

# **ESTIMACIÓN DE LA TASA DE ACTIVIDAD Y PARO A NIVEL DE SECCIÓN CENSAL**

**METODOLOGÍA**

# Índice

1.- Introducción.....	3
2.- Estimadores de áreas pequeñas aplicados al mercado de trabajo en España.....	3
3.- Estimación de la tasa de actividad a nivel de sección censal.....	6
4.- Estimación de la tasa de paro a nivel de sección censal.....	11
5.- Estimación de la varianza y del error cuadrático.....	14
6.- Resultados.....	16
7.- Cambio de base poblacional en las estimaciones de la EPA.....	19
Anexo.- Áreas Pequeñas.....	20
Bibliografía.....	22

# Estimación de la Tasa de Actividad y Paro a nivel de Sección Censal<sup>1</sup>

Parra Rodríguez, Francisco; Bedia Expósito, Raquel; Cobo Fernández, M<sup>a</sup> Isabel; Campo Moreno, Lorena

Instituto Cántabro de Estadística (ICANE)

## I- Introducción

La estimación de la tasa de actividad y paro a nivel de sección censal forma parte de un proyecto que está desarrollando el Instituto Cántabro de Estadística (ICANE) que responde a la encomienda que le realiza el Consejo Cántabro de Estadística sobre la comarcalización de las tasas de actividad y paro de la Encuesta de Población Activa (EPA). Para abordar este proyecto se creó en un grupo de trabajo que quedó constituido por representantes del INE, UGT, CCOO, CEOE, Ayuntamiento de Santander y Gobierno Regional. En la primera reunión del grupo de trabajo el ICANE presentó una metodología de estimación de ambas tasas a nivel municipal para la región de Cantabria basada en técnicas de estimación de áreas pequeñas para así facilitar una futura comarcalización de los resultados de la encuesta. A results de las deliberaciones en el grupo de trabajo se decidió realizar dicha estimación con la base de las secciones censales, estimación que es la que se presenta. El ejercicio se ha realizado para el periodo comprendido entre el primer trimestre de 2006 y el primer trimestre del 2012, habiéndose utilizado como datos base los microdatos de la EPA en Cantabria y los registros del Padrón Municipal de Habitantes y del colectivo DENO del Sistema de Información de los Servicios Públicos de Empleo (SISPE) que actúan como ficheros auxiliares. Dado que el colectivo DENO solo recoge la situación de la población entre 16 y 64 años el cálculo de las tasas se ha restringido a este estrato de población.

El documento se divide principalmente en cinco secciones, en la primera se resumen las experiencias de comarcalización de la EPA en España, en la segunda se expone el procedimiento seguido para estimar a nivel de sección censal la tasa de actividad, en la tercera se describe el procedimiento de estimación a nivel de sección censal de la tasa de paro, en el cuarto se explica la estimación de la varianza y del error cuadrático y en el quinto se analizan los resultados obtenidos.

## 2.- Estimadores de áreas pequeñas aplicados al mercado de trabajo en España.

Los estimadores de áreas pequeñas suelen clasificarse atendiendo a la procedencia de la información utilizada, distinguiendo:

- Estimadores directos: esta basado solamente en los datos muestrales del dominio específico. El problema de estos estimadores es cuando no hay observaciones de la muestra en alguna de las áreas de interés, entonces no se puede calcular. En otras ocasiones, aunque si se puede calcular, lo que suele ocurrir es que el tamaño muestral del dominio específico no es lo suficientemente grande como para soportar estimaciones de adecuada precisión.

Un estimador directo utilizado para estimar el número de parados en un área pequeña sería el siguiente:

$$\hat{y}_d^{directo} = \frac{\sum_{j=1}^{n_d} w_j y_j}{\sum_{j=1}^{n_d} w_j} N_d$$

Donde  $y_j = 1$  si la persona encuestada es un parado e  $y_j = 0$  si no esta en paro,  $N_d$  es el número de personas con edad >16 años en el área pequeña d,  $n_d$  el tamaño muestral del área d, y  $w_j$  el peso de diseño.

A partir de este podría obtenerse un estimador directo post-estratificado:

$$Y_d^{post} = \sum_g N_{dg} \hat{y}_{dg}$$

Donde  $\hat{y}_{dg}$  es la media calculada con el estimador directo anterior:

---

<sup>1</sup> Los autores agradecen su contribución a D. Juan Rodríguez Poo impulsor de este proyecto siendo Director del ICANE, y a los miembros del grupo de trabajo del Consejo Cántabro de Estadística: D. Fernando Silió Cervera, D. Miguel Ángel Cuesta Díez, Dña. Gema Díaz Domínguez, D. Gonzalo Sánchez-Crespo Benítez, D. Joaquín Incera Lavín y Dña. Dolores Ortiz Escribano.

$$\hat{y}_{dg} = \frac{\sum_{j \in s_{dg}} w_j y_j}{\sum_{j \in s_{df}} w_j} N_d$$

- Estimadores indirectos: cuando la muestra en cada dominio no es suficiente para obtener estimadores con la precisión deseada, entonces recurrimos a este tipo de estimadores. Se caracterizan porque toman prestada fuerza de la relación entre los valores de la variable de interés en áreas relacionadas o la información auxiliar disponible. Estos valores se incorporan al proceso de estimación a través de un modelo implícito que proporciona un link para toda la información auxiliar. Otros términos usados para denotar estimadores indirectos incluyen no tradicional y sintético.

Ejemplos de estos estimadores son:

- a) Si suponemos que la media del área pequeña es igual que la del área mayor que la contiene.

$$\hat{Y}_d^{\text{sin},m} = N_d \frac{\sum_{j=1}^n w_j y_j}{\sum_{j=1}^n w_j} = N_d \hat{y}$$

Siendo n el tamaño de la muestra en el área mayor.

- b) Si suponemos que determinados ratios son iguales en el área pequeña y el área grande.

$$\hat{Y}_d^{\text{sin},r} = X_d \frac{\hat{Y}}{\hat{X}}, \text{ con } \hat{Y} = \sum_{j=1}^n w_j y_j$$

Destacar que los estimadores sintéticos hacen un fuerte uso de la información de otras áreas, permitiendo por tanto muy poca variación local, lo que puede contribuir a un importante crecimiento del sesgo de no existir cierta homogeneidad entre estas áreas.

Dado que los estimadores sintéticos hacen un fuerte uso de la información de otras áreas, permitiendo por tanto muy poca variación local, lo que puede contribuir a un importante crecimiento del sesgo de no existir cierta homogeneidad entre estas áreas. Para evitar el potencial sesgo del estimador sintético y la inestabilidad del estimador directo, cabe construir un estimador a partir de una combinación lineal convexa de ambos, esto es:

$$\hat{Y}_d^{\text{comb}} = \lambda_d \hat{Y}_d^{\text{directo}} + (1 - \lambda_d) \hat{Y}_d^{\text{sin}}, \text{ con } \lambda_d \in [0,1]$$

se estima  $\lambda_d$  minimizando el error cuadrático medio, MSE, con los datos muestrales.

- Estimadores basados en el modelo: usan información auxiliar por medio de modelos de regresión como una manera de obtener estimadores consistentes de diseño, así se gana mayor eficiencia y se reduce el sesgo.

Estos modelos explícitos son útiles en la estimación de áreas pequeñas porque da al usuario una idea del proceso de generación de datos y de cómo las diferentes fuentes de información son incorporadas. Los modelos de áreas pequeñas se clasifican en dos grandes tipos:

(i) Modelo de niveles agregados o modelo de área: relacionan el estimador directo del área con una covariable específica del área. Tales modelos son necesarios cuando tenemos la información solo a nivel de área. Se asume que el estimador directo del área obedece a un modelo de población.

(ii) Modelo de errores anidado o modelo de individuo: relaciona los valores de la variable en estudio con el valor de la covariable medida en cada unidad. Estos modelos obedecen a un modelo de población, por lo tanto el sesgo bajo el diseño está ausente.

La inferencia de los estimadores basados en el modelo se refiere a la distribución dada por el modelo asumido. Por lo tanto, la selección del modelo y su validación juegan un papel vital en esta estimación.

Ejemplos de estimadores de modelos es el CREG lineal:

$$y_{jd} = x_{jd}^t \beta + \varepsilon_{jd}$$

$$\hat{Y}_d^{\text{GREClíneal}} = N_d \left( \hat{Y}_d^{\text{directo}} + \left( X_d - \hat{X}_d^{\text{directo}} \right)^t \beta \right)$$

Donde  $\bar{X}_d = (\bar{X}_{d1}, \bar{X}_{d2}, \dots, \bar{X}_{dp})^t$  es un vector de p medias de población auxiliar o covariables. El parámetro  $\beta$  se estima por mínimos cuadrados generalizados. En su versión logit, se formula:

$$\log\left(\frac{P_{jd}}{1-P_{jd}}\right) = x_{jd}^t \beta$$

$$\hat{Y}_d^{CREG \log it} = \sum_{j=1}^{N_d} \frac{\exp(x_{jd}^t \hat{\beta})}{1 + \exp(x_{jd}^t \hat{\beta})} + \frac{N_d}{\hat{N}_d} \sum_{j=1}^{n_d} w_{jd} \left( y_{jd} - \frac{\exp(x_{jd}^t \hat{\beta})}{1 + \exp(x_{jd}^t \hat{\beta})} \right)$$

Los estimadores de áreas pequeñas basados en modelos que se han construido para las variables de actividad, ocupación y/o paro suelen utilizar como covariables las siguientes:

- Grupos de edad y sexo
- Niveles educativos
- Relación con la actividad

En Euskadi (Eustat, 2008), Galicia (López et al, 2009), Canarias (López R., 2000) y Cataluña (Costa et al, 2007) se han probado distintos estimadores de áreas pequeñas y evaluado su sesgo y varianza, a nivel nacional habría que destacar el proyecto EURAREA.

