

**ASPECTS OF ESTIMATION PROCEDURES
AT EUROSTAT WITH SOME EMPHASIS ON OVER-SPACE
HARMONISATION**

John Panaretos & Nikolaos Tzavidis

Department of Statistics
Athens University of Economics & Business
76, Patission St. 104 34
Athens, GREECE
e-mail:jpan@aueb.gr

1. Introduction

In late September 1999 the Management Committee of Eurostat created a Task Force¹ in order to analyse the procedures that were being used for producing and publishing estimations at Eurostat. The present paper is based on the idea of this Task Force. Going beyond its descriptive character we give an analytical study of some specific estimation domains. The aim is to investigate how some aspects of the statistical science are applied in practice in order to produce results, which in many cases are being used for the adoption of policies.

Three estimation domains that either applied or discussed by Eurostat are presented: (1) the micro-aggregation techniques for creating confidential data, (2) the backward calculation techniques for obtaining homogeneous time series and (3) the sampling procedures in Eurostat and in Member States and especially a summary, in a theoretical and applied manner, of the most frequently used sample designs and weighting procedures. Due to the multinational character of the third domain and in order to trace our targets, we used as exploratory tools three sample surveys that are being applied in all Member States.

The main scopes of this paper can be summarised as follows: (1) to describe each estimation domain (2) to compare the alternative estimation procedures (3) to study criteria for assessing the quality of the results (4) to investigate the progress that has been made towards the over-space harmonisation of the estimation procedures that are being applied at the national level and the roles that Eurostat adopts with regard to this harmonisation procedure. Consequently, for the estimation domains that are applied in the country level i.e. the backward calculation methods and the sampling methods, we have studied the harmonisation attempts and the roles of Eurostat in relation to those attempts even if they refer to common concepts of measurement or to common methods of measurement.

¹ The term Task Force refers to groups that are formed internally to Eurostat for the study of a statistical problem.

2. Micro-aggregation Techniques

As the role of the European Community grows more complex, the demand for detailed information grows also. The Commission increasingly relies on statistics to allocate funds, promote competitiveness in industry and to evaluate the impact of the current Community programs. A colossal amount of data is needed some of which are highly sensitive or capable of causing harm to the respondent if disseminated in its original form. Hence, the Member States are not always willing to provide micro-data for confidentiality reasons. On the other hand, the European authorities must obtain sufficient information in order to carry out all relevant statistical analysis efficiently.

One possible way of obtaining confidential data is by using the micro-aggregation techniques. These methods are applied internally to Eurostat by Unit A4² to the micro data that are being transmitted by the Member States. Eurostat is responsible for transforming the transmitted data into confidential data sets ready for analysis and publication. In other words Eurostat is the intermediate between the Member States who wish their data to be confidential and the analysts who wish to have the “real” data.

The micro-aggregation techniques are a group of methods aiming at transforming the original data into a data set that respects a sufficient confidentiality level and at the same time making possible to carry out relevant analysis. The main idea of the micro-aggregation techniques is influenced by the work of Strudler, Lock, Oh and Scheuren (1986) on the Tax model at the US Internal Revenue Service. In the majority of the micro-aggregation methods, one sets up groups of fixed or variable size that are summarised by means of summary statistics. The various methods differ from each other by the method used for making the aggregation of the units (mean, median, mode) and especially the choice of the method for sorting the data. The sorting method has to set up groups as homogeneous as possible, in order not to lose too much information.

The micro-aggregation methods belong to different categories depending on the types of the variables that they are applied (numerical, ordinal or nominal). The methods applied to quantitative variables can be classified into the following categories:

² Unit A4 is responsible for Research & Development in Statistics

1. Single Axis Methods

- a) Single axis technique
- b) First principal component technique
- c) Sum of Z-scores technique

2. Classification Methods

- a) The adaptation of Hanani's algorithm
- b) The adaptation of Ward's method

3. Multivariate Methods

- a) The individual ranking technique
- b) The weighted moving average technique

The methods applied to qualitative variables are the following:

1. Method of Snake

2. Method of Entropy

3. Similarity of Distributions Method

2.1 Single Axis Methods

2.1.1 Sorting By a Single Axis

The simplest way of sorting the observations is to choose a variable and to sort the data set in an ascending or descending order according to this variable. After sorting the data, we create groups of k units (i.e. the threshold that is usually equal to 3) and we replace each individual value of each group with a within group summary statistic.

The advantages of this approach are obvious for units ranked as homogeneous as possible within each group and heterogeneous without. However, the single axis criterion is influenced by the choice of the ranking basis. In other words, if we change the sorting variable, we will derive different results.

2.1.2 Sorting By the First Principal Component

While the single axis method is based on only one variable for sorting the data, this method sorts the observations according to an axis which represents better all variables and not only one. More specifically after applying the principal components method to the standardised data set, the observations are ordered in descending order of their first principal component scores and then aggregated into groups of k units. Then each individual observation is replaced with a summary statistic of the group that it belongs.

By definition the principal component analysis studies the underlying multidimensional correlation structure of the data, to produce principal components. Methods of ranking based on such an approach, therefore, utilise the correlation matrix. These types of methods are very useful for data that are very highly correlated, and the higher the correlation, the lower will be the loss of information. However, not all data are highly correlated. This does not mean that these data sets do not contain useful information. Under such conditions, it is likely, that these types of procedures fail to retain the required level of information.

