
Estadística Oficial

Changes of classification in short term statistics: Calculation of retrospective series

Amelia Fresneda and María Ramos

S.G. Estadísticas del Mercado Laboral
Instituto Nacional de Estadística
✉ amefres@ine.es, mramochar@ine.es

Abstract

With the aim of showing the structural changes in economy an updated National Classification of Economic Activities is established, named from now on CNAE09 which, since 2009 replaces the previous one, CNAE93 Rev.1. 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.

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

1. Introducción

Cuando se trabaja con encuestas continuas, siempre que se produce un cambio de sistema, motivado por ejemplo por un cambio en las clasificaciones las estimaciones construidas a partir de la entrada en vigor del nuevo sistema no son comparables con las anteriores. Con frecuencia es deseable disponer de series temporales de los resultados con un horizonte temporal suficiente para permitir su análisis, por lo que resulta necesario elaborar series retrospectivas en el nuevo sistema.

La Encuesta Trimestral de Coste Laboral (ETCL) ([1]) que elabora el Instituto Nacional de Estadística (INE) permite atender la demanda de información

de indicadores coyunturales sobre costes laborales y sus componentes y horas trabajadas.

En el año 2009 la ETCL ha de adaptarse a la nueva clasificación de actividades económicas CNAE09 ([2],[3]) aprobada por el Real Decreto 475/2007 y que entró en vigor el 1 de enero de 2009. EL diseño muestral de esta encuesta se basa en un muestreo estratificado, en el que el código de CNAE es una de las variables de estratificación, por lo que el cambio de clasificación conlleva cambios tanto en la distribución de la población de referencia, como en el diseño muestral.

Con el fin de elaborar las series en CNAE09 para el período 2000-2008, se llevó a cabo un complejo proceso de reclasificación, estimación y enlace, de dos años de duración, consistente en varias fases.

2. Fase previa: Método de estimación

Para abordar un problema como el que se ha presentado, se plantean dos alternativas ([4],[5]):

- *Utilización de microdatos:*

Consiste en la reconstrucción del marco en los años previos al cambio, reclasificando las unidades poblacionales, y recalculando posteriormente las estimaciones.

La principal ventaja de este modo de actuación es que permite reconstruir la práctica totalidad de una encuesta bajo el nuevo sistema.

Como contrapartida, el principal inconveniente es que la calidad de las nuevas estimaciones obtenidas puede verse afectada. Dicho efecto puede ser paliado a través del uso de estimadores mejorados para el periodo que se pretende reconstruir, pero eso produciría una ruptura en la serie elaborada con los estimadores originales, que debería ser tenida en cuenta.

- *Elaboración de matrices de conversión:*

Consiste en la construcción de coeficientes de conversión a partir de las estimaciones obtenidas bajo ambas clasificaciones en un mismo periodo de referencia.

La principal ventaja es que implica un menor volumen de trabajo, al no precisar la reclasificación de las unidades poblacionales en los períodos anteriores.

El inconveniente es su mayor complejidad debido a que el número de matrices a elaborar aumenta con el número de variables a recalcular, y que dichas matrices deben asegurar la coherencia de las nuevas estimaciones.

Analizando ambas alternativas en el caso de la ETCL, respecto a la utilización de microdatos hay que tener en cuenta que se parte de muestras que se diseñaron para cubrir con precisión los estratos de la antigua clasificación, por lo que la cobertura que darán de la nueva clasificación podría resultar insuficiente, especialmente en los estratos de nueva creación. Dado que la CNAE09 se desagrega en un mayor número de estratos que la antigua CNAE93, el problema de cobertura parece evidente.

Sobre la construcción de matrices de conversión, debido a la gran diversidad de variables que proporciona la ETCL, dichas matrices deberían ser coherentes entre sí para proporcionar unas variables transformadas que respetaran las interrelaciones de las variables iniciales (estructura de distribución de los componentes del coste laboral y del tiempo de trabajo), manteniendo al mismo tiempo la coherencia temporal de las series.

