistics

**1. Coherence as a principle of quality**

The quality management system of National Statistics Institute of Spain (INE) is based on the European Statistics Code of Practice (CoP), which sets the standard for developing, producing and disseminating European statistics through fifteen principles.

In this document we will focus on the study of the principle 14 of the CoP which establishes coherence and comparability as a dimension of the quality of the statistics.

This principle of coherence defined in the CoP covers the internal consistency, *over time and regional comparability, but also the possibility of combining related data from different sources. The 14th principle is* developed through five indicators and we are going to focus on Indicator 14.4 which looks at how statistics from different sources and of different periodicity are compared and reconciled. The concept of coherence is further broken down into more specific ones, but we will concentrate on cross-domain coherence. The aim of cross-domain coherence is to compare the extent to which statistics are reconcilable with those obtained through other data sources or statistical domains.

According to the latest User Satisfaction Survey (USS2016), carried out by INE, coherence is the most difficult quality dimension for users to understand, which is reflected in the low response rate of the question related to this aspect of quality. So, it could be interesting to carry out projects analyzing coherence among different sources.

That being said, the coherence of the INE labour market statistics is rated positively or very positively by 94.9% of users.

**2. Different sources of labour market data**

Nowadays, official statistics has access to different data sources: surveys, administrate data, web data, etc.

On the one hand, there is labour market data from **surveys**. The information that comes from surveys, allows us to define concepts such as ***employment*** or ***unemployment*** according to the International Labour Organization (ILO) definitions.

On the other hand, there is data from **administrative registers**. The purpose of administrative registers is always administrative management, but data could be the information source of a statistic, as well. Therefore, although their concepts do not coincide with those defined by the ILO, they might be very similar.

* In the case of **employment**, there is the term ***“affiliated”***, which refers to those who pay social security contributions.
* In the field of **unemployment**, there is the term ***"job seeker"*** which refers to someone who is registered in a public employment office.

**3. Labour Market Statistics in Spain**

One of the main pillars of the Official Statistics in Spain is the four-year National Statistical Plan (PEN) which lists all statistical operations and statistical activities conducted by the institutions of the National Statistical System (NSS), and defines the dissemination of the results. All the statistics included in the PEN are considered statistics for state purposes and must be completed.

The sources of information on the labour market in Spain included in the PEN are numerous and diverse:

* **The Economically Active Population Survey (EAPS)**: is a continuous investigation of quarterly periodicity directed to families and has been carried out since 1964. Its main objective is to obtain data on the population in relation to the labour market: active, unemployed and inactive people.

The EAPS follows the concepts defined by the ILO and Eurostat. This allows EAPS to be internationally comparable and serves as a basis for the development of the Labour Force Survey (LFS).

* **Statistics on the affiliation of workers to Social Security:**Its objective is to obtain and disseminate data on affiliated workers and movements of entering and leaving of the social security system.

These statistics are based on the administrative register of **Social Security Affiliates**. This is where the additions and deletions of social insurance contributions are registered.

* **Registered Labour Movement Statistics:** Its objective is to obtain and disseminate data on the labour movement registered with the public employment services. These statistics are based on the administrative register of the **job seekers** **register.**

These three statistical operations measure similar realities (employment and unemployment) within the scope of the labour market. This means that they inevitably interact with each other, giving rise to a variety of situations within an employment and unemployment scenario:

Focusing on **employment:**

* Typically, employed persons, as defined by the ILO, are registered in a Social Security Administration.
* There are also people who are employed, as defined by the ILO, and are not registered with the Social Security Administration. We are referring, for example, to the black economy.
* Finally, there are people registered in the Social Security registers who are not considered employed, according to the ILO definition, e.g. people covered by Special Health Care Provision or people not working but contributing to Social Security in a special scheme in order to top up their state pension.

Focusing on **unemployment:**

* Most commonly, unemployed people, as defined by the ILO, actively seek employment and are registered as job seekers in a public employment office.
* However, many of the unemployed, as defined by the ILO, are not registered with public employment offices for various reasons. The most common is because they don't trust that it's a good way to find work.
* Finally, there are people who are registered as job seekers but who are not classified as unemployed according to the ILO definition, either because they do not actively seek employment or because they actually do work but in the black economy.