2.1.3 Sorting By the Sum of Z-Scores

This method resembles the principal component one as far as its purpose is to consider as many variables as possible. The difference between the two methods is that the sum of Z-scores gives equal importance to all variables whereas for the principal component method the weighting depends on the correlation structure of the data set.

In the sum of Z-scores method the variables are standardised and a standardised score is calculated by adding across all variables the standardised value of each observation. This score is then used for sorting the observations. The last two steps of the micro-aggregation procedure is to partition the data set into groups and replace each individual value with a summary statistic of the group that it belongs.

The sum of Z-scores method faces the problem of the correlation structure by using the same weights for all observations. However, this method suffers from the same impotencies as the rest of the single axis techniques since there is no gurantee that the

axis chosen will provide an optimal clustering and the methods are very sensitive to the set of variables chosen.

2.2 Classification Methods

The micro-aggregation techniques that belong to this category are derived from the automatic classification methods. The aim of the micro-aggregation techniques, which is to build groups of the k most homogeneous elements, approximates that of the techniques of automatic classification, which aim at partitioning a data-set in classes by using a homogeneity criterion. The criterion most usually used is the minimisation of the intra group variance.

2.2.1 The Hanani's Algorithm

The approach followed by Hanani uses the following idea. Let $G^{(0)}$ be an arbitrary partition of Ω obeying the equal size constraint, i.e. $G^{(0)} = \{G_1^{(0)}, G_2^{(0)}, \dots, G_n^{(0)}\}$ with $|G^{(0)}| = k$. Let $L_1^{(0)}$ be the set of n spheres, centered on the centers of gravity of the classes $G^{(0)}$ and having radius calculated to contain exactly k points of Ω i.e. $L_1^{(0)} = \{S_1^{(0)}, S_2^{(0)}, \dots, S_n^{(0)}\}$. Let $L_2^{(0)}$ be the set of centers of gravity $m_i^{(0)}$ of the members of the partition $G^{(0)}$ i.e. $L_2^{(0)} = \{m_1^{(0)}, m_2^{(0)}, \dots, m_n^{(0)}\}$. Define a new partition $G^{(1)}$ associated to $L_1^{(0)}$ and $L_2^{(0)}$ in the following way:

$x \in G_i^{(1)}$ if and only if

$$\begin{cases} \text{either } x \in S_i^{(0)} \text{ and } \|x - m_i^{(0)}\| \leq \|x - m_j^{(0)}\| \text{ for all } j \\ \text{or there is no } k \text{ such that } x \in S_k^{(0)} \text{ and } \|x - m_i^{(0)}\| \leq \|x - m_j^{(0)}\| \text{ for all } j \leq n \end{cases}$$

The new partition will not necessarily obey the equal size constraint. From $G^{(1)}$, two new sets, $L_1^{(1)}$ and $L_2^{(1)}$ are constructed. A new partition $G^{(2)}$ associated to $L_1^{(1)}$ and $L_2^{(1)}$ can be obtained. The algorithm stops at step r if $G^{(r)} = G^{(r+1)}$.

This method can be improved when it is realized that the regions that contain the clusters are not spheres but areas limited by hyperplanes. For instance, when $n=2$ $L_1^{(0)}$ is a partition defined by hyperplanes orthogonal to the line that joins $m_1^{(0)}$ with $m_2^{(0)}$ and which separates Ω in two groups of equal size. For $n>2$, this method can be generalised by considering each pair (G_i, G_j) defining a hyperplane orthogonal to the line joining the centers of gravity. The algorithm can be described by the following steps:

- (1) Set $r=0$
- (2) Draw a random partition $G^{(0)}$ in n groups of equal size.
- (3) Set $i=1, j=2$
- (4) Find the hyperplane orthogonal to the line joining $m_1^{(r)}$ with $m_2^{(r)}$ and separates G_i^r, G_j^r into two groups of equal size.
- (5) Replace G_i^r, G_j^r by the two groups separated by the hyperplane.
- (6) Set $j=j+1$
- (7) If $j \leq n$ go to (4) Else go to (8)
- (8) Set $i=i+1$
- (9) If $i \leq n-1$ set $j=j+1$ and go to (4). Else set $r=r+1$ and go to (3)
- (10) Stop when for the last $\binom{n}{2}$ iterations the classes have not been modified.

The present method compared to Hanani's algorithm has two advantages. Firstly, the solution will always obey to the equal size constraint. This is very important since in order to be in accordance with the confidentiality rules, clusters with less than k units are not allowed. Secondly, if one reaches at some stage the best solution the algorithm will be terminated. This is not the case in Hanani's algorithm which it will continue searching for a partition with a lower within-group variance breaking the equal size constraint. Finally, in both cases there is a danger of falling into local minimum. It is therefore recommended to start the algorithm from different initial partitions.

2.2.2 The Adaptation of Ward's Method

In contrast to the Hanani's algorithm, this method creates groups of variable size i.e. groups that consist of at least k and not exactly of k observations. The Ward's criterion forms clusters by maximising the within clusters homogeneity. The algorithm starts by creating at least two groups from the data set. Subsequently, using Ward's method we classify all units into groups of minimum size k and maximum size less than $2k$. If we are facing a multivariate problem, the only extra thing that we have to do is to define a multivariate criterion for specifying the initial partition. At this point we can combine the single axis techniques with the Ward's method. This means that we can use a single axis method to obtain an initial partition and then apply the Ward's method in this initial partition.

