SMALL-AREA ESTIMATION IN THE SURVEY OF THE POPULATION IN RELATION TO ACTIVITY IN THE A.C. OF THE BASQUE COUNTRY

EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA
BASQUE INSTITUTE OF STASTISTICS

Donostia-San Sebastián, 1
01010 VITORIA-GASTEIZ
Tel.: 945 01 75 00
Fax.: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

Presentation

Conscious of the growing demand for ever more disaggregated quality statistics, Eustat set up a research team in 2003 made up of members of Eustat and the University. The aim was to work on improving estimation techniques in different statistical operations, and to introduce small area estimation techniques based on models in the statistical production. One result of this project was the application of the small area estimation system to the Annual Industrial Statistics, published by Eustat in 2005 in a Technical Handbook.

This estimation methodology has been applied to another statistical operation which is equally relevant within Eustat's statistical production: the Survey of the Population in Relation to Activity (PRA), published for users with quarterly results referring to the labour market within the Autonomous Community of the Basque Country at province level. As with the Industrial Statistics, the estimations are based on models and provide information about the 20 statistical districts into which the Autonomous Community is divided.

The aim of this publication is to provide material of use to all interested users referring to knowledge and usage of methods for small areas.

This document is divided into two different parts. The first one covers the methodology used, together with certain aspects specific to the estimators and the auxiliary information used, and the second part is a presentation of the district-level results corresponding to 2005, 2006 and 2007.

Vitoria-Gasteiz, March 2008

JOSU IRADI ARRIETA

General Director

Index

Chapter
1

Introduction

Official statistics currently have to meet a demand for more and more disaggregated quality information relating to social and economic indicators.

One method of tackling the demand for more disaggregation is to increase the sample size, with the corresponding rise in costs, and to continue applying the design-based estimators currently used in official statistics.

Another alternative currently being researched is to use more complex estimation techniques, as the model-assisted and model-based estimators.

Conscious of this growing demand for ever more disaggregated quality statistics, Eustat set up a research team five years ago made up of members of Eustat and members of the University in order to work on improving estimation techniques in different statistical operations, and to introduce small area model-based estimation techniques.

The project for small area estimation began with a training course given by the University to the personnel of the Institute and the Basque Statistics Organisation. Over a series of sessions they covered both the theory of design-based estimation as well as the model-assisted and model-based estimation.

The first operation undertaken in the project was Industrial Statistics and the results were published in a technical handbook in 2005. That study gave rise to the periodic publication of annual estimations in the 20 statistical districts of the A.C. of the Basque Country on the main economic magnitudes of the survey.

This document aims to disseminate the results of Eustat's second operation undertaken using this methodology, the Survey of the Population in Relation to Activity (henceforward referred to as PRA).

A reference from the European sphere of the use of model-based estimation in official statistics is to be found in the UK's Office for National Statistics (ONS). Their model-based estimations of local authority unemployment figures, taken from the Labour Force Survey (LFS), have recently been accepted as a national statistic. This is the first time that the UK has given this rank to the model-based estimates (CLARKE et al. 2007).

Other small area estimations obtained in the ONS are still considered as experimental statistics, which means that they are still subject to methodological improvements.

In general, there is a move in the international arena towards accepting small area estimations as official statistics, considered as the ones that comply with all the requirements of the Code of Good Practice for official statistics. On one hand, this implies new challenges for research into these methods, and on the other hand, an adequate presentation and explanation of these results to the users.

This document considers various aspects. There are two sections in the theoretical part: the main characteristics of the Survey of the Population in Relation to Activity will first be presented alongside the error and result estimators used (Chapter 2), followed by a presentation of the system of small area estimation applied in the PRA (Chapter 3).

The section on application includes a commentary on the results obtained from said survey, using this methodology for the districts of the A.C. of the Basque Country. Results will be presented for the following concepts: rate of activity, unemployment rate, working and unemployed population – always based on population of 16 and over (Chapter 4). Finally, conclusions will be drawn on the project (Chapter 5) and a Bibliography is included. The Annex details the division of municipalities into districts in the A.C. of the Basque Country.

Chapter 2

Survey of the Population in Relation to Activity (PRA)

2.1 Description of the Survey of the Population in Relation to Activity in the A.C. of the Basque Country

The Survey of the Population in Relation to Activity (PRA) was begun in the 1980s in order to build a rich and detailed source of information about the labour market which would be comparable internationally.

More specifically, the aim of the operation is to produce continuous statistical information on the participation, or not, of the population in labour activities, with special attention being paid to economic activities. This survey produces quarterly and annual results on the volume and characteristics of the main collectives from the point of view of the labour market: the working population, employed and unemployed with their corresponding rates of activity, employment or unemployment.

This information is obtained for the main demographic characteristics across the Autonomous Community, as corresponds to its sampling design which will be described later in this paper.

The reference population in the PRA resides in family households in the Autonomous Community of the Basque Country. The framework for the survey is the Directory of Households and the Statistical Register of Population for the A.C. of the Basque Country. The first complete year of data for the PRA was 1985. The survey has since undergone various changes in the size and design of the sampling, as well as the sample framework and weight treatments.

The quarterly sample of the survey is a rotating panel of households. These households remain on the panel for 8 quarters and are contacted once every quarter. The panel is renewed by changing an eighth of the households each quarter.

The initial sample (the latest to be completed corresponding to the first quarter of 2005) was composed of 12 independent, systematic sub-samples, one for each week of the quarter. In total, this amounted to 5,088 households, distributed throughout the provinces in the following manner: 1,114 in Álava, 2,196 in Bizkaia and 1,690 in Gipuzkoa.

This distribution was carried out in proportion to the square root of the number of households in the provinces in order to reduce the differences in population sizes. To guarantee this distribution, the provinces make up the strata of the sample.

Every quarter the panel is partially renewed. This renewal consists of removing a sample equivalent to one eighth of the quarterly sample, and in the same way, making a systematic sampling of households throughout the strata for inclusion in the panel, and removing any households which has already spent 8 quarters in the survey.

There are two types of unit in the survey: the households which make up the panel and the individuals members of the households. During the fieldwork, information is gathered on all members of the household, normally by means of an informant in each house. The results of the survey are relative to the population and to the household or families.

The move from sample-based data to estimations is made after a process of weighting adjustments or calibration, where weighting is calculated for individuals or families, according to projections for both types.

2.2 Estimators used in the Survey of the Population in Relation to Activity in the A.C. of the Basque Country

2.2.1 Definition of the estimators and weighting formulas

2.2.1.1 Total number of people

In every stratum h determined by each province, calibrating estimators are obtained in two phases:

To begin with the Horvitz-Thompson estimator is applied, based on the calculation of the factor of design or inverse probability in the selection of the household (all households in the stratum has the same probability of being selected). Every member of the household is selected and so the design factor coincides with the household factor. There is no problem with partial responses, which is to say from the inhabitants of the household, and therefore no correction due to lack of response is needed.

$$\hat{X}_h = \sum_{i=1}^{v_h}\sum_{j=1}^{n_{v_i}} w_{hi} X_{hij} = \sum_{i=1}^{v_h} w_{hi} \sum_{j=1}^{n_{v_i}} X_{hij} = w_h \sum_{i=1}^{v_h}\sum_{j=1}^{n_{v_i}} X_{hij} = \frac{V_h}{v_h}\sum_{i=1}^{v_h}\sum_{j=1}^{n_{v_i}} X_{hij}$$

where:

v_h is the number of households in the effective sample from stratum h

n_{v_i} is the number of people sampled within household i

$w_{hi} = w_h$, design factor in household i

V_h is the amount of household in the population sector of stratum h

X_{hij} is the value of the characteristics to be estimated of person j in household i of stratum h

Next, the previous weights are adjusted by calibration.

$$\hat{X}_h^* = \sum_{i=1}^{v_h} \sum_{j=1}^{n_{v_i}} w_{hij}^* X_{hij}$$

with w_{hij}^* obtained from the initial factors $w_{hij} = w_{hi}$ and applying post-stratification according to the crossing of the following variables: province, age group (7 groups: <=15,16-24, 25-34,.., 55-64, >=65) and sex.