**4. Conciliation projects in Spain**

In Spain, the High Council on Statistics and the Working Group on Current (Short-term) Labour Market Statistics have supported and encouraged the study of the coherence between sources in this field.

In order to implement the mentioned study, the Working Group carries out two conciliation projects with the objective of studying the coherence between sources through the reconciliation of their figures. These projects are aimed at the end user and a final report and the results are published on INE’s website every two years.

The studies focus on two levels: Macro and Micro analysis; and on two concepts: Employment and Unemployment.

* **Macro reconciliation** compares the results of the statistics with each other, trying to create conceptual identities between them.
* **Micro-conciliation** consists in comparing the statistical data in order to check whether it is consistent with each other. If the data is consistent, it is foreseeable that the statistics created from it will also be coherent.
* The term **"employment"** is legally linked to the right to be affiliated to the Social Security or other equivalent mutual insurance companies (MUFACE, MUGEJU and ISFAS). This makes the concepts of occupation and affiliation similar and therefore, the reconciliation process possible.
* The term **“unemployment”** is not legally linked to any administrative register. According to the ILO's definition, "being unemployed" implies a lack of work, willingness to work and an active search of work. “Active job search” involves a subjective perception of the respondent and, in this sense, is very difficult to measure reliably.

The reports begin with the definitions and peculiarities between methodologies of the information sources involved. Then they explain the reconciliation methodology of the publication and, finally, they comment on the most important results. The publication provides data that crosses the following classification variables: sex, age, nationality, regional breakdown (NUTS2) and activity sector.

Currently, there are two publications: the analysis of unemployment at micro level (published the odd-numbered years) and the analysis of employment at macro level (published the even-numbered years). It is important to note that the reconciliation of the figures will always be done under the prism of the EAPS methodology. They are all available on the INE website in both English and Spanish.

***4.1. Unemployment – Micro Level***

The micro-conciliation of unemployment data is carried out by cross-checking individual data from the EAPS with the databases of the Public State Employment Service (PSES). In this publication we work with the concept of job seeker and unemployed person. For the matching process, we obtain the ID number of the EAPS sample though record linkage techniques.

The reconciliation process is structured as follows:

* Match the EAPS sample with the PSES database.
* From those who were found in the PSES database, we check if they are job seekers registered in the reference period.
* Match persons of the EAPS sample, who state in the survey that they are registered as job seekers in a Public Employment Office, with the job seekers registered of the PSES database.
* Match persons of EAPS sample, who are classified as unemployed person, with the job seekers registered who are consider as unemployed according to the Registered Labour Movement Statistics.
* Match persons of EAPS sample, who state in the survey that they receive unemployment benefits, with the job seekers registered that actually receive benefits.

In overall terms, despite the fact that the time evolution is coherent and the two concepts have a similar name, a direct comparison of the unemployment registered figures and the unemployment estimated by the EAPS is not possible, as they correspond to different concepts and come from different sources.

***4.2. Employment – Macro Level***

This study is carried out in even-numbered years with the aim of comparing, in a homogeneous way, the employment data obtained from the EAPS and the Social Security affiliation. The affiliation data comes from the statistical use of the databases maintained by the General Social Security Treasury (Ministry of Employment and Social Security-MEYSS).

Both Affiliation and EAPS data have their own particularities, derived from their own methodology and purposes, and offer a reflection of the evolution of employment in Spain filtered through their respective specificities. For a comparison of results, it is necessary to make adjustments to the groups covered in each case and to homogenise the reference period of the data.

These adjustments are made by adding other mutual societies such as MUFACE, ISFAS or MUGEJU to the Social Security schemes. The sum of Social Security affiliates and those affiliated to MUFACE, ISFAS and MUGEJU correspond to the term **Adjusted affiliation**. Adjusted affiliation is a similar concept to that of employed persons in the EAPS and will be used to reconcile the figures from both sources.

