

# **Indicadores Urbanos. Métodos de estimación de las variables sobre mercado de trabajo**

## **Estimación de la población activa municipal**

Las variables que tenemos que calcular en este dominio presentan una cierta complejidad ya que, en la mayoría de los casos requieren de estimaciones de activos, empleados y parados que, siendo proporcionados por la Encuesta de Población Activa del INE no llegan a tener los niveles de desagregación espacial deseados. Así, mientras que para las variables anteriormente especificadas el proyecto URBAN AUDIT requiere desagregación a nivel de ciudades, la EPA solo es capaz de dar estimaciones directas a nivel de provincias. Con el objeto de realizar la tarea requerida proponemos el siguiente procedimiento de estimación. Comenzamos con la variable población activa para más adelante proponer un procedimiento de cálculo de desempleados y ocupados.

La información muestral que se utiliza para la estimación son los microdatos de la Encuesta de Población Activa elaborada por el Instituto Nacional de Estadística (INE) para cada trimestre del año de referencia objeto de la estimación. Las variables que se utilizan de la EPA son la disposición del individuo a participar en el mercado de trabajo,  $y$ , el sexo y la edad del encuestado,  $X$ . Con estas variables se procedió a estimar los parámetros de interés de la ecuación

$$P\left(y = 1 / X = x, \theta_d\right) = \frac{\exp\left(\theta_d + \sum_p \beta_p x^p + \sum_k \delta_k (x - \tau_k)_+^p\right)}{1 + \exp\left(\theta_d + \sum_p \beta_p x^p + \sum_k \delta_k (x - \tau_k)_+^p\right)},$$

donde la cantidad  $P\left(y = 1 / X = x, \theta_d\right)$  denota la probabilidad o tasa para un individuo de la muestra de la EPA de participar en el mercado de trabajo teniendo la edad  $X$  y residiendo en la provincia  $d$ . Los parámetros de interés de este modelo se estiman a través de la técnica de estimación máximo verosímil penalizada (ver

Ruppert, Wand y Carroll, 2003). La función

$$f(x) = \theta_d + \sum_p \beta_p x^p + \sum_k \delta_k (x - \tau_k)_+^p$$

consiste en un spline polinomial de orden P que incorpora un efecto aleatorio,  $\theta_d$ , que varía a nivel provincial (ver Opsomer, Claeskens, Ranalli y Kauermann, 2008). En él distinguiremos una parte polinomial de orden 3 (P=3) y luego una parte que variará con los valores  $\{\tau_1, \tau_2, \dots, \tau_k\}$  denominados nodos que deben de ser elegidos por el usuario.

Una vez calculada la tasa de participación, el número de activos por municipio puede calcularse a través de la siguiente expresión

$$\hat{Y}_m = \sum_{i=1}^N 1_{\{X=x_i, m\}} \hat{P}\left(y = 1 / X = x_i, \theta_d\right),$$

donde N es el número de habitantes mayores de 16 años del municipio m, y  $1_{\{X=x_i, m\}}$  es una función indicatriz que toma valor 1 si el individuo i-ésimo tiene la edad  $x_i$  y 0 si no es así y

$$\hat{P}\left(y = 1 / X = x, \theta_d\right) = \frac{\exp(\theta_d + \sum_p \hat{\beta}_p x^p + \sum_k \hat{\delta}_k (x - \tau_k)_+^p)}{1 + \exp(\theta_d + \sum_p \hat{\beta}_p x^p + \sum_k \hat{\delta}_k (x - \tau_k)_+^p)},$$

Este estimador, de forma intuitiva viene a reflejar lo siguiente: si la tasa de actividad de los habitantes de un municipio (su probabilidad de que sean activos) se puede aproximar de manera correcta a través de su edad entonces, si conocemos la distribución por edades de los residentes de un municipio tendremos información suficiente para calcular el número de activos de ese municipio.