Eustat's population projections corresponding to these groupings by age, sex and provinces are used.

2.2.1.2 Total number of families

In each stratum, h determined by province, calibration estimators are obtained in two phases:

To begin with the Horvitz-Thompson estimator is applied, based on the calculation of the factor of design or inverse probability in the selection of the household (all the households in the stratum have the same probability of being selected).

$$\hat{X}_h = \sum_{i=1}^{v_h} w_{hi} X_{hi} = w_h \sum_{i=1}^{v_h} X_{hi} = \frac{V_h}{v_h} \sum_{i=1}^{v_h} X_{hi}$$

where:

v_h is the number of households in the effective sample from stratum h

$w_{hi} = w_h$, design factor of household i

V_h is the amount of household in the population sector of stratum h

X_{hi} is the value of the characteristic to be estimated in household i of stratum h

Next, these weights are adjusted by calibration.

$$\hat{X}_h^* = \sum_{i=1}^{v_h} w_{hi}^* X_{hi}$$

with w_{hi}^* obtained from the initial factors w_{hi} and applying a simultaneous adjustment of the marginal distribution of the following variables: province and family size (5 categories: 1, 2, 3,…, >=5 members).

Additional information about numbers of families by province and family size is obtained from population projections from Eustat, and from average family-size obtained from the Local Authority Register.

Both adjustments are made by means of the SAS (Institut National de la Statistique et des Études Économiques) macro Calmar, using the “raking ratio” method (INSEE 1993).

2.2.2 Method of estimation of sampling errors

Taylor’s Expansion Method is used. This permits the calculation of estimates of sampling error for totals, averages and ratios in samples with stratification, clusters and unequal probabilities. The method obtains linear approximations from the estimator and calculates the variance using this as an estimation of the sampling variance.

The expression used to calculate the estimated variance for population average is as follows:

$$\bar{V}\left(\hat{\bar{Y}}\right)=\sum_{h=1}^{H}\frac{n_h\left(1-f_h\right)}{n_h-1}\sum_{i=1}^{n_h}\left(e_{hi\cdot}-\bar{e}_{h\cdot\cdot}\right)^2 \tag{2}$$

Where:

$$e_{hi\cdot}=\left(\sum_{j=1}^{m_{hi}}w_{hij}\left(y_{hij}-\hat{\bar{Y}}\right)\right)\Big/w_{\cdots}$$

$$\bar{e}_{h\cdot\cdot}=\left(\sum_{i=1}^{n_h}e_{hi\cdot}\right)\ /\ n_h$$

and

$$w_{\cdots}=\sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j=1}^{m_{hi}}w_{hij}$$

<u>NB</u>:

$h = 1, 2, \ldots, H$ indicates stratum with a total of H strata.
$i = 1, 2, \ldots n_h$ indicates the number of clusters in stratum h, with a total of n_h clusters.
$j = 1, 2, \ldots, m_{hi}$ indicates the number of units within cluster i of stratum h, with a total of m_{hi} units.

$$n=\sum_{h=1}^{H}\sum_{i=1}^{n_h}m_{hi}$$

is the total number of observations in the sample.

w_{hij} indicates the weight of observation j in cluster i of stratum h

$Y_{hij} = (Y_{hij}(1), Y_{hij}(2), \ldots, Y_{hij}(P))$ are the observed values of variable Y in observation j of cluster i of stratum h. (numerical and category variables).

This calculation is made following the PROC SURVEYMEANS procedure from the SAS (Sas Institute Inc. 2004) statistics packet.

Chapter 3

Small-Area Estimation System in the PRA

3.1 Simulation Study

The methodology of estimation has been established using a simulation study as a base, carried out on population information available from the latest census. The output of various estimators was analysed, both classic and model-assisted and model-based used to estimate unemployment figures, by sex, in 20 and 40 districts in the A.C. of the Basque Country.

To do this, the mean squared error of the estimators was evaluated, calculated using simulation based on the 2001 Census on Population and Households (henceforward referred to as the Census or CPV) by means of the same sampling procedure used in the PRA from 2005 onwards.

500 samples have been obtained through simulation since the 2001 Census, using the PRA design.

3.1.1 Evaluation Indicators.

In order to evaluate the margin and root-mean-square error of the estimators, based on 500 simulated samples from the 2001 Census, the following indicators were calculated:

Absolute relative bias (ARB):

$$ARB_d(\hat{y}) = \frac{1}{K}\left|\sum_{k=1}^{K}\frac{\hat{y}_d(k)-Y_d}{Y_d}\right|100 \quad \text{with d as the zone or small area}$$

And its average: $ARBM(\hat{y}) = \frac{1}{D}\sum_d ARB_d(\hat{y})$

Square root of the relative mean squared error (RMSE):

$$RMSE_d(\hat{y}) = \left(\frac{1}{K}\sum_{k=1}^{K}\left(\frac{\hat{y}_d(k)-Y_d}{Y_d}\right)^2\right)^{\frac{1}{2}}100$$

And its average: $RMSEM(\hat{y}) = \frac{1}{D}\sum_d RMSE_d(\hat{y})$

The following were evaluated:

- Design-based estimators
- Model-assisted estimators
- Model-based estimators

3.1.2 Design-based estimators.

- Direct:

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

where $y_j = 1$ (unemployed) $yj = 0$ (not unemployed)

N_d number of people aged > 16 in zone d

n_d sample size in zone d

d is the small area(zone)

w_j design weight

- Post-stratification

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

where $\hat{\bar{y}}_{dg}$ is the average calculated by means of the aforementioned direct estimator,

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

- Synthetic

$$\hat{y}_d^{\sin t} = \sum_g \hat{\bar{y}}_g N_{dg}$$

Where $\hat{\bar{y}}_g$ is the average calculated by means of the aforementioned direct estimator

- Composite 1 - Composite 4

$$\hat{y}_d^{dep} = \lambda_d \hat{y}_d^{post} + (1 - \lambda_d)\hat{y}_d^{\sin t}$$

where $0 \leq \lambda_d \leq 1$ is represented by

$$\lambda_d = \begin{cases} 1 \ \text{ si } \ \hat{N}_d \geq \alpha N_d \\ \dfrac{\hat{N}_d}{\alpha N_d} \quad \text{n another case} \end{cases}$$

$\hat{N}_d = \sum_d w_j$ is the total estimated population in each area d and α is a parameter. The composite estimator is evaluated using different values of $\alpha = \frac{2}{3}, 1, 1.5 \text{ and } 2$.

The above estimators are calculated for the following groups g:

- Age group and sex. (8 categories)
- Age group and sex *Level of education (24 categories)
- Age group and sex * Relation with activity in the AC of the Basque Country 1996 (24 categories)
- Age group and sex * Level of education * Relation with activity in the AC of the Basque Country 1996 (72 categories)

3.1.3 Model- assisted estimators.

GREG estimator assisted by a linear model

The probability of being unemployed is assumed as part of the model and, to this end, the values obtained must be within the $[0,1]$ range. This is not monitored on a linear model, so if a negative prediction is obtained, it becomes 0 and if the prediction is greater than 1 it becomes 1.

The binary variables y_{id} indicate if the i th individual of area d is unemployed or not.