EAPS Data

Affiliation

Social Security Affiliates

Mutual insurance companies (MUFACE, MUGEJU,ISFAS)

Collectivities

**Data Reconciliation**

As stated above, the EAPS is the source of data that serves as a reference for the coherence study projects. In this sense, the figures used in this publication are the same as those used in the statistics. Only in the regional breakdown are there differences such as, the fact that instead of taking into account the place of residence of the respondent, the province of the place of work is considered. In this way we gain comparability.

On the other hand, when analysing the results, a distinction is made between the agricultural and non-agricultural sectors (including all mutualities).

In the comparative analysis, the information is broken down by regional scope, sex, age, professional situation, economic activity, type of employment relationship, type of working day and nationality.

The latest available publication compares data for the second quarter of 2016. The EAPS occupancy rate exceeds the adjusted affiliation rate (531,600 people in favor of the EAPS occupation). This situation is maintained when analysing the figures by sex and nationality (both for Spaniards and foreigners). By type of working day, there are major differences, which may be due to the fact that towork less than 40 hours per week is considered to be part-time work in affiliation data.

Overall, the employment study show more coherence among the two statistics studied.

***4.3. Employment – Micro Level***

As in the case of unemployment, we could also carry out a comparative analysis at the micro level, using the data from the EAPS as a basis and searching for each person surveyed in the register of Social Security affiliation.

This study is fundamentally concerned with the characterization of dissenting cases and with the analysis of the extent to which the relationship with the labour market obtained from the EAPS and the Social Security situation can be explained.

This micro analysis is more complex than the unemployment analysis because the administrative register has more functions than the management of the people who work. The register also includes non-working members, such as some special agreements aimed at supplementing pension rights.

On the other hand, as with the micro unemployment reconciliation, when the reference period is extended to more than the EAPS reference week, the percentage increases in both sources. In addition, the full micro study should also involve the members of the mutual insurance companies.

This project is still in progress, but the latest results indicate very high consistency between the two data sources. Regarding discrepancies, a high percentage of cases can be explained by specific situations.

Broadly speaking, among those employed persons in the EAPS and not found in the Social Security registers, we find a greater number of people who are self-employed or who are assited by a family member, with temporary contracts with fewer hours with elementary training and who are foreigners.

***4.4. Quarterly Coherence***

To complete the micro/macro reconciliation studies, a comparative study is carried out, each quarter, between the figures published in the EAPS, the average affiliation of Social Security Affiliates Statistics, and the registered unemployment of the Registered Labour Movement.

In this case, it is not intended to make a methodological approximation or reconciliation of figures, but rather to analyse the differences received by the end user, that is, to know the perception of cross-domain coherence that the user has.

At the same time, it allows us, as producers of statistics, to better understand the aspects of the labour market that EAPS cannot collect, and that an administrative register can better analyse.

**5. Conclusion**

As this document has shown, users of Spanish labour market statistics have not only information provided by the EAPS but also statistics based on administrative data. In this situation, users may think that the information that they receive is contradictory or inconsistent. Consequently, users may perceive that the quality of those statistics is lower. In this sense, these projects are fundamental, because they provide users a better comprehension of labour market statistics, their similarities and differences.

Furthermore, from the statistics producer perspective, the study of coherence improve the quality of their statistics as it allow them to better know the concepts and the reality that they study.

While the cross-domain coherence analysis of employment figures outcomes in a high consistent between the two sources, the unemployment analysis is more difficult. This discrepancy could be explained by the difference in the definition of the concept made by the ILO. The unemployment as is defined by the ILO, implies a lack of job and a willingness of working, which is more subjective and difficult to find in an administrative register.

**7. References**

European statistics Code of Practice — revised edition 2017.

Quality Assurance Framework of the European Statistical System – 1.2 version

Comparison of Employment according to The Economically Active Population Survey and to affiliation data – January 2018

Comparison of Unemployment according to The Economically Active Population Survey and to Public Employment Service data – September 2017