

# Urban Indicators. Methods of estimating labor market variables

## Estimation of active population at municipal level

The variables that we have to calculate in this domain offer a certain complexity as in the majority of cases they require estimations of the working-age, employed and unemployed persons which, as they are provided by the INE Labour Force Survey, do not finally possess the desired levels of spatial disaggregation. Thus, while for the above-specified variables the Urban Audit project requires disaggregation at city level, the Labour Force Survey (EPA) is only capable of giving direct estimations at provincial level.

In order to carry out the required task, we put forward the following estimation procedure. We start off with the active population variable so as to go on later to propose a procedure for calculation of persons unemployed and employed.

The sampling information that is used for the estimation is the microdata of the Labour Force Survey (EPA) produced by the National Institute of Statistics (INE) for each quarter of the reference year subject to estimation. The Labour Force Survey variables that are used are the individual's willingness to take part in the labour market,  $y$ , and the interviewee's gender and age,  $X$ . With these variables we proceeded to estimate the parameters of interest of the equation

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

where the quantity  $P\left(y = 1 / X = x, \theta_d\right)$  denotes the probability or rate for an individual in the Labour Force Survey sample of taking part in the labour market, taking the age  $X$  and living in province  $d$ . The parameters of interest of this model are estimated by means of the technique of maximum penalized likelihood estimation (see Ruppert, Wand and Carroll, 2003). The function

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

consists of a spline polynomial of order  $P$  including a random effect,  $\theta_d$ , which varies at provincial level (see Opsomer, Claeskens, Ranalli and Kauermann, 2008). In it we will distinguish a polynomial part of order 3 ( $P=3$ ) and then a part which will vary with the values  $\{\tau_1, \tau_2, \dots, \tau_k\}$  called nodes, which have to be chosen by the user.

Once the participation rate has been calculated, the number of working age persons per municipality may be calculated by way of the following expression

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

where  $N$  is the number of inhabitants of over 16 years of age in municipality  $m$ , and  $1_{\{x=x_i, m\}}$  is an indicator function that takes the value of 1 if the  $i$ -th individual is aged  $x_i$  and 0 if not, and

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

Intuitively, this estimator serves to reflect the following: if the employment rate of the inhabitants of a municipality (their probability of being in employment) may be approached correctly by way of their age, then, if we know that age distribution of the residents of a municipality, we will have sufficient information to calculate the number of persons in employment in that municipality.

To handle the gender variable, we opt for using two models separately, since we will assume intuitively that both sexes display a different behaviour pattern insofar as participation is concerned. We will treat the age variable as the main variable in the model specified. The parametric part will consist of a grade three polynomial, which is sufficient to achieve precision in the approach and not to forfeit smoothness in this. The nodes are chosen the same as the age bands which are offered in the results of the estimation. It is obvious that many other variables included in the Labour Force Survey (EPA) could appear as factors explaining the decision to take part in the labour market (see Fernandez and Rodriguez-Poo, 1997), yet the choice that we make here is predetermined by the aim of predicting the number of persons in employment per municipality. As may be seen in the last equation, in order to make the prediction at municipal level, we need to know the population over 16 years of age in each municipality,  $N$ , stratified by gender and age, and this item is not provided by the Labour Force Survey (EPA), which only offers estimations of population at provincial level. It is necessary, therefore, to seek an administrative register which will allow us to supplement and at the same time enrich the information supplied by the Labour Force Survey (EPA). The Municipal Roll of Inhabitants has much less information than the Labour Force Survey (EPA), but amongst this information we find all the citizens registered in Spain, characterized by gender and age, which are variables that they have in common with the Labour Force Survey (EPA).

Lastly, in order to make the results compatible with previous domains, we will use the data of the Population Figures 'distributed' on a national level, employing the Municipal Roll of Inhabitants stratified by gender and age.

# Estimation of unemployment

## Introduction

The methodology uses information available from the EPA microdata file and data provided by the SEPE (an autonomous body under the Ministry of Employment and Social Security). This methodology is based on the distribution among the municipalities of each province of the provincial EPA estimates by age and sex (8), according to data from the SEPE. In this way, it is ensured that finally, the sum of the thus obtained municipal estimates coincide with EPA estimates.

For each age group and sex, the general expression of the estimate in a municipality  $i$  is:

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

However, this formula may be more accurate if it is calculated at the stratum level, ie, for each age group and sex in stratum  $h$  of a province, the expression is:

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

Regarding to this formula, for a detailed methodology, see the annex of this document.

## Problems

According to the above formula, two problems arise:

- In some large cities, there are important differences between the EPA estimates and thus obtained.
- The order of the municipalities in a province established depending on the number of unemployed is not the same using data from the SEPE than with the estimates.