The linear model is as follows:

$$p_{id} = x_{id}^T \beta + \varepsilon_{id}$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = (x_{id,1}, x_{id,2}, \ldots, x_{id,p})^T$ is the vector of p co-variables where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

$\varepsilon_{id} \approx N(0, \sigma^2)$ are independent random variables.

The estimator of the total in each area appears as:

$$\hat{Y}_d^{GREG} = N_d\left(\frac{1}{\hat{N}_d}\sum_{i\in s_d} w_{id}\, y_{id} + \left(\bar{X}_d - \frac{1}{\hat{N}_d}\sum_{i\in s_d} w_{id}\, x_{id}\right)^T \hat{\beta}\right) = N_d\left(\bar{X}_d\hat{\beta} + \frac{1}{\hat{N}_d}\sum_{i\in s_d} w_{id}\left(y_{id} - x_{id}^T\hat{\beta}\right)\right)$$

where $\bar{X}_d = \left(\bar{X}_{d1}, \bar{X}_{d2}, \ldots, \bar{X}_{dp}\right)^T$ is the vector of population averages of the p co-variables and $\hat{\beta} = \left(\sum_{i\in s} w_i x_i x_i^T\right)^{-1} \sum_{i\in s} w_i x_i y_i$

NB: A GREG estimator assisted by a linear model with a fixed effect of the district produces the same response as the estimator based on a weighted linear model with a fixed effect of the district. For this reason the fixed district is not included here as an explicative value

GREG estimator assisted by a logit model.

The binary variables y_{id} indicate if the i th individual of area d is unemployed or not.

The logit model is as follows:

$$\log it(p_{id}) = \log\left(\frac{p_{id}}{1-p_{id}}\right) = x_{id}^T\beta$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = \left(x_{id,1}, x_{id,2}, \ldots, x_{id,p}\right)^T$ is the vector of p co-variables where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

The estimator of the total in each area appears as:

$$\hat{Y}_d^{GREG} = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T\hat{\beta}}}{1+e^{x_{id}^T\hat{\beta}}} + \frac{N_d}{\hat{N}_d}\sum_{i\in s_d} w_{id}\left(y_{id} - \frac{e^{x_{id}^T\hat{\beta}}}{1+e^{x_{id}^T\hat{\beta}}}\right)$$

where $\hat{\beta}$ is calculated with weights w_i.

NB: A GREG estimator assisted by a logit model with a fixed effect of the district produces the same response as the estimator based on a weighted logit model with a fixed effect of the district. For this reason the fixed district is not included here as an explicative value

GREG estimator assisted by a mixed logit model.

The binary variables y_{id} indicate if the i th individual of area d is unemployed or

not.

The logit model with the random effect of the area is as follows:

$$\log\left(\frac{p_{id}}{1-p_{id}}\right) = x_{id}^T\beta + u_d$$

where:

y_{id} verifies that $y_{id}\left|u_d \approx Binomial(1, p_{id})\right.$

p_{id} is the probability that the i th individual of area d is unemployed

u_d is the random effect of the area, $u_d \approx N\left(0, \sigma_u^2\right)$

$x_{id} = \left(x_{id,1}, x_{id,2}, \ldots, x_{id,p}\right)^T$ is the vector of co-variables, where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

The estimator of the total in each area appears as:

$$\hat{Y}_d^{GREGMixto} = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T\hat{\beta}+\hat{u}_d}}{1+e^{x_{id}^T\hat{\beta}+\hat{u}_d}} + \frac{N_d}{\hat{N}_d}\sum_{i\in s_d} w_{id}\left(y_{id} - \frac{e^{x_{id}^T\hat{\beta}+\hat{u}_d}}{1+e^{x_{id}^T\hat{\beta}+\hat{u}_d}}\right)$$

where $\hat{\beta}$ is calculated without weight.

On occasions, the variance component is estimated as zero, in fact when it is not estimated as zero it is not always statistically significant.

3.1.4 Model-based estimators.

3.1.4.1 Estimators based on a linear model.

When a linear model is attempted the prediction obtained is not monitored to check whether it is within the [0,1] range. What is assumed as part of the model is the probability of being unemployed and so the values obtained must be between zero and one. If a negative prediction is obtained, it becomes 0, and if the prediction is greater than 1 it becomes 1.

***Synthetic estimator based on a linear model*.**

The binary variables y_{id} indicate if the i th individual of area d is unemployed or not.

The model is as follows:

$$p_{id} = x_{id}^T \beta + \varepsilon_{id}$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = \left(x_{id,1}, x_{id,2}, \ldots, x_{id,p}\right)^T$ is the vector of p co-variables where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

$\varepsilon_{id} \approx N\left(0, \sigma^2\right)$ are independent random variables.

The estimator (projected) of the total in each area is represented by:

$$\hat{P}_d = X_d \hat{\beta}$$

where $X_d = \left(X_{d1}, X_{d2}, \ldots, X_{dp}\right)^T$ is the vector of the total population of the p co-variables, and

$$\hat{\beta} = \left(\sum_{i \in s} x_i x_i^T\right)^{-1} \sum_{i \in s} x_i y_i$$

***Synthetic estimator based on a weighted linear model*.**

The synthetic estimator based on a linear model with weights produces the same response as the previous estimator described above, the difference being the estimator of $\hat{\beta}$. This is calculated with weights w_i.

$$\hat{\beta} = \left(\sum_{i \in s} w_i x_i x_i^T\right)^{-1} \sum_{i \in s} w_i x_i y_i$$

***Estimator based on a linear model with a fixed effect of the area*.**

The binary variables y_{id} indicate if the i th individual of area d is unemployed or not.

The model is as follows:

$$p_{id} = x_{id}^T \beta + \varepsilon_{id}$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = \left(x_{id,1}, x_{id,2}, \ldots, x_{id,p}\right)^T$ is the vector of p co-variables where $x_{id,1}$ is the area included as fixed effect and may also include any of the co-variables mentioned in the previous section.

The estimator (projected) of the total in each area is represented by:

$$\hat{P}_d = \sum_{i=1}^{N_d} \hat{p}_{id} = X_d \hat{\beta}$$

where $X_d = \left(X_{d1}, X_{d2}, \ldots, X_{dp}\right)^T$ is the vector of the total populations of the p co-variables, and

$$\hat{\beta} = \left(\sum_{i \in s} x_i x_i^T\right)^{-1} \sum_{i \in s} x_i y_i$$

Estimator based on a linear model with weights and a fixed effect of the area.

The synthetic estimator that is based on a linear model with weights and a fixed effect of the area produces the same response as the previous estimator described above, the difference being the estimator of $\hat{\beta}$. This is calculated with weights w_i.

$$\hat{\beta} = \left(\sum_{i \in s} w_i x_i x_i^T\right)^{-1} \sum_{i \in s} w_i x_i y_i$$

Synthetic estimator based on a linear model with a random effect of the area

Various simulations were carried out and in some it was impossible to adjust a linear model with uncertain area effect. Therefore, the possibility of using this model has been abandoned.

3.1.4.2 Estimators based on a logit model.

Synthetic estimator based on a logit model.

The binary variables y_{id} indicate whether the i th individual in area d is unemployed or not.

The logit model is as follows:

$$\log it(p_{id}) = \log\left(\frac{p_{id}}{1-p_{id}}\right) = x_{id}^T \beta$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = (x_{id,1}, x_{id,2}, \ldots, x_{id,p})^T$ is the vector of p co-variables where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

The estimator of the total in each area is represented by:

$$\hat{P}_d = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T \hat{\beta}}}{1+e^{x_{id}^T \hat{\beta}}}$$

Synthetic estimator based on a weighted logit model

The synthetic estimator based on a logit model with weights produces the same response as the previous estimator described above, the difference being the estimator of $\hat{\beta}$. This is calculated with weights w_i.

