

Integration of LFS with Spanish Affiliation Data

Sara Carrascosa García
National Statistics Institute of Spain (INE)
sara.carrascosa.garcia@ine.es

Miguel Ángel García Martínez
National Statistics Institute of Spain (INE)
miguelangel.garcia.martinez@ine.es

1. INTRODUCTION

One of the key aspects of statistical product quality is its internal coherence as well as its coherence with other statistical operations within the same domain. This quality principle is more straightforward to study and ensure in a market with few statistical products. However, given the increased availability of data, a more in-depth analysis is required to guarantee this critical aspect of the quality of our statistics.

In this work, we will analyse the coherence between two statistics on the labour market in Spain. Furthermore, we will present an integration of the data from both statistics in order to study the coherence between both statistics and between the two statistical concepts that measure these statistics.

2. LABOUR FORCE SURVEY AND AFFILIATION STATISTICS IN SPAIN

In Spain, there are two significant labour market statistics: the Labour Force Survey (LFS) and the Statistics on Workers' Affiliation to Social Security. While both provide employment data, they differ in terms of periodicity, data collection, breakdown of information, international comparability, internal coherence, and conceptual definitions.

When analysing the coherence between these statistics, the key factor explaining the differences lies in the nature of the data used. The LFS collects data through household surveys, allowing for targeted data collection aligned with research objectives. In contrast, the Affiliation Statistics rely on administrative records, requiring statisticians to extract statistical definitions from predetermined variables.

The LFS, conducted quarterly by the National Statistics Institute of Spain (INE), follows the International Labour Organization (ILO) guidelines and covers employment, unemployment, and individuals outside the labour force. Conversely, the Affiliation Statistics, compiled monthly by the General Social Security Treasury, records the number of workers affiliated to the system, excluding certain groups like civil servants affiliated with other mutualities.

The LFS excels in international comparability, as it is conducted across all EU Member States, following a standardized methodology and questionnaire. On the other hand, the Affiliation Statistics are specific to Spain and not directly comparable to other countries.

Regarding internal coherence, differences arise in the definition of employment. The LFS adheres to internationally established definitions by the ILO, which remain relatively stable over time. In contrast, statistics based solely on administrative records, like the average affiliation, are more susceptible to changes in regulatory frameworks. These statistics rely heavily on administrative processes and are influenced by modifications in labour market laws and regulations. Consequently, the definitions and variables used in these statistics may undergo frequent changes, leading to potential inconsistencies and making direct comparisons more challenging.

3. INTEGRATION PROCESS

The study of coherence between the two statistics has been conducted through the integration of information from both sources, starting with the quarterly sample of the LFS and including, for each surveyed individual, the information from the affiliation register in the reference week. The study has undergone the following phases:

- Obtaining the identifier of each interviewed person in the LFS using the personal data collected in the questionnaire and the municipal population register.
- Requesting the registration information on employment affiliations for the identifiers obtained in the previous step.
- Linking the LFS sample with affiliation data
- Obtaining the variable "affiliated during the reference week."
- Creating coherence tables using the LFS weighting factor.
- Analysing the non-coherent groups and identifying possible causes of such inconsistencies.

To properly interpret the results obtained in this study, it is important to consider the potential errors and incidents that may have occurred during the process:

- Missing or incorrect assignment of identifiers in the LFS sample.
- Errors in the identifiers from the registration information.
- Mistakes in the information regarding affiliations from the registration source.
- Errors in the information collected in the LFS questionnaire.

Since in this analysis we restrict ourselves to the LFS sample, we will only compare the population residing in family households. On the other hand, residents in Spain who work abroad are included in the scope of the LFS and may be affiliated with social security systems of other countries, and therefore may not be appear in the affiliation data.

4. RESULTS OF COHERENCE ANALYSIS

After integrating the data from the LFS with the Social Security affiliation records, we are now able to conduct a coherence study of the two definitions. To do this, we need to assign the variable "affiliated" to each person in the LFS sample, trying to emulate as closely as possible the definition used in the Affiliation statistics.

