



Working Papers

03/2010

Changes of classification in continuous statistics: Calculation of retrospective series

Application to the Quarterly Labour Cost Survey

Amelia Fresneda Pacheco

María Ramos Charbonnier

The views expressed in this working paper are those of the authors and do not necessarily reflect the views of the Instituto Nacional de Estadística of Spain

First draft: December 2010

This draft: December 2010

Changes of classification in continuous statistics: Calculation of retrospective series. Application to the Quarterly Labour Cost Survey

Abstract

With the aim of showing the structural changes in economy an updated National Classification of Economic Activities is established, named from now on CNAE091 which, since January 2009 replaces the previous one, CNAE931. This classification allows enterprises, financial organizations and governments to have at their disposal comparable and reliable data.

The transition to that new classification requires the producers of official statistics to adapt their statistical systems before implementing the new classification.

This paper tries to explain the procedure applied to adjust the estimations about Labour Costs to the new classification

Keywords

Backcasting, Calibration, Post-stratified estimator, Stratified Random Sampling, National Classification of Economic Activities, Quarterly Labour Cost Survey

Authors and Affiliations

Amelia Fresneda Pacheco and Maria Ramos Charbonnier

Labour Market Statistics. National Statistics Office of Spain

Changes of classification in continuous statistics: Calculation of retrospective series

Application to the Quarterly Labour Cost Survey

Amelia Fresneda Pacheco and María Ramos Charbonnier

Labour Market Statistics

National Statistics Office of Spain

Abstract

With the aim of showing the structural changes in economy an updated National Classification of Economic Activities is established, named from now on CNAE09¹ which, since January 2009 replaces the previous one, CNAE93¹. This classification allows enterprises, financial organizations and governments to have at their disposal comparable and reliable data.

The transition to that new classification requires the producers of official statistics to adapt their statistical systems before implementing the new classification.

This paper tries to explain the procedure applied to adjust the estimations about Labour Costs to the new classification.

Key Words: Backcasting, Calibration, Post-stratified estimator, Stratified Random Sampling, National Classification of Economic Activities, Quarterly Labour Cost Survey.

AMS Subject classifications: 62P25, 62M10, 62D05.

1. Introduction

When working with short term statistics, whenever a change in the system is produced, motivated for example by a change in the classifications, the estimations built from the coming into effect of the new system are not comparable to the previous ones. Frequently it is desirable to provide temporal series of the results with a temporal horizon enough to allow its analysis, and that's why it is necessary to elaborate retrospective series in the new system.

The Quarterly Labour Cost Survey (QLCS) [1] which is elaborated by National Statistics Office of Spain (NSO) allows us to attend to the demands of information of time series about labour cost and its main components and worked hours.

¹ The National Classification of Economic Activities, CNAE, is the spanish version of the EU Classification of Economic Activities, NACE.

In 2009 the QLCS must be adapted to the new classification of economic activities CNAE09 ([2],[3]) which came into effect on January the 1st 2009. QLCS sampling design is based on stratified sampling , and CNAE09 is one of the stratifying variables and that's why the change of classification entails changes both in the distribution of population of reference as in the sampling design.

With the aim of elaborating the series in CNAE09 for the period of 2000-2008 a complex process of reclassification, estimate and linkage was carried out for two years, consisting of different phases.

2. Previous phase: Estimation method

To approach a problem as the one arisen here two alternatives are set out ([4],[5])

- *Micro approach*: reconstructing the frame during the years previous to the change, reclassifying population unities and recalculating estimations later on.

The main advantage of that way of actuating is to allow reconstructing practically the whole of a survey under the new system. In return the main disadvantage is that the quality of the new obtained estimations could be affected. That effect could be alleviated using improved estimators for the period to be reconstructed, but that could produce a break in the series elaborated with the original estimators, which should be taken into account.

- *Elaboration of conversion matrixes (Macro approach)*: constructing conversion coefficients from the estimations obtained under both classifications in the same period of reference.

The main advantage is that it implies a lesser volume of work because it doesn't need reclassification of population units in previous periods. The disadvantage is the bigger complexity due to the number of matrixes to be elaborated which increase with the number of variables to be recalculated and those matrixes should secure the coherence of the new estimations.