En general se afirma que el mayor problema que se plantea es la existencia o no de una mínima representación muestral del área pequeña en base al diseño original. Si no hay muestra ó es demasiado exigua, el estimador a posteriori no se puede obtener o es inoperante. Tampoco se va a poder calcular la estimación del sesgo para el estimador sintético, ya que la varianza de éste, por sí sola, no da una buena idea de su precisión al ser el estimador sesgado (López R., 2000), en Eustat (2008) y López R. (2000) se considera que la disponibilidad de una información auxiliar adecuada es muy determinante a la hora de obtener las estimaciones y sus varianzas, López R. (2000) muestra que la variable población por grupos de edad y sexo puede ser más idónea para la variable ocupados que para la variable parados, y que la varianza de las estimaciones así como las propias estimaciones obtenidas a partir de los estimadores de regresión, vendrán influidos por la diferencia entre los vectores de efectivos poblacionales que proporciona la EPA y la fuente estadística auxiliar. Para las comarcas-EURAREA, se muestra que comportamiento de los estimadores sintéticos es considerablemente mejor que el de los otros estimadores ya que los tamaños muestrales en éstas son, generalmente, muy pequeños, y no encuentran diferencias significativas entre el comportamiento del estimador GREG lineal y GREG Logit. Costa et al (2007) afirman que aunque la utilización de un estimador sintético para la actividad y la aproximación de la tasa de paro como una proporción son fácilmente criticables desde un punto de vista metodológico dan lugar a resultados razonablemente fiables a partir de procedimientos sencillos y objetivos.

En el ámbito de Cantabria, Calvo G (2010) realizó un ejercicio para estimar la EPA a nivel municipal utilizando un estimador de áreas pequeñas no paramétrico a partir de un Modelo Logit Mixto, con la siguiente formulación:

$$Y_{id}^+ = \xi_d + m_o(X_i) + V_{id}, i = 1, \dots, N_d; d = 1, \dots, D$$

Donde  $Y_{id}^+$  es una variable aleatoria no observable relativa a la decisión del individuo i del área pequeña de participar o no en el mercado de trabajo;  $\xi_d$  desconocido  $\approx NID(0, \sigma_d)$ ;  $m_o(\cdot)$  es una función spline de orden p+1 con nodos  $\eta_1, \dots, \eta_p$  de un conjunto de variables X de efecto municipio (Fernández y Rodríguez-Poo, 1997) y  $V_{id}$  un error aleatorio.

La información muestral que se utilizó para la estimación fue la EPA del cuarto trimestre de 2009, la variable Y era la disposición del individuo a participar en el mercado de trabajo, el sexo y la edad del encuestado, fueron las covariables, X, si bien la variable del sexo, se introdujo en modelos separado, asumiéndose que ambos géneros presentan un comportamiento muy diferenciado en lo que respecta a la participación. Las variables explicativas estaban condicionadas a la información auxiliar disponible a nivel municipal que era el Padrón Municipal de habitante. Los resultados obtenidos estaban muy próximos a las estimaciones regionales de la EPA (Tabla 2.1), y se proponía una alternativa de calibrado.

**Tabla 2.1: Estimación del número de activos a través de un modelo de función spline**

<b>Estimación Población Activa (Estimación en pequeñas áreas)</b>						
	<b>16-24</b>	<b>25-34</b>	<b>35-44</b>	<b>45-44</b>	<b>55 y mas</b>	<b>Total</b>
Varones	10,3	41,6	44,2	38,5	19,3	153,9
Mujeres	9,9	37,7	35,6	28,2	11,6	123,0

<b>Estimación Población Activa (EPA-INE)</b>						
	<b>16-24</b>	<b>25-34</b>	<b>35-44</b>	<b>45-44</b>	<b>55 y mas</b>	<b>Total</b>
Varones	10,9	42,2	44,1	38,6	18,6	154,4
Mujeres	9,8	38,1	35,4	28,1	11,3	122,7

Fuente: Calvo (2010)

### 3.- Estimación de la tasa de actividad a nivel de sección censal

A la hora de estimar la tasa de actividad a niveles de área pequeña se ha considerado un estimador de modelo de niveles agregados o de área, cuya formulación sería la siguiente:

$$Y_{id} = \xi_d + m_o(x_i) + V_{id}, i = 16, \dots, 65; d = 1, \dots, D$$

Donde  $Y_{id}$  es una variable aleatoria no observable relativa al número de individuos de edad  $i$  del área pequeña de participar o no en el mercado de trabajo;  $\xi_d$  desconocido  $\approx NID(0, \sigma_d)$ ,  $m_o(\cdot)$  es una función transformada rápida de Fourier (FFF) del conjunto de edades  $x_i$ , que van desde los 16 a los 65 años, y  $V_{id}$  un error aleatorio.

La aproximación a una función no periódica  $g(x)$  por una serie de expansión de Fourier se realiza en Gallant (1981) añadiendo es esta un término lineal y cuadrático. De esta forma que la aproximación univariada se escribe como:

$$m(x/\theta) = a + bx + \frac{1}{2}cx^2 + 2\sum_{j=1}^J u_j \cos(jx) - v_j \sin(jx)$$

El vector de parámetros es  $\theta = (a, b, c, u_1, v_1, \dots, u_J, v_J)$  de longitud  $K = 3 + 2J$ .

Suponiendo que los datos siguieran el modelo  $y_i = g(x_i) + e_i$  para  $i=1, 2, \dots, n$  estimamos  $\theta$  por mínimos cuadrados,

$$s_n(\theta) = \left(\frac{1}{n}\right) \sum_{i=1}^n [y_i - g_K(x_i/\theta)]^2$$

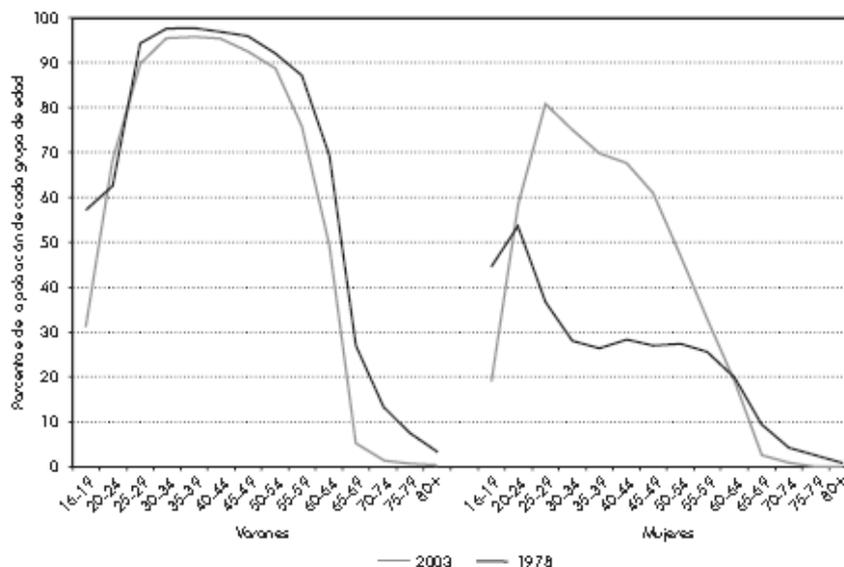
minimizando

Dado que la variable exógena  $x_i$  no esta expresada en forma periódica, debe de transformarse o normalizarse en un intervalo de longitud menor que  $2\pi$ ,  $[0, 2\pi]$ .

Efectuada la descripción del modelo estadístico a estimar, comentamos su puesta en práctica. La información muestral que se utiliza para la estimación son los microdatos de Cantabria de la EPA del primer trimestre del 2006 al primer trimestre del 2012. Las variables que se utilizan de la EPA son la tasa de actividad de los individuos de edad  $i$  mayores de 16 años y menores de 65,  $Y_i$ , el sexo y las  $i$  edades que van de los 16 a 65 años serían las variables  $X_i$ . Con estas variables se procede a estimar los parámetros de interés de la ecuación. Para tratar la variable del sexo, optamos al igual que Calvo (2010) por emplear dos modelos por separado, ya que como se muestra en la figura nº 1 ambos géneros presentan un comportamiento diferente en lo que respecta a tasa de actividad (Toharia L., 2003; Cuadrado et al, 2007). La estimación de la población de edad  $i$  en la sección censal que desea participar en el mercado de trabajo  $Y_{id}$ , precisa de conocer la población de edad  $i$  residente en la sección censal,  $N_{id}$ , que se obtiene de la explotación de padrón de habitantes (datos provisionales) correspondiente a 1 de enero, que se suponen reflejan la población residente en la sección censal en el primer y segundo trimestre del año, y padrón de habitantes correspondiente a 1 de julio, que se supone reflejan la población del tercer y cuarto trimestre del año.

El cálculo de la tasa de actividad se realizaría a partir de los activos estimados en cada sección censal y la población empadronada en la sección.

