Eustat
EUSKAL ESTATISTIKA ERAKUNDA
INSTITUTO VASCO DE ESTADÍSTICA

Estimation of variation coefficients in the Industrial Survey

Eustat
EUSKAL ESTATISTIKA ERAKUNDA
INSTITUTO VASCO DE ESTADÍSTICA

Contents

1. Brief description of the Industrial Survey

1.1 Objectives of the operation

The main objectives of the Industrial Survey of the A.C. of the Basque Country can be summarised as follows:

- to get to know the structural characteristics of Basque industry and its evolution;
- to estimate the main economic macro-magnitudes of the industrial sector, an essential basis for the production of Input-Output tables and Economic Accounts;
- to develop a set of indicators determining the relevant aspects of the sector, which can act as a basis for obtaining the ongoing indicators produced by EUSTAT, such as the Industrial Production Index and the Industrial Price Index;
- to provide an instrument for social agents that is useful and relevant for economic analysis, which is of great importance in times of fluctuating cycles;
- to use the results of the operation in the production of the Economic Accounts Advance.

1.2 Scope of the operation

Poputlation scope: The population scope is confined to those establishments whose main activity, measured in terms of value added generated, is industrial. It includes, in accordance with the 1993 National Classification of Economic Activities (from now on CNAE-93), the following sections:

- Section C: Mining and quarrying industries
- Section D: Manufacturing industry
- Section E: Production and distribution of electric energy, gas and water

Geographical scope: The statistical units that are located in the goegraphical area of the Autonomous Community of the Basque Country, even though the head office or management is outside the A.C. of the Basque Country.

Time scope: The reference period is the financial calendar year before the calendar year when the data was collected. Exceptionally, when dealing with establishments whose accountancy refers to periods that do not coincide with the calendar year, the information may refer to financial periods that end during the corresponding year.

1.3 Sampling Design

Survey frame. The basis for establishing the survey frame is the Eustat Directory of Economic Activities, which allows the production of a probabilistic sample, limiting sampling errors.

A probabilistic sampling is made in two phases: a first one in which all the units with more than 19 employees are selected with probability "one" (also included among self-weighted are new entries in the Directory in year *t*); in the second phase a random stratified sample is made, also in two sub-phases:

- **Sub-phase 1**. The stratification variables used are:

  - Province: Araba, Bizkaia y Gipuzkoa.
  - Activity: EUSTAT normalized classification A-84. The optimum allocation of each stratum is calculated with the corresponding budgetary restriction for each year, which in this case is that the global sample size does not exceed 3,500 statistical units. The optimum allocation is calculated, over-representing Araba with an allocation proportional to the square root of the number of establishments in each Province.

    The sectorial distribution within each province is made with the following formula:

$$n_{s,h} = n_h \frac{T_{s,h}}{\sum T_{s,h}}, \qquad (1)$$

    With $T_{s,h} = N_{s,h}\sqrt{E_{s,h}}$, where $N_{s,h}$ is the number of establishments in sector *s* in Province *h*, $E_{s,h}$ is average employment in sector *s* in Province *h* and n_h is the sampling size in Province *h*, with $n = \sum n_h$ being the total sampling size.

- **Sub-phase 2**. The optimum allocation distribution in each sector, following optimum allocation criteria, is made proportional to the weight of each sub-class in the sector total. The Sectorial Analysis Sector of the Department of Agriculture and Fishing in activities 15.1, 15.5, 15.321, 15.333, 15.411, 15.413, 15.7, 15.93, 15.94, 15.95 y 20.201 over-represents the sample. Those establishments that carry out R&D activities and belong to the stratum with less than 20 employees are automatically included in the sample. In this second sub-phase, in order to diminish the rate of non-replies, the random selection made in 1993 has been substituted for a panel of establishments

that had replied to the survey of the previous year plus a random section to complete the optimum allocation.

1.4 Post-stratification

One the information is collected, the statistical units are post-stratified and the populations recalculated, since the sampling design has been made with the Directory of year *t* - 1.

The populations are calculated:

1. For the employment strata 20 - 49, 50 - 99, 100 - 499 and ≥ 500, tabulating the Economic Activities Directory for the year *t*.

2. For the stratum of < 20 employees, a projection of the population *t* is made for each stratum, multiplying the population in year *t* – 1 by the employment variation of the sample establishments that remain in the stratum and updated between year *t* and *t* - 1.

2. System of estimation in the Industrial Survey

2.1 Introduction

As previously stated, all the statistical units with more than 19 employees are self-weighted, so that the main interest lies with estimations within the 1-19 employee employment stratum.

The EUSTAT economic surveys use different types of estimators when extrapolating the sample information to the population. On the one hand, direct estimators based on the sample design (Horvitz-Thompson estimator, a reason estimator, …) are used and on the other hand, model-assisted estimators that use ancillary information from other fields for areas where the sample is scarce. The latter have the advantage of reducing the sample error as the estimate is in a small area, but can also introduce a signifacant bias if the ancillary information in the different fields (or stratums) is not homogenous. Therefore, an optimum solution is to use estimators that on the one hand, offset the instability of the direct estimators and on the other hand, the bias of the indirect ones. See [1] and [2].

2.2 Composite estimators and their variance

The Economic Survey of the Industrial Survey uses the type of estimators referred to in the last part of hte above paragraph. They are known as composite estimators and have the following generic expression:

$$\hat{\theta}_{COMPOSITE} = \phi\ \hat{\theta}_{DIRECT} + (1-\phi)\ \hat{\theta}_{INDIRECT} \qquad \text{with } 0 \leq \phi \leq 1 \qquad \textbf{(2)}$$

The expression of the Mean Quadratic Error for this type of estimator is not simple and an approximation is proposed as follows:

$$MSE(\hat{\theta}_{COMPOSITE}) = \phi^2 MSE(\hat{\theta}_{DIRECT}) + (1-\phi)^2 MSE(\hat{\theta}_{INDIRECT}) - 2\phi(1-\phi)[MSE(\hat{\theta}_{DIRECT}) - \hat{\theta}_{INDIRECT*} Biais] \quad \textbf{(3)}$$

Both the expression of the estimator and of its mean quadratic error are implemented in a SAS macro programmed for that purpose. Further details about the origin and calculation of the above expression can be consulted in reference of the bibliography.

Bibliografía

[1] EUSTAT (2005). "*Cálculo de coeficientes de variación para diferentes estimadores directos e indirectos utilizados en las encuestas económicas de Eustat*". http://www.eustat.es/document/datos/Errores_c.pdf

[2] EUSTAT (2005). "*Estimación de Áreas Pequeñas en la Encuesta Industrial de la C.A. de Euskadi*". http://www.eustat.es/document/datos/ct_14_.pdf