SMALL-AREA ESTIMATION IN THE SURVEY ON THE INFORMATION SOCIETY - FAMILIES OF THE BASQUE COUNTRY

EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

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

In 2003, Eustat, aware of the growing demand for increasingly disaggregated quality statistics, formed a research team made up of members of Eustat and the University. The objective was to work on the improvement of estimation techniques in various statistical operations and to introduce small-area estimation techniques based on statistical production models. This work resulted in the application of the small-area estimation system to the annually-produced Industrial Statistics, published by Eustat in a Technical Notebook in 2005 and to the Survey on the Population in Relation to Activity, published by Eustat in a Technical Notebook in 2008.

This estimation methodology has been applied to another statistical operation which is equally relevant within Eustat's statistical production, the Survey on the Information Society - Families, which offers users annual results on the access and use of the Internet, as well as other areas of information technology in the Basque Country at Province level. The estimations based on small-area methods provide information on the 20 statistical districts into which the Basque Country 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, 2007 and 2008.

Vitoria-Gasteiz, May 2009

JOSU IRADI ARRIETA

General Director

Index

PRESENTATION 3

INDEX 3

INTRODUCTION 4

THE SURVEY ON THE INFORMATION SOCIETY - FAMILIES (ESIF) 5

2.1 DESCRIPTION OF THE SURVEY ON THE INFORMATION SOCIETY - FAMILIES OF THE BASQUE COUNTRY 5

2.2 ESTIMATORS USED IN THE SURVEY ON THE INFORMATION SOCIETY - FAMILIES (ESIF) OF THE BASQUE COUNTRY 6

SMALL-AREA ESTIMATION SYSTEM IN THE ESIF 9

3.1 STUDY OF ESTIMATORS 9

3.2 ESTIMATION OF THE MEAN SQUARED ERROR 13

3.3 SOFTWARE USED 18

DISTRICT-LEVEL ESTIMATIONS 2005-2008 19

4.1 DEFINITIONS 19

4.2 RESULTS 21

CONCLUSIONS 26

BIBLIOGRAPHY 27

ANNEX 29

Chapter 1

Introduction

Official statistics currently have to meet a demand for increasingly disaggregated quality information on the main social and economic indicators.

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

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

The aim of this document is to disseminate the results of the third operation undertaken using this methodology in EUSTAT, the Survey on the Information Society - Families (hereafter ESIf).

A reference in the European sphere for the use of model-based estimation in official statistics can be found in the United Kingdom Office for National Statistics, (ONS). There, the model-based estimations for local authority unemployment figures, taken from the Labour Force Survey (LFS), have recently been accepted as a national statistic. It is the first time that model-based estimations have been given this status in the UK (CLARKE et al, 2007).

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

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

This document will present various aspects. On the theoretical side, there are two parts: firstly, the main characteristics of the Survey on the Information Society will be presented, with the estimators of results and errors which are used (chapter 2), followed by a presentation of the small-area estimation system applied to the ESIf (chapter 3).

The section on application includes a commentary on the results obtained from the aforementioned survey, using this methodology, for the districts of the Basque Country. Results are shown for the following concepts: Internet users, in all cases for the group aged 15 and over (chapter 4) and divided by sex. Finally, the conclusions on the project will be drawn (chapter 5) and the Bibliography is given. The Annex details the division of municipalities into districts in the Basque Country.

Chapter 2

The Survey on the Information Society - families (ESIf)

2.1 Description of the Survey on the Information Society - Families of the Basque Country

The Survey on the Information Society - Families (ESIf) got underway in 2000, with the aim of making rich and detailed information available for the growing demand from the public for greater degrees of information. The rise of different access channels, in which the Internet occupies pride of place, with ample opportunities for future development, drives Eustat's involvement in this area.

More specifically, the aim of the operation is to produce continuous statistical information on Internet access and use, as well as other types of information technology. Access to information is joined to the very concept of welfare itself and sooner or later will be considered a Fundamental Right.

This information is obtained for the main characteristics at Province level, as corresponds to its sample design, which will be commentated on later.

The reference population of the ESIf is that resident in family dwellings in the Basque Country. The sample frame is the Directory of Housing and the Statistical Population Register of the Basque Country. The first year for which the ESIf has complete data is 2000. Since then, the survey has undergone some changes in the sample size and design, as well as in the sample frame and the treatments of weighting. The data of the ESIf is obtained using a specific questionnaire and is carried out by sampling on the population of the Basque Country. It is a survey linked to the sample and collection processes of the Eustat Survey on the Population in Relation to Activity (PRA), which has been carried out since 1985. At present, two surveys are carried out per year: one in the 2nd quarter and another in the 4th quarter.

This distribution of dwellings is carried out proportionally to the square root of the number of dwellings of the Provinces to reduce the differences between them in population size. To guarantee this distribution, the Provinces form the sample strata.

In each household of the sample the selection of the first person is made randomly using Kish tables (based on the survey week and the number of individuals, students and employed persons respectively in the household); additionally, in the event of students and/or employed persons residing in the dwelling, one of each group is selected by the same procedure. Since 2003 the sample has been completed with all the children aged between 6 and 14 resident in the family unit.

The step from the sample data to the estimations is taken following a process of weighting or calibration. In this process weights are calculated for individuals and families according to the projections of both types.

2.2 Estimators used in the Survey on the Information Society - Families (ESIf) of the Basque Country

2.2.1 Definition of the estimators and weighting formulas

To estimate the characteristics of the survey the following estimator is considered and the formulas used in the calculation are detailed:

2.2.1.1 Estimator for people aged over 15

In each stratum h (Province), calibrated estimators are also obtained in two phases:

Phase 1: Horvitz Thompson Estimator, based on the calculation of design factor or inverse probability in the selection of each person. There is no lack of partial response, which is to say from people within the household, and so there is no correction for nonresponse.

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

where:

n_{v_i} is the number of people sampled in household *i*.

$w_{hij} = w_{hi} p^*_{hij}$, design factor of person *j* from household *i*, where w_{hi} is the design factor of household *i*

$p^*_{hij} = e_{hi}$ if person *j* is a student.

$= o_{hi}$ if person *j* is employed.

$= p_{hi}$ in another case.

where:

e_{hi} number of students in household *i*.

o_{hi} number of employed people in household *i*.

p_{hi} number of sampleable people in household *i.*

X_{hij} is the value of the characteristic to be estimated of person *j* of household *i* of stratum *h*.

Phase 2: Calibration adjustment.

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

with w_{hij}^* obtained from the initial factors w_{hij} applying post-stratification according to the variable cross of province, age group (6 groups: 15-24, 25-34, 35-44, 45-54, 55-64, >=65) and sex. This is the projected population by Province, age group and sex with reference to day 15 of the central month of the quarter.