2.3 Multivariate Methods

In contrast to the previously refereed methods, this kind of methods treats each variable in a separate way.

The **individual ranking method** is based on sorting and aggregating the observations according to each variable separately.

For the **weighted moving averages** method initially we sort the observations according to each variable separately and we form clusters of k units. Subsequently, we compute a weighted moving average in each group. As a result of this method we have to define k weights (depending on the number of units in each group) for calculating the weighted averages.

2.4 Methods Applicable to Qualitative Variables

In practice it is rare to find a data set consisting only of metric variables. More common is to find all three types of variables, metric, ordinal and nominal ones coexisting in a single set of data. In the sequel we examine micro-aggregation techniques applied to qualitative variables. Such techniques are applied only to ordinal variables or only to nominal variables or simultaneously to ordinal and nominal variables under suitable definitions.

2.4.1 Method of Snake

This technique is appropriate for multidimensional ordinal variables. Assume that we have two ordinal variables each consisting of five levels and which are represented by an appropriate matrix. In case that we have more than two variables we separate the variables in segments of two. The sorting of the observations is done on a basis of a starting point, and by taking a relatively arbitrary route into this matrix .

The ranking in the case of the method of snake is based on the computation of the so-called snake variable and on a recursive algorithm in which the results from the last level are transferred into the next level. More specifically, the snake variable is calculated in each step, the results are ranked in descending or ascending, and groups of units are constructed. Finally, each individual value is replaced by the group mode or another statistic of central tendency.

The formulas for calculating the snake variable in the case that the range of variables is between 2 and 5 and the recursive algorithm are as follows:

$$\begin{cases} v(l, m) = 6 * 1 + m^{(-1)^{l+1}} \\ w(k, l, m) = 6^2 * 1 + [v(l, m)]^{(-1)^{k+1}} \\ x(j, k, l, m) = 6^3 * 1 + [w(k, l, m)]^{(-1)^{j+1}} \\ y(i, j, k, l, m) = 6^4 * 1 + [x(j, k, l, m)]^{(-1)^{i+1}} \end{cases}$$

where y is the resulting snake variable on which the sorting of the observations is based, i is the index of the first variable, j the index of the second variable and m is the index of the fifth variable. In this method it is important to realise that the indexes i , j , m are defined by the arbitrary route that we choose to follow.

2.4.2 Method of Entropy

This method is defined for both nominal and ordinal variables by adopting the appropriate definitions. More specifically, the entropy, H , in the case of a nominal

variable is given by $H_i = \frac{\left(- \sum_{i=1}^L p_i \text{ld} p_i \right)}{\text{ld} L}$ where p_i is the probability that an observation

belongs to category i , ld is the logarithm base 2 and L is the number of categories. In

the case of an ordinal variable the entropy is given by

$$H_i = \frac{\left(- \sum_{i=1}^{L-1} \left(\left(p_i' \text{ld} p_i' \right) + \left(1 - p_i' \right) \text{ld} \left(1 - p_i' \right) \right) \right)}{L-1}$$

where p_i' is the probability that an

observation belongs to category i or to a higher category, ld is the logarithm base 2 and L is the number of categories.

The entropy can be characterized as a measure of homogeneity and the entropy of a data set that consists of nominal and ordinal variables is the sum of the entropies of each individual variable. Describing the entropy algorithm in the case that we have one variable we can say the following. Firstly, we create all possible pairs of observations and the entropies are calculated. The pair with the least entropy is matched with all other observations in order to create all possible triads (if we want to have groups of three units). The entropies are recalculated, the triad with the least entropy constitutes the first group and the observations of this group are removed.

2.4.3 Similarity of Distributions Method

This method is appropriate for nominal variables and uses a similarity or dissimilarity index as a basis for sorting and grouping the observations.

2.5 Evaluation Criteria

The most important part of the micro-aggregation procedure is the quality of the derived results. We have already stressed the effort for keeping a balance between data confidentiality and data quality. The data quality that is translated into the preservation of the structure of the original data ensures that the statistical analysis based on the modified data set is valuable. The data confidentiality ensures that the original data set has been modified in such a way that ensures the confidentiality of the respondent. Consequently, the evaluation criteria serve a dual purpose. On the one hand they measure the data quality and on the other hand the confidentiality of the modified data set. The quality criteria can be classified into two categories:

1. **Criteria of data confidentiality**
2. **Criteria of data quality**

In the first category we have the following indicators:

- a) **Value of threshold 'k':** The threshold is the minimum number of units that the micro-aggregation cluster contains. The higher the threshold the higher the data confidentiality and the higher the data perturbation.
- b) **Concentration or predominance rule:** A cluster is defined as confidential if the n largest elements contribute more than k percent to the cluster total. The rule if the threshold is three is that the two largest units must contribute more than 85% of the cell total.
- c) **Indicator of data perturbation:** This indicator compares the value of an observation before and after micro-aggregation for assessing the degree of modification. If the result is lower than a specific bound for a large number of observations, we assume that the micro-aggregation method does not sufficiently maintain the confidentiality of the data set.