En consecuencia, se opta por la opción de estimar las series retrospectivas en CNAE09 a través de microdatos a fin de obtener directamente todas las estimaciones requeridas y mantener la coherencia y comparabilidad del sistema de información de Costes Laborales del INE.

3. Primera fase: Año 2008, coexistencia CNAE93 y CNAE09

Cuando se cambia de sistema se precisa de un periodo común de convivencia entre la antigua y la nueva estructura que permita enlazar ambas. En el caso de la ETCL, el año 2008 es el periodo común elegido por ser el previo a la entrada en vigor de la nueva CNAE y por ser el año en que el INE dispone ya de un Directorio Central de Empresas (DIRCE) (marco de la mayoría de las encuestas a empresas de esta institución, [6]) codificado en esta nueva clasificación.

El primer paso es tener el marco poblacional definido en esta nueva clasificación, que en el caso de la ETCL es el directorio de Cuentas de Cotización de la Seguridad Social, pero en 2008 aún no estaba disponible en CNAE09, por lo que es reclasificado provisionalmente según el siguiente método, cuyos resultados se muestran en el Cuadro 1:

- Los códigos CNAE09 que tienen una relación biunívoca con códigos CNAE93 se reclasifican directamente.
- Para el resto de unidades se efectúa un cruce por NIF entre el directorio de Seguridad Social y el DIRCE, asignando la CNAE09 del DIRCE a las Cuentas de Cotización (unidades básicas de Seguridad Social) unilocalizadas, es decir, cuyo NIF aparece una sola vez en el directorio.

- Finalmente, para el resto de unidades se realiza una asignación probabilística de CNAE09, en función de la distribución dada por la matriz de correspondencias (M) CNAE09-CNAE93 obtenida del DIRCE.

Procedimiento	Registros	Porcentaje
P1. Conversión directa 4 dígitos	160.305	10.51 %
P2. Conversión directa 5 dígitos	1.266.929	83.07 %
P3. Utilización DIRCE a doble codificación	66.313	4.35 %
P4. Conversión probabilística	31.533	2.07 %
Total de unidades	1.525.080	100 %

Cuadro 1: Reclasificación previa del marco poblacional ETCL-2008

Una vez disponible este marco poblacional provisional en CNAE09, se diseña una ampliación de la muestra ETCL-2008 para que, la muestra total, sea representativa de la nueva clasificación. Dado que el cambio de CNAE supone pasar de 54 divisiones de actividad a 82, se produce un incremento muestral de 1/3 aproximadamente, como se puede observar en el Cuadro 2.

	Muestra 2008 en CNAE09	Muestra 2008 en CNAE93	Incremento muestral	
Total	28.432	20.902	7.530	36 %
Industria	8.269	6.962	1.307	19 %
Construcción	3.019	2.141	878	41 %
Servicios	17.144	11.799	5.345	45 %

Cuadro 2: Ampliación muestral ETCL-2008

Esta ampliación muestral permite obtener ya estimaciones en CNAE09 que coexisten durante 2008 con las estimaciones en CNAE93, efectuadas con la muestra tradicional.

La comparación entre ambas permitirá obtener el enlace entre las series elaboradas bajo la nueva CNAE09 a partir de 2008, y las estimaciones retrospectivas en la nueva CNAE de los datos anteriores a 2008.

4. Segunda fase: Disponibilidad del directorio 2009 en CNAE09 y reclasificación de directorios anteriores.

A finales de 2008, la Seguridad Social dispone ya de un directorio de Cuentas de Cotización codificado en CNAE09. Con él se corrige la asignación provisional de CNAE09 utilizada en la primera fase, a fin de evitar divergencias en los datos debidas a criterios de clasificación. Asimismo, se reclasifican los directorios poblacionales para los años 2007-2000 para poder estimar la ETCL retrospectivamente bajo la nueva codificación.