2.2.2 Method of estimation of sampling errors

The Taylor Expansion Method was used. It allows sampling error estimations to be calculated for totals, means and ratios in samples with stratification, clustering and unequal probabilities. The method obtains linear approximations of the estimator and calculates its variance using this as the estimation of the sampling variance.

The expression for the calculation of the variance estimated for the population mean is as follows:

(2)

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

Where:

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

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

y

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

Notes:

$h = 1, 2, \ldots, H$ indicates the 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 the unit 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 weighting of observation j in the cluster i of stratum h

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

This calculation was made using the PROC SURVEYMEANS procedure from the SAS statistical package (Sas Institute Inc. 2004).

Chapter 3

Small-Area Estimation System in the ESIf

3.1 Study of estimators

The estimation methodology was established using the analysis of various estimators, both classic and model-assisted and model-based to estimate the percentage of Internet users aged over 15, by sex, in 20 districts of the Basque Country and in the 3 capitals of the Provinces.

To do so, the mean squared error of the calculated estimators was evaluated.

The following were evaluated:

- Design-based estimators
- Model-based estimators

3.1.1 Design-based estimators.

3.1.1.1 Direct:

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

where $y_j = 1$ (Internet user) $yj = 0$ (non-user)

N_d number of people aged > 15 in zone d.

n_d sample size in zone d

d is the small area(zone)

$\tilde{w}_j$ calibrated weight using the design weight

3.1.1.2 Post-stratified

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

where

N_{dg} number of people aged > 15 in zone d and in group g.

$\hat{\bar{y}}_{dg}$ is the mean calculated with the previous direct estimator, $\hat{\bar{y}}_{dg} = \frac{\sum_{j \in s_{dg}} \tilde{w}_j y_j}{\sum_{j \in s_{dg}} \tilde{w}_j}$

3.1.1.3 Synthetic

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

where

N_{dg} number of people aged > 15 in zone d and in group g.

$\hat{\bar{y}}_g$ is the mean calculated with the previous direct estimator

3.1.1.4 Composite

$\hat{y}_d^{dep} = \lambda_d \hat{y}_d^{post} + (1-\lambda_d)\hat{y}_d^{\sin t}$ where $\hat{y}_d^{post}$ is the post-stratified estimator and $\hat{y}_d^{\sin t}$ is the synthetic estimator calculated with the group mean of the total in the area calculated with the direct estimator.

and where $0 \le \lambda_d \le 1$ is given by

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

$\hat{N}_d = \sum_d w_j$ is the population total estimated in each area d and α is a parameter. The composite estimator is evaluated for different values of $\alpha = 2/3, 1, 1.5$ y 2. Hereafter composite1 ($\alpha = 2/3$), composite2 ($\alpha = 1$), composite3 ($\alpha = 1.5$) and composite4 ($\alpha = 2$).

The previous estimators are calculated for the following groups g:

- Age group and sex. (12 categories, 6 by sex: 15-24, 25-34, 35-44, 45-54, 55-64 and ≥ 65)
- Age group and sex (12 categories, 6 by sex: 15-24, 25-34, 35-44, 45-54, 55-64 and ≥ 65) * Education level regrouped (2 categories: “Higher-middle and Higher” and “Secondary and Vocational and Primary or Less”).

3.1.2 Model-based estimators.

3.1.2.1 Estimators based on a mixed linear model

When a linear model is proposed the prediction obtained is not monitored to check that it is within the interval [0,1]. The probability of being an Internet user is being modelled, and so the values obtained should be between zero and one. In the event of obtaining a negative prediction it becomes 0 and in the event of it being greater than 1 it becomes 1.

A mixed linear model disaggregated by sex was studied, which is to say one model for women and another for men, with random area effect. The auxiliary information used was age groups and sex (6 categories by sex).

$$ydj = \beta 1xdj1 + \beta 2xdj2 + \beta 3xdj3 + \beta 4xdj4 + \beta 5xdj5 + \beta 6xdj6 + vd + edj$$

with $d = 1, \ldots, Ds$, y $j = 1, \ldots, nd$,

Where:

ydj is the variable of interest for individual j of district d,

$xdj1, \ldots, xdj6$ values of the auxiliary variable (categories of age-sex),

$\beta 1, \ldots, \beta 6$ are the fixed effects of the model,

vd common random effect for all the individuals of district d,

edj are the specific random errors of each individual.

The super-population model in matrix form would be the following:

$$Y = X\beta + Zv + \varepsilon, \quad v \sim N(0, \sigma v2ID), \quad \varepsilon \sim N(0, \sigma e2),$$

Once the projective version of the model with the sampled and the non-sampled part is developed, the model predictors take the following form:

The mean: $\hat{\bar{y}}_d^* = \bar{X}'_{d(p)}\hat{\beta} + \hat{\gamma}_d(\bar{y}_d - \bar{x}'_d\hat{\beta})$ with d=1,…,D and where $\hat{\beta} = \tilde{\beta}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ has been evaluated with the estimations of the variance components $\bar{X}'_{d(p)} = (1, \bar{x}_{d(p)})$ and $\bar{x}_{d(p)} = \dfrac{\sum_{j \in N_d} x_{dj}}{N_d}$ is the population mean of the auxiliary variable for a district d (sampled or otherwise).

The total: $\hat{t}_d^* = X'_{d(p)}\hat{\beta} + N_d\hat{\gamma}_d(\bar{y}_d - \bar{x}'_d\hat{\beta})$ with d=1,…,D where $X'_{d(p)} = (N_d, X_{d(p)})$, N_d is the total number of individuals in district d. If this

model-based version of the total predictor is developed we arrive at a composite estimator as follows:

$$\hat{\bar{y}}_d = \hat{\gamma}_d[\bar{y}_d + (\bar{X}_{d(p)} - \bar{x}_d)'\hat{\beta}] + (1-\gamma_d)\bar{X}'_{d(p)}\hat{\beta} \text{ with } d\text{=1,...,D } \hat{\gamma}_d = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \frac{\hat{\sigma}_e^2}{n_d}}$$

It is a composite estimator where the first part is the weighted sum of the generalised regression estimator and the second part is a synthetic model-based estimator. In this case, $\hat{\gamma}_d$ (which is the equivalent of λ_d for the composite) depends on the relative variance $\hat{\sigma}_v^2$, the total variance $\hat{\sigma}_v^2 + \hat{\sigma}_e^2$ and the size of the sample in area n_d.