In the second category we have the following indicators:

- a) **Use of summary statistics:** One way for examining if the modified data set preserves the characteristics of the original data set is by using summary statistics of position or dispersion. In other words by comparing summary statistics of the original and the micro-aggregated data set we can evaluate the closeness of the two distributions and consequently the data quality. Apart from the univariate structure we can evaluate the preservation of the multivariate structure by comparing the correlation coefficients or other measures of data association.
- b) **Loss of information criteria:** These criteria are usually described by means of analysis of variance. In the case of metric variables the loss of information is defined as the ratio of the within clusters variance to the total variance. A low ratio implies closeness between the original and the modified data set. In the case of qualitative

variables we can use compare the entropy of the original and the modified data set. Again the lower the ratio the closer the two data sets are.

- c) Further processing ability:** This indicator represents the ability of the transformed data set to be used for making new analysis and for obtaining results that are sufficiently close to those that we would have obtained if instead we had used the original data. The closer the results the better the preservation of the structure of the original data.

2.6 Procedures in Eurostat and in Member States

The micro-aggregation procedures are applied in Eurostat in the context of the first and the second innovation survey. With regard to the first innovation survey, quantitative variables are modified according to the individual ranking method using the mean as an aggregation statistic. Ordinal variables are modified using the snake method and the mode as an aggregation statistic. Nominal variables are modified according to the similarity of distributions method. In the second innovation survey some minor changes have taken place regarding the sorting of the data and the arbitrary route that we follow in the snake method.

Concerning the quality criteria, the use of summary statistics, the loss of information criteria and the index of data perturbation have been used.

The following table describes briefly the most frequently used micro-aggregation procedures in Eurostat, Italy, Spain and Russia

Table2.1: Procedures in Eurostat and in European countries concerning the micro-aggregation techniques

	Eurostat (CIS) ³ 12	Italy	Russia	Spain
Method for numeric variables	Individual ranking	Sum of z-scores	Individual ranking	Adaptation of Ward's method
Aggregation statistic	Mean	Mean	Mean	
Method for ordinal variables	-Snake -Snail	Not treated	Not treated	
Aggregation statistic	-Mode -Weighted mode			
Method for nominal variables	Similarities of distributions	Not treated	Not treated	
Aggregation statistic	-Mode -Weighted mode			
Threshold	3	3,4,5 (3 by default)		
Measure of quality and confidentiality	-Deciles, -variances, -marginal distributions, -correlation coefficient			

3. Backward Calculation Techniques

Data currently produced by the National Statistical Institutes (NSI's) are often subject to a revision process. The revision process can be viewed according to each double dimension: routine revisions and occasional revisions. For both of them the main purpose is to achieve better quality of the published data. While the former are regularly made to incorporate new available information, the latter occur at irregular intervals depending on major economic events.

Occasional revisions are produced by NSIs at longer and infrequent intervals. The nature of such revisions may be statistical, when results from changes in surveys or in

³ Community Innovation Survey

estimation procedures, or conceptual, when results from changes in concepts, definitions or classifications. NSIs often face occasional revisions. For example, when they change the reference year for constant price figures, they entail a revision process that involves the entire national accounting framework. From a conceptual point of view, occasional revisions are different because generated by different causes. Causes of this type might be

- Changes due to new surveys
- Changes due to modifications in definitions or interpretations of the European System of Accounts (ESA) and the System of National Accounts (SNA)
- Introduction of new calculation methods

Important economic events that have a big impact on the national accounting system. An example of such an economic event is the introduction of EURO.

The main effect of these revisions is to affect all national accounts. Time series associated with the national accounts aggregates have to be revised according to the new changes in order to be homogeneous. The reconstruction of the national accounts time series is related to a revision process usually referred as backward calculation or retropolation.

Eurostat's Unit B2⁴ is now revising the backward calculation methodologies adopted by some Member States. The reason for that is that Eurostat wants to make suggestions to other Member States that have not developed methodology on the backward calculation techniques to develop one. This is very important especially at this time that there is an increasing demand for homogeneous time-series inside the European Union. Furthermore, Eurostat foresees the development of the same methodology internally. It is apparent that the role of Eurostat in this estimation domain aims at the over space harmonization of the backward calculation methods.

3.1 An Overview of the Backward Calculation Methods

In the backward calculation methods of national accounting data, two archetypes can be distinguished:

⁴ Unit B2 is responsible for Economic Accounts, International Markets; Production and Analyses; Quarterly Accounts and Statistical Coordination of Euro-Indicators.

- **Annual backward calculation**
- **Benchmark years and interpolation**

In both methods several variants are possible and also a combination of both methods is thinkable. The former class of methods is very well known in NSI's and methods belonging to them are currently used. The latter has not been applied intensively till now to revise national accounts series.

3.1.2 Annual Backward Calculation

The annual backward calculation is based on the principle that the retroplated figures are calculated year by year back in time. Several methods can be used to obtain such results. The differences between them depend on the accuracy and on the intensive use of statistical techniques. The most well known techniques belonging to this class of methods are the following:

- Full revision method:** This is a very complete one. Figures to be revised are estimated year by year back in time by applying the same revision rules. Due to its detailed level of accuracy this method is very time consuming and requires much staff.
- Revision by superposition of corrections:** Time series figures concerning the years of the backward calculation period are determined by superposing corrections on the figures before revision. Starting point is the consistent data set of national account, which was compiled in the past. Corrections resulting from the revision process are added to this basic set.
- Growth rates method:** Starting from the balanced set of national accounts figures for the revision year, time-series figures for the past are determined by applying backwards the growth rates associated to the time series before revision.
- Simple proportional method:** The simple proportional method is a simplified version of the annual backward calculation method. The revision year is expressed both under the new and the old accounting system rules. Then in order to reconstruct the past revised values of the series, a simple proportional rule is applied to the old time series values. The simple proportional method offers an easy technique to carry

out backward calculation, especially in a first attempt to determine the new path of the involved time series. Clearly, it is an approximate solution that does not analyze in a very deep way the revision effects on time-series but on the contrary is a low resource and time consumption approach to the backward calculation.