**Figura 3.1: Tasa de actividad por género y grupos de edad en España.**

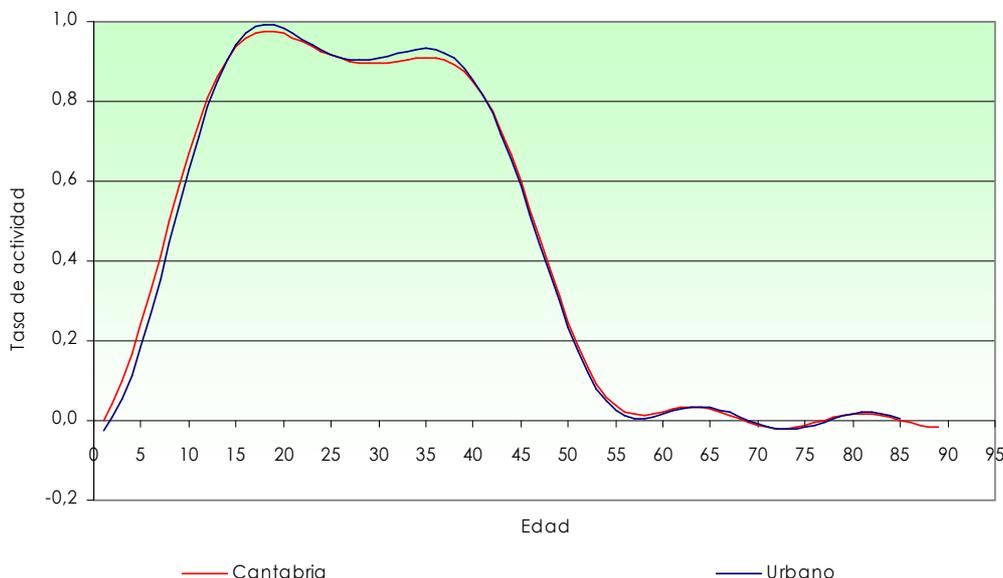


Fuente: Toharia, 2003

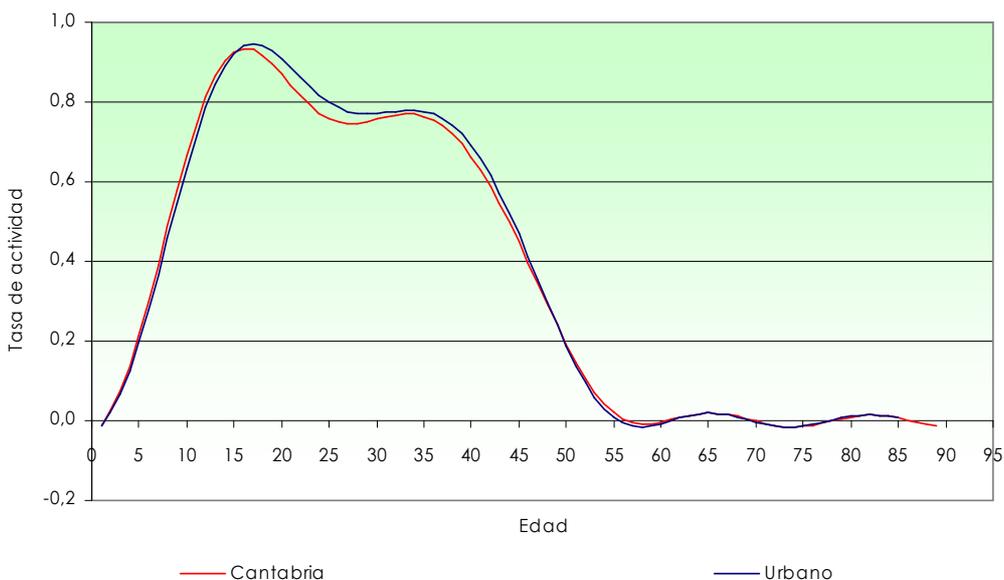
En los estudios de las tasas de actividad y en las experiencias realizadas en áreas pequeñas no se incluyen covariables de carácter territorial, en el sentido de que el medio urbano y rural conlleva una diferenciación en la evolución de dichas tasas más allá de la estructura poblacional dominante en cada medio.

En el caso de Cantabria, hemos hecho un estudio del comportamiento de las tasas de actividad de la EPA edad a edad por sexo y poblamiento, distinguiendo entre el medio urbano, intermedio y rural. El medio urbano lo comprenden aquellos municipios que tienen 10.000 ó más habitantes, el medio intermedio aquellos que tienen entre 2.001 y 10.000 habitantes y por último el medio rural los municipio de 2.000 o menos habitantes. El primer problema que nos hemos encontrado es que la muestra de la EPA, no es suficiente para poder hacer este análisis. Por ejemplo, en el medio rural para hombres y mujeres y en el medio intermedio de las mujeres no hay muestra en muchas edades, por lo que no se tienen tasas de actividad. Por otro lado las muestras no son suficientemente representativas en los medios rurales e intermedios, ya que en el caso rural femenino se dispone de una muestra de 75 individuos, en el rural de varones de 92, en el intermedio de mujeres 304 y en el intermedio de varones de 330. Solo en el caso del grupo de poblamiento urbano se dispone de una muestra suficiente, ya que en el caso de los varones esta es de 930 individuos y en el caso de las mujeres de 967. Como puede observarse en las figuras 3.2 y 3.3 el comportamiento de la tasa de actividad de los núcleos urbanos es similar al del total de Cantabria. Por estos motivos, finalmente hemos concluido que lo más recomendable es utilizar las mismas probabilidades de estar activo, diferenciando por sexo, independientemente del grupo de poblamiento.

**Figura 3.2. Aproximación FFF de la tasa de actividad masculina edad a edad. 2012 -1T**

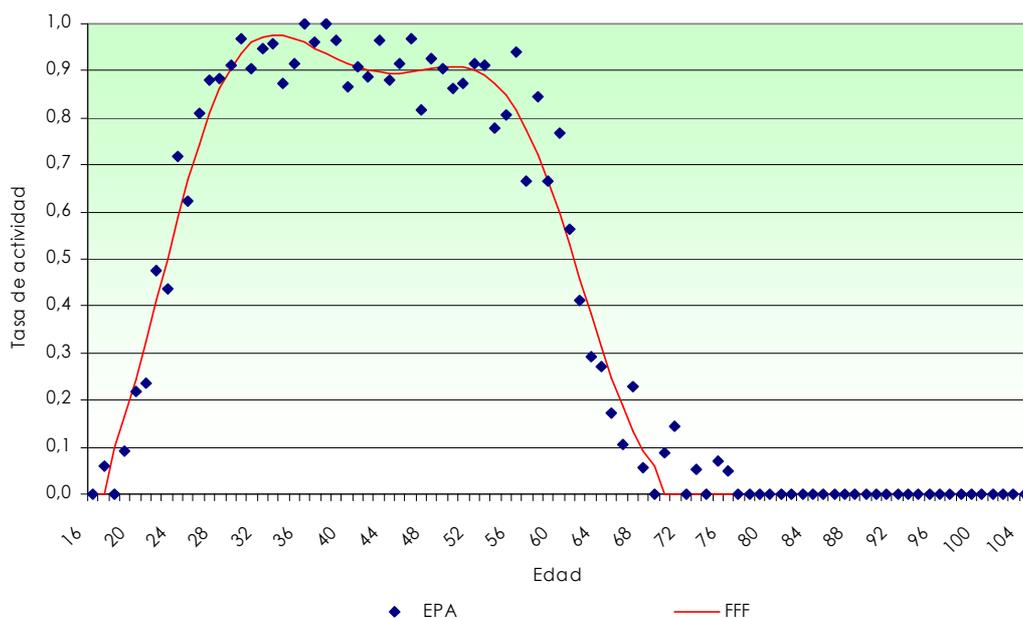


**Figura 3.3. Aproximación FFF de la tasa de actividad femenina edad a edad. 2012 -1T**

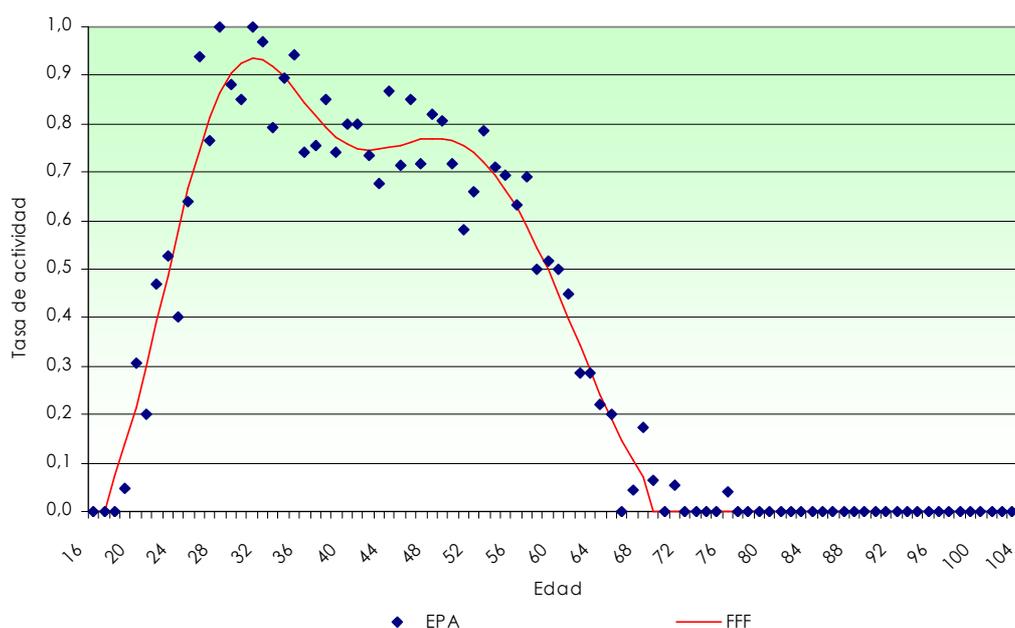


De hecho, los análisis del componente generacional de la tasa de actividad en Cantabria para cada edad y sexo en el primer trimestre del 2012 (figura nº 3.4 y figura nº 3.5), que se representan junto a la aproximación de la función generacional de tasa de actividad utilizando la forma de una función de flexibilidad global (FFF) (Gallant, 1981), conduce a semejantes conclusiones; un aumento de la actividad masculina a partir del momento en que se finaliza la educación a los distintos niveles, que aumentaría desde un valor medio del 40% en los estratos más jóvenes hasta alcanzar valores cercanos al 90% que se mantienen durante toda la edad adulta 25-55 años, y una disminución de dicha actividad a medida que se acerca la edad de jubilación; y una actividad femenina que en los estratos más jóvenes se asemeja a la masculina, un posterior descenso hasta niveles del 80% determinado por la natalidad, y una progresiva caída a partir de las generaciones con más de 45 años, generaciones de mujeres que en general se han mostrado menos activas en toda su vida laboral.

**Figura 3.4. Tasa de actividad masculina de la EPA en Cantabria y su aproximación con una función de flexibilidad global (FFF). 2012-1T**



**Figura 3.5. Tasa de actividad femenina de la EPA en Cantabria y su aproximación con una función de flexibilidad global (FFF). 2012-1T**

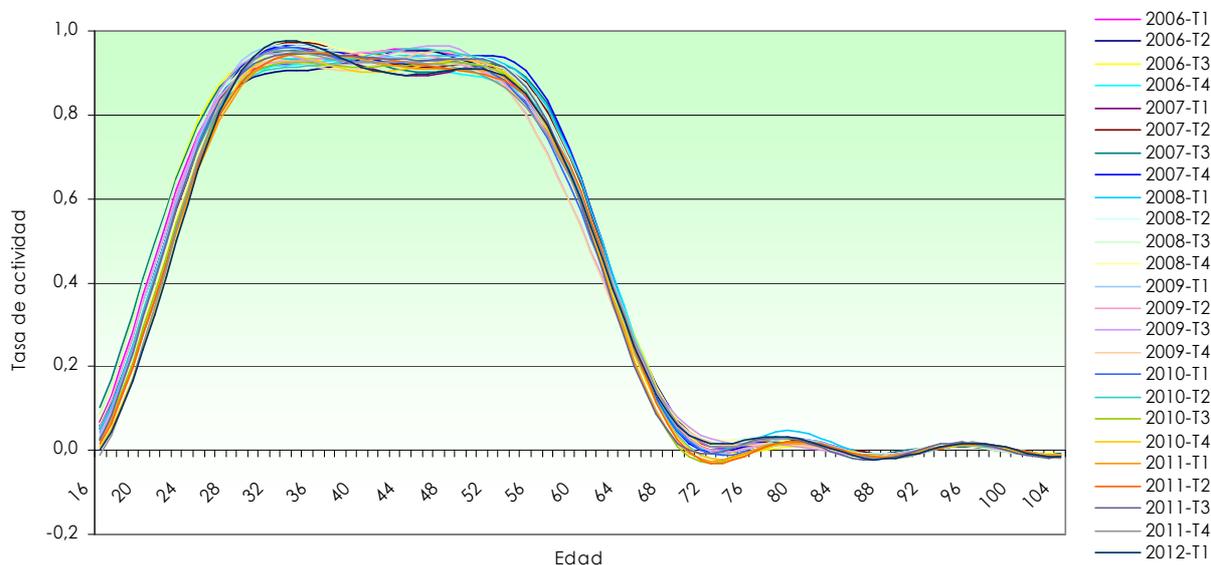


Las altas tasas de actividad de los varones y su vinculación con los procesos de escolarización y abandono de la vida laboral, al igual que las altas tasas de actividad de las mujeres su vinculo con la natalidad y con la presencia de generaciones menos activas, no creemos que presente una lectura diferenciada en el medio territorial y sobre todo en una región como la cántabra en el que la población activa de 16 a 64 años agraria apenas representó un 2,7% en la EPA del primer trimestre de 2012.