Analyzing both alternatives in the case of QLCS, regarding the use of microdata , it must be taken into account that the starting point are the samples designed to cover accurately the strata of the previous classification, and that's why the coverage given to the previous classification could be insufficient specially in newly created strata. Given that the CNAE09 is disaggregated in a bigger number of strata than the previous CNAE93 the problem of coverage seems evident.

About the construction of conversion matrixes, due to the great diversity of variables that QLCS provides, they should be coherent among them to provide transformed variables that should respect the interrelationships of the initial variables (distribution structure of the components of labour cost and time of work) maintaining simultaneously the time coherence of the series.

Consequently it was chosen to estimate the retrospective series in CNAE09 through microdata to obtain directly all the required estimations and maintain the coherence and comparability of the information system of Labour Cost elaborated by NSO.

3. First phase: year 2008, coexisting CNAE93 and CNAE09

When a system is changed it is necessary a common period of coexistence between the old and the new structure that enables the linking of both of them. In the case of QLCS, the year 2008 is the common period chosen as it is the year previous to the coming into effect of the new CNAE and also as it is the year when NSO already has a Central Business Register (CBR) (frame for the majority of business surveys elaborated by NSO and where business are identified by a Tax Identification Number (TIN), [6]) codified in this new classification.

The first step is to have the population frame codified in this new classification, which, in the case of QLCS, is the Social Security Contribution Accounts Register (which is a local unit register that includes among its variables the tax identification number of the business), but in 2008 it was not still available in CNAE09, and that is why it is provisionally reclassified according to the following method:

- CNAE09 codes having a biunivocal relationship with CNAE93 codes are reclassified directly.
- For the rest of the units, a crossing by TIN between the Social Security Register and CBR is carried out, assigning the CNAE code of CBR to the contribution accounts (basic units of the Social Security) whose TIN appear only once in the CBR.
- Finally, for the rest of the units, a probabilistic assignation of CNAE is done, depending on the distribution given by the matrix of correspondences (M) CNAE09 – CNAE93 obtained from CBR.

Procedure	Record	Percentage
P1. Direct conversion (4digits)	160.305	10.51%
P2. Direct conversion (5digits)	1.266.929	83.07%
P3. Code from Business Register	66.313	4.35%
P4. Probabilistic conversion	31.533	2.07%
Total Units	1.525.080	100%

Table 1: Previous reclassification of population frame QLCS-2008

Once available this provisional population frame in CNAE09, an enlargement of the sample QLCS-2008 is designed so that, the whole sampling, is representative of the new classification. Given that the change means passing from 54 divisions to 82, a sampling increase of approximately 1/3 is produced.

	Sample 2008 in CNAE-09	Sample 2008 in CNAE-93	Sample enlargement	
Total	28.432	20.902	7.530	36%
Industry	8.269	6.962	1.307	19%
Construction	3.019	2.141	878	41%
Services	17.144	11.799	5.345	45%

Table 2: Sampling enlargement QLCS-2008.

This sampling enlargement makes it possible to already obtain estimations in CNAE09 which coexist during 2008 with estimations in CNAE93 made with the traditional sample.

The comparison between them will make it possible to obtain the linkage among the series elaborated under the new CNAE from 2008 on and the retrospective estimations in the new CNAE of the data previous to 2008.

4. Second phase: Availability of the population frame 2009 in CNAE09 and reclassification of previous population frame.

At the end of 2008 the Social Security already has a Contribution Accounts Register codified in CNAE09. By means of it the provisional assignation of CNAE used in the first phase is corrected, in order to avoid divergences in data due to classification criteria. In the same way, population frames for the years 2007-2000 are reclassified to be able to estimate QLCS retrospectively under the new classification.

To carry out these reclassifications the following sequential procedure is followed, and its results are shown in Table 3.