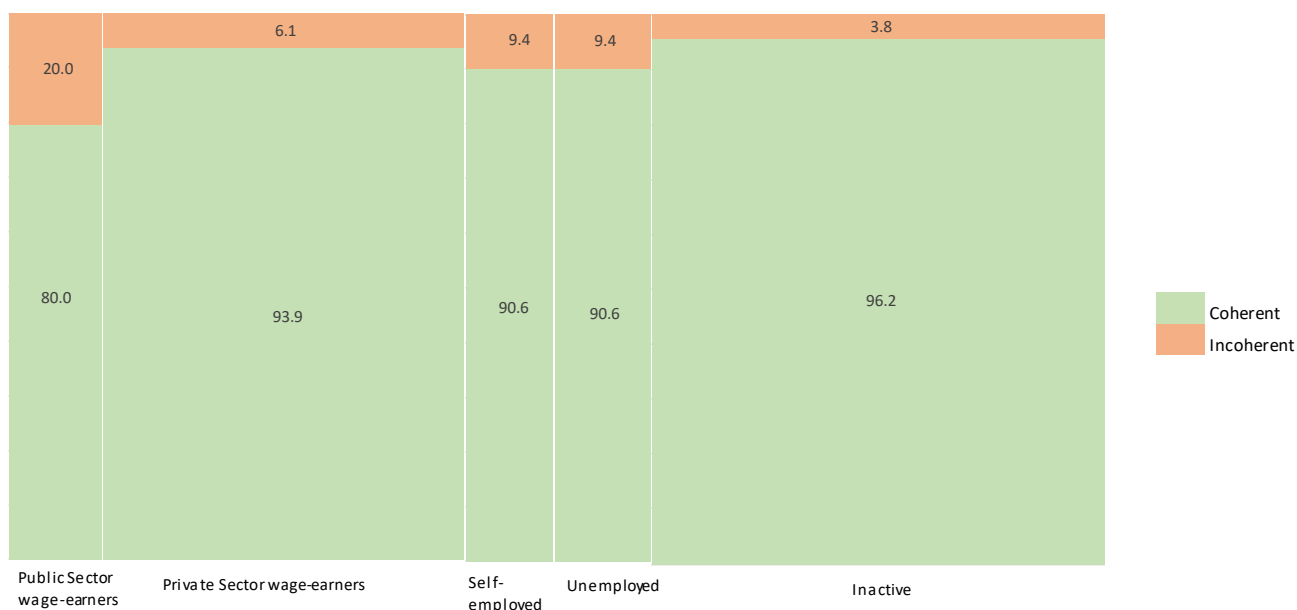
The results of this integration process for the second quarter of the year 2021 are as follows:

Table 1. Population aged 16 and over by economic activity status in LFS and coherence with the affiliation status during the reference week. Breakdown for Q2-2021. Initial situation.

| Q2- 2021 RELATIONSHIP WITH ECONOMIC ACTIVITY (LFS) | COHERENCE WITH AFFILIATION STATUS DURING THE REFERENCE WEEK | | | | | |
|--|---|-------|------------|------|--------------|------|
| | Total | | COHERENT | | NOT COHERENT | |
| | Value | % | Value | % | Value | % |
| TOTAL (POB > 16 YEARS) | 39,633,100 | 100.0 | 36,899,000 | 93.1 | 2,734,100 | 6.9 |
| EMPLOYED | 19,671,700 | 100.0 | 17,896,700 | 91.0 | 1,775,000 | 9.0 |
| Public Sector wage-earners | 3,439,600 | 100.0 | 2,752,600 | 80.0 | 687,000 | 20.0 |
| Private Sector wage-earners | 13,067,900 | 100.0 | 12,276,000 | 93.9 | 791,800 | 6.1 |
| Self-employed | 3,164,200 | 100.0 | 2,868,100 | 90.6 | 296,100 | 9.4 |
| NON-EMPLOYED | 19,961,500 | 100.0 | 19,002,300 | 95.2 | 959,200 | 4.8 |
| UNEMPLOYED | 3,543,800 | 100.0 | 3,212,100 | 90.6 | 331,800 | 9.4 |
| INACTIVE | 16,417,600 | 100.0 | 15,790,200 | 96.2 | 627,400 | 3.8 |

Note: preliminary data under revision

Graph 1. Distribution of population aged 16 and over by relationship with activity and affiliation status. Percentages relative to each relationship with activity. Initial situation.