***Estimator based on a logit model with the area as fixed effect*.**

The binary variables y_{id} indicate whether the i th individual in area d is unemployed or not.

The logit model is as follows:

$$\log it(p_{id}) = \log\left(\frac{p_{id}}{1-p_{id}}\right) = x_{id}^T \beta$$

where:

p_{id} is the probability that the i th individual of area d is unemployed.

$x_{id} = (x_{id,1}, x_{id,2}, \ldots, x_{id,p})^T$ is the vector of p co-variables where $x_{id,1}$ is the area included as fixed effect and may also include any of the co-variables mentioned in the previous section.

The estimator of the total in each area is represented by:

$$\hat{P}_d = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T \hat{\beta}}}{1+e^{x_{id}^T \hat{\beta}}}$$

Estimator based on a logit model with weights and the area as fixed effect.

The synthetic estimator based on a logit model with weights and the area as fixed effect produces the same response as the previous estimator described above, the difference being the estimator of $\hat{\beta}$. This is calculated with weights w_i.

Estimator based on a logit model with the area as random effect (EB Mixed Logit).

The binary variables y_{id} indicate whether the i th individual in area d is unemployed or not.

The logit model with the area as random effect is as follows:

$$\log it(p_{id}) = \log\left(\frac{p_{id}}{1-p_{id}}\right) = x_{id}^T \beta + u_d$$

where:

y_{id} verifies that $y_{id} | u_d \approx Binomial(1, p_{id})$

p_{id} is the probability that the i th individual in area d in unemployed

u_d is the random effect of the area, $u_d \approx N(0, \sigma_u^2)$

$x_{id} = (x_{id,1}, x_{id,2}, ..., x_{id,p})^T$ is the vector of p co-variables, where said co-variables $x_{id,k}$ may be the following:

Age group and sex

Level of education

Relation with activity in the AC of the Basque Country 1996

The estimator (EB Empiric-Bayesian) of the total in each area is:

$$\hat{P}_d = \sum_{i=1}^{N_d} \frac{e^{x_{id}^T \hat{\beta} + \hat{u}_d}}{1+e^{x_{id}^T \hat{\beta} + \hat{u}_d}}$$

3.1.5 Conclusions

Once the study was completed using all of the aforementioned estimators, the most appropriate estimator for districts appears to be, in quarters of the evaluation of described indicators, composite 4, which includes the following additional variables: sex and age, level of education and the relation with the activity in the previous period.

The results obtained for 40 and for 20 districts do not differ greatly, as the most sensitive zones are classified in both sectors.

3.2 Calculation of estimations in real PRA samples

The selected estimator, composite 4, has been applied to real PRA samples from every quarter.

In order to do this, it was necessary to have all of the additional information about every one of the units in the sample. In order to obtain the relation with activity relative to the previous period – information which was available in this case from the Census on Population and Households (henceforward referred to as the Census) of 2001 – it was necessary to merge the PRA registers with those of the Census. At this stage it was not possible to correctly identify approximately 12% of the sample.

Other information needed for estimation of districts includes population projections made at this level according to the additional variables used: age group, sex, level of education and relation with activity relative to the previous period. Population sizes were calculated using the 2001 Census, and calibrated to population projections by age group, sex and province, calculated quarterly by Eustat.

The estimations by district obtained for each of the estimators were again calibrated to the estimations made by the direct estimator for each province. In this way, estimations were obtained which had been calibrated to data already published. Additionally, and in order to obtain more stable estimations over a period of time, moving averages have been calculated every four months.

3.3 Estimation of the mean squared error

3.3.1 Procedures in calculating the mean squared error (MSE)

A simulation study has also been made to analyse the behaviour of three methods of estimation for the corresponding mean squared errors. The linearization method and the following re-sampling methods: the Jackknife method and the Bootstrap method. The behaviour of the MSE estimator was studied using the three methods. In order to do this, it was investigated whether said MSE estimators produced results similar to those results obtained in the simulation study.

The three methods were evaluated with reference to composite 4 and the post-stratification and synthetic estimators.

The linearization method consists of the application of a Taylor development in series.

The re-sampling methods were based on the evaluation of statistics from re-sampling or sub-sampling obtained from original data, and by means of these values it was possible to obtain estimators of the sampling distribution of the statistics.

In the case of the Jackknife method, we dispose as many sub-samples as clusters has the original sample, obtained through the successive elimination of clusters from the original sample. New weights were defined for each sub-sample and the estimator calculated (post-stratification, synthetic or composite). The variance and bias of the estimators were then obtained, as detailed below.

In the Bootstrap method the sub-samples are obtained by means of simple random sampling but it is important to determine the necessary number of samples required. Similarly, new weights are defined for each sub-sample and each estimator calculated. The mean squared error, described below, was obtained with these estimations.

The following indices are used:

- h is the stratum number, where $h = 1,2,..,H$
- i is the i th cluster in stratum h where $i = 1,2,..,n_h$
- j is the j th unit of cluster i in stratum h, where $j = 1,2,..,m_{hi}$

In the PRA, stratum h is the province; cluster i is the household and j is the j th person in the household.

3.3.1.1 Variance linearization method.

The linearization or delta method consists of the application of a Taylor development in series.

The following indicator variables are defined for each dominium D:

$$I_D(h,i,j) = \begin{cases} 1 & \text{if } (h,i,j) is \text{ in } D \\ 0 & \text{any other case} \end{cases}$$

$$z_{hij} = y_{hij} I_D(h,i,j) = \begin{cases} y_{hij} & \text{if } (h,i,j) \text{ is in } D \\ 0 & \text{any other case} \end{cases}$$

$$v_{hij} = w_{hij} I_D(h,i,j) = \begin{cases} w_{hij} & \text{if } (h,i,j) \text{ is in } D \\ 0 & \text{any other case} \end{cases}$$

The estimator of the average in dominium D is referred to as:

$$\hat{\bar{Y}}_D = \left(\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} v_{hij} z_{hij} \right) \Bigg/ v... \text{, where } v... = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} v_{hij}$$

The linearization variance of the estimator of the average in dominium D is referred to as:

$$Var_L\left(\hat{\bar{Y}}_D\right) = \sum_{h=1}^{H} Var_h\left(\hat{\bar{Y}}_D\right) \quad \text{where} \quad Var_h\left(\hat{\bar{Y}}_D\right) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(U_{hi.} - \bar{U}_{h..}\right)^2$$

$$U_{hi.} = \frac{1}{v...} \sum_{j=1}^{m_{hi}} v_{hij} \left(z_{hij} - \hat{\bar{Y}}_D\right) \quad \text{and} \quad \bar{U}_{h..} = \frac{1}{n_h} \sum_{i=1}^{n_h} U_{hi.}$$

3.3.1.2 Jackknife Method for estimating variance and bias.

In order to apply the Jackknife method to the scheme of sampling used in the PRA a cluster (household) must be eliminated each time. New weights are defined, represented as follows:

$$w_{j(hi)} = \begin{cases} w_{hij} & \text{if unit j is not in stratum h} \\ 0 & \text{if unit j is in cluster i of stratum h} \\ \dfrac{n_h}{n_h - 1} w_{hij} & \text{if unit j is in stratum h but not in cluster i} \end{cases}$$

Then :

- $\hat{\theta}$ the composite 4 estimator obtained from data from a simulation using weights w_{hij}
- $\hat{\theta}_{(hi)}$ the composite 4 estimator obtained from data from the sub-sample resulting from the elimination of cluster i (household) from stratum h (th) in said simulation, and using weights $w_{j(hi)}$