This calculation is made with the PROC MIXED procedure from the SAS statistics package (Sas Institute Inc. 2004). By default this procedure uses the restricted maximum likelihood method (REML) for the estimation of the variance components $\sigma^2 = (\sigma_e^2, \sigma_v^2)$

3.1.3 Conclusions

Once the study had been made with all the design-based estimators, the most suitable estimator by district turned out to be composite 4, which includes sex and age as auxiliary variables. Composite 4, with α=2, is, among the four composites, the one which gives most weight to the synthetic part, which is to say the part taken from the other domains in the form of proportions of Internet users. It is the estimator with the least error and the most stable in its estimations.

The results obtained for 20 districts and the 3 capitals did not differ considerably between the Model and Composite 4 and were very similar. In the largest districts, the model-based estimation is slightly larger than that of the composite and in the other districts is slightly smaller than that of the composite. The estimations given by the model are more unstable in the districts with a smaller sample. The estimations of the Composite 4 estimator are more stable even in the districts where the sample is not well-distributed among the population.

As regards errors, the errors of both estimators are low but are not comparable since they were calculated using different methods.

Having analysed the stability and errors of the different estimations, the decision was taken to use the composite 4 estimator.

3.2 Estimation of the mean squared error

3.2.1 Procedures for calculating the mean squared error (MSE)

Three estimation methods of the corresponding mean squared errors were used, the variance linearization method and the following resampling methods: the jacknife method and the bootstrap method. The performance of the MSE estimator was studied using the three methods.

The 3 methods were evaluated for the composite estimators and the post-stratified and synthetic estimators.

Resampling methods are based on the evaluation of the statistics in resamples or sub-samples obtained from the original data and using these values estimators are obtained of the measurement of accuracy or of the sampling distribution of the statistics.

In the case of the jacknife method as many sub-samples are available as there are clusters in the sample, since they are obtained by successive eliminations of clusters in the original sample. For each sub-sample new weights are defined and the estimator (post-stratified, synthetic or composite) is calculated. Subsequently the variance and bias of the estimators are obtained as detailed later.

In the bootstrap method, the sub-samples are obtained using simple random sampling, but it must be determined how many are necessary. In a similar way, for each sub-sample new weights are defined and each estimator is calculated. With these estimations the mean squared error is obtained as detailed later.

Below 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 $j = 1,2,..,m_{hi}$

It was considered that stratum h is the province (th), cluster i is the household (identified by the variable nt1_idev) and finally j is the surveyed person in said household.

The errors of the estimator based on the Mixed linear model are calculated based on the Prasad and Rao formula (1990), which provides an estimator of the mean squared error (MSE) of the predictor of the mean in the projective version, valid when the estimators of the variance components have been obtained by REML or by the method of moments.

3.2.1.1 Variance linearization method.

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

The following indicator variables are defined for each domain D:

$$I_D(h,i,j)=\begin{cases}1 & \text{si } (h,i,j) \text{ está en } D\\ 0 & \text{en otro caso}\end{cases}$$

$$z_{hij}=y_{hij}I_D(h,i,j)=\begin{cases}y_{hij} & \text{si } (h,i,j) \text{ está en } D\\ 0 & \text{en otro caso}\end{cases}$$

$$v_{hij}=w_{hij}I_D(h,i,j)=\begin{cases}w_{hij} & \text{si } (h,i,j) \text{ está en } D\\ 0 & \text{en otro caso}\end{cases}$$

The estimator of the mean in domain D is given by the expression:

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

The linearized variance of the estimator of the mean in domain D is given by:

$$\hat{Var}_L\left(\hat{\bar{Y}}_D\right)=\sum_{h=1}^{H}\hat{Var}_h\left(\hat{\bar{Y}}_D\right) \quad \text{where} \quad \hat{Var}_h\left(\hat{\bar{Y}}_D\right)=\frac{n_h}{n_h-1}\sum\nolimits_{i=1}^{n_h}\left(U_{hi.}-\overline{U}_{h..}\right)^2$$

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

3.2.1.2 Jacknife method for the estimation of variance and bias.

To apply the jacknife method to the sampling scheme used in the ESIf, one cluster (household) must be eliminated each time. New weights are defined, given by:

$$w_{j(hi)}=\begin{cases}w_{hij} & \text{si la unidad j no está en el estrato h}\\ 0 & \text{si la unidad j está en el cluster i del estrato h}\\ \dfrac{n_h}{n_h-1}w_{hij} & \text{si la unidad j está en el estrato h pero no en el cluster i}\end{cases}$$

Then:

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

The jacknife variance estimator in stratum h is given by:

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

Where:

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

The jacknife estimator of the bias of an estimator in stratum h is given by:

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

The jacknife estimator of the MSE in stratum h:

$$\hat{MSE}_{JK(h)}\left(\hat{\theta}\right)=\hat{Var}_{JK(h)}\left(\hat{\theta}\right)+\hat{Sesgo}^2_{JK(h)}\left(\hat{\theta}\right)$$

As the strata are independent, the MSE of the estimator is given by:

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

3.2.1.3 Bootstrap estimation of the mean squared error.

There follows a description of the steps to be taken in constructing the version of the rescaled 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 extract a sub-sample with n_h-1 clusters by simple random sampling with replacement of

the sample from stratum h. This process is repeated independently for each stratum.

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

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

4. To obtain the bootstrap estimator of the mean squared error we carry out the following: $\hat{MSE}_B(\hat{\theta}) = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\theta}_r^* - \hat{\theta})^2$ where:

- $\hat{\theta}$ is the composite 4 estimator obtained with the data of a simulation using weights w_{hij}.
- $\hat{\theta}_r^*$ is the composite 4 estimator obtained with the data of sub-sample r using weights $w_{hij}(r)$.

One of the issues to be decided is the size of R so that the method works correctly. Different values of R were considered, from a certain size of R stabilizes. In the light of the results, it was decided that R=200.

3.2.1.4 Estimation of the mean squared error for the Model.

The mean squared error is calculated as the sum of the approximations of the contributions of the estimation of the fixed and random effects of the model.

Prasad and Rao (1990) offer an estimator of the mean squared error (MSE) of the predictor of the mean in the projective version, valid when the estimators of the variance components were obtained by REML or by the method of moments. It is given by:

$\hat{MSE}[\hat{\bar{y}}_{d(p)}] = g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2) + 2g_{3d}(\hat{\sigma}^2)$ and as the third term is very small, representing the error due to the estimation of the variance components, it can be discounted and in practice the following is calculated:

$$\hat{MSE}[\hat{\bar{y}}_{d(p)}] \approx g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2)$$

Where:

$$g_{2d}(\hat{\sigma}^2) = (\bar{X}_{d(p)} - \hat{\gamma}_d\bar{x}_d)'\hat{\Phi}_s(\bar{X}_{d(p)} - \hat{\gamma}_d\bar{x}_d)$$

$$g_{1d}(\hat{\sigma}^2) = (1 - \hat{\gamma}_{dc})\hat{\sigma}_v^2$$

$$\hat{\Phi}_s = (X'_s \hat{V}_s^{-1} X_s)^{-1}$$

The MSE of the predictor of the total by district is estimated by multiplying the estimator of the MSE of the mean by the square of the population size of district N_d^2. Effectively

$$\hat{MSE}[\hat{t}_d] \approx N_d^2[g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2)]$$

$g_{1d}(\hat{\sigma}^2)$ represents the error committed in the estimation (assuming all the parameters, the β and the variances are known), which is to say the error on estimating the random effects.