Among the employed, we have an initial coherence of 91.0%, and among the non-employed, it is 95.2%. Overall, combining the found employed and the unfound non-employed people in Affiliation, we have a total initial coherence of 93.1%. Analysing the groups that show inconsistencies between the two data sources, we can identify certain collectives whose differences can be explained by specific circumstances:

- Many civil servants and public employees are not found in the Social Security Affiliation File because they are covered by specific mutual insurance schemes (MUFACE, MUGEJU, ISFAS).
- Some individuals with regulated professions are also not included in the FGA as they belong to professional associations with alternative social security mutual insurance, distinct from the general Social Security System. This includes professions such as lawyers, architects, doctors, administrative managers, chemists, etc.
- Some employed individuals in the LFS who work abroad are not found in the FGA as they are affiliated with social security systems of other countries and are not registered in the Spanish Social Security System.

Alternatively, some people are found in Affiliation who are non-employed according to the LFS definition:

- There is a large group who are members of the Special System for Employed Agricultural Workers who are in a situation of inactivity.
- During the analysed period (second quarter of 2021), it was possible to be affiliated through a partial retirement contract, which may result in individuals not working during the reference week.

Considering these explained groups, we can calculate successive coherence indexes and achieve a final coherence of 95.9% in the total population (95.3% for employed people and 96.6% for non-employed people).

The groups that remain without explanation are:

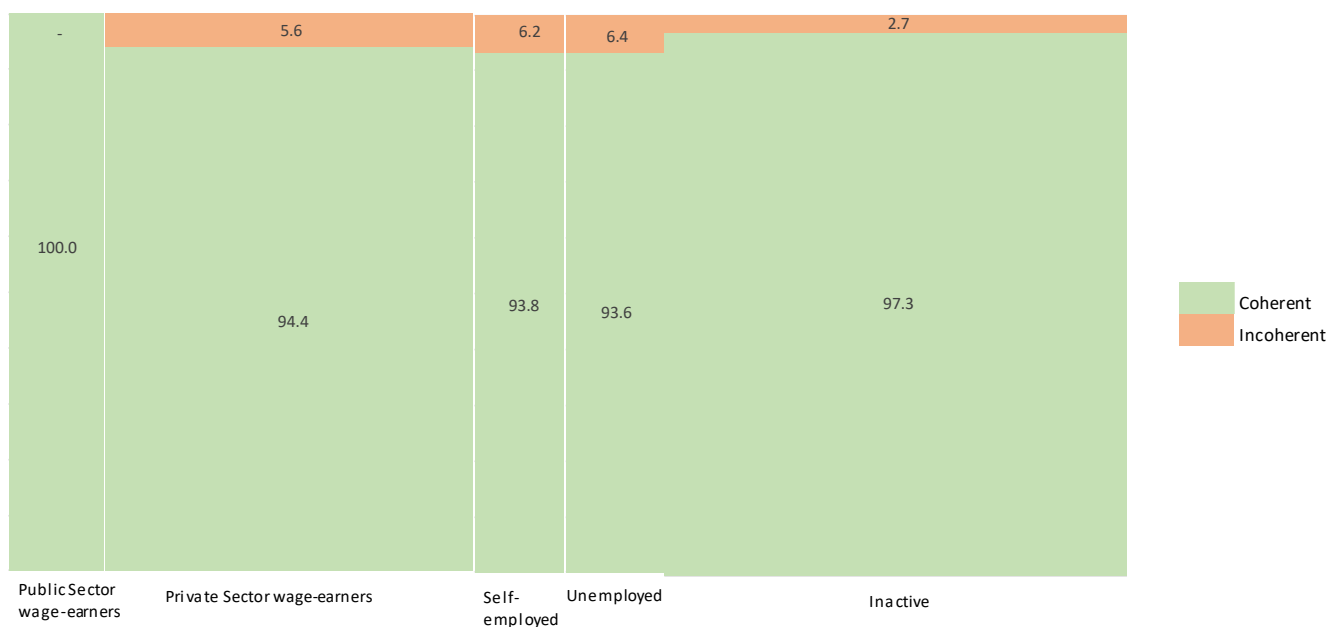
- A total of 933,000 employed people in the LFS (737,100 employees in the private sector plus 195,900 self-employees), who are not registered in the Affiliation.
- A total of 677,200 non-employed people in the LFS (227,100 unemployed people and 450,000 inactive people) who are registered in Affiliation with an employment relationship equivalent to 'work'.