Sequential procedure:

1. For a year T, for $2009 > T \geq 2000$, given $DIR(T)$, $DIR(T-1)$ the probabilistic conversion matrix is obtained between both classifications at a 2 digits level.
2. CNAE09 code is assigned according to the situation of each unit in the population frame of the consecutive year:
 - a) $DIR(T) \cap DIR(T-1)$, with the same CNAE93 code $\Rightarrow CNAE09(T-1) = CNAE09(T)$
 - b) $(DIR(T-1) \not\subset DIR(T))$ or $(DIR(T) \cap DIR(T-1))$, with different CNAE93 code:
 - i. If a biunivocal relationship exists between CNAE93 and CNAE09 at 2 or more digits $\Rightarrow CNAE09(T-1)$ obtained by direct correspondence.
 - ii. If a direct correspondence does not exist $\Rightarrow CNAE09(T-1)$ is obtained from the probabilistic conversion matrix.

	Units				Percentage		
	Total	Common units	Direct conversion	Probabilistic conversion	Common units	Direct conversion	Probabilistic conversion
2008	1.525.080	1.262.346	221.595	41.139	82,77%	14,53%	2,70%
2007	1.485.807	1.239.342	209.858	36.607	83,41%	14,12%	2,46%
2006	1.442.343	1.234.679	176.388	31.276	85,60%	12,23%	2,17%
2005	1.330.785	1.156.466	148.917	25.402	86,90%	11,19%	1,91%
2004	1.283.352	1.108.959	147.005	27.388	86,41%	11,45%	2,13%
2003	1.241.146	1.072.496	142.320	26.330	86,41%	11,47%	2,12%
2002	1.203.872	1.036.656	140.914	26.302	86,11%	11,71%	2,18%
2001¹	1.160.121	999.549	8.294	152.278	86,16%	0,71%	13,13%
2000¹	1.119.843	943.501	15.452	160.890	84,25%	1,38%	14,37%

Table 3: Reclassification procedures of the annual population frame of QLCS

5. Third phase: Estimators in CNAE09

5.1. Traditional estimators

QLCS sampling design is based on stratified sampling ([7]) attending to three variables. Each stratum is defined as the crossing of the following variables: Region (Autonomous Communities in Spain) Economic activity (according to CNAE valid classification) and the number of employees of the units (8 strata or size groups)

The estimator used in QLCS is a *Separated Ratio Estimator* (see 8) using as auxiliary variable the number of employees in the Social Security Contribution Accounts Register.

$$\text{For the stratum h: } \hat{Y}_h = \frac{\sum_{k=1}^{nh} Y_{hk}}{\sum_{k=1}^{nh} D_{hk}} \cdot D_h = D_h \cdot \frac{\sum_{k=1}^{nh} Y_{hk}}{d_h} = F_h \cdot \sum_{k=1}^{nh} Y_{hk}$$

$$\text{For the whole population: } \hat{Y} = \sum_h \hat{Y}_h = \sum_h \left(F_h \cdot \sum_{k=1}^{nh} Y_{hk} \right)$$

Where:

Y_{hk} = value of the variable Labour Cost in sampling units from stratum h.

Y_h ; Y = totals of the variable Labour Cost in stratum h and in the whole population, respectively.

¹ The increase of units reclassified with the probabilistic conversion matrix during the years 2000 and 2001 is due to the fact that during those years population frames are only coded at a 2 digits level.

D_h = total of employees in stratum h in the Social Security Contribution Accounts Register.

D_{hk} = number of employees in the Social Security Contribution Accounts Register, from sampling units from stratum h.

d_h = total of employees from sampling units from stratum h

$F_h = D_h / d_h$ weighting factor from sampling units from stratum h.

Through them “The Estimation of monthly Labour Cost Per Employee” and “The Hourly Labour Cost” are obtained as ratio of the Total Labour Cost and the Total of Employees, and as ratio of the Total Labour Cost and or the Total Hours Worked, respectively:

$$\hat{CT} = \frac{\hat{Y}}{\hat{T}} = \frac{\sum_h \left(F_h \cdot \sum_{k=1}^{nh} Y_{hk} \right)}{\sum_h \left(F_h \cdot \sum_{k=1}^{nh} T_{hk} \right)}, \quad \hat{CH} = \frac{\hat{Y}}{\hat{H}} = \frac{\sum_h \left(F_h \cdot \sum_{k=1}^{nh} Y_{hk} \right)}{\sum_h \left(F_h \cdot \sum_{k=1}^{nh} H_{hk} \right)}$$

5.2. Estimators for 2008-2009

During 2008 and 2009 samples designed under the new classification CNAE09 are available what enables to construct the usual QLCS estimators, but under the new classification.