3.1.3 Benchmark Years and Interpolation

The second group of basic revision techniques is based on a two-stage procedure. In the first step detailed estimates are calculated for one or more benchmark years. In the second step, figures for the remaining years are determined by means of interpolation.

The benchmark years are crucial time points where much information is contained. The revision year (i.e. the year that the revision starts) is an example of a benchmark year. The corrections that carried out for the revision year have to be determined also for the other benchmark years. Corrections for the intermediate years are calculated by means of interpolation. However, the estimation procedure for the intermediate years differs from Member State to Member State.

Two methods can be classified into the broader class of the benchmark years and interpolation revision processes.

- a) Full benchmark year:** In this method figures for the benchmark years are estimated in a detailed way using the new definitions and classifications. The figures for the intermediate years are interpolated. The level of detail is chosen according to the level of detail of the revision of the benchmark years.
- b) Layer correction method:** Revisions for the benchmark years are determined. Corrections resulting from revisions are balanced and then superposed to the basic set of original data. In this way layers of correction matrices become available. Corrections for the intermediate years are determined by means of interpolation of the correction matrices. Afterwards, figures for the intermediate years are determined by integrating the original data set and the corrections.

3.2 The Backward Calculation Methods in Member States

The Dutch case: In the Netherlands the various methods for backward calculation have been used extensively. The annual backward calculation has been used for

compiling time-series 1977-1985, consistent with the 1987 revision. The revision by superposition of corrections method was used for compiling the time-series 1969-1977, following the 1977 revision. The growth rates method was used for compiling the time-series 1969-1976 following the 1987 revision. The simple proportional method was originally planned for the time-series 1948-68 according to the 1987 revision. The full benchmark year method was originally planned for compiling time-series for the years 1969-1986 following the 1987 revision. However, in the end the annual backward calculation method and especially the growth rates method were used.

Recently, Statistics Netherlands decided to change the method of backward calculation. The new method is a variant of the layer correction method belonging to the benchmark years and interpolation category.

In the Dutch case the corrections for the intermediate years are calculated with the help of an interpolation procedure that has been developed for this purpose Kazemier (1997). The interpolation is carried out within the framework of the input-output tables. To start with, the corrections for the benchmark years are expressed as a percentage of the values before the correction. This means, if we take a column of the input-output table as an example, that for each cell of the column a correction percentage is determined. These percentages are interpolated between the benchmark years by assuming a linear pattern. Next, the percentages are applied to the values before the corrections. In this way the before revision structures are kept and the developments of the variables before revision are taken into account.

The French case: The French approach to the estimation of the intermediate years is based on a linear model that links the variable in the new accounting system and the variables of the accounting system before revision. The estimates of this linear model are obtained by applying the Kalman -filter algorithm.

The principle underlying the retropolation in the French case is as follows. We are trying to work out for a period of T years the measurement of an item that is known for only k years ($k < T$) in two accounting systems simultaneously and for T years in one of the systems. For this purpose France and more specifically the Institut National de la Statistique et des Etudes Economiques (INSEE) proposes to use a linear model linking the variables in the two accounting systems and from this to calculate the conditional (or

linear) expectation of the non-observed variables as a function of the set of values available under the former system.

Let x_t be a vector of the observed values available at date t of N accounting magnitudes obtained in the former base. Assume now that a new measure X_t is introduced. Our objective is to construct the time-series for the past according to the new measure X . In other words what we are looking for, is to calculate the following conditional expectation $E(x_t | x_1, x_2, x_T, \dots, x_{T+k}, X_{T+1}, \dots, X_{T+k})$.

A retropolation program was developed to provide automated estimates. French retropolation program is supported by SAS and IML⁵ and it is developed for MVS operating system. The software provides a number of statistics for the evaluation of the candidate models. Statistics of this type are the standard deviations of the parameter estimations (the smaller it is the better), the standard deviation of the period of analysis (the smaller it is the smaller the degree of imprecision), the Bayes information Criterion for assessing how good is the fit of the estimated models as well as a number of other criteria for evaluating the convergence of the algorithm. Moreover the software produces graphs of the retropolated series and the old base series which provide an additional help in the evaluation of the candidate models. The big advantage of this software is that by using it we are in the position to evaluate all candidate models and decide which is the optimal one. However, we must be very careful when speaking for maximization of the likelihood or for convergence because instead of a global maximum we might find a local one. Also the evaluation of the candidate models is based on some statistical criteria but these are not the only ones. The best solution in this case is to adopt a model choice approach i.e. run different models and decide which is the best.

4. Aspects of Sampling Procedures

The main scopes of the analysis of this estimation domain are the following: (1) to give a description of the most frequently used sample designs and weighting procedures and (2) to investigate the progress that has been made towards the harmonisation of the measurement procedures. The first scope is connected with estimation procedures that

⁵ Interactive Matrix Language

are being applied at the national level but are also supervised by Eurostat. The second scope is connected with the harmonisation attempts that are co-ordinated both by Eurostat and the Member States. The complexity of this domain is its multinational character due to the fact that the sampling procedures are applied at the national level. Consequently, in order to achieve the targets of our analysis, we have decided to use as exploratory tools three important sample surveys that are applied in every Member State. These are the labour force survey, the European Community household panel survey and the household budget survey.