Para realizar estas reclasificaciones se sigue el siguiente procedimiento secuencial, cuyos resultados se muestran en el Cuadro 3:

Para el año T tal que $2009 > T \geq 2000$, dado $DIR(T)$ se obtiene $DIR(T-1)$:

1. Se obtiene la matriz de correspondencias probabilística (M_T) entre CNAE93 a 2 dígitos y CNAE09 a 2 dígitos.
2. Se asigna CNAE09 según la situación de cada unidad en el directorio del año consecutivo:
 - (a) $DIR(T) \cap DIR(T-1)$, con idéntica CNAE93 \Rightarrow $CNAE09(T-1) = CNAE09(T)$
 - (b) $[DIR(T-1) \notin DIR(T)]$ ó $[DIR(T) \cap DIR(T-1)$, con distinta CNAE93]:
 - i. Si existe una relación biunívoca entre CNAE93 y CNAE09 a 2 ó más dígitos \Rightarrow CNAE09(T-1) obtenida por correspondencia directa.
 - ii. Si no existe correspondencia directa \Rightarrow CNAE09(T-1) a partir de la matriz de correspondencias probabilística (M_T).

Año	Total	Unidades			Porcentaje		
		Comunes idéntica	Relación biunívoca	Relación probabilística	Comunes idéntica	Relación biunívoca	Relación probabilística
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 %

Cuadro 3: Procedimiento de reclasificación de los directorios poblacionales anuales de ETCL

5. Tercera fase: Estimadores en CNAE09

5.1. Estimadores tradicionales

El diseño muestral de la encuesta ETCL se corresponde con un muestreo estratificado ([7]) atendiendo a tres variables. Cada estrato se define como el cruce

¹El aumento de unidades reclasificadas por probabilidad en los años 2000 y 2001 se debe a que en dichos años en los directorios poblacionales solo se dispone de la clasificación CNAE93 a 2 dígitos.

de las siguientes variables: la región (Comunidades Autónomas de España), la actividad económica (según la clasificación CNAE vigente) y el número de trabajadores de las unidades (8 estratos o grupos de tamaño).

Los estimadores para totales utilizados en la ETCL son *Estimadores Separados de Razón* ([8]), usando como variable auxiliar el número de trabajadores en el Directorio de Cuentas de Cotización a la Seguridad Social:

$$\text{Para el estrato } h \rightarrow \hat{Y}_h = \frac{\sum_{k=1}^{n_h} Y_{hk}}{\sum_{k=1}^{n_h} D_{hk}} \cdot D_h = D_h \cdot \frac{\sum_{k=1}^{n_h} Y_{hk}}{d_h} = F_h \cdot \sum_{k=1}^{n_h} Y_{hk}$$

$$\text{Para el total poblacional } \rightarrow \hat{Y} = \sum_h \hat{Y}_h = \sum_h \left(F_h \cdot \sum_{k=1}^{n_h} Y_{hk} \right)$$

siendo:

Y_{hk} = valor de la variable *Coste* de la unidades muestrales del estrato h .

Y_h, Y = totales de la variable *Coste* en el estrato h y en la población, respectivamente.

D_h = total de trabajadores del estrato h , en el Directorio de Cuentas de Cotización a la Seguridad Social.

D_{hk} = número de trabajadores en el Directorio, de las unidades muestrales del estrato h .

$d_h = \sum_{k=1}^{n_h} D_{hk}$ total de trabajadores de las unidades muestrales del estrato h .

$F_h = \frac{D_h}{d_h}$ factor de elevación de las unidades muestrales del estrato h .

Los estimadores de *Coste por trabajador y mes* y *Coste por hora efectiva* se obtienen como cocientes entre el estimador del *Total del Coste* y los estimadores del *Total de Trabajadores* y *Total de Horas*, respectivamente:

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

5.2. Estimadores para 2008-2009

Durante 2008 y 2009 se dispone de muestras diseñadas bajo la nueva clasificación CNAE09 lo que permite construir los estimadores tradicionales de ETCL pero bajo la nueva clasificación.