En este sentido y dado que la aproximación FFF de las tasas de actividad, representa en definitiva una estimación de estar activo para cada edad, se ha utilizado esta función para obtener una estimación en cada sección censal de la tasa de actividad, utilizando como variable auxiliar las estructuras demográficas del padrón municipal de habitantes.

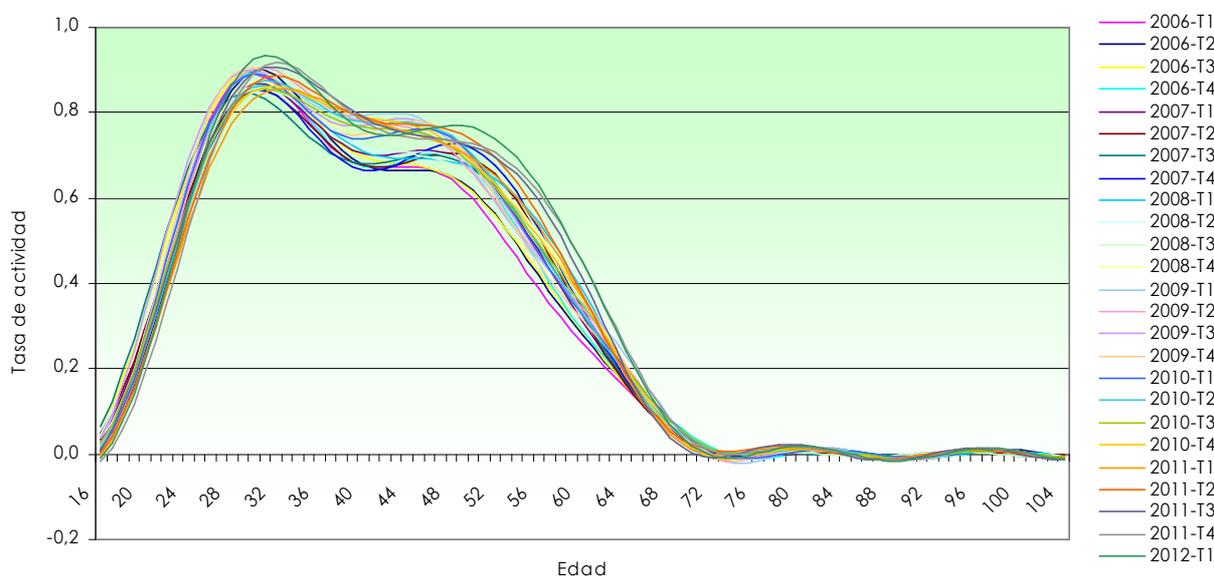
La aproximación a las tasas de actividad de varones realizada por este método para el periodo que va desde el primer trimestre de 2006 al primer trimestre de 2012 presenta una gran estabilidad (figura 3.6), si bien en los extremos de la función (edades por encima de los 70 años), se puede observar la oscilación armónica asociada a este tipo de funciones, esta oscilación se denomina efecto Gibbs. Este efecto es también el causante de que en ocasiones se obtengan valores negativos en la tasa de actividad en los niveles de 16 años. Para obviar dicho inconveniente, la función únicamente se utiliza cuando se obtienen valores positivos en el entorno de edades [106;16] y tomando valor cero en el resto de edades.

**Figura 3.6. Aproximación FFF a la tasa de actividad masculina por edad. 2006-1T a 2012-1T.**



La aproximación FFF a la tasa de actividad femenina edad a edad (figura 3.7), lógicamente no tiene la estabilidad de la tasa de los varones, y esto creemos que es por dos motivos. Uno la sustitución en el tiempo de generaciones de mujeres más jóvenes y con tasas de actividad más altas por generaciones de mujeres con tasa de actividad más bajas, de forma que en el tiempo la forma funcional de la función se aproxima a la de los varones. El otro motivo, y que también se percibe con claridad en esta figura, es el desanimo que las crisis provoca en la participación de la población femenina sobre todo entre las más activas, las mujeres adultas de 23 a 30 años, que entre 2006 y 2011 han reducido su tasa de actividad.

**Figura 3.7. Aproximación FFF a la tasa de actividad femenina por edad. 2006-1T a 2012-1T.**



Dado que el ejercicio de aproximación presenta una dinámica comprensible y que además está dentro de los patrones que caracterizan la actividad entre géneros y edades, aplicamos estas tasas o si se prefiere la probabilidad de ser activo a las estructuras demográficas de las secciones censales de Cantabria.

Los activos así obtenidos se han comparado con los activos de 16 a 64 años que la EPA proporciona para Cantabria, estimación que es insesgada para los grupos de edad que figuran en la tabla 3.2. Se comprueba que la aproximación es bastante razonable, si bien presenta generalmente un sesgo al alza que se evalúa en el primer trimestre del 2012 en 6.052 activos varones de más y 3.900 mujeres activas. Este sesgo, sin pasar por alto los problemas estimativos asociados a la aproximación FFF que, como ya se ha dicho, es imprecisa en los valores extremos, no se concentra en algún grupo específico de población.

**Tabla 3.2. Sesgo de las estimaciones preliminares de actividad en las secciones censales de Cantabria. 2012-1T**

**Estimación Población Activa (Estimación en pequeñas áreas)**

	16-24	25-34	35-44	45-54	55-64	Total
Varones	8,7	39,8	46,2	41,1	20,8	156,5
Mujeres	7,9	36,9	37,4	34,0	16,2	132,5

**Estimación Población Activa (EPA-INE)**

	16-24	25-34	35-44	45-44	55 -64	Total
Varones	8,6	38,2	45,4	38,1	20,1	150,5
Mujeres	7,2	36,8	35,9	32,3	16,4	128,6

Una vez obtenidas las estimaciones  $\hat{Y}_d$  para cada uno de las secciones censales, estas deben ser corregidas con la finalidad de que sean consistentes con los datos publicados por el INE para Cantabria. Para ello se utiliza el factor de calibrado:

$$F_{yr} = Y_r^{epa} / \sum_{d \in r} \hat{Y}_d$$

Siendo  $Y_r^{epa}$  la estimación de actividad o población activad de la EPA en Cantabria. Los resultados obtenidos se calibran para cada grupo de edad y sexo, para los que la EPA presenta resultados estadísticamente significativos.

## 4.- Estimación de la tasa de paro a nivel de sección censal

De todos es conocido que la definición de paro que hace el registro del Servicio Público de Empleo Estatal (SEPE) y la estadística de la EPA es discordante. Pérez J.I.( 2008) considera que la cifra de paro registrado es más baja que la de paro de la EPA, por las exclusiones de determinados colectivos de demandantes de empleo, que, pese a que puedan cumplir los criterios de la OIT para ser considerados como parados y, en consecuencia, ser computados como tales en la EPA, se excluyen del paro registrado. Los colectivos excluidos serían los siguientes:

a) Demandantes que solicitan exclusivamente un empleo de determinadas características:

- Demandantes de un empleo coyuntural por un período de tiempo inferior a tres meses.
- Solicitantes de un trabajo de jornada reducida, inferior a veinte horas semanales.
- Personas inscritas como demandante de empleo como requisito previo para participar en un proceso de selección para un puesto de trabajo específico.
- Solicitantes de un empleo para el extranjero o a domicilio.

b) Demandantes que se consideran que no tienen una disponibilidad inmediata para trabajar o que se encuentran en situación incompatible con el trabajo:

- Personas inscritas que son pensionistas de jubilación o que tienen 65 o más años.
- Personas que están cursando estudios de enseñanza oficial reglada, siempre que sean menores de 25 años o que superando esa edad sea demandantes de primer empleo.
- Asistentes a cursos de formación profesional ocupacional, cuando sus horas lectivas superan las veinte horas semanales, tengan una beca de manutención y sean demandantes de primer empleo.
- Demandantes en situación de incapacidad temporal o baja médica.
- Pensionistas por invalidez absoluta o gran invalidez.
- Demandantes que rechacen acciones de inserción laboral adecuadas a sus características.

c) Trabajadores beneficiarios de prestaciones por desempleo:

- Demandantes inscritos que están percibiendo el subsidio por desempleo de los trabajadores agrarios de carácter eventual de Andalucía y Extremadura o que agotándolo no haya transcurrido un período superior a un año desde el día del nacimiento del derecho.

La estimación del paro a nivel de sección censal puede acometerse con el auxilio del fichero del SISPE del SEPE utilizando un estimador sintético de ratio, si bien conviene acercar lo más posible ambas definiciones de paro.

En el SISPE existe nuevo concepto de DENOs, que surge con la necesidad de identificar a los beneficiarios de las acciones de las políticas de empleo, algo que con el anterior registro de paro registrado no se podía realizar.