Table 2. Population aged 16 and over by economic activity status in LFS and coherence with the affiliation status during the reference week. Breakdown for Q2-2021. Final situation.

| Q2- 2021 RELATIONSHIP WITH ECONOMIC ACTIVITY (LFS) | COHERENCE WITH AFFILIATION STATUS DURING THE REFERENCE WEEK | | | | | |
|--|---|-------|------------|-------|--------------|-----|
| | Total | | COHERENT | | NOT COHERENT | |
| | Value | % | Value | % | Value | % |
| TOTAL (POB > 16 YEARS) | 39,633,100 | 100.0 | 38,023,000 | 95.9 | 1,610,100 | 4.1 |
| EMPLOYED | 19,671,700 | 100.0 | 18,738,700 | 95.3 | 933,000 | 4.7 |
| Public Sector wage-earners | 3,439,600 | 100.0 | 3,439,600 | 100.0 | . | . |
| Private Sector wage-earners | 13,067,900 | 100.0 | 12,330,800 | 94.4 | 737,100 | 5.6 |
| Self-employed | 3,164,200 | 100.0 | 2,968,400 | 93.8 | 195,900 | 6.2 |
| NON-EMPLOYED | 19,961,500 | 100.0 | 19,284,300 | 96.6 | 677,200 | 3.4 |
| UNEMPLOYED | 3,543,800 | 100.0 | 3,316,700 | 93.6 | 227,100 | 6.4 |
| INACTIVE | 16,417,600 | 100.0 | 15,967,600 | 97.3 | 450,000 | 2.7 |

Note: preliminary data under revision

Graph 2. Distribution of population aged 16 and over by relationship with activity and affiliation status. Percentages relative to each relationship with activity. Final situation.



While part of this unexplained inconsistency may be due to the difficulty of perfectly framing short work periods in time (both in terms of the response collected in the survey associated with the reference week and the statuses recorded in the Affiliation), there are indications that a main share of the occupation not registered in Affiliation and registered in LFS (higher ratio of temporary work

relations, in specific sectors and occupations, etc.) could be cases of informal employment. On the other hand, the cases that correspond to people who pay Social Security contributions without working or who simply do not respond to the survey truthfully are also significant and deserve to be analysed carefully.

5. CONCLUSIONS

The realization of these coherence studies is increasingly necessary in an environment where there is growing availability of data and, consequently, new statistics. Moreover, these exercises are highly cost-effective as studying the coherence between two statistics enhances the quality of both statistical products. Users of these statistics have access to an explanation of possible observed inconsistencies.

From the producer's perspective, we also improve the quality as this contrast of concepts and figures enables us to understand and measure in more detail the strengths and weaknesses of our statistics. In other words, we have a tool for constant quality assessment, including the measurement of coverage.

Furthermore, in the process of conducting these studies, efficient communication among the producers of both statistics, as well as data providers, is necessary. This also contributes to higher quality as it aligns with the new principle of coordination outlined in the Code of Good Practices.

On the other hand, the unexplained groups can serve as a starting point for further analysis of the informal economy, its characteristics, and its evolution over time.

Lastly, this exercise can be the starting point for the development of new statistical products that are based on the survey and include registry information, or alternatively, use registry data and leverage the cross-referencing with the survey to obtain a statistical definition that is more aligned with the reality we aim to measure.

6. REFERENCES

- 'Micro' conciliation of employment as measured by the Economically Active Population Survey and Social Security Affiliation". INE report. Available at: https://www.ine.es/en/daco/daco42/daco4211/compa_empleo_bp11_2016_microdatos_en.pdf
- Comparison of the statistical magnitudes of employment according to The Economically Active Population Survey and to Affiliation data. INE report. Available at: https://www.ine.es/en/daco/daco42/daco4211/compa_empleo_bp11_2018_en.pdf
- Methodology of the Economically Active Population Survey. INE report. Available at: https://www.ine.es/en/inebaseDYN/epa30308/docs/resumetepa21_en.pdf
- Methodology of the Statistics on Workers' Affiliation to Social Security. General Social Security Treasury report. Available at: https://www.seg-social.es/wps/wcm/connect/wss/b7e3b61f-4001-404f-b7e0-b2aaf91a785d/2023-02+IME.pdf?MOD=AJPERES&CONVERT_TO=linktext&CACHEID=ROOTWORKSPACE.Z18_81D21J401P5L40QTIT61G41000-b7e3b61f-4001-404f-b7e0-b2aaf91a785d-oomPec7