5.3. Estimadores post-estratificados para las series retrospectivas 2000-2007

Una vez reclasificados todos los directorios y muestras de ETCL desde 2007 hasta 2000 debe tomarse otra importante decisión: qué estimador utilizar para obtener los datos de la encuesta en CNAE09. Se valoran tres posibles métodos de estimación:

- *Método 1*: Estimador de razón con los factores de elevación originales, que se calcularon en CNAE93 aplicados al estimador separado de razón utilizado en ETCL.
- *Método 2*: Estimador de razón con los factores de elevación recalculados, obtenidos para la reclasificación efectuada en CNAE09 y aplicados al estimador separado de razón.
- *Método 3*: Estimador post-estratificado, cuyo factor de elevación tenga en cuenta tanto el diseño muestral original en CNAE93, como el efecto de la reclasificación y redistribución de la muestra a CNAE09.

Para evaluarlos, se utilizan las tasas de variación interanuales de las dos variables principales de la encuesta: “*Coste salarial ordinario por trabajador*” y “*Coste total por trabajador*”.

	Coste salarial ordinario por trabajador				Coste total neto por trabajador			
	CNAE09			CNAE93	CNAE09			CNAE93
	Métodos			Tasa	Métodos			Tasa
	1	2	3	publicada	1	2	3	publicada
Total	4,29 %	4,19 %	4,07 %	4,4 %	4,62 %	3,90 %	4,07 %	4,0 %
Industria	4,34 %	4,14 %	4,16 %	4,1 %	4,80 %	3,85 %	4,16 %	4,2 %
Construcción	5,48 %	4,58 %	3,91 %	5,0 %	5,56 %	4,71 %	4,14 %	5,0 %
Servicios	4,19 %	4,24 %	4,21 %	4,5 %	4,49 %	3,90 %	3,71 %	4,0 %

Cuadro 4: Evaluación del método de estimación retrospectiva para ETCL (1)

En el Cuadro 4 se observa que las tasas obtenidas utilizando el método 1 (estimación con los factores originales) son más elevadas que las calculadas con los otros métodos, por lo que cabe suponer que el método 1 sobre-estima los resultados en CNAE09.

Para confirmarlo, se comparan las estimaciones de la variable “*Total de Trabajadores*” obtenidas con los distintos métodos en CNAE09 frente a las obtenidas con los estimadores en CNAE93².

²La variable “Total de Trabajadores” no es una variable objetivo de la ETCL, pero por disponer de una cifra comparable derivada del marco poblacional, resulta interesante su estudio.

Total de Trabajadores				
	CNAE09			CNAE93 Estimación
	Métodos			
	1	2	3	
Total	14.006.716	13.391.444	13.058.553	13.114.080
Industria	2.420.865	2.312.457	2.215.098	2.366.345
Construcción	1.918.150	1.930.571	1.908.049	1.835.722
Servicios	9.667.701	9.148.416	8.935.406	8.912.013

Cuadro 5: Evaluación del método de estimación retrospectiva para ETCL (2)

Como se observa en el Cuadro 5, a través del método 1 se obtiene un número total de trabajadores que excede en casi 1 millón a los resultados obtenidos en CNAE93. Por sectores de actividad también se aprecia esta tendencia, si bien la conclusión resulta menos obvia porque los sectores de actividad económica han sufrido modificaciones en su nueva definición en CNAE09 respecto de su definición en CNAE93.

Finalmente, se recurre a la comparación de las estimaciones obtenidas frente a los valores poblacionales (Cuadro 6). Dado que la muestra para el año 2007 es insuficiente para dar cobertura a la nueva CNAE09, y la falta de muestra repercute notablemente en la estimaciones de los totales, se realiza la comparación para las secciones de actividad cuyo tamaño muestral no sufre incrementos con el cambio de CNAE.

Total de Trabajadores				
	CNAE09			Directorio Poblacional
	Métodos			
	1	2	3	
Industrias extractivas	44.376	40.955	41.210	42.850
Suministro de energía eléctrica, gas, vapor y aire acondicionado	40.463	38.422	38.176	38.267
Actividades financieras y de seguros	408.303	404.335	403.151	394.184
Administración Pública y Defensa	896.215	970.967	966.040	974.956
Educación	589.719	587.358	585.726	559.096

Cuadro 6: Evaluación del método de estimación retrospectiva para ETCL (3)

Los resultados obtenidos con el método 1 difieren más de los valores poblacionales que los obtenidos con los métodos 2 y 3.