5.3. Post-stratified estimators for the retrospective series 2000-2007

Once reclassified all the QLCS population frame and samples from 2007 to 2000 another important decision should be taken: which estimator should be used to get the survey results in CNAE09.

Three possible methods of estimation are considered:

- *Method 1:* calculate separate ratio estimator using the original weighting factors calculated in CNAE93.
- *Method 2:* calculate separate ratio estimator using the weighting factors recalculated, obtained for the reclassification carried out in CNAE09
- *Method 3:* calculate post-stratified estimator, whose weighting factor takes into account both the original sampling design in CNAE93 and the effect of reclassification and redistribution of the sample in CNAE09

To assess them the annual growth rate of the two main variables of the survey are used: “Wage and Salary Costs per Employee” and “Total Labour Cost per Employee”.

	Wages and Salary Costs per Employee				Total Labour Cost per Employee			
	CNAE09			CNAE93	CNAE09			CNAE93
	1	Method 2	3	Growth rate published	1	Method 2	3	Growth rate published
Total	4,29%	4,19%	4,07%	4,40%	4,62%	3,90%	4,07%	4,00%
Industry	4,34%	4,14%	4,16%	4,10%	4,80%	3,85%	4,16%	4,20%
Construction	5,48%	4,58%	3,91%	5,00%	5,56%	4,71%	4,14%	5,00%
Services	4,19%	4,24%	4,21%	4,50%	4,49%	3,90%	3,71%	4,00%

Table 4: Assessment of retrospective estimation method for QLCS (1)

It is noticed that the rates obtained using method 1 (estimations with the original weighting factors) are higher than the ones calculated using the other methods, and that is why we may assume that method 1 over-estimates the results in CNAE09.

To confirm the above, estimations of “Total of employees”¹ obtained with the different methods in CNAE09 are compared to the ones obtained in CNAE93.

	Total of Employees			
	CNAE09			CNAE93
	1	Method 2	3	Growth rate published
Total	14.084.804	13.342.423	13.099.401	13.111.939
Industry	2.383.897	2.324.564	2.231.228	2.362.751
Construction	1.864.348	1.856.047	1.839.146	1.770.008
Services	9.836.560	9.161.812	9.029.027	8.979.180

Table 5: Assessment of retrospective estimation method for QLCS (2)

Through method 1 a total number of workers, exceeding in almost 1 million the results obtained in CNAE93, is obtained. By sectors of activity this tendency is also noticed, although the conclusion is less obvious because sectors of economic activity have undergone modifications in their new definition in CNAE09 with regard to its definition in CNAE93.

Finally, the comparison between the estimations obtained and the population data (table 6) is done. Due to the fact that the sample designed in 2007 doesn't cover the CNAE09, and the lack of sample has an important effect on the estimations of the totals, only the comparison for the sections of activities whose sampling size does not increase with the change in NACE is carried out.

¹ “Total of Employees” is not a QLCS objective variable, but it is interesting because is a comparable data obtained from the population frame.

	CNAE09			Population frame
	Methods			
	1	2	3	
Mining and quarrying	44.376	40.954	41.210	42.850
Electricity, gas, steam and air conditioning supply	40.463	38.422	38.176	38.267
Financial and insurance activities	408.303	404.335	403.151	394.184
Public administration and defence; compulsory social security	896.215	970.967	966.040	974.956
Education	589.719	587.357	585.726	559.096

Table 6: Assessment of retrospective estimation method for QLCS (3)

The results obtained with method 1 differ more in the population values than the ones obtained by methods 2 and 3

Definitely, those tests let us conclude that method 1 give us results which differ more from the results in ETCL than the results obtained through methods 2 and 3 and, a consequently, it is rejected.