Los DENOs son los parados registrados más los siguientes colectivos:

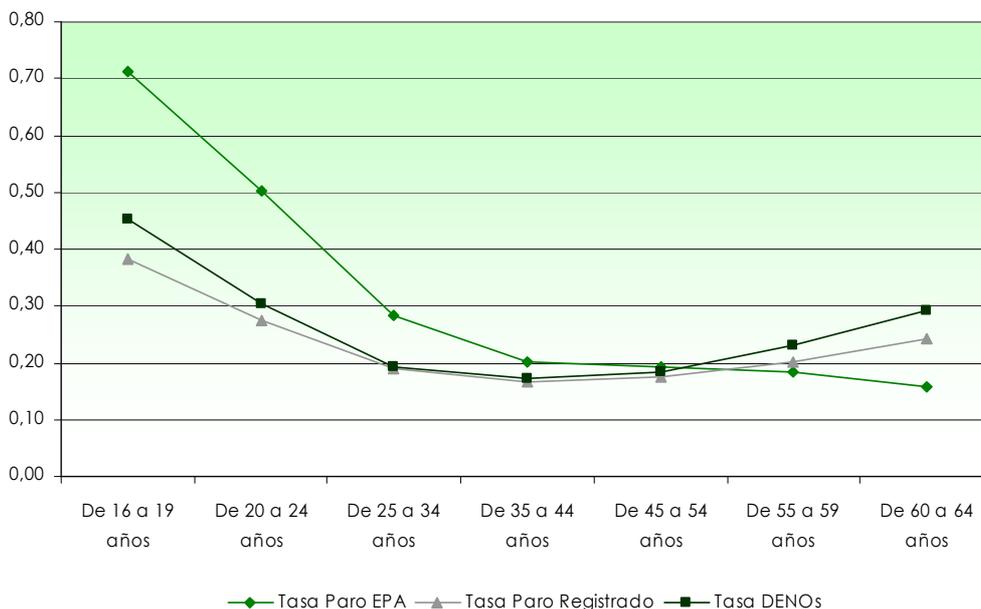
- Estudiantes
- Personas que han rechazado una acción de inserción laboral
- Demandantes de empleo coyuntural
- Demandantes de empleo de jornada reducida (menor de 20 horas)
- Trabajadores eventuales agrarios subsidiados.

Como ven se trata de algunos de colectivos considerados parados por la EPA y que se excluyen de la definición de paro registrado.

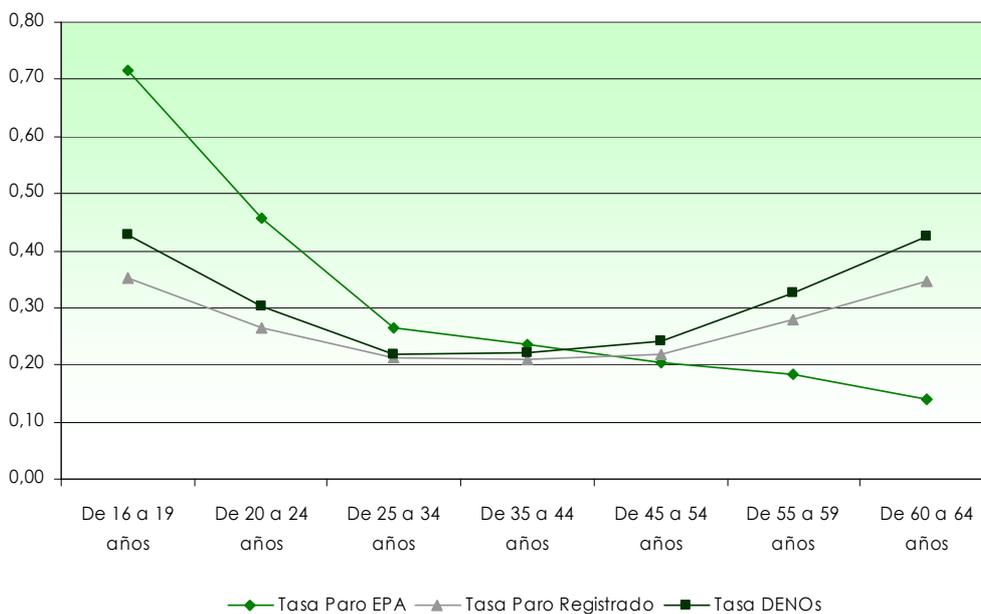
A la vista de dichas exclusiones cabe suponer que en las segmentaciones de paro registrado y el paro EPA según edad, las discrepancias han de concentrarse entre los colectivos de personas más jóvenes y más ancianas. Las figuras 4.1a 4.4 , en donde se recogen los datos del registro SISPE del SEPE y de la EPA en España y Cantabria, respectivamente confirman dicha suposición. El paro EPA es más elevado en los colectivos jóvenes, en tanto que entre la población de más edad se registra un mayor número de personas que las que estima la EPA como parados. Por otro lado, los datos regionales llevan a las mismas conclusiones que los nacionales.

Es pues este fichero de DENO el que se va a utilizar como variable auxiliar para estimar el paro EPA a nivel de sección censal.

**Figura 4.1. Tasas de Paro EPA y Paro Registrado masculino por grupos de edad en España. (% de población activa de la EPA). 2012-1T**



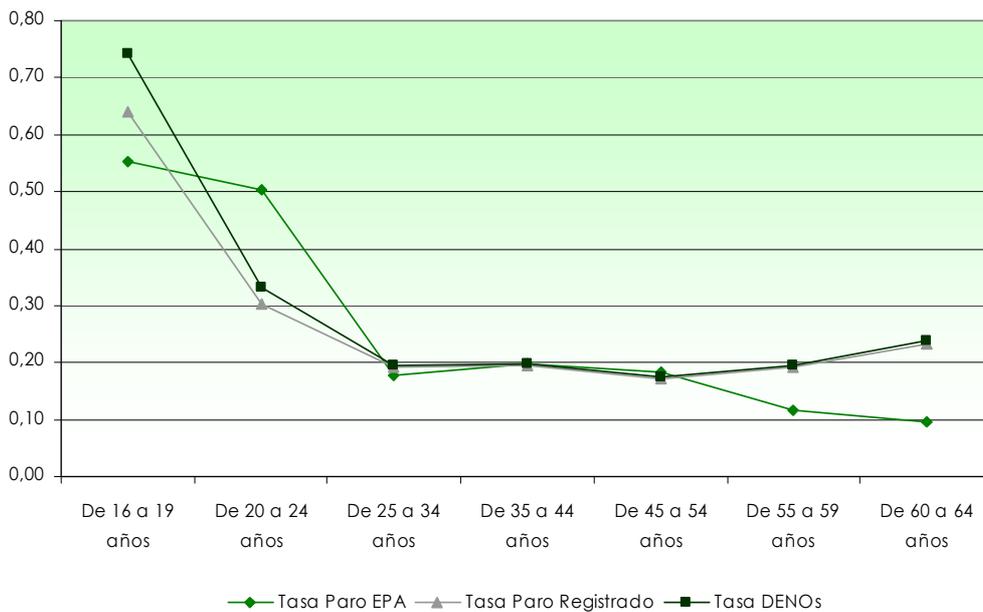
**Figura 4.2. Tasas de Paro EPA y Paro Registrado femenino por grupos de edad en España. (% de población activa de la EPA). 2012-1T**



**Figura 4.3. Tasas de Paro EPA y Paro Registrado masculino por grupos de edad en Cantabria. (% de población activa de la EPA). 2012-1T**



**Figura 4.4. Tasas de Paro EPA y Paro Registrado femenino por grupos de edad en Cantabria. (% de población activa de la EPA). 2012-1T**



Un inconveniente que presenta el fichero DENO's es la inadecuada codificación del domicilio del demandante de empleo que dificulta la adscripción de dicha persona a la sección censal de residencia, para solventar este inconveniente se han vinculado los ficheros del SISPE y el padrón municipal, lo que permite identificar la sección censal en la que residen a aproximadamente el 95% de los DENO's. El 5% de los demandantes de empleo que no han podido identificarse están dispersos entre los diferentes códigos postales de la región.

Para estimar el desempleo se va a utilizar entonces estimador sintético de ratio:

$$\hat{Y}_d = \sum_{g=1}^n \left( X_{gd} \frac{\hat{Y}_g}{X_g} \right), \text{ con } \hat{Y}_g = \sum_{j \in g} w_j y_j$$

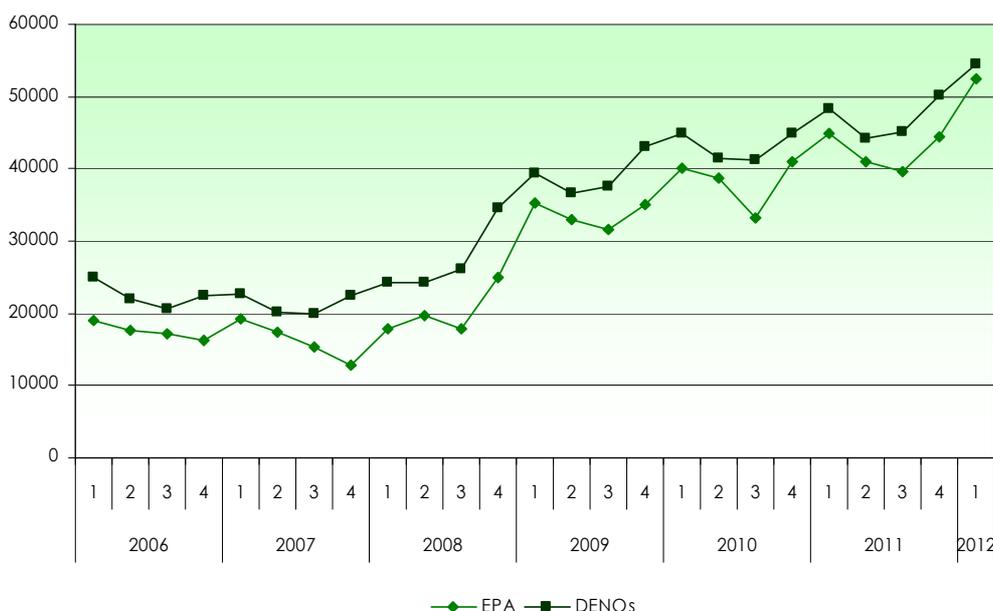
Donde  $\hat{Y}_d$  sería el número de población parada en el área pequeña,  $g$  los grupos de edades siguientes: ,  $\hat{Y}_g$  la población parada estimada en el grupo de edad  $g$  en el área grande obtenida el peso de diseño,  $w_j$ , y  $X_{gd}$  e  $X_g$  los

parados registrados como DENO en los registros del sistema público de empleo en el área pequeña  $d$  y el área grande, respectivamente.

Esta estimación se basa en la suposición que la ratio paro estimado entre paro registrado se mantiene estable para cada grupo de edad en el área grande y el área pequeña.

En Cantabria es mayor el colectivo DENO que el de parados de la EPA. Pero como se puede apreciar en la figura 4.5 las propiedades dinámicas, en cuanto a estacionalidad y tendencia, de ambas series son comparables lo que posibilita su utilización como fichero auxiliar para el objetivo propuesto.