En definitiva, estos ensayos permiten concluir que el método 1 produce resultados que difieren más de los resultados de ETCL que los obtenidos a través de los métodos 2 y 3, como consecuencia, es desechado.

Respecto a la elección entre los métodos 2 y 3, los resultados bajo ambas opciones son similares y no permiten concluir. Ante la falta de evidencia en contra, se opta por utilizar estimadores post-estratificados debido a que en sus factores de elevación interviene tanto el diseño muestral original en CNAE93, que fue con el que se extrajo la muestra, como el efecto de la reclasificación y redistribución de la muestra a CNAE09. Así, por ejemplo, en la división 47 de la CNAE09, que proviene únicamente de las divisiones 50 y 52 de CNAE93, la expresión del estimador post-estratificado viene dada por:

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

siendo:

$n_{k,47}$ = número de unidades de la muestra que se clasifican en la división k de CNAE93 y en la división 47 de CNAE09.

$D_{k,47}$ = número de trabajadores, según directorio que están asociados a la división k de CNAE93 y en la división 47 de CNAE09.

$d_{k,47}$ = número de trabajadores de la muestra asociados a la división k de CNAE93 y en la división 47 de CNAE09.

$y_{k,47,i}$ = valor de la variable y en la unidad i -ésima de las unidades de la muestra que se clasifican en la división k de CNAE93 y en la división 47 de CNAE09.

Esta manera de construir las estimaciones en CNAE09 partiendo de una muestra extraída en CNAE93 permite mantener los perfiles de las series publicadas en la clasificación inicial.

6. Cuarta fase: Enlace entre series 2000-2007 y 2008-2009

La ampliación muestral llevada a cabo durante 2008 para dar cobertura a la nueva CNAE09 supone un tercio de la muestra utilizada hasta este año. Además, este incremento muestral no se distribuye equitativamente entre los diferentes estratos, sino que se concentra, principalmente, en las divisiones de actividad nuevas.

Como consecuencia, las nuevas estimaciones obtenidas desde 2008 en CNAE09 con la muestra ampliada producen resultados que difieren de los obtenidos hasta 2007 con los estimadores post-estratificados sobre la muestra tradicional. Surge así la necesidad de efectuar un enlace entre las series reconstruidas para 2007-2000 y las nuevas series que arrancan en 2008 (véase Figura 1).

Para realizar este enlace, se hace uso de la doble estimación en CNAE09 y CNAE93 durante el periodo común (2008) de convivencia de ambos sistemas.

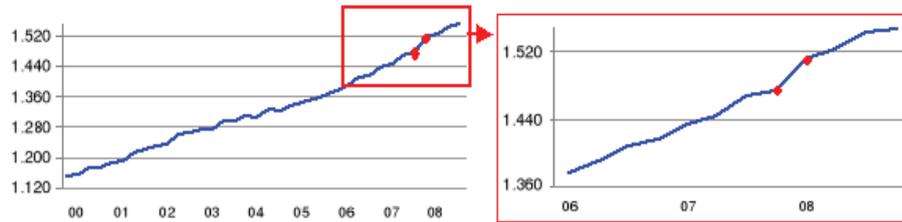


Figura 1: Serie 2007-2000 de la variable Coste Salarial Ordinario para las secciones comparables (excluye la sección Administración Pública)

El cambio de clasificación implica una redistribución de unidades según su actividad económica pero evidentemente la distribución de los datos por regiones no se ve alterada. Además, en términos generales, aunque las divisiones y secciones de actividad sufren importantes cambios respecto a la anterior clasificación, su agrupación en los tres grandes sectores de actividad (industria, construcción y servicios) en esencia mantiene sus características, por lo que es de esperar que los resultados obtenidos para estos sectores principales no presenten desviaciones respecto a los resultados que se obtuvieron con la clasificación anterior.