$g_{2d}(\hat{\sigma}^2)$ represents the error committed due to the estimation of the β.

$g_{3d}(\hat{\sigma}^2)$ represents the estimation of the other parameters (estimation of the variance parameters) and diminishes as the number of small areas grows (When the number of areas of the study is small, it is not recommendable to use mixed random effect models because the errors increase).

The greatest contribution to error is usually the estimation of fixed effects g2, followed by the estimation of random effects g1.

3.2.2 Conclusions

The analyses carried out show that the Bootstrap estimator is the most suitable for calculating the mean squared error of the composite estimators and the post-stratified and synthetic estimators. The variance linearization method and the jacknife method for estimating the variance and bias show very similar errors to Bootstrap. Computationally, Bootstrap is the most efficient.

In the estimation of the error of the model, the coefficient $g_{3d}(\hat{\sigma}^2)$ the second order coefficient was calculated extraordinarily and effectively, it did not make a significant contribution to the error and its calculation complicated the process too much.

Both the errors of the Composite 4 estimator and those of the Mixed linear model with random effects are low. These errors obtained with such different methods cannot be compared. One is design-based and the other is model-based.

3.3 Software used

In the study of this methodology and the application of the aforementioned estimators a computer programme based on SAS was used. Specific macro programmes were written which execute the different tasks outlined: production of estimations by district and calculation of the mean squared errors for the different methods.

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

Other entry parameters of this macro are: the variable to be estimated, the auxiliary variables to be used, the option of calibrating the district-level estimations to the provincial ones through the direct estimation of the survey, the mean squared error estimation method (and in the case of bootstrap, the value of R, the number of sub-samples).

The programme also offers the possibility of obtaining moving averages from various consecutive quarters. The mean squared errors and their corresponding variation coefficients (quotient of the square root of the mean squared error and the estimation) are also calculated using the three aforementioned methods.

This macro is applied quarterly to the ESIf samples, obtaining, with the parameters deemed as the best, the estimations of the aforementioned figures and their mean squared errors.

Chapter 4

District-level estimations 2005-2008

4.1 Definitions

There follows a presentation of the estimations obtained using the aforementioned estimation system in the Survey on the Information Society - Families (ESIf), for 2005-2008.

The estimations refer to the percentage of Internet users aged 15 or over who have connected to the Internet, whether in the home, the workplace, the place of study or another place, in the 20 districts of the Basque Country and in the capitals of the Provinces, by sex.

Together with the estimations, the tables of their variation coefficients (VC) are also offered.

The official division 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 the list of municipalities within districts)

The most important results are detailed below.

4.2 Results

Table 1. Internet user population aged 15 and over by sex, Province and distric (%). 2005-2008

Source: EUSTAT. Survey on the Information Society - Familias (ESIF)

	2005			2006			2007			2008		
	Estimations			Estimations			Estimations			Estimations		
	Total	Men	Women	Total	Men	Women	Total	Men	Women	Total	Men	Women
Basque Country	**37,8**	**41,6**	**34,2**	**41,3**	**44,6**	**38,2**	**46,7**	**49,7**	**43,8**	**49,9**	**54,0**	**46,0**
Alava	**38,3**	**39,5**	**37,2**	**40,7**	**42,1**	**39,5**	**45,6**	**47,9**	**43,3**	**50,6**	**54,1**	**47,0**
Valles Alaveses	35,2	37,1	32,9	35,0	36,6	33,0	42,1	41,9	42,4	45,0	46,2	43,4
Llanada Alavesa	39,0	40,4	37,7	41,5	43,0	40,1	46,4	48,7	44,2	51,5	55,1	48,0
Montaña Alavesa	29,6	31,9	26,6	30,0	32,3	27,2	35,7	40,2	29,8	34,9	42,3	24,8
Rioja Alavesa	35,9	37,7	33,9	37,4	36,7	38,1	40,4	42,2	38,5	44,6	48,8	39,9
Estribaciones del Gorbea	36,1	37,7	34,4	39,4	41,0	37,6	44,1	44,5	43,6	47,9	51,3	44,3
Cantábrica Alavesa	35,8	35,5	36,2	38,6	39,8	37,4	43,4	47,3	39,6	48,6	52,5	44,9
Bizkaia	**36,1**	**41,1**	**31,5**	**40,7**	**45,1**	**36,6**	**47,1**	**50,4**	**44,0**	**49,5**	**54,1**	**45,3**
Arratia-Nervión	35,2	37,0	33,3	39,2	42,9	35,4	48,5	50,5	46,4	49,4	53,3	45,6
Gran Bilbao	36,0	41,2	31,1	40,6	45,3	36,4	46,9	50,5	43,6	49,3	54,0	45,0
Duranguesado	37,3	41,2	33,5	41,7	44,8	38,6	47,9	50,2	45,6	51,3	55,3	47,4
Encartaciones	36,1	39,6	32,6	39,2	41,4	37,0	45,6	49,6	41,8	47,8	52,7	42,9
Gernika-Bermeo	33,7	38,7	29,0	38,7	42,3	35,2	44,6	45,3	44,0	47,0	49,8	44,2
Markina-Ondarroa	35,1	40,0	30,2	39,9	44,0	35,6	45,5	48,6	42,2	49,5	53,3	45,5
Plentzia-Mungia	39,2	43,7	34,8	44,3	49,9	38,9	52,3	55,5	49,1	54,3	58,7	49,9
Gipuzkoa	**40,4**	**43,3**	**37,7**	**42,6**	**44,9**	**40,4**	**46,4**	**49,3**	**43,6**	**50,1**	**53,7**	**46,7**
Bajo Bidasoa	41,3	44,4	38,4	44,5	47,9	41,2	49,0	52,3	45,8	51,6	55,9	47,4
Bajo Deba	38,0	41,5	34,6	39,7	42,6	36,9	43,6	46,6	40,6	46,9	50,0	43,9
Alto Deba	39,8	43,2	36,3	41,2	43,9	38,4	44,4	46,7	42,1	48,5	51,2	45,7
Donostialdea	40,9	43,9	38,1	43,2	45,4	41,1	47,0	50,3	44,1	50,9	55,2	47,0
Goierri	39,0	41,1	36,8	41,8	43,6	40,0	45,9	49,3	42,4	48,9	52,4	45,4
Tolosa	39,3	42,0	36,6	40,6	43,1	38,1	43,7	47,0	40,4	47,6	50,4	44,7
Urola Costa	42,1	44,3	39,9	43,2	44,5	42,0	46,8	47,6	46,0	52,0	53,8	50,2