**Figura 4.5 DENO SISPE y paro EPA de 16 a 64 años**



En consecuencia, se han estimado los parados a nivel de sección censal con el auxilio del registro de DENO, repartiendo la discrepancia estadística también de manera uniforme en cada grupo de edad y sexo, es decir que las diferencias encontradas entre parados de la EPA y DENO a nivel regional sirven para estimar a nivel de sección censal el paro EPA, en base a los DENO registrados en cada grupo de edad y sexo.

Las población parada se calibra finalmente utilizando el factor de calibrado descrito en el apartado anterior.

**Tabla 4.1 Sesgo de las estimaciones preliminares de paro en las secciones censales de Cantabria. 2012-1T**

**Estimación Población Parada (Estimación en pequeñas áreas)**

	16-24	25-34	35-44	45-54	55-64	Total
Varones	3,2	7,4	7,1	6,0	3,8	27,5
Mujeres	2,6	6,9	6,9	5,5	3,4	25,3

**Estimación Población Parada (EPA-INE)**

	16-24	25-34	35-44	45-44	55 -64	Total
Varones	3,5	9,1	7,1	4,8	2,9	27,4
Mujeres	3,7	6,5	7,1	5,9	1,8	25,0

## 5.- Estimación de la varianza y del error cuadrático

Las estimaciones de la varianza y del error cuadrático se ha realizado en base al método de remuestreo Jack-knife. Este método consiste en tomar tantas submuestras como clusters -secciones censales en el caso de la EPA- existen en la muestra (Tukey, 1958). Cada submuestra se forma así con la muestra original al dejar fuera una sección censal cada vez. En

el ejercicio realizado al partir de 92 secciones censales se obtienen 91 muestras. Denotando  $\hat{Y}_{d(hi)}^{\sin r_d}$  el estimador sintético-

región obtenido con la submuestra que elimina la sección censal  $i$  del estrato  $h$ , y donde los pesos han sido modificados de la siguiente forma:

$$w_{j(hi)} \begin{cases} w_j \text{ si el individuo } j \text{ no está en el estrato } h \\ 0 \text{ si el individuo } j \text{ está en la sección } i \text{ del estrato } h \\ \frac{n_h}{n_h - 1} w_j \text{ si el individuo } j \text{ no está en la sección } i \text{ pero si en el estrato } h \end{cases}$$

Con  $n_h$  el número de secciones en el estrato  $h$  ( $h=1, \dots, H$ ). El estimador Jack-knife del error cuadrático medio del estimador sintético-región es:

$$MS\hat{E}(\hat{Y}_d^{\sin r_d}) = \sum_{h=1}^H \left( \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{Y}_{d(hi)}^{\sin r_d} - \hat{Y}_{d(h)}^{\sin r_d})^2 + \left( (n_h - 1) (\hat{Y}_{d(h)}^{\sin r_d} - \hat{Y}_d^{\sin r_d})^2 \right) \right)$$

Con 
$$\hat{Y}_{d(h)}^{\sin r_d} = \frac{\sum_{i=1}^{n_h} \hat{Y}_{d(hi)}^{\sin r_d}}{n_h}$$

El cociente entre la raíz cuadrada de la estimación del error cuadrático medio obtenido utilizando el método Jack-knife y la estimación de la tasa en cada sección censal sería su coeficientes de variación. Los coeficientes de variación mayores y menores calculados para el primer trimestre de 2012 de se presentan en la tabla 5.1.

**Tabla 5.1. Coeficientes de variación por secciones censales. 2012-1T**

**Varianzas más altas**

COD_SECCION	Tasa de Actividad Hombres	COD_SECCION	Tasa de Actividad Mujeres	COD_SECCION	Tasa de Paro Hombres	COD_SECCION	Tasa de Paro Mujeres
3907507010	0,0608032	3905101001	0,1022690	3908801001	0,1054661	3908801001	0,1275740
3900801007	0,0519279	3908801001	0,0957651	3905101001	0,0853469	3905901002	0,1186609
3907503004	0,0512988	3902002012	0,0920501	3907507010	0,0842039	3907504011	0,1180222
3907502014	0,0502420	3907508028	0,0918862	3909301001	0,0755561	3905101001	0,1153474
3908703012	0,0497654	3901601013	0,0917658	3903401001	0,0743715	3903502005	0,1101523
3907507006	0,0494204	3908703010	0,0914541	3907502014	0,0735400	3907507022	0,1096665
3908701012	0,0491199	3904501001	0,0899538	3908701012	0,0725314	3907508006	0,1085674
3907901002	0,0475783	3901601014	0,0893521	3908703012	0,0722520	3907501004	0,1076082
3907508001	0,0466150	3901801004	0,0886388	3903901001	0,0713245	3907301002	0,1063291
3908704003	0,0459612	3907502001	0,0885362	3907507006	0,0709933	3905901003	0,1053962
3908001001	0,0459359	3909101003	0,0874995	3905901002	0,0707203	3907502001	0,1052198
3908705003	0,0456648	3900901001	0,0869574	3909501002	0,0702860	3907505015	0,1047472

**Varianzas más bajas**

COD_SECCION	Tasa de Actividad Hombres	COD_SECCION	Tasa de Actividad Mujeres	COD_SECCION	Tasa de Paro Hombres	COD_SECCION	Tasa de Paro Mujeres
3907502012	0,0170863	3901601012	0,0610564	3900802004	0,0439120	3902002012	0,0743533
3907507013	0,0170379	3908701012	0,0602580	3902002010	0,0438002	3905201003	0,0743505
3907507015	0,0170044	3900801007	0,0601938	3901801001	0,0436116	3901901001	0,0742291
3906601001	0,0170031	3907901002	0,0601564	3900801008	0,0435451	3907401006	0,0738161
3908705001	0,0169562	3908001001	0,0598721	3908704009	0,0433539	3902002013	0,0732410
3903301001	0,0169511	3908703012	0,0595353	3907508028	0,0425301	3902801003	0,0728332
3907502021	0,0169215	3907503004	0,0595314	3901601013	0,0424759	3908703002	0,0726091
3909101003	0,0167397	3900301001	0,0595230	3902002003	0,0423515	3907508028	0,0725854
3907501007	0,0167241	3907508019	0,0592350	3901801004	0,0418606	3904701002	0,0725325
3906201002	0,0166666	3907507006	0,0591755	3901601014	0,0417602	3902002011	0,0723369
3901801002	0,0166569	3909101002	0,0577511	3902002011	0,0411175	3905201005	0,0716566
3903601001	0,0166521	3907507010	0,0548536	3902002012	0,0408679	3907401005	0,0714258

## 6.- Resultados

Para evaluar los resultados obtenidos, se ha hecho una comparativa de las estimaciones municipales de las tasas de actividad (tabla 6.1) y paro (tabla 6.2) con las que resultan en la EPA para aquellos municipios que tienen mayor muestra: Castro Urdiales (20), Santander (75), y Torrelavega (87). Las muestras de la EPA en estos municipios son de 89 personas en Castro Urdiales, de 956 en Santander y de 323 en Torrelavega. A la hora de valorar estos resultados hay que señalar que el tamaño muestral de Santander es el único adecuado a estos análisis, ya que en las estimaciones por grupos de edades y sexos de las tasas de actividad y paro de Torrelavega y Castro Urdiales en la EPA están sujetas a elevados errores muestrales. De hecho los problemas muestrales de las estimaciones EPA, en el caso de las tasas de actividad de Castro Urdiales y Torrelavega llevan aparejado algún anacronismo, en particular el que encontremos estratos en los que todos los individuos están activos (mujeres de 55-59 años, varones de 20 a 24 años, etc) o no haya ninguno (mujeres de 16 a 19 años, mujeres de 60 a 64, etc). El estimador auxiliado por la población padronal como se puede apreciar en la tabla no incurre en estas inconsistencias.

**Tabla 6.1. Tasas de actividad por grupos de edad y sexos, estimadas con variable auxiliar y con los factores de elevación de la EPA en Castro Urdiales (20), Santander (75) y Torrelavega (87). 2012-1T**

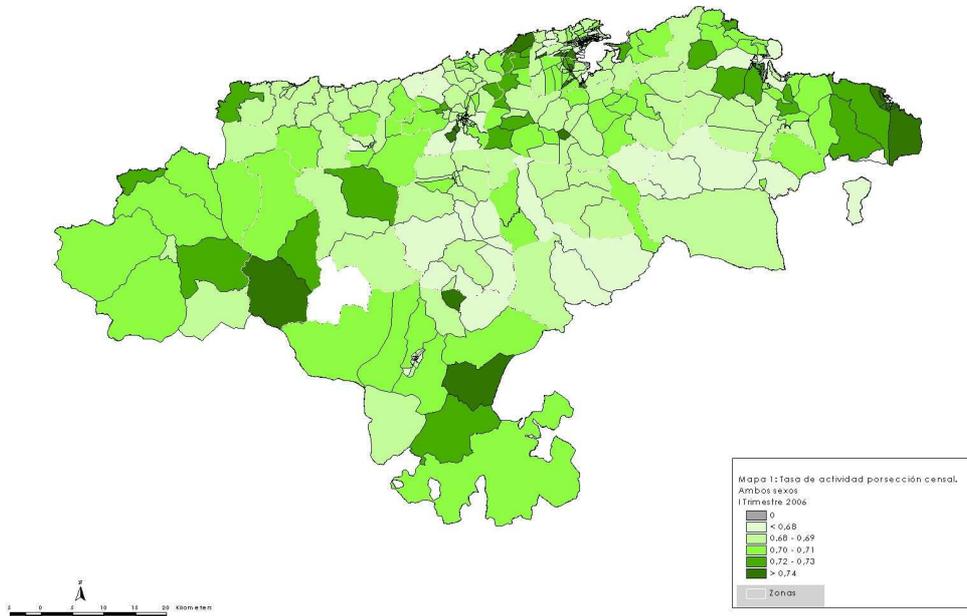
Edad	Sexo	Estimadas			EPA		
		20	75	87	20	75	87
Varones	16-19	10,1	10,6	10,5	0,0	7,3	27,5
	20-24	54,4	54,6	54,6	100,0	57,7	25,2
	25-34	91,7	90,6	90,9	100,0	91,9	94,3
	35-44	93,3	93,3	93,4	100,0	95,8	84,8
	45-54	88,0	87,8	87,8	100,0	90,7	89,9
	55-59	79,2	78,4	78,3	86,3	72,8	83,1
	60-64	32,2	31,9	32,1	32,4	41,0	40,9
Mujeres	16-19	6,9	7,2	7,3	35,1	20,0	0,0
	20-24	48,8	48,8	48,6	0,0	32,8	49,3
	25-34	89,9	89,4	89,4	90,1	88,8	92,9
	35-44	76,3	76,3	76,4	67,7	85,6	77,7
	45-54	73,3	72,9	72,8	59,7	76,8	66,0
	55-59	56,5	56,5	56,6	100,0	55,1	79,9
	60-64	29,7	29,6	29,8	0,0	33,8	29,1