Se establece así la siguiente hipótesis básica:

“Las tasas de variación, regionales y por sectores de actividad, entre los datos en CNAE09 obtenidos con estimadores post-estratificados y los nuevos datos que desde 2008 se elaboran en CNAE09 con muestra ampliada, no deben divergir de las que se publicaron en 2008 para las series tradicionales en CNAE93”.

Se consideran, además, dos hipótesis adicionales:

- *La distribución de la variable “Trabajadores (T)” obtenida con el post-estratificado es correcta y se trata de encontrar coeficientes correctores para aplicar a los estimadores de las variables de “Costes” y “Tiempo de Trabajo”.*
- *Para ajustarse lo máximo posible a los estimadores originales se exige que, para cada división d_s , el estimador resultante no difiera más de un 15% del estimador original.*

Se obtiene el siguiente sistema de ecuaciones:

Para cada sector $S \in \{\text{Industria, Construcción, Servicios}\}$ y cada región $R \in \{01,02,\dots,17\}$, se buscan coeficientes A_{Rd_s} que verifiquen:

$$\frac{\sum_{d_s \in S} A_{Rd_s} \cdot C_{Rd_s}^{07}}{T_{RS}^{07}} = \frac{CT_{RS}^{08}}{(1 + t_{RS}^{07,08})} \quad \text{sujeto a } 0,85 \leq A_{Rd_s} \leq 1,15$$

siendo:

$t_{RS}^{07,08}$ = tasa de variación del coste por trabajador en CNAE93 publicada en 2008.

CT_{RS}^{08} = coste neto por trabajador en 2008 calculado con el estimador tradicional.

T_{RS}^{07} = trabajadores en 2007 calculado con el estimador post-estratificado.

$C_{Rd_s}^{07}$ = coste neto en 2007 calculado con el estimador post-estratificado.

La solución a este sistema proporciona coeficientes A_{Rd_s} que permiten corregir la serie hacia atrás y garantizan que la serie corregida para la variable “*Coste Total Neto por Trabajador (CT)*” presente un perfil paralelo al de la serie original.

Pero la variable “*Coste Total Neto*” no es sino la variable resultante de agregar los diversos componentes que forman el coste laboral. De estos componentes el más destacado es el “*Coste Salarial (ST)*” que supone el 80% del coste total y que tiene un marcado comportamiento estacional que incorpora al perfil de la serie coste total.

Para mantener el perfil de esta componente se replica el razonamiento anterior, obteniéndose un segundo sistema idéntico al anterior pero planteado sobre la variable “*Coste Salarial (ST)*”, del que se obtienen los coeficientes de ajuste B_{Rd_s} .

Las soluciones de ambos sistemas permiten replicar el perfil de las series publicadas para las variables “*Coste Total Neto*” y “*Coste Salarial Total*” respectivamente, pero ninguna de ellas proporciona una solución que ajuste simultáneamente ambos perfiles. Este ajuste simultáneo se consigue a través de la media geométrica de ambas soluciones:

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

Así pues, W_{Rd_s} es considerado como el coeficiente corrector único para todos los componentes del coste.

Pero aún quedan por enlazar las variables referentes al “*Tiempo de Trabajo*”. Para hacerlo, se aplica de nuevo la hipótesis básica, aunque ahora sobre la varia-

ble “*Coste por hora efectiva*” y se hace utilizando la variable “*Coste total neto*” ya ajustada con los coeficientes correctores W_{Rd_s} .

Los sistemas propuestos se replican para cada uno de los cuatro trimestres de 2008, produciendo 4 coeficientes de corrección W_{Rd_s} para las variables económicas de la encuesta, y 4 coeficientes de corrección K_{Rd_s} para las variables de tiempo de trabajo.

Finalmente, se aplican de nuevo estimadores post-estratificados sobre los microdatos ajustados con estos coeficientes correctores. Se obtienen así series retrospectivas de costes laborales bajo la nueva clasificación CNAE09 cuyo perfil de crecimiento, en grupos comparables, resulta similar al anteriormente publicado en CNAE93, como se observa en la Figura 2.