About the choice between methods 2 and 3, the results obtained by both methods are similar and no conclusion is possible. Because of the lack of evidence against, it is decided to use post-stratified estimators, due to the fact that in their weighting factors take part as much the original sampling design in CNAE93, used to obtain the sample, as the redistribution and reclassification effect of the sample in CNAE09. For example, in CNAE09 division 47, coming only from CNAE93 divisions 50 and 52, its post-stratified estimator expression is given by:

$$\hat{Y}_{47} = \sum_{k \in \{50,52\}} \left(\frac{D_{kj}}{d_{kj}} \right) \cdot \hat{Y}_{kj} = D_{50\ 47} \sum_{i=1}^{n_{50\ 47}} \frac{y_{50\ 47\ i}}{d_{50\ 47}} + D_{52\ 47} \sum_{i=1}^{n_{52\ 47}} \frac{y_{52\ 47\ i}}{d_{52\ 47}}$$

Where:

$n_{k,47}$ = number of units in the sample coded in code k from CNAE93 and in code 47 from CNAE09.

$D_{k,47}$ = number of units in the population frame coded in code k from CNAE93 and in code 47 from CNAE09.

$d_{k,47}$ = number of employees in the sample asociated in code k from CNAE93 and in code 47 from CNAE09.

$y_{k,47,i}$ = value for the variable y from the unit "i" in the sample asociated in code k from CNAE93 and in code 47 from CNAE09.

This way of constructing estimations in CNAE09 starting from a sample taken in CNAE93 enables to maintain the profiles of the series published in the initial classification.

6. Forth phase: Linking between 2000-2007 and 2008-2009 series.

The sampling enlargement carried out during 2008 to cover the new CNAE means a third of the sampling used until this year. Besides, this sampling enlargement is not distributed in an equitable way among the different strata but it is mainly concentrated on the divisions of new activities.

As a result, the new estimations obtained since 2008 in CNAE09 with the enlarged sample produce results that differ from the ones obtained till 2007 with post-stratified estimators on the traditional sampling. And so the need to link the reconstructed series for 2007-2000 and the new series starting in 2008 arises.

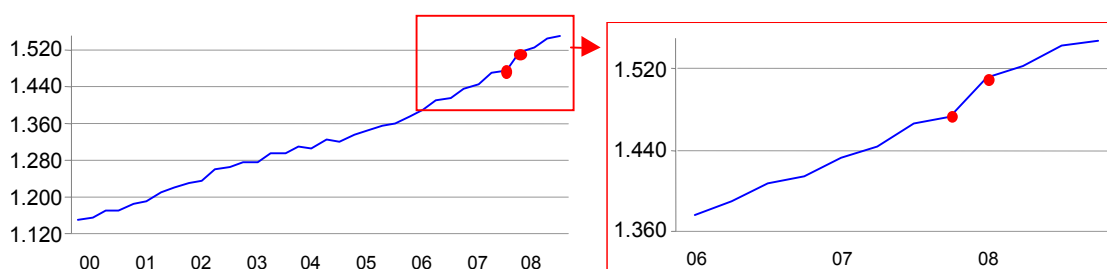


Figure 1: Wage and Salary Costs Series for 2007-2000 for the comparable sections (excluding the Public Administration section)

To carry out this linking the double estimation in CNAE09 and CNAE93 for the common period (2008) of coexistence of both systems is used.

The change of classification implies a redistribution of units according to their economic activity, but obviously the distribution of data by regions does not get altered. Furthermore, although the divisions and sections of activity undergo important changes with regard to the previous classification, its grouping into the three big sectors of activity (industry, building and services) keep their characteristics, and that is why it is expected that the results obtained for those main sectors do not show deviations regarding the results that were obtained with the previous classification.