4.1 The Labour Force Survey

The labour force survey is an inquiry designed to obtain information on the labour market and related issues by collecting data through administrative records and sample surveys of households and enterprises.

Concerning this survey the majority of the Member States use multistage stratified clustered sample designs. Examples of countries that use such designs are Belgium, Greece, Spain, Germany, Ireland, Italy France and Netherlands. However there are cases where element sample designs are used. This happens in Finland, Sweden and Denmark. More specifically Finland and Sweden use a stratified sample design with systematic selection of units while Denmark uses a stratified sample design with random selection of the units. As far as it concerns the weighting procedures we have found that all countries except Greece use the post-stratification method.

Perfect comparability is difficult to be achieved. The over-space and over-time comparability is achieved mainly in terms of the measured concepts and not of the measurement procedures. In other words, the record of the same characteristics, the close correspondence between the list of questions, the use of the same definitions and classifications for all countries, the grade stability of the content of the survey and the higher frequency of the surveys have attributed the maximum to both dimensions of comparability. However, there are some problems mainly due to modifications in the sample designs and in the order of the questions or due to changes in the reference period and in the adjustment procedures.

4.2 The European Community Household Panel Survey

The European Community household panel is a standardised survey conducted in Member States of the European Union under auspices of Eurostat i.e. Unit E2 (living conditions and social protection). The survey involves annual interviewing of a representative panel of households and individuals in each country, covering a wide range of topics of living conditions. It was established in response to the increasing demand for comparable information across the Member States on income, work and employment, poverty and social exclusion, housing, health and many other diverse social indicators concerning living conditions of private households and persons.

A major aim of the survey is to provide an up-to-date and comparable data source on personal incomes. Information on incomes of households and persons is indispensable for policy-makers at the national and European levels. The survey provides detailed information at the individual and household levels on a variety of income sources: wage income, rent subsidies, unemployment and sickness benefits, social assistant benefits, occupational and private pensions. Hence the strength of the ECHP is its focus on the income received from all sources by every member of the household.

Three characteristics make this survey a unique source of information. (a) Its' multidimensional character which covers a range of topics simultaneously, (b) a standardised methodology and procedures yielding comparable information across countries and (c) a longitudinal design in which information on the same sample of households is gathered in order to study changes over time at the micro-level. Especially for the second characteristic, comparability is achieved through a standardised design and common technical implementation of the procedures with a centralised support by Eurostat. More specifically these procedures are: (1) a common survey structure and procedures, (2) common standards for data processing and statistical analysis including editing, variable construction, weighting, imputation and variance computation, (3) common sampling requirements and standards, (4) common frameworks of analysis through a collaborative network of researchers and (5) common "blue-print" questionnaire which serve as the point of departure for all national surveys.

All surveys in ECHP are based on probability sampling. Most of the surveys are based on a two-stage sampling i.e. a selection of sample areas in the first stage, followed

by the selection of a small number of addresses or households at the second stage within each selected area. However, there are cases where a single stage sample is drawn or even a three-stage sample (i.e. large areas → smaller clusters → addresses of households). Diverse criteria are used for the stratification of area units before selection. The most common criterion is the geographical region and/or urban-rural classification. Stratification by population size and other social indicators is also used in some countries. Within explicit strata, areas can be selected systematically or randomly.

The recommended cross sectional weighting procedure is step by step one that involves the following steps: **(a) Design weights:** The design weights are introduced to compensate for differences in the probabilities of selection into the sample. The weight given to each household is inversely proportional to its probability of selection. With multi-stage sample design, the design weight refers to the overall selection probabilities of the households. **(b) Non response weights:** These weights are introduced in order to reduce the effect of differences in unit responses rates in different parts of the sample and are based on characteristics which are known for responding as well for non-responding households. Weighting for non-response involves the division of the sample into certain appropriate weighting classes, and the weighting of the responding units inversely to their response rate. More specifically, the following steps are involved:

- Division of the sample groups into groups of weighting classes, showing the number of households that selected and the number of households that interviews were completed in each class.

- Computation of the response rate for each category in the classification.

- Assignment of a uniform weight to all households in a category, in inverse proportion to the response rate in this category.

- Normalisation the weights.

(c) Weights correcting for the distribution of households: The third step in the weighting procedure has the purpose of adjusting the distribution of households by various characteristics, so as to make the sample distribution agree with the same from some more reliable external sources. The distribution in various categories may involve

numbers of households and/or aggregates of some variables measured on each household. In order to achieve this, a raking or a generalised raking technique is used.

(d) Weights correcting for the distribution of persons: The next step in the ECHP weighting procedure involves the adjustment of the household-level weights so as to make the distribution according to certain characteristics covered in the sample agree with the same from some more reliable external source. More specifically, this step controls the distribution of persons (with completed interview) according to external information. The most important controls that are applied in the ECHP are by age (normally in 5-year groups) and sex.

(e) Final weights: At the end of the weighting procedure, each household receives a weight which, is the product of the weights assigned sequentially to it at the four previously described steps. Any extreme weights are trimmed so as to avoid large inflation in variances. Finally, after each step, the weights are normalised.