The jackknife variance estimator in stratum h is referred to as:

$$Var_{JK(h)}\left(\hat{\theta}\right) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left[\hat{\theta}_{(hi)} - \hat{\theta}_{(h.)}\right]^2$$

where:

$$\hat{\theta}_{(h.)} = \frac{1}{n_h}\sum_{i=1}^{n_h}\hat{\theta}_{(hi)}$$

The Jackknife estimator of the bias of an estimator in stratum h is referred to as:

$$Bias_{JK(h)}\left(\hat{\theta}\right) = \left(n_h - 1\right)\left(\hat{\theta}_{(h.)} - \hat{\theta}\right)$$

The Jackknife estimator of the MSE in stratum h:

$$M\hat{S}E_{JK(h)}\left(\hat{\theta}\right) = V\hat{a}r_{JK(h)}\left(\hat{\theta}\right) + Bias^2_{JK(h)}\left(\hat{\theta}\right)$$

As the strata are independent the MSE of the estimator is referred to as:

$$M\hat{S}E_{JK}\left(\hat{\theta}\right) = \sum_{h=1}^{H}\left[\frac{n_h - 1}{n_h}\sum_{i=1}^{n_h}\left[\hat{\theta}_{(hi)} - \hat{\theta}_{(h.)}\right]^2 + \left(\left(n_h - 1\right)\left(\hat{\theta}_{(h.)} - \hat{\theta}\right)\right)^2\right]$$

The MSE has also been calculated directly as follows:

$$M\hat{S}E_{JK}\left(\hat{\theta}\right) = \sum_{h=1}^{H}\left[\frac{n_h - 1}{n_h}\sum_{i=1}^{n_h}\left[\hat{\theta}_{(hi)} - \hat{\theta}\ \right]^2\right]$$

No appreciable differences have been observed.

3.3.1.3 Bootstrap estimation of the mean squared error

There follows a description of the steps to take to construct a version of the re-scaled Bootstrap in a simple stratified sampling proposed by Rao and Wu (1988).

1. Having fixed stratum h, we have a sample with n_h clusters. We take a sub-sample with $n_h - 1$ clusters by means of simple random sampling with replacement from the sample from stratum h. The process is repeated independently for each stratum

2. We construct a new weight for each sub-sample r (r=1,2,..R): $w_{hij}\left(r\right) = w_{hij}\dfrac{n_h}{n_h - 1}m_i\left(r\right)$ where $m_i\left(r\right)$ is the number of times cluster i is selected in the sub-sample, and we calculate $\hat{\theta}^*_r$ using the new weights $w_{hij}\left(r\right)$.

3. We repeat steps 1 and 2, R times.

4. In order to obtain the Bootstrap estimator for the mean squared error, we carry out the following:: $M\hat{S}E_B\left(\hat{\theta}\right) = \dfrac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\theta}^*_r - \hat{\theta}\right)^2$ where:

- $\hat{\theta}$ is the composite 4 estimator obtained from data from a simulation using weights w_{hij}

- $\hat{\theta}_{(hi)}$ is the composite 4 estimator obtained from data from sub-sample r using weights $w_{hij}(r)$.

One of the questions to respond to is the size of R so that the method works correctly. In order to do this, 5 simulations with the post-stratification, synthetic and composite 4 estimators were carried out, using age group, sex and relation to activity in the previous period as additional variables. Different values of R were considered, specifically R took

the values of 200, 500, 1000 and 2000. It was observed that there were no differences in behavior according to the size of R. After having seen the results, it was decided that R =200.

3.3.2 Comparison of the results obtained with those obtained from the simulation.

The behaviour of the estimator of the MSE was studied by means of the three methods. Said estimators of the MSE were analysed to see whether they produced similar results to those obtained in the simulation study.

The evaluation indicator used in the simulation to evaluate the mean squared error is referred to as:

$$RMSE_d(\hat{y}_d) = \left(\frac{1}{K} \sum_{k=1}^{K} \left(\frac{\hat{y}_d(k) - Y_d}{Y_d} \right)^2 \right)^{\frac{1}{2}} 100$$

This RMSE indicator (square root of the relative mean squared error) was obtained from 500 simulated samples taken from the 2001 Census. Using this indicator a new indicator was defined which approximates the square root of the mean squared error $MSE\_S_d(\hat{y}_d)$ (square root of the mean squared error obtained by simulation). This is referred to as:

$$MSE\_S_d(\hat{y}_d) = \frac{Y_d}{100} RMSE_d(\hat{y}_d)$$

This error is an approximation of Monte Carlo to the true error.

In order to evaluate the behavior of any of the MSE estimation methods, the MSE of the estimator is calculated using the same 500 samples obtained through simulation.

3.3.3 Application of MSE estimation methods to real PRA samples. Calibration.

Working with each of the real samples from the PRA, it was possible to obtain estimations by district for each one of the estimators. These estimations are calibrated to the estimations given by the direct estimator in each province which Eustat publishes quarterly.

The calculation methods for the MSE of the calibrated estimations are:

Bootstrap: to obtain the Bootstrap estimation of the mean squared error, the following was applied: $M\hat{S}E_B\left(\hat{y}_d\right)=\dfrac{1}{R-1}\sum_{r=1}^{R}\left(\hat{y}^*_{d(r)}-\hat{y}_d\right)^2$ where:

- $\hat{y}_d$ is the estimator obtained from data from a sample of a specific quarter, calibrating the estimations to values given by the direct estimator by province (th).

- $\hat{y}^*_{d(r)}$ is the estimator obtained from data from a sub-sample r with weights $w_{hij}(r)$, calibrating the estimations to values given by the direct estimator by th if the data from this sub-sample r were used.

- **Jackknife**: in order to obtain the Jacknife estimation of the mean squared error the following should be applied:

$$M\hat{S}E_{JK}\left(\hat{y}_d\right)=\sum_{h=1}^{H}\left[\frac{n_h-1}{n_h}\sum_{i=1}^{n_h}\left[\hat{y}_{d(hi)}-\hat{y}_{d(h.)}\right]^2+\left(\left(n_h-1\right)\left(\hat{y}_{d(h.)}-\hat{y}_d\right)\right)^2\right]$$

where:

- $\hat{y}_d$ is the estimator obtained from data from a sample of a specific quarter, calibrating the estimations to values given by the direct estimator by th.

- $\hat{y}^*_{d(r)}$ is the estimator obtained from data from the sub-sample resulting from the elimination of cluster i (household) from stratum h (th) with weights $w_{hij}(r)$, calibrating the estimations to values given by the direct estimator by th if the data from this sub-sample r were used.

- $\hat{y}_{d(h.)}=\dfrac{1}{n_h}\sum_{i=1}^{n_h}\hat{y}_{d(hi)}$

- **Linearization:** The estimation's MSE is applied without being calibrated.

3.3.4 Conclusions

The simulation studies carried out demonstrate that the Bootstrap estimator is the most adequate estimator for calculating the mean squared error of the composite 4 estimator.

3.4 Software used

In the study of this methodology and the applications of the estimators described above, a computer program based on SAS was used. Specific macro programs have been written which execute the following different tasks: making estimations (number of unemployed, employed, and activity, employment and unemployment rates) by districts, and the calculation of mean squared errors for the different methods.

The macro provides estimations calculated by means of the composite estimator (the alpha parameter is an entry parameter) and, as this is a combination of a post-stratified estimator and a synthetic estimator, it also provides estimations calculated using these estimators.

Other entry parameters of this macro are: the variable to be estimated, the auxiliary variables to be used, the calibration option of the district estimations to the provincial estimations obtained by means of direct estimation from the survey, the mean squared error estimation method (and, in the case of the Bootstrap method, the value of R).