So the following basic hypothesis is established:

“The growth rates, regional and by economic sectors, between the data in CNAE09 obtained through post-stratified estimators and the new data, which since 2008 are elaborated in CNAE09 with an enlarged sample, should not differ from the ones published in 2008 for the traditional series in CNAE93”

Moreover, two other additional hypothesis are considered:

- *The distribution of the variable “Employees (T)” obtained through the post-stratified estimator is correct and corrector coefficients are tried to be searched to apply to the estimators of Total Labour Cost and Worked Hours”*
- *To adjust to the maximum to the original estimators it is required that, for each division ds, the resulting estimator should not differ more than 15% of the original estimator.*

The following system of equations is obtained:

For each sector S (Industry, Construction, Services) and each region $R \in \{01,02,\dots,17\}$ coefficients $A_{R ds}$ are sought to check

$$\frac{\sum_{ds \in S} A_{R ds} \cdot C_{R ds}^{07}}{T_{RS}^{07}} = \frac{CT_{RS}^{08}}{(1 + t_{RS}^{07/08})}, \text{ with the constraint } 0,85 \leq A_{R ds} \leq 1,15$$

Where:

$t_{RS}^{07/08}$ = growth rates for Cost per employee in NACE Rev. 1.1 published in 2008.

CT_{RS}^{08} = Total Labour Cost per employee in 2008 calculated with the traditional estimator.

T_{RS}^{07} = number of employees in 2007 calculated with the post-stratified estimator

$C_{R ds}^{07}$ = Total Labour Cost in 2007 calculated with the post-stratified estimator

The solution to this system provides coefficients $A_{R ds}$ that makes it possible to correct the series backwards and guarantees that the corrected series for the variable “Total Labour Cost per Employee (CT)” shows a parallel profile to the original series.

But the variable “Total Labour Cost” is nothing but the variable resulting of adding the different components that form labour cost. From those components the most outstanding is “Wage and Salary Costs (ST)” which is the 80% of the Total Labour Cost and which has a marked seasonal behaviour that incorporates the total labour cost to the series profile.

To maintain the profile of this component the system previously stated is replicated, but on the variable “Wage and Salary Costs (ST)”, obtaining the adjust coefficients $B_{R ds}$.

The solutions of both systems make it possible to replicate the profile of the series published for the variables “Total Labour Cost” and “Wage and Salary Costs” respectively, but none of them provides a solution that adjusts simultaneously both profiles. This simultaneous adjustment is obtained through a geometric mean of both solutions:

$$W_{R ds} = \sqrt[2]{A_{R ds} \cdot B_{R ds}}$$

Thus, $W_{R ds}$ is considered as the only corrector coefficient for every labour cost component.

But there are still to link the variables referring to “*Worked Hours*”. To do that the basic hypothesis is applied again, though this time on the variable “*Cost per Worked Hour*” and it is done by using the variable “*Total Labour Cost*”, already adjusted with corrector coefficients $W_{R ds}$.

The proposed systems are replicated for each one of the four quarter of 2008, producing 4 correction coefficients $W_{R ds}$ for the economic variables of the survey and 4 correction coefficients $K_{R ds}$ for the variable Hours Worked.

Finally post-stratified estimators on adjusted microdata with those corrector coefficients are applied once more. Thus retrospective series of labour cost under the new classification CNAE09 are obtained, whose growing profile, in comparable groups, are similar to that previously published in CNAE93