**Tabla 6.2. Tasas de paro por grupos de edad y sexos, estimadas con variable auxiliar y con los factores de elevación de la EPA en Castro Urdiales (20), Santander (75) y Torrelavega (87). 2012-1T**

Sexo	Edad	Estimadas			EPA		
		20	75	87	20	75	87
Varones	16-19	41,9	42,8	70,0	0,0	66,4	100,0
	20-24	34,6	30,1	46,0	0,0	34,7	52,7
	25-34	19,0	24,3	28,0	34,3	34,2	33,7
	35-44	12,0	18,2	19,6	4,7	17,4	14,0
	45-54	10,1	12,9	15,2	11,5	15,6	14,5
	55-59	12,8	15,3	15,1	0,0	3,5	0,0
	60-64	11,1	14,3	11,4	100,0	6,5	0,0
Mujeres	16-19	37,1	47,8	82,5	0,0	76,2	0,0
	20-24	40,3	43,3	58,9	0,0	52,9	46,7
	25-34	13,6	17,2	19,9	0,0	20,1	24,4
	35-44	20,3	19,1	24,3	30,5	17,4	20,6
	45-54	23,7	18,3	22,4	21,3	15,7	24,2
	55-59	17,1	12,4	13,5	0,0	3,5	29,8
	60-64	12,4	11,3	11,8	0,0	17,9	0,0

En las figuras 6.1 a 6.4 se presentan los mapas a nivel de secciones censales de las tasas de actividad y paro estimadas, correspondientes al primer trimestre de 2006 y de 2012. En todos los mapas es manifiesta la continuidad territorial de las tasas representadas.

**Figura 6.1. Tasa de actividad. Primer trimestre de 2006.**



**Figura 6.2. Tasa de actividad. Primer trimestre de 2012.**

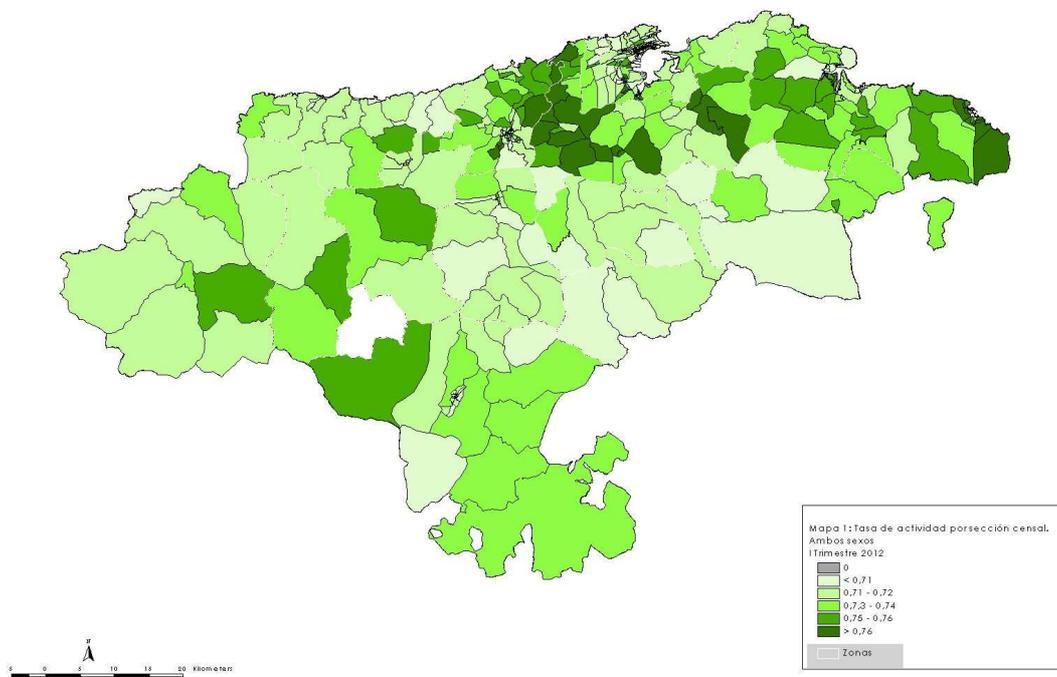


Figura 6.3. Tasa de Paro. Primer trimestre de 2006

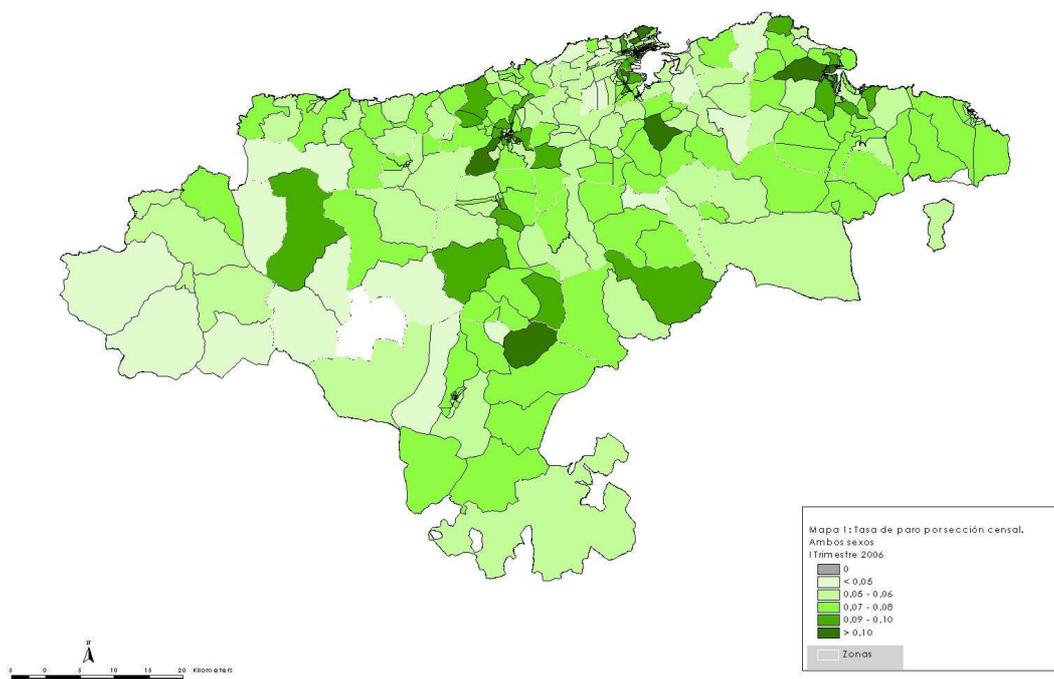
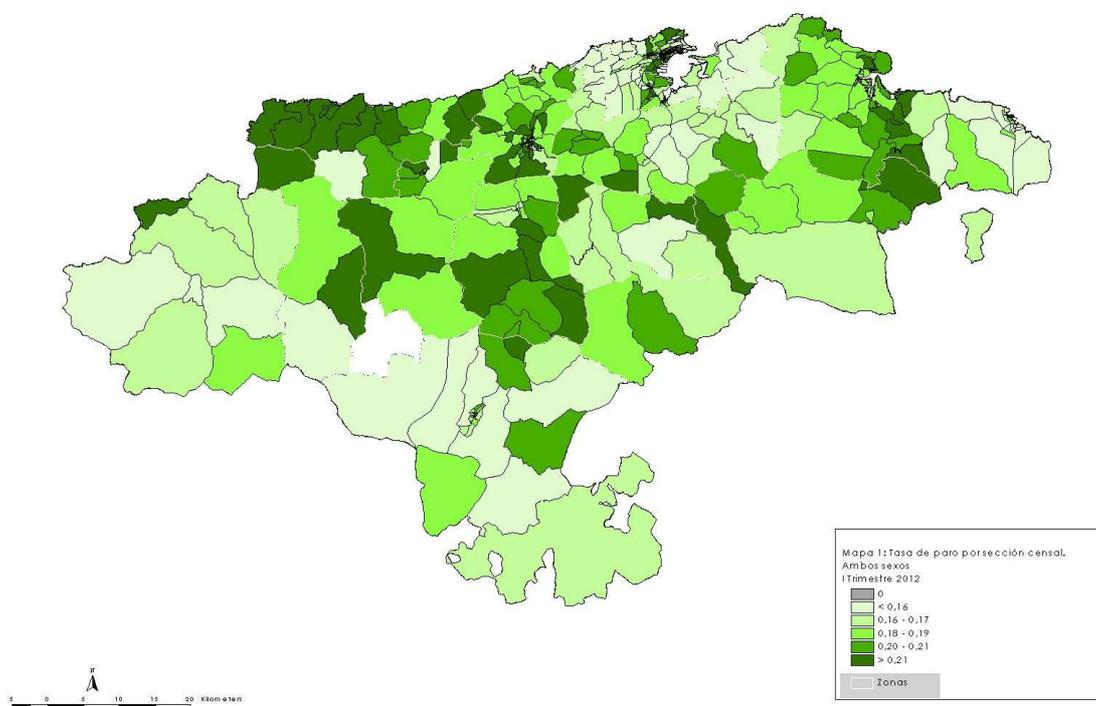


Figura 6.4. Tasa de Paro. Primer trimestre de 2012



## 7.- Cambio de base poblacional en las estimaciones de la EPA

Todos los datos utilizados en esta metodología estaban basados en la base poblacional del Censo de 2001, pero tras la revisión por parte del INE de los factores de elevación al cambiar la base poblacional de la EPA al Censo 2011 a partir del primer trimestre de 2014, se han recalculado las estimaciones de las áreas pequeñas de la EPA en Cantabria para poder ofrecer una continuidad de las series.

El cambio de base poblacional en la Encuesta de Población Activa (EPA) consiste en la incorporación y actualización de las nuevas series de población y hogares derivadas del Censo de Población y Viviendas de 2011, en sustitución de las que se venían utilizando hasta ahora basadas en el Censo de 2001.

La actualización de las cifras de población es un procedimiento habitual y obligado a realizar en todas las encuestas a hogares, y por lo tanto también en la EPA, cuando surgen variaciones significativas en las cifras de población. Este procedimiento de actualización está previsto en la propia metodología de la Encuesta y está recomendado por la Oficina de Estadística de la Unión Europea (Eurostat). En este caso, además, se ha revisado el método de calibrado de los factores.