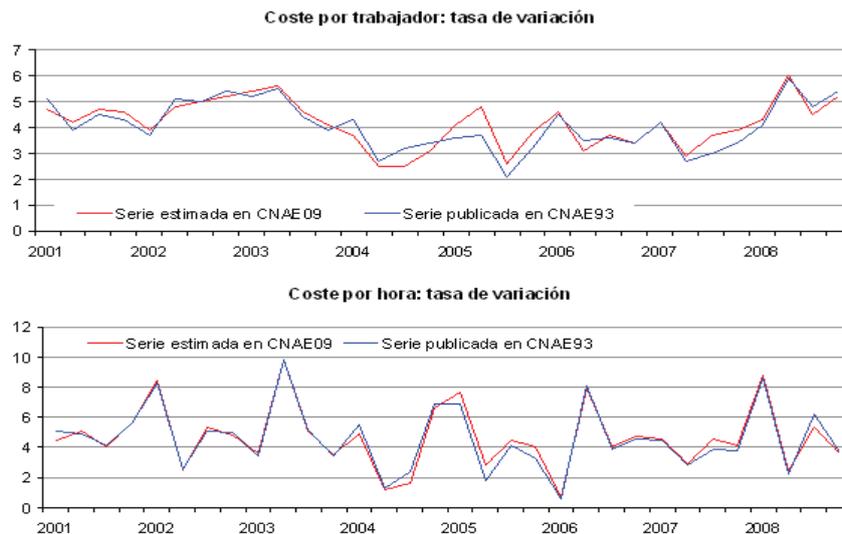


Figura 2: Comparación de las series 2008-2000 en ambas clasificaciones. Sector: Industria

7. Quinta fase: Revisiones de la nueva clasificación

Dado que la CNAE09 es una clasificación nueva, durante los dos primeros años de su implantación (2008 y 2009) se investiga con detenimiento la asignación de los nuevos códigos CNAE09 en las unidades de la muestra que colabora en ETCL. Con las rectificaciones que se obtienen de esta investigación se revisan y corrigen los directorios y muestras iniciales:

$N_h, n_h, D_h, d_h \Rightarrow$ tamaños del estrato h previo a la revisión

$N'_h, n'_h, D'_h, d'_h \Rightarrow$ tamaños del estrato h tras la revisión

Con la revisión de los tamaños poblacionales y muestrales se revisa por completo el proceso expuesto hasta aquí. Además, para los períodos 2008 y 2009 hay que ajustar los factores de elevación para obtener estimaciones representativas para cada actividad económica. Para resolver este problema se consideran varios procedimientos de refactorización:

- *Método 1*: Utilización de técnicas de calibrado ([9]) $\Rightarrow F_{calh} = F_h \cdot \frac{N'_h}{N_h}$
- *Método 2*: Utilización de los factores originales $\Rightarrow F_h = \frac{D_h}{d_h}$
- *Método 3*: Recálculo de factores tras la revisión $\Rightarrow F'_h = \frac{D'_h}{d'_h}$

El método 2 es desechado porque distorsiona los resultados ya que, al aplicar un cambio de clasificación a una unidad muestral, si no se rectifican los factores de elevación el efecto inmediato es que la unidad en cuestión queda sobrerrepresentada o infravalorada en la rama de actividad nueva en la que ha sido clasificada, como se ilustra en el ejemplo del Cuadro 7:

Unidad	CNAE inicial	Factor inicial	CNAE revisada	Factor revisado	Salario
i	Comercio	133	Minería	1	610 Euros/trabajador
Salario medio obtenido en la CNAE=Minería					
Inicialmente (unidad i fuera de la minería)		\Rightarrow 2.297	Al no rectificar el factor, la unidad representa a 133 trabajadores con salario 610 Euros, cuando realmente debería representar a 1. Como consecuencia, el salario medio en el estrato queda infravalorado.		
Con la unidad i revisada		\Rightarrow 1.352			