As we have already mentioned, the unique feature of the ECHP is its panel design. Within each country, the original sample of households and persons is followed over time at annual intervals. In this manner, the sample reflect demographic changes in the population and continues to remain representative of the population over time. At each wave t , the basic weight $u_{i,k}^t$ is obtained by adjusting the basic weight of the previous wave $u_{i,k}^{t-1}$. This adjustment is made by using an adjustment factor f that depends on changes in the population and the samples between waves $t-1$ and t : $u_{i,k}^t = u_{i,k}^{t-1} f_{i,k}^{(t-1 \rightarrow t)}$

The following adjustments are involved: (1) Adjustments for non-response. Weights for wave $t-1$ are multiplied by a factor inversely proportional to the person's propensity to respond at wave t after having responded to wave $t-1$. (2) At each wave the weights may be also adjusted for consistency with external information on the distribution of the population. (3) Large variations in the weights is avoided by constrained the adjustment factor within practical limits. (4) Children born between $t-1$ and t are given appropriate weights determined by the basic weights of their parents. Sample persons from wave $t-1$ no longer alive are excluded from the survey while all non-sample persons have a zero basic weight.

4.3 The Household Budget Survey

The household budget surveys are sample surveys carried out in the Member States of the European Union. The information collected through these surveys covers a wide variety of items concerning consumption expenditure, income of private households, savings and indebtedness, and other household characteristics.

The narrow purpose of the conducting household budget surveys is to obtain the weights that are used for the construction of the consumer price indices and the cost of living indices. The purpose of conducting a household budget survey in a more broad sense is to give a picture of living conditions of private households in a defined area and time. In this sense, the aim of the survey is to give a precise picture of the household's total consumption broken down in sufficient detail as a function of household characteristics such as income, social and economic characteristics, size and composition, degree of urbanisation, and region. The surveys also provide information about the living and expenditure levels.

Despite the common focus of the surveys on the study of the patterns of consumption of private households, the national household budget surveys present a diversity of structures and designs as well as differences in the topics covered. In order to reduce this diversity Eurostat gives specific recommendations on many technical issues. Harmonisation attempts of this type are: (1) the emphasis that is given on defining a common approach to measure the consumption expenditure, and the income of the households, (2) the agreement for the use of a common set of variables and (3) the agreement for the use of a harmonised nomenclature regarding consumption.

Eurostat is committed in assisting Member States, as well as other interested countries, for the improvement of their survey methods through the provision of guidelines and direct technical support. Generally, Eurostat does not emphasize the use of same questions or survey structure. The major importance is put into the use of harmonising concepts and definitions. There are two reasons for this. Firstly it could be argued that the adoption of certain definitions and concepts is not specifically based on national circumstances but has a more universal character and secondly the use of certain concepts and definitions can be justified by the nature of the survey. Examples of

concepts that Eurostat seeks to harmonise are the consumption expenditure and the income of the households.

Probability sampling is used in the majority of the household budget surveys. Most sample designs involve the selection of the sample in multiple stages. There are a number of advantages in using multistage sampling. By concentrating the units to be enumerated, the travel costs and other costs of data collection are reduced. For the same reason, it can improve the coverage, supervision, control, follow up and other aspects determining the response rates and the quality of the data collected. The major disadvantage is the complexity of design and analysis. The most common practice is the use of a two-stage design. First, a stratified sample of suitable area units is selected, typically with probabilities proportional to size after stratification by region, social-economic status of the reference person and household type or size. The second stage consists of the selection, within each sample area, of households and addresses for inclusion in the survey. However, there are cases where a three-stage sampling is performed. This design involves a two area stages: first some larger areas such as localities are selected as the primary sampling units; then one or more smaller areas are selected from the primary sampling units and lastly, the household or addresses are selected.

In most cases the entire population is sampled uniformly and the sample is distributed proportionately across different areas and population groups in the country. However, in some surveys, population groups of special interest such as farmers or poorer households are over-sampled.

The need to weight data from the household budget survey is generally recognised. Two types of weighting procedures are proposed. (1) Spatial weighting and (2) temporal weighting.

The spatial weighting aims at improving the quality of the sample in relation to the size, distribution and characteristics of the population, selection probabilities and survey non-response. The procedure recommended by Eurostat is a multi-stage one similar to the one described for the European Community household panel survey. Firstly, weights inversely proportional to the probability of selection are calculated, next weights that take into account the non-response are introduced and finally weights correcting for the

distribution of households are computed. The final weights are worked out as the product of the weights of the previous stages.

The temporal weighting of the data stems from the fact that the household observation period is often different from the reference period.

5. Conclusions

There are two features that this paper aims to study. The first and most apparent target is to describe and compare theoretically some of the estimation procedures discussed or used by Eurostat and to make proposals, wherever this is possible, for the adoption of better practices. The second target is less evident but also of great importance. This target concerns the over space harmonisation of the estimation procedures. The idea of harmonisation mainly refers to those procedures that are being applied at the Member State level.

As far as the first target is concerned, in this paper we have described three estimation domains. These are: the micro-aggregation techniques for producing confidential data, the backward calculation methods for constructing homogeneous time series and some aspects of the sampling techniques discussed by Eurostat and applied by the Member States.