Table 2. Internet user population aged 15 and over by sex, Province and capital (%). 2005-2008

Source: EUSTAT. Survey on the Information Society - Familias (ESIF)

	2005			2006			2007			2008		
	Estimations			Estimations			Estimations			Estimations		
	Total	Men	Women	Total	Men	Women	Total	Men	Women	Total	Men	Women
Basque Country	**37,8**	**41,6**	**34,2**	**41,3**	**44,6**	**38,2**	**46,7**	**49,7**	**43,8**	**49,9**	**54,0**	**46,0**
Alava	**38,3**	**39,5**	**37,2**	**40,7**	**42,1**	**39,5**	**45,6**	**47,9**	**43,3**	**50,6**	**54,1**	**47,0**
Vitoria-Gasteiz	39,0	40,4	37,7	41,5	43,1	39,9	46,3	48,8	44,0	51,4	55,1	47,8
Bizkaia	**36,1**	**41,1**	**31,5**	**40,7**	**45,1**	**36,6**	**47,1**	**50,4**	**44,0**	**49,5**	**54,1**	**45,3**
Bilbao	35,8	41,6	30,6	40,5	45,8	35,9	46,0	50,1	42,4	48,5	54,0	43,6
Gipuzkoa	**40,4**	**43,3**	**37,7**	**42,6**	**44,9**	**40,4**	**46,4**	**49,3**	**43,6**	**50,1**	**53,7**	**46,7**
Donostia-San Sebastián	42,4	45,6	39,6	44,6	47,6	42,0	48,1	52,4	44,5	51,8	56,6	47,8

Table 3. Variation coefficients of the Internet user population aged 15 and over by sex, Province and district (%). 2005-2008

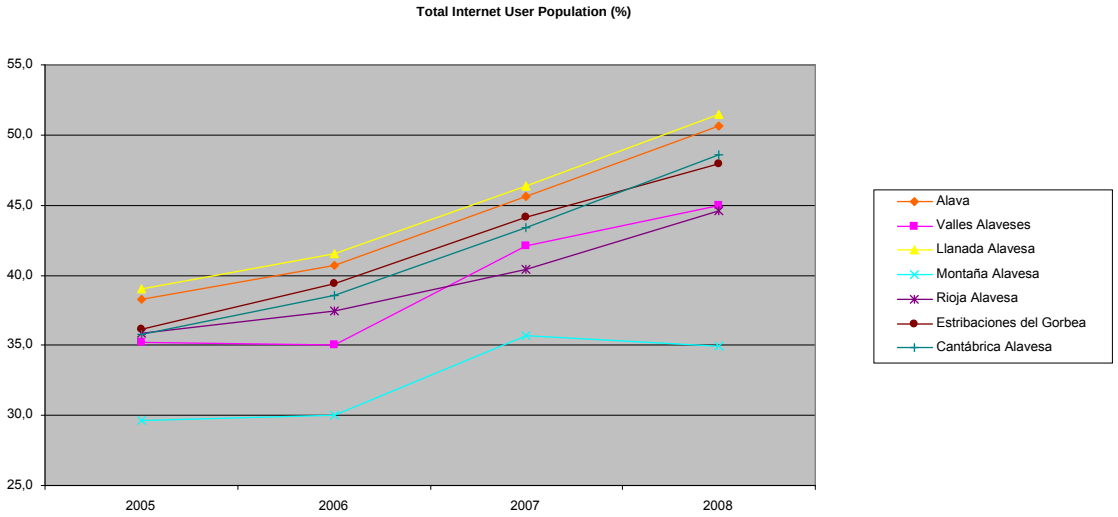
Source: EUSTAT. Survey on the Information Society - Familias (ESIF)