In view of these problems, the following considerations should be taken into account:

A. The EPA estimates by stratum show three aspects :

1. May have some instability derived from small samples and two-stage sampling (cluster effect).

2. Estimating grouping municipalities into strata, it is possible to change order of unemployment in the municipalities of a province with respect to that provided by the SEPE.
3. Given that the EPA is a continuous survey, the stratum to which a municipality is associated not change until the availability of a new census information. Thus is very likely to find certain outdated stratum for some municipalities

B. It is not obvious that the variable measured by the EPA, unemployed population by place of residence, is the same as that provided by the SEPE, registered in the unemployment register as possibly workplace.

### Proposal

Given that the only variable available to distribute the provincial unemployment EPA among the municipalities, it is from the SEPE, the final proposal focuses on the grouping of strata.

Having into account that in the provincial capitals with major population, the EPA sample is large enough, the following is proposed:

1. In the provinces of Madrid, Barcelona, Sevilla, Zaragoza, Valencia and Malaga, we consider two strata: Capital and other municipalities, and the formula used available:

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

The formula is applied to each of the eight age groups and sex.

2. In the missing provinces, strata are not considered, and therefore it is used the expression:

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

The latter expression ignores strata and simply distributes the unemployment EPA of each age group and sex among municipalities according to data from the SEPE.

The most positive aspects of this method are:

- The method is simple to implement, and in many cases replicable.
- The municipal ranking of unemployed is not altered, except regarding to the six capitals mentioned.
- It is a stable method, and it does not depend on the stratification of the survey.

## Estimation of employment

The employed population is calculated as the difference between the active population and the unemployed population.

### Annex

The number of unemployed population at municipal level will be estimated by means of registered unemployment, data originating from the State Public Employment Service. The basic idea is to consider as the starting point the estimations of unemployment at autonomous community stratum level stemming from the Labour Force Survey (EPA) and by means of these distribute the Labour Force Survey unemployment data by municipality, using registered unemployment as an auxiliary variable. The strata that we use for calibration purposes are

Stratum 1	Municipality, provincial capital
Stratum 2	Self-represented municipalities, important in relation to the capital
Stratum 3	Other self-represented municipalities, important in relation to the capital or municipalities of more than 100,000 inhabitants
Stratum 4	Municipalities with between 50,000 – 100,000 inhabitants
Stratum 5	Municipalities with between 20,000 – 50,000 inhabitants
Stratum 6	Municipalities between 10,000 – 20,000 inhabitants
Stratum 7	Municipalities between 5,000 – 10,000 inhabitants
Stratum 8	Municipalities between 2,000 – 5,000 inhabitants
Stratum 9	Municipalities of less than 2,000 inhabitants

The estimation procedure is as follows. We are going to consider that for a given period of time we have  $N$  unemployed persons registered at the public employment offices nationwide. Of these  $N$  unemployed persons we have information on a set of characteristics, such as gender, age and municipality of residence. We are going to denote all these characteristics with the vector  $X$ . In this way, the characteristics vector  $X_i$  will correspond to the  $i$ -th registered unemployed person and those characteristics may be combined to form different strata, which we are going to refer to as  $\{X_1, X_2, \dots, X_j\}$ . Similarly, we are going to define as  $n_E(X_j)$  the number of unemployed estimated by way of the labour force survey who are assigned to the  $j$ -th stratum. For example,  $n_E(X_1)$  could consist of the number of labour force survey (EPA) unemployed persons of masculine gender between 16-25 years of age, who live in the provincial capitals of a given autonomous

community. Applying the same logic, we are going to define as  $n_p(X_j)$  the number of unemployed persons registered at the public employment offices assigned to the  $j$ -th stratum. Lastly, we denote with  $w(X_j) = \frac{n_E(X_j)}{n_P(X_j)}$  for  $j=1, \dots, J$  stratum.

We propose the following estimator for calculating the total number of unemployed persons in municipality  $m$ :

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

where  $\mathbf{1}(X_j = X_i)$  is a function that takes the value of 1 if the characteristics of the  $i$ -th individual match up with those of the  $j$ -th stratum and 0 if not. In other words, if the individual is of the male sex, is between 16-25 years old, and lives in a given autonomous community (all these characteristics of one of the stratum), then this indicator function will take the value of 1. The function  $\mathbf{1}_i(m)$  takes the value of 1 if the  $i$ -th individual lives in municipality  $m$  and 0 if not.

As may be observed by its construction, this estimator presents useful properties in terms of consistency with the direct estimations made through the Labour Force Survey (EPA). In this way, it may be confirmed that the estimations of unemployment made at municipal level match up with the EPA unemployment figures stratified by gender and age band at both autonomous community and national level.