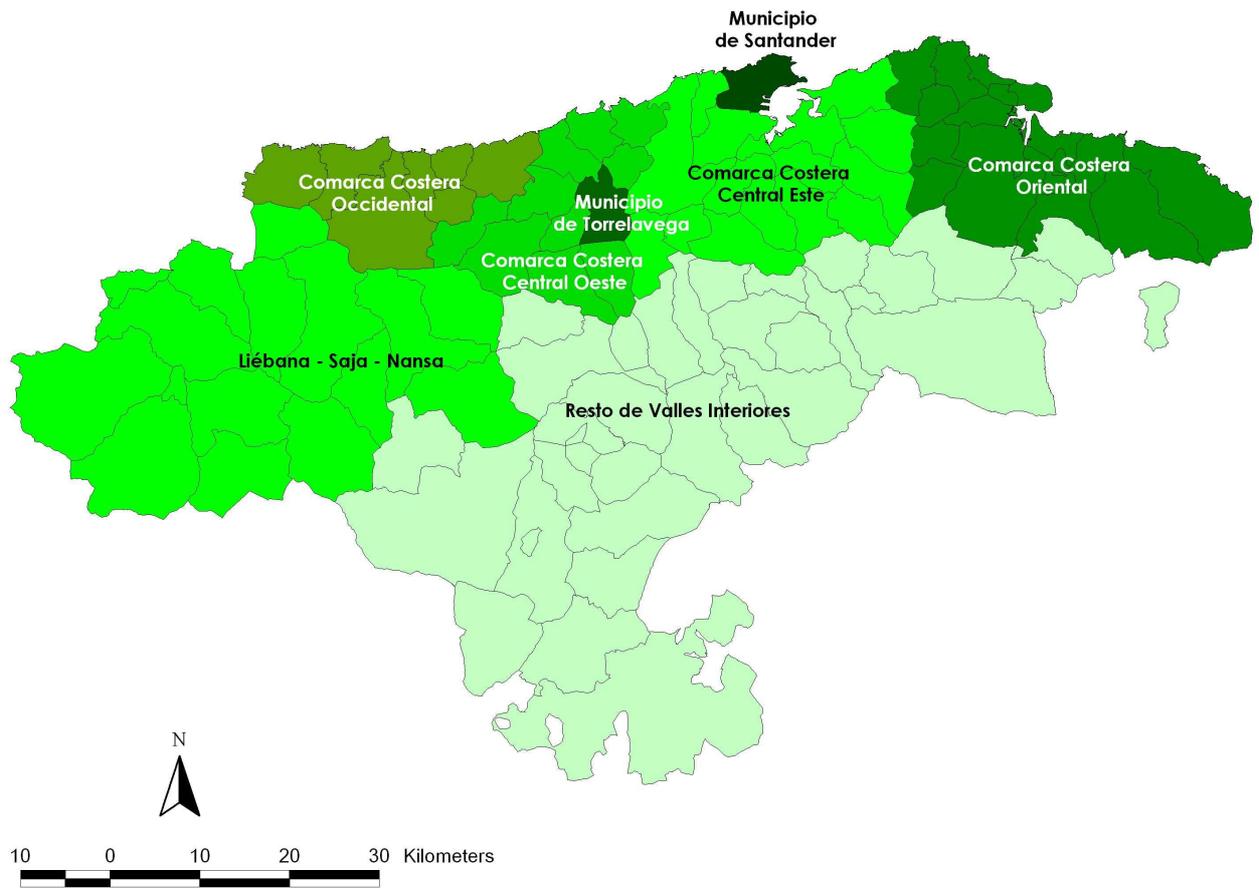
## Anexo.- Áreas Pequeñas

Finalmente, aunque los cálculos se han realizado a nivel de sección censal, en la publicación se presentan los resultados agrupados según las siguientes áreas pequeñas (Tabla Anexo.1), representadas como aparece en la figura Anexo.1:

**Tabla Anexo.1. Correspondencia entre Áreas Pequeñas de la EPA y Municipios de Cantabria**

Áreas Pequeñas	Municipio	Áreas Pequeñas	Municipio
1 Santander	75 Santander	13 Cabezón de Liébana	
2 Torrelavega	87 Torrelavega	14 Cabuérniga	
3 Costera Central Este	8 Astillero, El	15 Camaleño	
	16 Camargo	22 Cillorigo de Liébana	
	19 Castañeda	33 Herrerías	
	28 Entrambasaguas	34 Lamasón	
	37 Liérganes	49 Peñarrubia	
	40 Marina de Cudeyo	50 Pesaguero	
	42 Medio Cudeyo	53 Polaciones	
	48 Penagos	55 Potes	
	52 Piélagos	63 Rionansa	
	56 Puente Viesgo	66 Ruento	
	61 Ribamontán al Mar	86 Tojos, Los	
	62 Ribamontán al Monte	88 Tresviso	
	64 Riotuerto	89 Tudanca	
73 Santa Cruz de Bezana	96 Vega de Liébana		
74 Santa María de Cayón			
99 Villaescusa			
4 Costera Central Oeste	12 Cabezón de la Sal	3 Anievas	
	18 Cartes	4 Arenas de Iguña	
	25 Corrales de Buelna, Los	7 Arredondo	
	41 Mazcuerras	10 Bárcena de Pie de Concha	
	44 Miengo	17 Campoó de Yuso	
	54 Polanco	21 Cieza	
	60 Reocín	26 Corvera de Toranzo	
	69 San Felices de Buelna	27 Enmedio, Campoó de	
	76 Santillana del Mar	32 Hermandad de Campoó de Suso	
	85 Suances	39 Luena	
5 Costera Oriental	2 Ampuero	45 Miera	
	5 Argoños	46 Molledo	
	6 Arnauero	51 Pesquera	
	9 Bárcena de Cicero	57 Ramales de la Victoria	
	11 Bareyo	58 Rasines	
	20 Castro-Urdiales	59 Reinosa	
	23 Colindres	65 Rozas de Valdearroyo, Las	
	29 Escalante	67 Ruesga	
	30 Guriezo	70 San Miguel de Aguayo	
	31 Hazas de Cesto	71 San Pedro del Romeral	
	35 Laredo	72 San Roque del Río Miera	
	36 Liendo	77 Santiurde de Reinosa	
	38 Limpias	78 Santiurde de Toranzo	
	43 Meruelo	81 Saro	
	47 Noja	82 Selaya	
	79 Santoña	83 Soba	
84 Solórzano	92 Valdeolea		
102 Voto	93 Valdeprado del Río		
6 Costera Occidental	1 Alfoz de Lloredo	94 Valderredible	
	24 Comillas	97 Vega de Pas	
	68 Ruiloba	98 Villacarriedo	
	80 San Vicente de la Barquera	100 Villafufre	
	90 Udías	101 Valle de Villaverde	
	91 Valdáliga		
95 Val de San Vicente			
		7 Liébana-Saja-Nansa	
		8 Resto de Valles Interiores	

Figura Anexo.1. Áreas Pequeñas.



## Bibliografía

- Calvo G. (2010): ESTIMACIÓN EN ÁREAS PEQUEÑAS DE LA POBLACIÓN ACTIVA DE LOS MUNICIPIOS DE CANTABRIA.. Universidad de Cantabria.
- Costa A, García M, Pardal M. (2009): ESTIMANDO EL PARO COMARCAL DE CATALUNYA CON ESTIMADORES COMPUESTOS DE PEQUEÑA ÁREA. XXX Congreso Nacional de Estadística e Investigación Operativa.
- Cuadrado P, Lacuesta A, Martínez JA y Pérez E. (2007) EL FUTURO DE LA TASA DE ACTIVIDAD ESPAÑOLA: UN ENFOQUE GENERACIONAL. Documentos de Trabajo. N.º 0732. Banco de España.
- EUSTAT (2008): ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA DE POBLACIÓN EN RELACIÓN CON LA ACTIVIDAD DE LA C.A. DE EUSKADI. 2008
- Fernández, A.I., Rodríguez-Poo, J.M. (1997) ESTIMATION AND SPECIFICATION TESTING IN FEMALE LABOR PARTICIPATION MODELS: PARAMETRIC AND SEMIPARAMETRIC METHODS. *Econometric Reviews*, vol. 16 n. 2, págs 229-251.
- Gallant, A. R.(1981) ON THE BIAS IN FLEXIBLE FUNCTIONAL FORMS AND AN ESSENTIALLY UNBIASED FORM. *J. Econometrics* 15. págs 211-45.
- Gallant, A. R.(1984) THE FOURIER FLEXIBLE FORM. *Amer. J. Agr. Econ.* 66. págs, 204-15.
- López, R. (2000): ESTIMACIONES PARA ÁREAS PEQUEÑAS. *Estadística Española* Vol. 42, Núm. 146, 2000, págs. 291 a 338.
- ISTAC (2008): ENCUESTA DE POBLACIÓN ACTIVA. METODOLOGÍA PARA LA ESTIMACIÓN EN PEQUEÑAS ÁREAS DE CANARIAS.
- López, M E., Sánchez P. y Vaamonde A.: ESTIMACIÓN LOCAL DE VARIABLES LABORALES. UNA APLICACIÓN A LAS COMARCAS DE GALICIA (2009). IX Congreso Galego de Estatística e Investigación de Operacións. Ourense, 12-13-14 de novembro de 2009
- Morales, D., Esteban, M. D., Sánchez, A. Santamaría, L., Marhuenda, Y., Pérez, A., Saralegui, J. y Herrador, M. (2007): ESTIMACIÓN EN ÁREAS PEQUEÑAS CON DATOS DE LA ENCUESTA DE POBLACIÓN ACTIVA EN CANARIAS. *Estadística Española*, 49, 2007, págs 301-332.
- Pérez, J.I. (2008): ¿POR QUÉ EL PARO REGISTRADO SUPERA AL PARO DE LA EPA?. *Revista del Ministerio de Trabajo y Asuntos Sociales* nº71, 2008 págs. 53 a 65.
- Rao, J.N.K. *Small Area Estimation*. New York: Wiley. 2003.
- Sánchez-Crespo, G. (1983): METODOLOGÍA PARA LA ESTIMACIÓN EN DOMINIOS DE ESTUDIOS PEQUEÑOS. Universidad Autónoma de Madrid, 1983.
- Toharia Cortés, L. (2003): EL MERCADO DE TRABAJO EN ESPAÑA, 1978-2003. ICE. diciembre 2003 número 811 págs 204 a 220