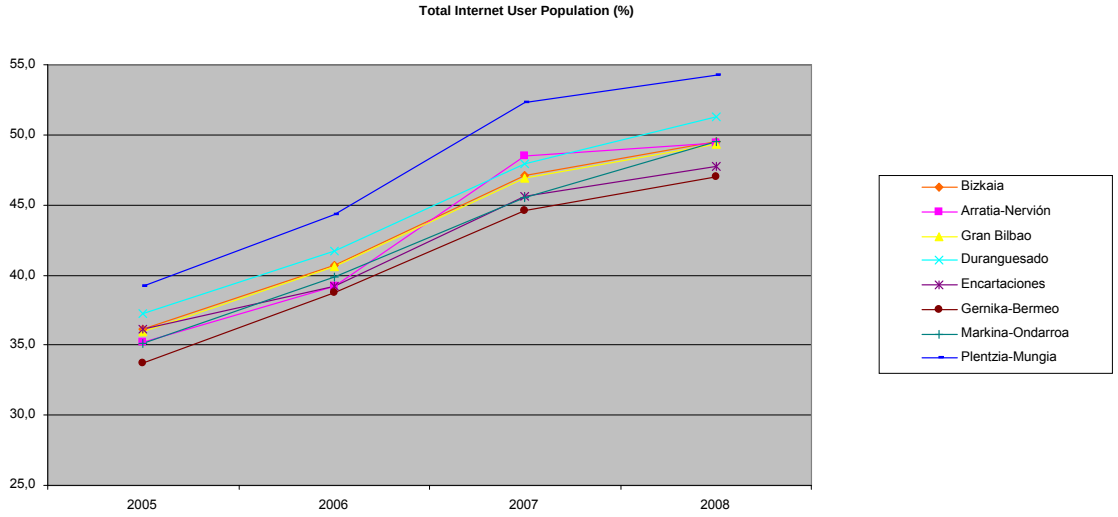
	2005			2006			2007			2008		
	CV			CV			CV			CV		
	Total	Men	Women	Total	Men	Women	Total	Men	Women	Total	Men	Women
Basque Country	**1,7**	**2,4**	**2,5**	**1,7**	**2,3**	**2,4**	**1,6**	**2,2**	**2,2**	**1,5**	**2,1**	**2,2**
Alava	**3,5**	**4,7**	**5,1**	**3,4**	**4,7**	**5,0**	**3,3**	**4,6**	**4,7**	**3,1**	**4,4**	**4,5**
Valles Alaveses	3,4	4,4	5,3	4,1	5,4	6,5	3,7	5,2	5,1	3,5	5,2	4,0
Llanada Alavesa	3,3	4,2	5,0	3,1	4,3	4,6	3,0	3,9	4,6	2,9	3,9	4,1
Montaña Alavesa	4,7	5,9	7,7	5,4	7,7	7,4	5,5	7,1	8,9	7,0	7,3	16,1
Rioja Alavesa	4,6	6,7	6,0	5,1	7,4	7,0	5,0	6,7	7,5	4,3	5,6	6,7
Estribaciones del Gorbea	3,7	5,1	5,5	3,6	4,6	5,7	4,5	6,5	6,1	3,9	4,8	6,4
Cantábrica Alavesa	3,6	4,6	5,5	3,5	4,9	5,1	3,3	4,3	5,1	3,0	4,0	4,5
Bizkaia	**2,6**	**3,5**	**3,9**	**2,5**	**3,4**	**3,7**	**2,3**	**3,2**	**3,3**	**2,3**	**3,0**	**3,4**
Arratia-Nervión	4,0	6,0	5,0	4,4	6,5	5,8	3,2	3,5	5,5	3,4	3,7	5,8
Gran Bilbao	2,4	3,1	3,8	2,3	3,0	3,6	2,1	2,8	3,2	2,0	2,7	3,1
Duranguesado	2,9	3,7	4,5	2,8	4,0	4,0	2,3	3,0	3,6	2,2	3,0	3,3
Encartaciones	3,4	4,0	5,8	4,0	4,8	6,4	3,9	5,6	5,3	3,7	5,7	4,3
Gernika-Bermeo	3,4	4,2	5,4	3,1	4,5	4,0	3,1	4,2	4,4	2,6	3,6	3,9
Markina-Ondarroa	3,3	4,0	5,6	3,5	4,2	5,8	3,4	4,2	5,6	2,9	3,5	4,9
Plentzia-Mungia	3,4	4,7	4,7	3,3	4,5	4,8	2,6	3,5	3,8	2,2	3,0	3,3
Gipuzkoa	**2,8**	**3,9**	**4,1**	**2,7**	**3,9**	**3,9**	**2,6**	**3,6**	**3,6**	**2,5**	**3,5**	**3,6**
Bajo Bidasoa	3,1	4,2	4,6	3,1	4,3	4,5	2,7	3,9	3,8	2,8	4,0	3,8
Bajo Deba	3,4	4,5	5,2	3,2	4,2	5,0	2,8	4,0	3,9	2,5	3,7	3,4
Alto Deba	3,2	4,3	4,7	3,1	4,4	4,5	2,8	3,8	4,0	2,6	3,7	3,7
Donostialdea	2,8	3,9	4,0	2,7	3,7	3,9	2,5	3,6	3,4	2,4	3,3	3,4
Goierri	3,2	4,5	4,5	3,1	4,4	4,3	2,6	3,5	3,9	2,6	3,6	3,6
Tolosa	3,3	4,9	4,3	3,5	4,9	5,1	3,3	4,7	4,6	3,3	4,6	4,7
Urola Costa	3,2	4,5	4,6	3,1	4,3	4,4	2,8	4,0	3,9	2,4	3,6	3,1

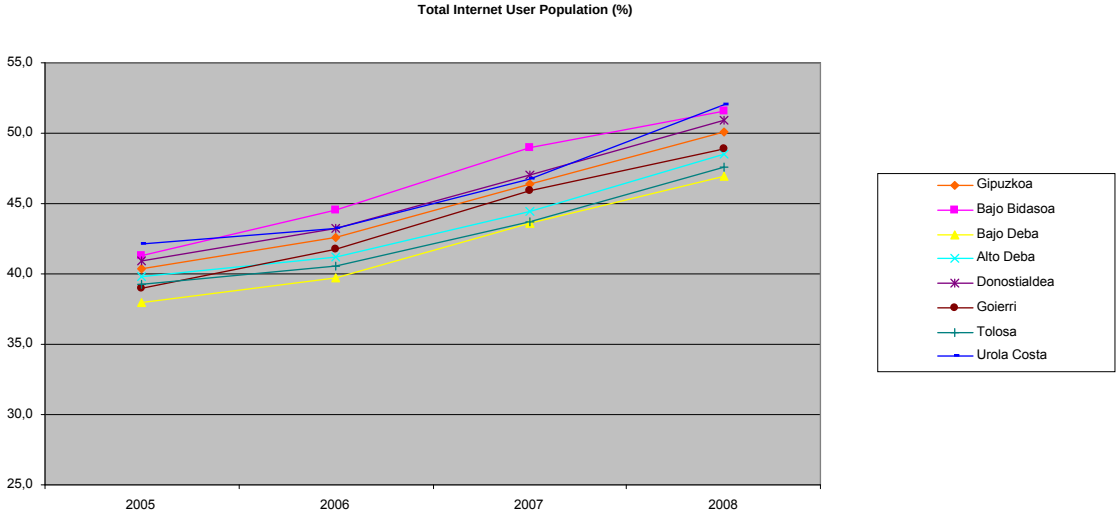
Tabla 4. Variation coefficients of the Internet user population aged 15 and over by sex, Province and district (%). 2005-2008

Source: EUSTAT. Survey on the Information Society - Familias (ESIF)