The program also offers the possibility of obtaining mobile averages for several consecutive quarters. In addition, it calculates the mean squared errors and the corresponding variation coefficients, using the three methods described here.

This macro is applied quarterly to the samples from the PRA and produces, within the optimum parameters determined by the simulation study, estimations of the aforementioned magnitudes and corresponding mean squared errors.

Chapter 4

District-level estimations 2005-2007

4.1 Definitions

This section presents the estimations obtained by using the estimation system described in this paper in the Survey of the Population in Relation to Activity (PRA), for 2005-2007.

The magnitudes, always referring to the population of 16 and over, chosen for publication were the following: (The definitions from the survey have been included here)

- Activity rate: proportion forming part of the active population. Normally expressed in percentages. In the PRA it is calculated on the population over 16.

$$Activity\_Rate = \frac{\sum activepopulation}{\sum population \quad over16} \times 100$$

- Unemployment rate: the proportion of active population currently unemployed. Normally expressed in percentages.

$$Unemployment\_Rate = \frac{\sum unemployedpopulation}{\sum activepopulation} \times 100$$

- Employed population: Refers to all those citizens who have paid work or exercise an independent activity and are working or else maintain a formal link with their job but are temporarily absent because of holidays, illness, labour conflicts, technical incidents, etc.

- Unemployed population: Refers to all those citizens who do not have paid work or independent work and are currently looking for work and are available to work.

Since 2002, and according to the European Commission Regulation 1897/2000, the unemployed are all those citizens who not only fulfil the aforementioned conditions, but that in the 4 previous weeks have been actively involved in looking for work in any of the types considered by said Regulation. The act of renewing one's search for work ("stamping the dole card") or contacting the employment office looking for information about training courses is not deemed to be active involvement in looking for work.

In addition to the estimations, this document includes the tables of variation coefficients of the same.

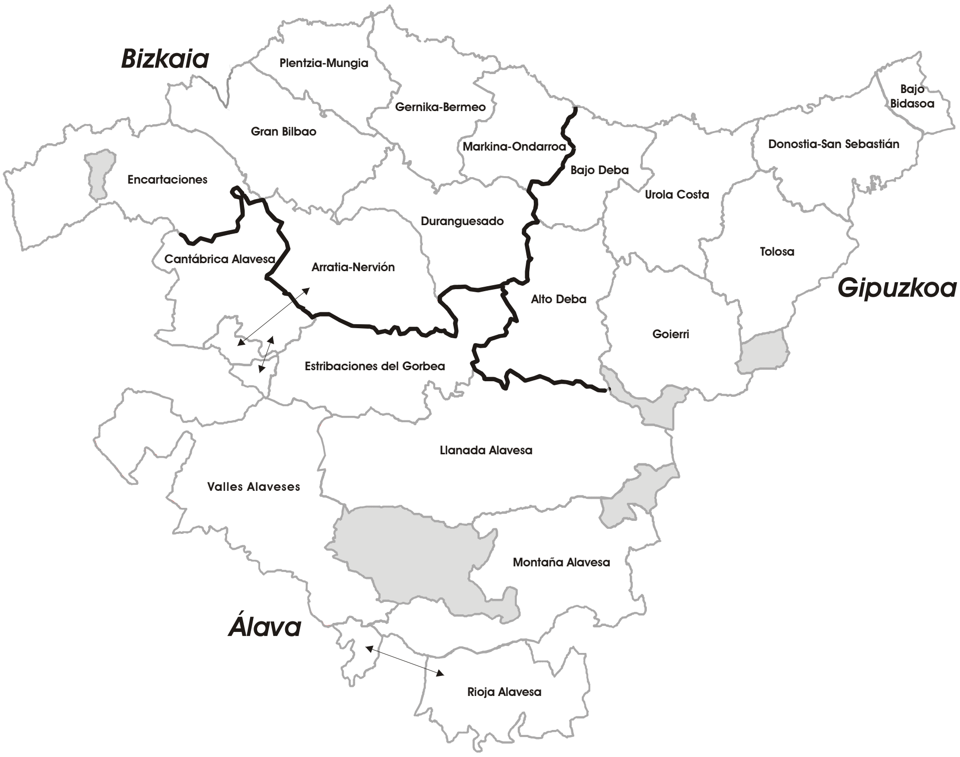
The official division of the A.C. of the Basque Country into districts is as follows:

Alava: Valles Alaveses, Llanada Alavesa, Montaña Alavesa, Rioja Alavesa, Estribaciones del Gorbea and Cantábrica Alavesa:

Bizkaia: Arratia-Nervión, Gran Bilbao, Duranguesado, Encartaciones, Gernika-Bermeo, Markina-Ondarroa and Plentzia-Mungia

Gipuzkoa: Bajo Bidasoa, Bajo Deba, Alto Deba, Donostia-San Sebastián, Goierri, Tolosa and Urola Costa

(See Annex for districts and municipalities)



The most important results are set out below.

4.2 Results

Table 1. Activity rate of the populatin of 16 and above by province and district. Estimation and Coefficients of Variation (in percentages).

Source: EUSTAT. Survey of the Population in Relation to Activity (PRA).

	2005		2006		2007	
	Estimation	CV	Estimation	CV	Estimation	CV
A.C. of the Basque Country	**54,8**	**0,35**	**54,6**	**0,37**	**54,6**	**0,35**
Alava	**57,0**	**0,68**	**57,6**	**0,68**	**56,8**	**0,66**
Valles Alaveses	45,0	1,79	45,6	1,89	54,0	1,70
Llanada Alavesa	58,6	0,92	59,2	0,88	58,2	0,87
Montaña Alavesa	41,0	2,32	35,0	2,84	42,2	3,39
Rioja Alavesa	54,7	2,15	55,0	2,36	50,5	2,20
Estribaciones del Gorbea	51,4	1,41	49,7	2,02	51,7	2,71
Cantábrica Alavesa	48,8	1,34	49,6	1,47	49,8	1,49
Bizkaia	**53,1**	**0,54**	**52,6**	**0,55**	**52,9**	**0,56**
Arratia-Nervión	43,6	1,91	44,5	1,94	47,5	1,86
Gran Bilbao	53,1	0,74	52,6	0,74	52,8	0,76
Duranguesado	55,6	1,13	54,2	1,07	54,7	1,43
Encartaciones	44,8	1,55	43,7	1,92	46,6	1,84
Gernika-Bermeo	49,0	1,28	48,8	1,37	49,1	1,25
Markina-Ondarroa	45,7	1,54	46,6	1,38	47,9	1,81
Plentzia-Mungia	59,1	1,18	57,6	1,47	58,0	1,35
Gipuzkoa	**56,7**	**0,58**	**56,6**	**0,58**	**56,7**	**0,59**
Bajo Bidasoa	55,9	1,13	55,7	1,09	57,5	0,98
Bajo Deba	49,4	1,30	49,3	1,31	49,2	1,17
Alto Deba	57,6	1,11	57,0	1,15	53,1	1,14
Donostialdea	58,8	0,83	58,4	0,88	59,0	0,88
Goierri	52,4	1,35	52,8	1,38	51,7	1,24
Tolosa	52,5	1,16	54,4	1,22	55,2	1,36
Urola Costa	55,8	1,12	57,3	1,07	57,4	1,08

Table 2. Unemployment rate of the population of 16 and above by province and district. Estimation and Coefficients of Variation (in percentages).

Source: EUSTAT. Survey of the Population in Relation to Activity (PRA).