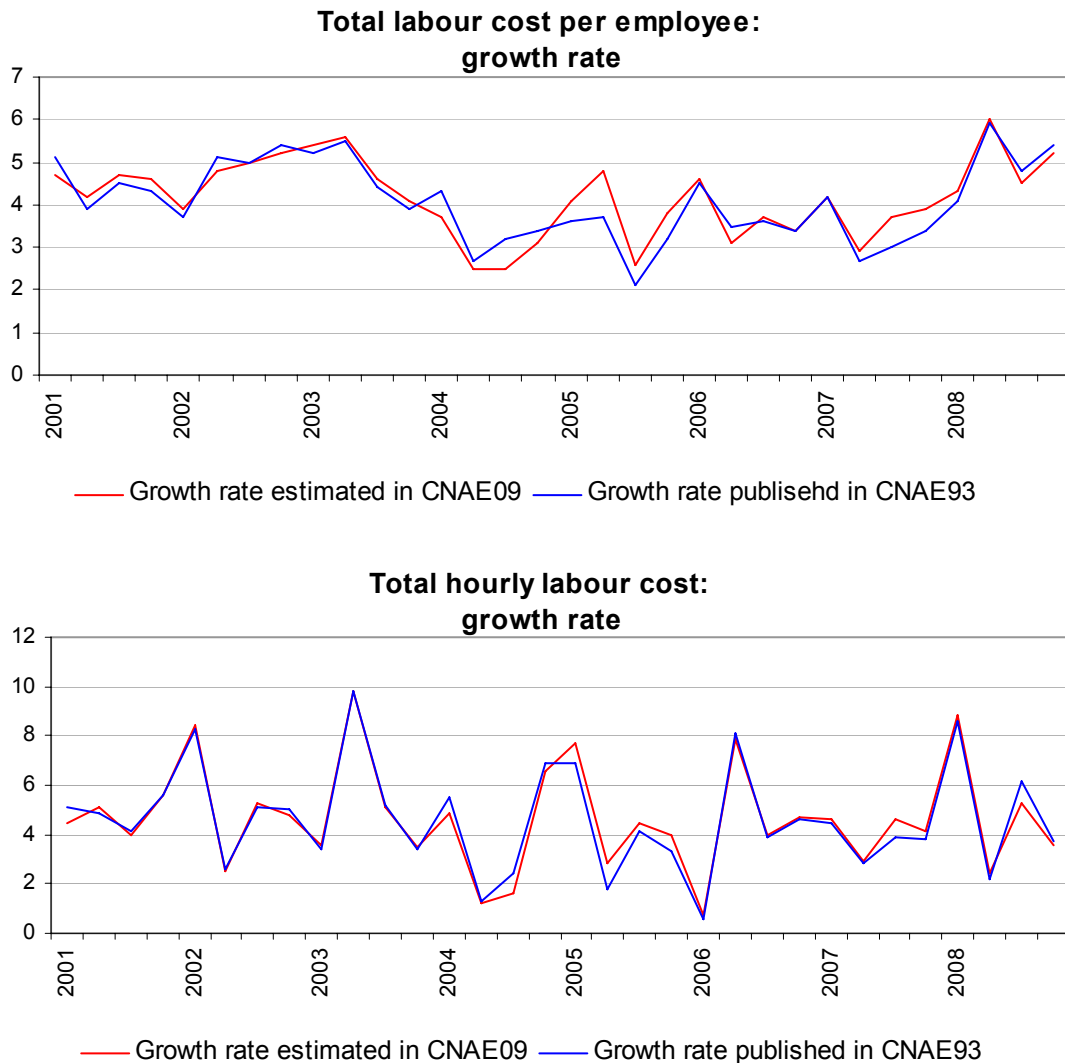


Figure 2: Comparison between series from 2008-2000 in both classifications. Sector: Industry.

7. Fifth Phase: Revision of the New Classification

Given that CNAE09 is a new classification, for the two first years of its implementation (2008 and 2009), the assignation of the new CNAE09 codes on the sample units that collaborate on QLCS is carefully researched. Through the rectifications obtained from this research the population frame and reclassified samples are revised and corrected.

$N_h, n_h, D_h, d_h \Rightarrow$ startum h sizes before the review.

$N'_h, n'_h, D'_h, d'_h \Rightarrow$ startum h sizes after the review

The whole procedure explained? above is reviewed? because of the revision of population frame and sample sizes. Moreover, for years 2008 and 2009 weighting factors should be adjusted to obtain representative estimations for each economic activity. To solve this problem different refactorization procedures are considered:

- *Method 1:* Using recalibrating techniques (see 9) $\Rightarrow F_{Calh} = F_h \cdot \frac{N'_h}{N_h}$
- *Method 2:* Using original weighting factors $\Rightarrow F_h = \frac{D_h}{d_h}$
- *Method 3:* Recalculating weighting factors after the revision $\Rightarrow F'_h = \frac{D'_h}{d'_h}$

Method 2 is rejected because it distorts the results since, when applying a change of classification to a sample unit, if the weighting factors are not rectified, the immediate effect is that the particular unit results over-represented or undervalued in the division of the new activity where it has been classified.