Para tratar la variable del sexo, optamos por emplear dos modelos por separado, dado que de manera intuitiva asumiremos que ambos géneros presentan un comportamiento diferente en lo que respecta a la participación. La variable edad la trataremos como la variable principal del modelo especificado. La parte paramétrica consistirá en un polinomio de grado tres, que es suficiente para tener precisión en la aproximación y no

perder suavidad en ésta. Los nodos se eligen iguales a los tramos de edad que se presentan en los resultados de la estimación. Es evidente que otras muchas variables que aparecen en la EPA podrían aparecer como explicativas de la decisión de participar en el mercado laboral (ver Fernandez y Rodriguez-Poo, 1997), sin embargo, la elección que aquí hacemos está condicionada por el objetivo de la predicción del número de activos por municipio. Como puede observarse en la última ecuación, con el objeto de efectuar la predicción a nivel municipal necesitamos conocer la población mayor de 16 años de cada municipio,  $N$ , estratificada por sexo y por edad, y este dato no nos lo proporciona la EPA, que solo da estimaciones de población a nivel provincial. Es necesario pues, buscar un registro administrativo que nos permita complementar y a la vez enriquecer la información aportada por la EPA. El Padrón Municipal de Habitantes dispone de mucha menor información que la EPA, pero entre esta información encontramos a todos los ciudadanos empadronados en España caracterizados por sexo y por edad. Variables que tienen en común con la EPA.

Finalmente, con el objetivo de dar coherencia a los resultados con respecto a dominios anteriores utilizaremos los datos de Cifras de Población ‘repartidos’ a nivel municipal utilizando el padrón municipal de habitantes estratificado por sexo y por edad.

## **Estimación del desempleo**

### **Introducción**

La metodología disponible utiliza información procedente del fichero de microdatos EPA y de los datos proporcionados por el SEPE.

Esta metodología se basa en el reparto entre los municipios de cada provincia de las estimaciones EPA provinciales por grupos de edad y sexo (8), según los datos del SEPE. De esta forma se garantiza que al final, la suma de las estimaciones municipales así obtenidas coincide con las estimaciones EPA.

Para cada grupo de edad y sexo, la expresión general de la estimación en un municipio  $i$  es:

$$\widehat{Y}_i = \frac{SEPE_i}{SEPE} \cdot \widehat{EPA}$$

No obstante, esta forma general puede resultar más precisa al calcularla al nivel de estrato, es decir para cada grupo de edad y sexo del estrato  $h$  de una provincia, la expresión es:

$$\widehat{Y}_{i,h} = \frac{SEPE_{i,h}}{SEPE_h} \cdot \widehat{EPA}_h$$

En relación a esta fórmula, véase más detalladamente la metodología al respecto en el anexo de este documento.

## Problemas

De acuerdo con la fórmula anterior, se presentan dos tipos de problemas:

- En algunos municipios grandes, hay diferencias importantes entre la estimación EPA y la así obtenida.
- El orden de los municipios de una provincia establecido en función del número de desempleados no es el mismo con las estimaciones, que al utilizar los datos del SEPE.

A la vista de estos problemas, se plantean las siguientes consideraciones:

1. En las estimaciones EPA por estratos hay que considerar tres aspectos:
  - Es posible que tengan cierta inestabilidad derivada de muestras pequeñas y del muestreo en dos etapas (efecto conglomerado)
  - Al estimar agrupando los municipios en estratos, es posible el cambio de orden según paro de los municipios de una provincia, respecto del proporcionado por el SEPE.
  - Como la EPA es una encuesta continua, el estrato al que se asocia un municipio no cambia, hasta que se disponga de la información de un nuevo censo. Por tanto es posible cierta desactualización del estrato en algunos municipios.

2. No es obvio que la variable medida por la EPA, parados según lugar de residencia, sea la misma que la que proporciona el SEPE, inscritos en el registro de desempleo posiblemente según lugar de trabajo.

## Propuesta

Dado que la única variable disponible para repartir el paro provincial EPA entre los municipios, es la procedente del SEPE, se soslaya la consideración 2, y la propuesta final se centra en la agrupación de estratos.

Teniendo en cuenta que en las capitales de provincia de mayor población, la muestra EPA es suficientemente grande, se propone lo siguiente:

1. En las provincias de Madrid, Barcelona, Sevilla, Zaragoza, Valencia y Málaga, se consideran dos estratos: Capital y resto de municipios, y se utiliza la fórmula disponible:

$$\widehat{Y}_{i,h} = \frac{SEPE_{i,h}}{SEPE_h} \cdot \widehat{EPA}_h$$

La fórmula se aplica en cada uno de los grupos ocho grupos de edad y sexo.