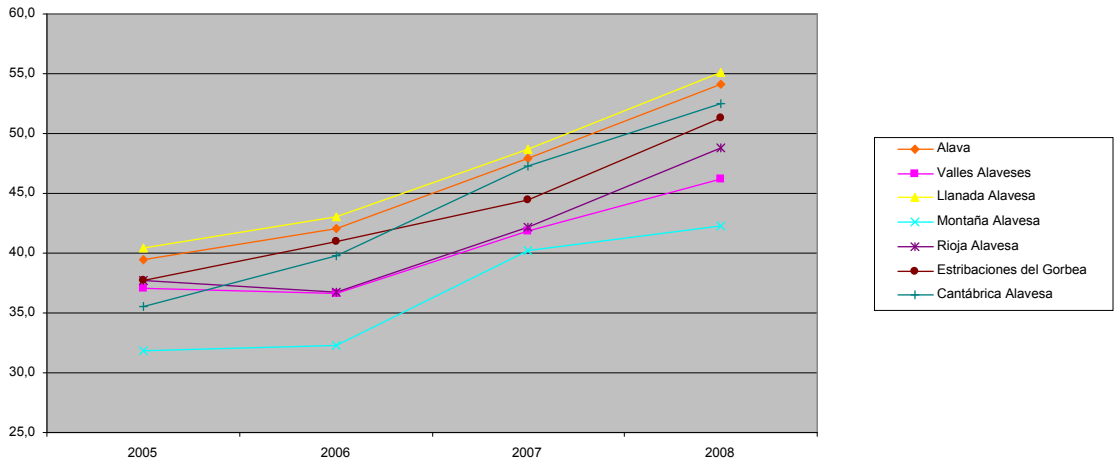
	2005			2006			2007			2008		
	vc			vc			vc			vc		
	Total	Men	Women	Total	Men	Women	Total	Men	Women	Total	Men	Women
Basque Country	**1,7**	**2,4**	**2,5**	**1,7**	**2,3**	**2,4**	**1,6**	**2,2**	**2,2**	**1,5**	**2,1**	**2,2**
Alava	**3,5**	**4,7**	**5,2**	**3,4**	**4,7**	**5,0**	**3,3**	**4,6**	**4,7**	**3,1**	**4,4**	**4,5**
Vitoria-Gasteiz	3,3	4,3	5,1	3,2	4,3	4,6	3,0	4,0	4,6	2,9	4,0	4,2
Bizkaia	**2,6**	**3,5**	**3,9**	**2,5**	**3,4**	**3,7**	**2,3**	**3,2**	**3,3**	**2,3**	**3,0**	**3,4**
Bilbao	2,7	3,5	4,2	2,6	3,4	3,9	2,3	3,2	3,4	2,3	3,0	3,5
Gipuzkoa	**2,8**	**3,9**	**4,1**	**2,7**	**3,9**	**3,9**	**2,6**	**3,6**	**3,6**	**2,5**	**3,5**	**3,6**
Donostia-San Sebastián	3,1	4,3	4,4	2,8	3,9	4,1	2,6	3,7	3,7	2,6	3,6	3,6


Total Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Alava
Valles Alaveses
Llanada Alavesa
Montaña Alavesa
Rioja Alavesa
Estribaciones del Gorbea
Cantábrica Alavesa

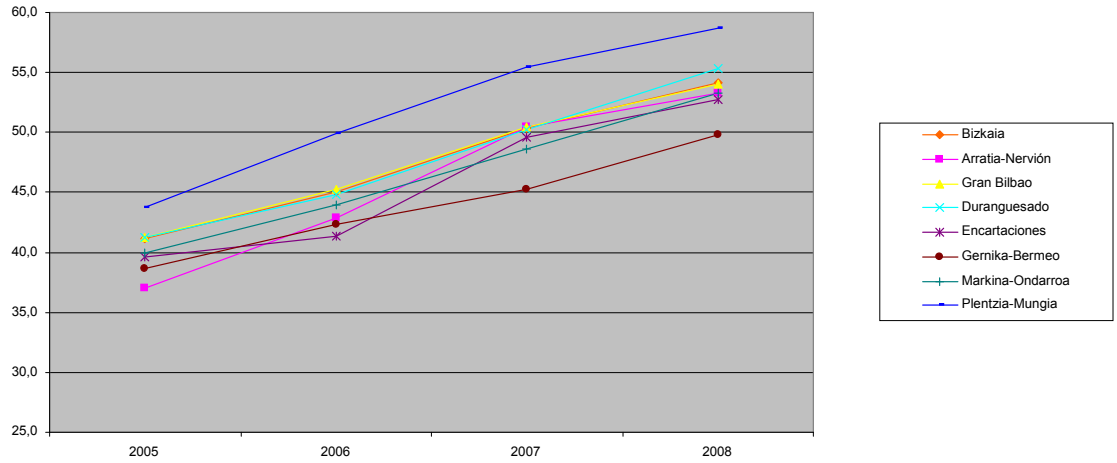

Total Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Bizkaia
Arratia-Nervión
Gran Bilbao
Duranguesado
Encartaciones
Gernika-Bermeo
Markina-Ondarroa
Plentzia-Mungia


Total Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Gipuzkoa
Bajo Bidasoa
Bajo Deba
Alto Deba
Donostialdea
Goierri
Tolosa
Urola Costa

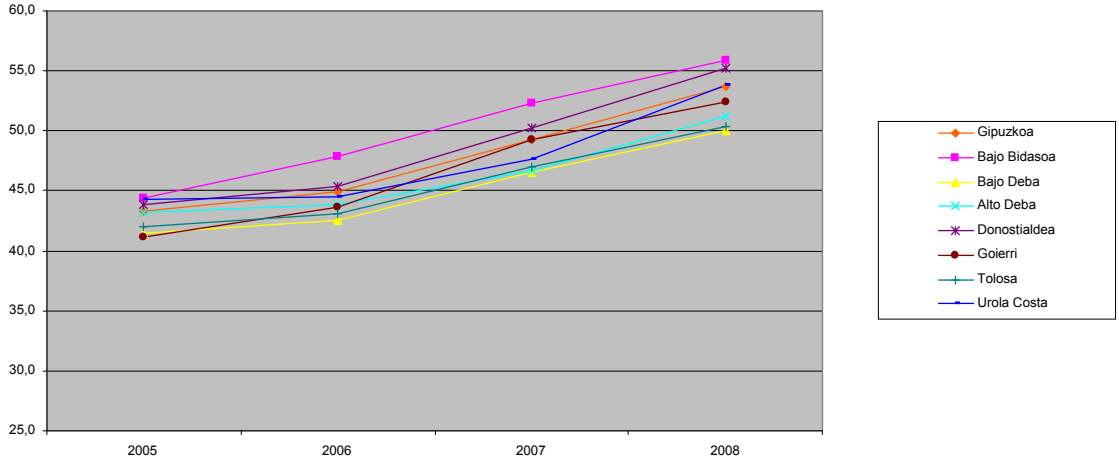
Male Internet User Population (%)