Unit	Initial CNAE	Initial Factor	Revised cNAE	Revised Factor	Wage
i	Retail	133	Quarryng	1	610 €/worker
Average Wage obtained for CNAE= Quarryng					
Initially (Unit "i" out of quarryng)		=> 2.297	Without revision of the factor, the unit "i" represents 100 workers with wage 610 €, but in fact it should represent only one worker. As consequence, the average wage in the stratum is underestimated.		
With the revision of the unit "i"		=> 1.352			

Table 7: Effects of method 2 in results

To evaluate methods 1 and 3 the results obtained and survey data (table 8), and the population frame (table 9) are compared:

Wages Excluding Bonuses by Employee Total Labour Cost per Employee

	CNAE09			CNAE09		
	Method		Growth rate published	Method		Growth rate published
	1	3		1	3	
Total	4,68%	5,17%	5,0%	4,49%	4,96%	5,1%
Industry	5,04%	4,95%	4,9%	3,37%	3,30%	4,1%
Construction	7,54%	7,35%	4,7%	7,96%	8,01%	5,8%
Services	3,99%	4,67%	5,0%	4,13%	4,75%	5,2%

Table 8: Effects of methods 1 and 3 in results of variable "Total Labour Cost".

Total of Employees

	CNAE09		Population frame
	Methods		
	1	3	
Total	16.432.111	14.473.682	14.567.499
Industry	2.738.422	2.413.453	2.457.847
Construction	2.302.353	1.972.279	1.982.571
Services	11.391.336	10.087.951	10.127.081

Table 9: Effects of methods 1 and 3 in results of variable "Number of employees".

In every case similar growth rates of Total Labour Cost are obtained. Nevertheless, method 1 (calibrating) overestimates the data from "Total of Employees", both in the total as in the main economic sectors. As a consequence it is concluded that the most suitable method to apply the CNAE09 changes is factor recalculation (method 3)

Conclusion

The CNAE change of classification has meant a substantial change in QLCS survey due to the fact that its sampling design is stratified according to this classification.

The further disaggregation in the new classification has meant a considerable sampling enlargement and a break in data series. In order to be able to obtain short-term data series in CNAE09 of QLCS, since the beginning of the survey, a complex process of re-classification and tabulation has been established.

The keys to the process have been the conversion of the whole of the survey to the new CNAE09, the use of post-stratified estimators considering CNAE double codification and the availability of a common year of coexistence of both classifications to be able to obtain corrector factors enabling the linking between the previous series and the subsequent ones to the change of classification.

References:

- [1] Quarterly Labour Cost Survey. INE
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft22/p187&file=inebase&L=0>
- [2] Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20 December 2006 establishing the statistical classification of economic activities NACE Revision 2 and amending Council Regulation (EEC) No 3037/90 as well as certain EC Regulations on specific statistical domains.
- [3] National Classification Of Economic Activities: CNAE-2009
Royal Decree 475/2007 of April 13th 2007.
- [4] Fortier, S.(2003) La conversion des données historiques selon un nouveau système de classification pour l'enquête mensuelle sur le commerce de gros et de détail". Proceedings Papers. 2003 Annual meeting of Statistical Society of Canada.
- [5] Hidioglou, M. Quenneville, B. and Huot, G. (2001) Methodological problems and options for SIC-NAICS conversion. Working Papers. Statistics Canada.
- [6] Central Business Register. INE
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t37/p201&file=inebase&L=0>
- [7] Särndal, C.E., Swensson, B. and Wretman J. (1992). Model Assisted Survey Sampling, Springer-Verlag.
- [8] Cochran, W.G. (1980). Técnicas de muestreo. Compañía Editorial Continental. México. Cochran, W.G. .
- [9] Deville, J.C., Särndal,C.E. (1991). Calibration Estimators in Survey Sampling.. Journal of the American Statistical Association (JASA)

About the authors

Amelia Fresneda Pacheco is Head of Unit of the Labour Cost Area in the National Statistics Office of Spain. Among her duties are the research and development of statistical techniques to obtain series about labour cost and for the integration of structural and temporary /short-term results.

María Ramos Charbonnier is Senior Statistician in the Labour Cost Area of the National Statistics Office of Spain . Up till now her research has been focussed on the development of the implementation procedure and the transition of the NSO labour cost integrated statistical system to the new National Classification of Economic Activities CNAE09