

Preparation of the retrospective series in CNAE-09

1. Introduction

The information on labour cost produced by the INE since the year 2000 is:

- the Quarterly Labour Cost Survey (QLCS)
- the Harmonised Labour Cost Index (HLCI)
- the Annual Labour Cost Survey (ALCS).

These three statistics comprise an integrated system of data regarding labour costs, with coherent estimates under any of the three perspectives, the QLCS being the basis that sustains this integrated system.

Given that both the HLCI and the ALCS are compiled based on the QLCS, the estimation of labour cost data in CNAE-09, of a retrospective nature, is resolved by obtaining retrospective data for the QLCS statistical base.

Section L (Public Administration) of the old CNAE, included as of the first quarter of 2009, uses different sources of information for the compilation of the HLCI, but does not pose problems with regard to the classification of activities on a section level (section L changes to be called section O in the new CNAE-09).

This document describes the procedure applied to obtain the retrospective estimates in CNAE-09 for the QLCS.

2. Availability of the directory in the new classification

The QLCS uses the Social Security Contribution Accounts File as a framework, that is, it is based on an administrative source that includes, among others, the economic activity classified according to the National classification of economic activities in force at any given time. The breakdown level used in the sample design are the two digits of the classification.

In order to comply with the European agreements regarding dates and periods of information in the new classification, it was necessary to have the Social Security Contribution Accounts File classified in the new CNAE at the end of the year 2007, for the purpose of using said classification in the sample selection of the QLCS during the year 2008. During this year, the results of the QLCS sample were obtained in both classifications (CNAE-93 and CNAE-09), and this has been the starting point for the compilation of the retrospective series.

At that time, the file with the new classification was not available. However, the Central Companies Directory (CCD) that the INE compiles and that serves as the framework for most of the INE surveys of companies, in the carrying out of its implementation of the new CNAE, had a new version available in June 2007, referring to 1 January 2006, with a double encoding.

The basic unit of the CCD is the company, classified by its main activity, and identified by its Tax Identification Number (NIF). The Social Security directory uses the local unit, but also includes amongst its variables the Tax Identification Number (NIF) of the company.

Therefore, the INE decided to use the information available in the CCD to carry out a rapid re-encoding of the Social Security Contribution Accounts File to two digits, which would allow for fulfilling the commitments of the HLCl while carrying out the work of the Social Security institution itself.

Starting from the following available elements:

- CCD with double encoding from the CNAE and NIF (Tax Identification Number)
- Correspondence matrix between CNAE-93 and CNAE-09 to two digits
- Probabilistic correspondence matrix obtained from the re-encoding work of the CCD
- Social Security file with the CCC (contribution account), NIF, CNAE-93 and number of workers

The Social Security file was crossed with the CCD, and a directory encoded to two digits of CNAE-09 was obtained that was compatible with the CCD.

As of January 2009, the Social Security file is available, classified in CNAE-09. During this year 2009, a revision of the CNAE-09 encoding of the sample units is being carried out. The quarterly results of this revision are included in the estimates from 2008, and thus the results obtained are of a provisional nature until the completion of the revision of the CNAE-09 encoding with the publication of the fourth quarter of 2009.

3. Sample design in the new classification

Once the survey directory was available in CNAE-09, we studied the sample sizes necessary to obtain estimates, to two digits, of the new classification. We must remember that, with the new classification, the 54 divisions increase to 83 divisions, indicating an increase of 50% in activities.

This increase has had an important impact on the sample size of the QLCS, with a 30% increase in the sample units, which in addition, has not occurred homogeneously in all of the economic activities. The main increases have taken place in the services sector, where the number of classes has grown more substantially.

4. Retrospective series

For the purpose of having time series available of labour costs, with a sufficient temporal horizon to allow for their analysis, it was agreed at the heart of the European Union to compile retrospective series in CNAE-09, on an activity section level, from the year 2000 onwards.

To this end, a complex linking system has been prepared, which is described below:

Firstly, it is necessary to have a common period, during which both classifications have existed, which, as previously mentioned, is the year 2008.

During the year 2008, quarterly estimates were available in CNAE-09:

$$\hat{X}(CNAE09) = \sum_{h \in CNAE09} F_h \sum_{i=1}^{nh} X_{hi} \quad , \text{ with } h=\text{strata defined in CNAE-09.}$$

Using the same procedure described in the reclassification of the unit directory for the year 2008, we have reclassified to the new CNAE, those units of the directories used in the QLCS corresponding to the years 2000 to 2007.

From these directories, estimates have been compiled in CNAE-09 for the periods comprised between the first quarter of 2000 and the fourth quarter of 2007, based on post-stratified estimators.

$$\hat{X}(POST) = \sum_{h \in CNAE09} \sum_{H \in CNAE93} F_{hH} \sum_{i=1}^{nh} X_{hi} \quad , \text{ with } h=\text{strata defined in CNAE-09, from the}$$

strata H defined in CNAE-93.

The result obtained presents a gap between the data prior to 2008, constructed with the post-stratified estimator, and the data from 2008, constructed with the habitual estimator, due to the sample increase as of 2008.

To bridge this gap, some linking coefficients have been calculated, constructed under the following premise:

For each region (REG) and activity sector (SECT= industry, construction and services), the values of the new estimator in CNAE-09 should present interannual variations similar to those published.

That is, for each branch of activity R, belonging to activity sector SECT, a corrector coefficient $Y_{REG R}$ is calculated, in such a way that

$$\left(\sum_{R \in SECT} (Y_{REG R} \cdot \hat{X}^{07}(POST)) \right) \cdot (1 + t_{REG SECT}^{08/07}) = \hat{X}^{08}(CNAE09)$$

where $(1 + t_{REG SECT}^{08/07})$ is the variation rate obtained between the years 2007 and 2008, published in CNAE-93.

Labour cost is comprised of wage cost and other costs, which in turn, are broken down into other variables (ordinary wage cost, extraordinary payments, delayed payments, obligatory contributions, etc.). Linking coefficients are obtained only for the two main variables: net total labour cost and wage cost, given that these are the variables that mark the temporal evolution of the series. Used as the final linking coefficient is the geometric average of the two previous linking coefficients, and it is applied to all of the economic variables of the units from 2000 to 2007, thus linking all of the labour cost series (irregardless of their breakdown level).

Likewise, a linking coefficient $H_{REG R}$ is calculated, with makes the labour cost per hour worked in each region and activity sector respect the variation rate published in 2008, and with this coefficient, the series of effective hours worked is linked.

Lastly, we must reiterate that, during the year 2009, the results obtained will be of a provisional nature, until the conclusion of the revision of the CNAE-09 encoding with the publication of the fourth quarter of 2009.