2. En el resto de provincias, se prescinde de estratos, y se utiliza, por tanto, la expresión:

$$\widehat{Y}_i = \frac{SEPE_i}{SEPE} \cdot \widehat{EPA}$$

Esta última expresión prescinde de los estratos y, sencillamente, reparte el paro EPA de cada grupo de edad y sexo entre municipios según los datos del SEPE.

Los aspectos más positivos de este método son:

- El método es sencillo de aplicar, y en muchos casos replicable
- No se altera el orden municipal de parados, excepto en lo que se refiere a las seis capitales señaladas.

- Es un método estable en cuanto a que no depende de la estratificación de la encuesta.

## Población ocupada municipal

La población ocupada se obtiene por diferencia entre la población activa y la población desempleada o parada.

## Anexo

La estimación de parados a nivel municipal se efectuará a través del paro registrado, datos provenientes del Servicio de Empleo Público Estatal. La idea fundamental es considerar como punto de partida las estimaciones de paro a nivel de comunidad autónoma y de estrato provenientes de la Encuesta de Población Activa y utilizando éstas, repartir los datos de paro EPA por municipio utilizando como variable auxiliar el paro registrado. Los estratos que utilizamos para la calibración son

Estrato 1	Municipio capital de provincia
Estrato 2	Municipios autorrepresentados, importantes en relación con la capital
Estrato 3	Otros municipios autorrepresentados, importantes en relación con la capital o municipios mayores de 100.000 habitantes
Estrato 4	Municipios entre 50.000 y 100.000 habitantes
Estrato 5	Municipios entre 20.000 y 50.000 habitantes
Estrato 6	Municipios entre 10.000 y 20.000 habitantes
Estrato 7	Municipios entre 5.000 y 10.000 habitantes
Estrato 8	Municipios entre 2.000 y 5.000 habitantes
Estrato 9	Municipios menores de 2.000 habitantes

El procedimiento de estimación es el siguiente. Vamos a considerar que disponemos para un determinado periodo de tiempo de  $N$  parados registrados en las oficinas públicas de empleo a nivel nacional. De estos  $N$  parados disponemos de información sobre un conjunto de características tales como sexo, edad y municipio de residencia. Todas estas características vamos a denotarlas por el vector  $X$ . De esta manera, al parado registrado  $i$ -ésimo le corresponderá el vector de características  $X_i$  y esas características pueden combinarse formando diferentes estratos que vamos a denominar  $\{X_1, X_2, \dots, X_J\}$ . Así mismo, vamos a definir como  $n_E(X_j)$  al número de parados estimados a través de la encuesta de población activa que están adscritos al estrato  $j$ -ésimo. Por ejemplo,  $n_E(X_1)$  podría consistir en el número de parados EPA de sexo masculino, edad comprendida entre los 16 y 25 años, que residen en las capitales de provincia de una determinada comunidad autónoma. Siguiendo la misma lógica vamos a definir por  $n_P(X_j)$  al número de parados registrados en las oficinas públicas de empleo adscritos al estrato  $j$ -ésimo. Finalmente denotamos por  $w(X_j) = \frac{n_E(X_j)}{n_P(X_j)}$

para  $j = 1, \dots, J$  estratos.

Proponemos el siguiente estimador para calcular el número total de parados en el municipio  $m$ :

$$\hat{y}_m = \sum_{i=1}^N \sum_{j=1}^J w(X_j) 1(X_j = X_i) 1_i(m),$$

Donde  $1(X_j = X_i)$  es una función que toma valor 1 si las características del individuo  $i$ -ésimo coinciden con las del estrato  $j$ -ésimo y 0 en caso contrario. Esto es, si el individuo es de sexo varón, tiene una edad entre 16 y 25 años, y reside en una determinada comunidad autónoma (todas estas características de uno de los estratos) entonces esta función indicatriz tomará valor 1. La función  $1_i(m)$  toma valor 1 si el individuo  $i$ -ésimo reside en el municipio  $m$  y 0 si lo contrario.

Como puede observarse por su construcción, este estimador presenta interesantes propiedades en términos de coherencia con las estimaciones directas de la Encuesta de Población Activa (EPA). Así, puede comprobarse que las estimaciones realizadas a nivel municipal de paro coinciden con las cifras de paro EPA estratificadas por sexo y tramo de edad a nivel de comunidad autónoma y a nivel nacional.