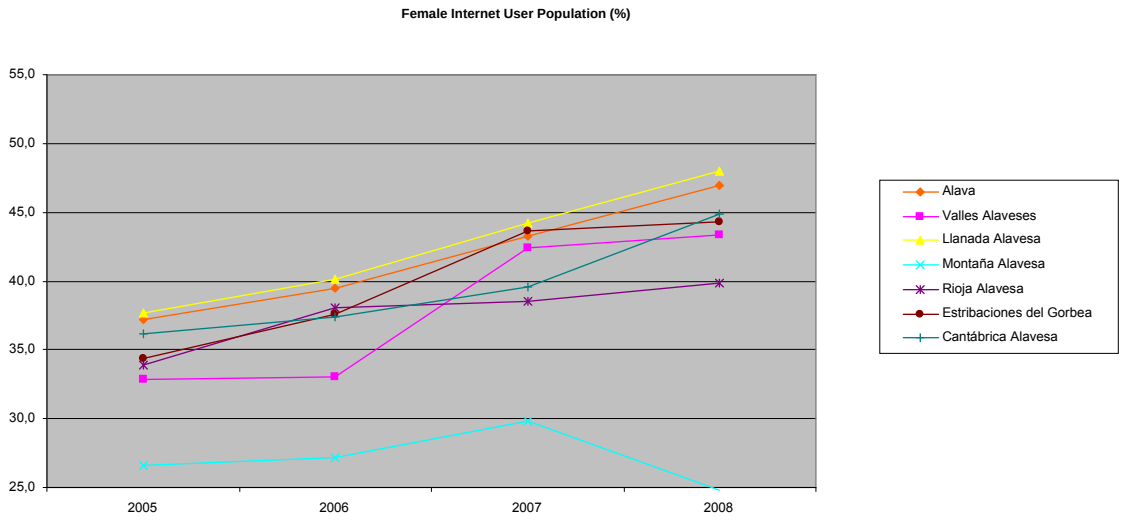


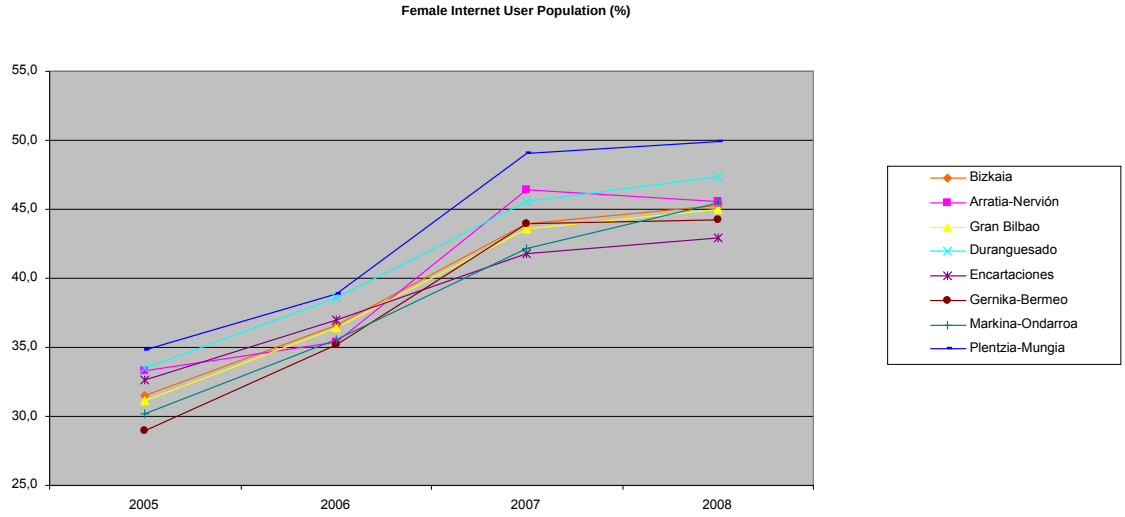
Male Internet User Population (%)

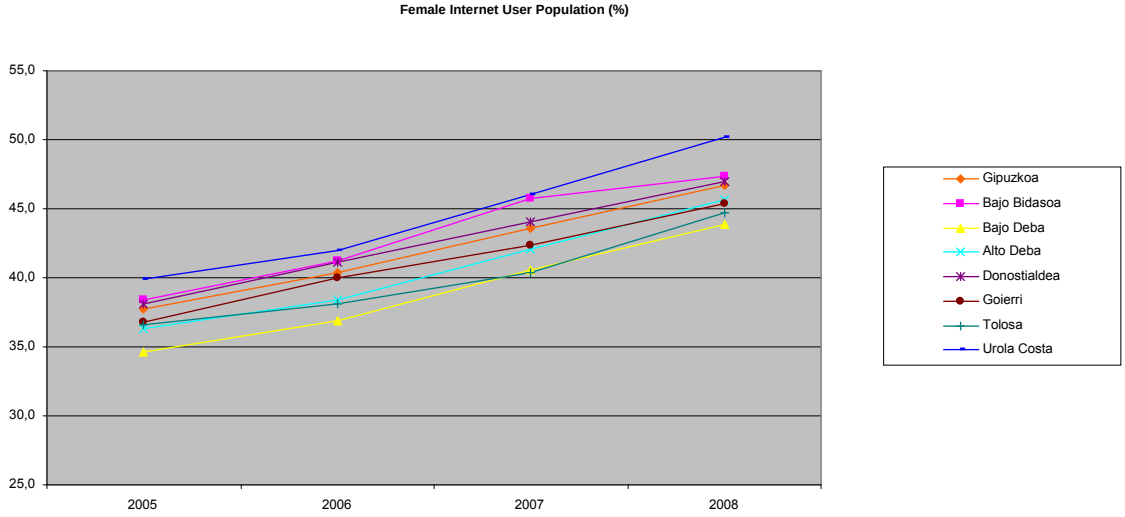


Male Internet User Population (%)




Female Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Alava
Valles Alaveses
Llanada Alavesa
Montaña Alavesa
Rioja Alavesa
Estribaciones del Gorbea
Cantábrica Alavesa


Female Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Bizkaia
Arratia-Nervión
Gran Bilbao
Duranguesado
Encartaciones
Gernika-Bermeo
Markina-Ondarroa
Plentzia-Mungia


Female Internet User Population (%)
55,0
50,0
45,0
40,0
35,0
30,0
25,0
2005
2006
2007
2008
Gipuzkoa
Bajo Bidasoa
Bajo Deba
Alto Deba
Donostialdea
Goierri
Tolosa
Urola Costa

Chapter 5

Conclusions

The growing demand for disaggregated information and the need to avoid placing too great a burden on informants means an progressively greater use of model-based estimation methods in official statistics.

Obtaining estimations of the figures related with activity in small areas such as districts, as described in this paper, is a step forward in the application by the Statistics Office of the new estimation model-based methodologies.

The results presented in this paper offer acceptable levels of quality in terms of accuracy. The majority of the variation coefficients (VC) obtained in the estimations do not surpass 10% and many do not pass the 5% mark.

From now on, Eustat, will be able to offer district-level estimations based on a conjunctural survey with the increased operation efficiency that this implies.

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

Eustat aims to carry on making progress in the study and application of model-based estimation methodology so as to be able to offer increasingly disaggregated quality information.

Chapter 6

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 on the Information Society - families (ESIf).
http://www.eustat.es/document/datos/Calculo_errores_ESI_i.pdf

EUSTAT (2008)

Proyecto Técnico de la Operación Encuesta sobre Sociedad de la Información - familias. (ESIf).

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