	2005		2006		2007	
	Estimation	CV	Estimation	CV	Estimation	CV
A.C. of the Basque Country	**5,7**	**2,30**	**4,1**	**2,63**	**3,3**	**2,90**
Alava	**3,0**	**5,70**	**3,5**	**5,39**	**2,3**	**6,90**
Valles Alaveses	2,8	8,70	3,4	9,06	2,1	9,10
Llanada Alavesa	3,0	7,60	3,5	7,26	2,2	9,50
Montaña Alavesa	3,1	10,00	5,8	31,37	2,9	18,60
Rioja Alavesa	2,3	10,10	3,0	16,82	2,9	20,50
Estribaciones del Gorbea	3,0	7,50	3,1	13,38	1,9	11,60
Cantábrica Alavesa	3,6	9,90	4,1	10,86	2,5	14,20
Bizkaia	**7,4**	**2,90**	**5,0**	**3,55**	**4,0**	**4,00**
Arratia-Nervión	7,2	14,50	6,5	20,38	3,5	17,80
Gran Bilbao	7,8	4,00	5,2	4,86	4,2	5,40
Duranguesado	5,6	7,30	3,4	8,77	2,6	9,20
Encartaciones	5,4	8,10	4,6	17,51	3,7	18,50
Gernika-Bermeo	7,5	10,90	4,1	12,00	4,0	13,70
Markina-Ondarroa	4,8	8,70	3,2	11,54	2,4	9,40
Plentzia-Mungia	5,8	7,90	4,6	12,05	3,6	13,50
Gipuzkoa	**4,2**	**4,10**	**2,9**	**5,05**	**2,6**	**5,30**
Bajo Bidasoa	4,4	7,70	2,8	9,44	3,0	11,00
Bajo Deba	4,2	9,40	3,3	12,95	2,6	11,80
Alto Deba	3,1	8,10	2,2	8,78	2,0	9,10
Donostialdea	4,7	6,10	3,2	7,58	2,7	8,30
Goierri	3,7	9,00	2,5	10,18	2,6	12,00
Tolosa	4,1	9,70	2,4	8,73	2,2	13,00
Urola Costa	3,8	7,80	2,9	9,40	2,3	10,40

Table 3. Working population of 16 and above by province and district. Estimation (in thousands) and Coefficients of Variation (in percentages).

Source: EUSTAT. Survey of the Population in Relation to Activity (PRA).

	2005		2006		2007	
	Estimation	**CV**	**Estimation**	**CV**	**Estimation**	**CV**
A.C. of the Basque Country	**941,2**	**0,78**	**954,2**	**0,78**	**964,8**	**0,77**
Alava	**141,1**	**1,53**	**142,4**	**1,54**	**143,5**	**1,55**
Valles Alaveses	1,9	2,24	1,9	2,76	2,4	1,99
Llanada Alavesa	116,1	1,16	117,5	1,17	117,8	1,18
Montaña Alavesa	1,2	3,05	1,0	4,19	1,2	3,55
Rioja Alavesa	4,8	3,22	4,8	3,44	4,5	2,57
Estribaciones del Gorbea	3,0	1,92	3,0	2,45	3,1	2,68
Cantábrica Alavesa	13,8	1,68	14,1	1,80	14,5	1,89
Bizkaia	**484,7**	**1,17**	**491,9**	**1,17**	**500,0**	**1,16**
Arratia-Nervión	7,6	2,24	7,8	2,41	8,6	2,05
Gran Bilbao	375,2	0,95	382,0	0,94	387,0	0,94
Duranguesado	41,4	1,36	41,2	1,28	41,9	1,63
Encartaciones	11,2	2,03	11,0	2,44	11,9	2,26
Gernika-Bermeo	17,7	1,64	18,3	1,82	18,5	1,61
Markina-Ondarroa	10,0	2,05	10,4	2,19	10,8	2,36
Plentzia-Mungia	21,4	1,59	21,1	1,75	21,5	1,59
Gipuzkoa	**315,5**	**1,28**	**319,8**	**1,28**	**321,3**	**1,26**
Bajo Bidasoa	33,0	1,36	33,4	1,23	34,4	1,28
Bajo Deba	22,7	1,51	22,9	1,46	23,0	1,27
Alto Deba	30,3	1,35	30,2	1,37	28,2	1,29
Donostialdea	152,2	1,11	153,5	1,12	155,9	1,07
Goierri	27,8	1,76	28,3	1,66	27,7	1,46
Tolosa	19,2	1,62	20,2	1,61	20,5	1,53
Urola Costa	30,1	1,37	31,2	1,41	31,4	1,49

Table 4. Unemployed population of 16 and above by province and district. Estimation (in thousands) and Coefficients of Variation (in percentages).

Source: EUSTAT. Survey of the Population in Relation to Activity (PRA).

	2005		2006		2007	
	Estimation	**CV**	**Estimation**	**CV**	**Estimation**	**CV**
A.C. of the Basque Country	**57,0**	**3,01**	**40,5**	**3,53**	**32,5**	**3,88**
Alava	**4,4**	**7,76**	**5,2**	**7,17**	**3,4**	**9,03**
Valles Alaveses	0,1	7,83	0,1	8,82	0,1	8,97
Llanada Alavesa	3,6	8,08	4,3	7,48	2,7	9,40
Montaña Alavesa	0,0	9,86	0,1	28,88	0,0	17,93
Rioja Alavesa	0,1	8,44	0,1	18,12	0,1	20,89
Estribaciones del Gorbea	0,1	7,45	0,1	12,25	0,1	11,51
Cantábrica Alavesa	0,5	10,36	0,6	12,23	0,4	15,11
Bizkaia	**38,6**	**3,86**	**25,7**	**4,75**	**20,7**	**5,17**
Arratia-Nervión	0,6	15,22	0,5	20,36	0,3	17,83
Gran Bilbao	31,6	4,01	21,0	5,02	17,0	5,49
Duranguesado	2,5	8,66	1,4	9,68	1,1	8,59
Encartaciones	0,6	8,79	0,5	17,59	0,5	20,58
Gernika-Bermeo	1,4	11,60	0,8	12,17	0,8	15,39
Markina-Ondarroa	0,5	9,05	0,3	10,86	0,3	8,15
Plentzia-Mungia	1,3	8,93	1,0	14,25	0,8	14,81
Gipuzkoa	**14,0**	**5,53**	**9,6**	**6,72**	**8,5**	**7,01**
Bajo Bidasoa	1,5	9,10	1,0	10,34	1,1	11,58
Bajo Deba	1,0	10,48	0,8	13,80	0,6	13,25
Alto Deba	1,0	8,75	0,7	9,46	0,6	9,30
Donostialdea	7,4	6,65	5,0	8,03	4,3	8,63
Goierri	1,1	9,91	0,7	11,04	0,7	13,51
Tolosa	0,8	10,25	0,5	8,80	0,5	14,73
Urola Costa	1,2	8,90	0,9	10,46	0,7	11,70

Chapter 5

Conclusions

The growing demand for disaggregated information and the need to not overload the informants has made for progressively greater use in official statistics of model based estimation methods .

The fact of obtaining estimations of the magnitudes relative to activity in small areas such as districts, as presented in this paper, is a step forward in the application of new methodologies for model based estimation in the Institute.

The results presented in this document offer an acceptable level of quality in terms of accuracy. The majority of the coefficients of variation (CV) obtained in the estimations was less than 15%, with only a few passing the 20% mark.

From now on, Eustat is able to offer district estimations based on a short-term survey which will improve the operation's efficiency.

The estimations will be able to be improved insofar as better auxiliary information is available. The availability of adequate auxiliary information is fundamental to the models and it is therefore important to be able to count on adequate frameworks and to have access to information from administrative files.