Cuadro 7: Efectos sobre los resultados obtenidos con el método 2

Para evaluar los métodos 1 y 3 se comparan los resultados obtenidos con los resultados de la encuesta (Cuadro 8) y con el marco poblacional (Cuadro 9):

	Coste salarial ordinario por trabajador			Coste total neto por trabajador		
	CNAE09 Métodos		CNAE93 Tasa	CNAE09 Métodos		CNAE93 Tasa
	1	3	publicada	1	3	publicada
Total	4,68 %	5,17 %	5,0 %	4,49 %	4,96 %	5,1 %
Industria	5,04 %	4,95 %	4,9 %	3,37 %	3,30 %	4,1 %
Construcción	7,54 %	7,35 %	4,7 %	7,96 %	8,01 %	5,8 %
Servicios	3,99 %	4,67 %	5,0 %	4,13 %	4,75 %	5,2 %

Cuadro 8: Efectos sobre los resultados de “Costes” obtenidos con los métodos 1 y 3

	Total de Trabajadores		
	CNAE09		Directorio
	Método 1	Método 3	
Total	16.432.111	14.473.682	14.567.499
Industria	2.738.422	2.413.453	2.457.847
Construcción	2.302.353	1.972.279	1.982.571
Servicios	11.391.336	10.087.951	10.127.081

Cuadro 9: Efectos sobre los resultados del “Total de Trabajadores” obtenidos con los métodos 1 y 3

En todos los casos se obtienen tasas de crecimiento de los costes similares. Sin embargo, el método 1 (calibrado) sobre-estima los datos del “Total de Trabajadores”, tanto en el total como en los principales sectores económicos. Como consecuencia se concluye que el método más adecuado para aplicar los cambios de CNAE09 es el recálculo de factores (método 3).

8. Conclusión

El cambio de CNAE ha supuesto un cambio rotundo en la encuesta ETCL, debido a que su diseño muestral está estratificado según esta clasificación.

La mayor desagregación de la nueva clasificación ha supuesto un aumento muestral considerable y una ruptura en las series de datos. Para poder obtener series de datos continuas en CNAE09 de la encuesta ETCL desde su comienzo, se ha establecido un complejo proceso de reclasificación y tabulación.

Las claves del mismo han sido la conversión de la totalidad de la encuesta a CNAE09, la utilización de estimadores post-estratificados que tienen en cuenta la doble codificación CNAE y la disponibilidad de un año común de convivencia de ambas clasificaciones para poder obtener factores correctores que permitan realizar un enlace entre las series previas y posteriores al cambio de clasificación.

Referencias

- [1] Encuesta Trimestral de Coste Laboral. INE.
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft22/p187&file=inebase&L=0>
- [2] Reglamento (CE) No 1893/2006 del Parlamento Europeo y del Consejo de 20 de diciembre de 2006 por el que se establece la nomenclatura estadística de actividades económicas NACE Revisión 2 y por el que se modifica el Reglamento (CEE) n° 3037/90 del Consejo y determinados Reglamentos de la CE sobre aspectos estadísticos específicos.
- [3] Clasificación Nacional de Actividades Económicas. CNAE-2009. Real Decreto de 475/2007. *BOE 102, Sáb. 28 de abril de 2007.*
- [4] Fortier, S. (2003) La conversion des donnés 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] Directorio Central de Empresas. 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. Mexico.
- [9] Deville, J.C., Särndal, C.E. (1991). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association (JASA)*.

Sobre las autoras

Amelia Fresneda Pacheco es Jefa del Área de Costes Laborales del Instituto Nacional de Estadística de España. Entre sus funciones se encuentran la investigación y desarrollo de técnicas estadísticas para la obtención de series sobre costes laborales y para la integración de resultados estructurales y coyunturales. **María Ramos Charbonnier** es Directora de Programa en el Área de Costes Laborales del Instituto Nacional de Estadística de España. Hasta la fecha, su labor de investigación se ha centrado en el desarrollo del procedimiento de implantación y transición del sistema estadístico integrado de costes laborales del INE a la nueva Clasificación Nacional de Actividades Económicas CNAE09.