Eustat aims to continue improving in the study and application of model based estimation methodology, in order to be able to offer quality, disaggregated information.

Bibliography

CLARKE, PHILIP; MCGRATH, KEVIN; HUKUM, CHANDRA AND TZAVIDIS, NIKOS (2007)

Developments in Small Area Estimation in UK with focus on current research activities. IASS Satellite Meeting on Small Area Estimation

EUSTAT (2005)

Report on the Calculation of Sampling Errors. Survey of the Population in Relation to Activity. (PRA). http://www.eustat.es/document/datos/Calculo_errores_PRA_i.pdf

EUSTAT (2007)

Proyecto Técnico de la Operación Encuesta de Población en Relación con la Actividad. (PRA).

GHOSH, M. AND RAO, J.N.K.,(1994)

Small Area Estimation: An Appraisal. Statistical Science, 9, 55-93.

GHOSH, N. AND SÄRNDAL, C.E. (2001)

Lecture Notes for Estimation for Population Domains and Small Areas. Statistics Finland ., vol. 48.

INSEE INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES (1993)

"La macro Calmar, Redressement d'un échantillon par calage sur marges", Document nº F9310 25/11/1993, Olivier Sautory. Série des documents de travail de la Direction des Statistiques Démographiques et Sociales.. Insee - La macro SAS Calmar.

QUENOUILLE, M. (1949)

Approximate tests of correlation in time series. J. Roy Statist. Soc. Ser. B, 11, 18-84.

QUENOUILLE, M. (1956)

Notes on bias in estimation. Biometrika, 43 pp. 353-360.

RAO, J.N.K. AND WU, C.F.J. (1988)

Resampling Inference with Complex Survey Data. Journal of the American Statistical Association , 83, 231-241

SÄRNDAL, C.E. SWENSSON, B. AND WRETMAN J. (1992)

Model Assisted Survey Sampling. Springer-Verlag

SAS INSTITUTE INC., "SAS/STAT® 9. (2004)

"User's Guide". Copyright © 2004, Cary, NC, USA. ISBN

TUKEY, J. (1958)

Bias and confidence in not quite large samples. Abstract, Ann. Math. Statist.., 29, 614

WOODRUFF, R.S., (1971)

A Simple Method for Approximating the Variance of a Complicated Estimate. Journal of The American Statistical Association. 66(334), 411-414

Annex

ALAVA/ARABA

Arabako Ibarrak / Valles Alaveses: Añana, Armiñón, Berantevilla, Kuartango, Lantarón, Ribera Alta, Ribera Baja/Erribera Beitia, Valdegovía/Gaubea, Zambrana

Arabako Lautada / Llanada Alavesa: Alegría-Dulantzi, Arrazua-Ubarrundia Asparrena, Barrundia, Elburgo/Burgelu, Iruña Oka/Iruña de Oca, Iruraiz-Gauna, Salvatierra/Agurain, San Millán/Donemiliaga, Vitoria-Gasteiz, Zalduondo

Arabako Mendialdea / Montaña Alavesa: Arraia-Maeztu, Bernedo, Campezo/Kanpezu, Harana/Valle de Arana, Lagrán, Peñacerrada-Urizaharra

Errioxa Arabarra / Rioja Alavesa: Baños de Ebro/Mañueta, Elciego, Elvillar/Bilar, Kripan, Labastida/Bastida, Laguardia, Lanciego/Lantziego, Lapuebla de Labarca, Leza, Moreda de Álava, Navaridas, Oyón-Oion, Samaniego, Villabuena de Alava/Eskuernaga, Yécora/Iekora

Gorbeia Inguruak / Estribaciones del Gorbea: Aramaio, Legutiano, Urkabustaiz, Zigoitia, Zuia

Kantauri Arabarra / Cantábrica Alavesa: Amurrio, Artziniega, Ayala/Aiara, Laudio/Llodio, Okondo

BIZKAIA

Arratia Nerbioi / Arratia-Nervión: Arakaldo, Arantzazu, Areatza, Arrankudiaga, Artea, Dima, Igorre, Orozko, Otxandio, Ubide, Ugao-Miraballes, Urduña-Orduña, Zeanuri, Zeberio

Bilbo Handia / Gran Bilbao: Abanto y Ciérvana-Abanto Zierbena, Alonsotegi, Arrigorriaga, Barakaldo, Basauri, Berango, Bilbao, Derio, Erandio, Etxebarri, Galdakao, Getxo, Larrabetzu, Leioa, Lezama, Loiu, Muskiz, Ortuella, Portugalete, Santurtzi, Sestao, Sondika, Valle de Trápaga-Trapagaran, Zamudio, Zaratamo, Zierbena

Durangaldea / Duranguesado: Abadiño, Amorebieta-Etxano, Atxondo, Bedia, Berriz, Durango, Elorrio, Ermua, Garai, Iurreta, Izurtza, Lemoa, Mallabia, Mañaria, Zaldibar

Enkartazioak / Encartaciones: Artzentales, Balmaseda, Galdames, Gordexola, Güeñes, Karrantza Harana/Valle de Carranza, Lanestosa, Sopuerta, Trucios-Turtzioz, Zalla

Gernika-Bermeo: Ajangiz, Arratzu, Bermeo, Busturia, Ea, Elantxobe, Ereño, Errigoiti, Forua, Gautegiz Arteaga, Gernika-Lumo, Ibarrangelu, Kortezubi, Mendata, Morga, Mundaka, Murueta, Muxika, Nabarniz, Sukarrieta

Markina-Ondarroa: Amoroto, Aulesti, Berriatua, Etxebarria, Gizaburuaga, Ispaster, Lekeitio, Markina-Xemein, Mendexa, Munitibar-Arbatzegi Gerrikaitz-, Ondarroa, Ziortza-Bolibar

Plentzia-Mungia: Arrieta, Bakio, Barrika, Fruiz, Gamiz-Fika, Gatika, Gorliz, Laukiz, Lemoiz, Maruri-Jatabe, Meñaka, Mungia, Plentzia, Sopelana, Urduliz

GIPUZKOA

Bidasoa Beherea / Bajo Bidasoa: Hondarribia, Irun

Deba Beherea / Bajo Deba: Deba, Eibar, Elgoibar, Mendaro, Mutriku, Soraluze-Placencia de las Armas

Deba Garaia / Alto Deba: Antzuola, Aretxabaleta, Arrasate/Mondragón, Bergara, Elgeta, Eskoriatza, Leintz-Gatzaga, Oñati

Donostialdea / Donostia-San Sebastián: Andoain, Astigarraga, Donostia-San Sebastián, Errenteria, Hernani, Lasarte-Oria, Lezo, Oiartzun, Pasaia, Urnieta, Usurbil

Goierri: Altzaga, Arama, Ataun, Beasain, Ezkio-Itsaso, Gabiria, Gaintza, Idiazabal, Itsasondo, Lazkao, Legazpi, Mutiloa, Olaberria, Ordizia, Ormaiztegi, Segura, Urretxu, Zaldibia, Zegama, Zerain, Zumarraga

Tolosaldea / Tolosa: Abaltzisketa, Aduna, Albiztur, Alegia, Alkiza, Altzo, Amezketa, Anoeta, Asteasu, Baliarrain, Belauntza, Berastegi, Berrobi, Bidegoian, Elduain, Gaztelu, Hernialde, Ibarra, Ikaztegieta, Irura, Larraul, Leaburu, Legorreta, Lizartza, Orendain, Orexa, Tolosa, Villabona, Zizurkil

Urola-Kostaldea / Urola Costa: Aia, Aizarnazabal, Azkoitia, Azpeitia, Beizama, Errezil, Getaria, Orio, Zarautz, Zestoa, Zumaia