For the micro-aggregation techniques we have concluded that there is a need for the adoption of more complex methods like the improvement of the Hanani's algorithm and the modified Ward's criterion. These methods must be incorporated into the micro-aggregation software that Eurostat develops at this time point. Furthermore, we have argued that an enrichment of the evaluation criteria is required. This means that more criteria aim at assessing the confidentiality aspect of the modified data set as well as the violation of the structure of the original data set must also be incorporated into the new micro-aggregation software. Criteria of both types are the indicator of data perturbation, the examination of the correlation coefficient of the initial and the modified data set and the further processing ability. Finally, we have proposed the adoption of a model choice approach for those that are being involved into the micro-aggregation process. This means that they have to apply different micro-aggregation techniques and different evaluation criteria in order to arrive at an "optimal" solution. Using the term "optimal"

we mean a solution that is acceptable, from the confidentiality point of view, by the Member States and useful, from quality point of view, for the independent analysts.

The second domain of estimation concerns the backward calculation techniques. For this domain we have described generally the backward calculation techniques and we have mainly focused in the models applied by the Netherlands and France. We recognised that both countries apply the benchmark/interpolation technique. The difference among the two approaches is in the backward calculation of the intermediate years. While the Netherlands use a linear pattern for between years interpolation, France uses the Kalman filter algorithm. It is common belief that the best method for the backward calculation is the annual backward calculation. However, given the time and staff constraints, the Member States can resort to the benchmark years and interpolation method but only under certain quality requirements.

The third domain of estimation concerns some aspects of the sampling techniques used by the Member States. The complexity of this domain is connected with its the multinational character. As a result, the three surveys that we have chosen to describe make easier the study of this complex and multivariate domain. Two aspects of the surveys have been selected for description i.e. the sample designs and the weighting procedures. In other words, we have tried to identify the most frequently used sample designs and weighting procedures. With respect to the sample designs we have found that in general the Member States are in favor of clustered samples. More specifically, the Member States usually use multistage stratified cluster samples with probability proportional to size selection for the primary units. However, there are cases where one-stage element samples are used. These designs involve stratified samples with random or systematic selection of the units within each stratum. As far as it concerns the weighting procedures, we have found that the most frequently used are the post-stratification and the raking ratio adjustment procedures.

As it has already been stated, the second feature that this paper aims to describe, is the progress in the domain of the over space harmonisation of the estimation procedures. First of all, we have to stress the importance of harmonisation as an indicator of quality. This means that the more harmonised the methods used by the Member States are, the

easiest the comparisons between the different results and consequently the better the evaluation of the quality of the results is.

The role of Eurostat regarding the harmonisation procedure can be viewed in two dimensions. The first dimension is this under which Eurostat revises the methodology on a specific topic, makes proposals to the Member States that have no methodology to develop one on that topic and at the same time sets internally the bases for applying the methodology. An example is the role of Eurostat in the domain of backward calculation methods. We have already said that Eurostat revises the backward calculation techniques and the existing models (Dutch and French backward calculation models) in order to be in a position to make proposals to countries that have no methodology to develop one. At the same time Eurostat acts so as to develop internally the same methodology. It is apparent that the role of Eurostat aims at the better harmonisation of the backward calculation procedures. It is also apparent that the internal development of the same methodology aims at the better evaluation of the quality of the micro results that are transmitted to Eurostat.

Under the second dimension, every Member State has already developed an estimation procedure on a topic. In this case Eurostat receives micro (individual) results which have to be assessed for their quality, and to be composed in order to produce the macro (aggregated results). It is apparent that the evaluation of data quality as well as the composition of the individual results become more difficult due to the multinational character of these estimation domains. Consequently, the harmonisation in terms of using common definitions, tools and measurement procedures is important. The role of Eurostat in this case is to study and describe the different definitions, tools and methods and to be able to propose the adoption of better and more harmonised practices. Some examples of this role of Eurostat have been studied in the part concerning the sampling techniques.

Under the first dimension the role of Eurostat has more degrees of freedom. Eurostat should identify estimation domains where Member States have not developed yet a methodological approach, study the existed estimation procedures and make proposals in the direction of harmonisation.

Under the second dimension the role of Eurostat is more restricted. In this case, the study of the different methodologies plays a central role. Only by analysing the different

methodologies Eurostat will be in the position to identify those cases where the heterogeneity problem is severe and to make proposals in the direction of the better harmonisation.

This paper aims at the evaluation of the effectiveness of the estimation procedures, the comparison of the different procedures, the development of quality criteria for the evaluation of the results in each estimation domain as well as the investigation of the progress concerning the harmonisation of the estimation methods. Setting the same scopes in the future, this work must be also expanded into the other domains of estimation in Eurostat.

References

- Anwar, N. (1993).** Micro-aggregation: The small aggregates method. Internal report, *Direction D, Eurostat, Luxembourg.*
- Anwar, N. and Defays, D. (1997).** Masking microdata using micro-aggregation. *Proceedings of the Seminar on Data Protection, Luxembourg.*
- Anwar, N. and Defays, D. (1995).** Micro-aggregation: A generic method. *Proceedings of the 2nd International Symposium on Statistical Confidentiality, Eurostat, Luxembourg.*
- Ariane II (1999).** Study report for the development of a system of micro-aggregation of confidential data. *Luxembourg.*
- Barcellan, R.** Backward calculation of national accounts data. Internal Report, *Direction B, Eurostat, Luxembourg.*
- Defays, D. (1997).** Protecting microdata by micro-aggregation: The experience in Eurostat. Internal Report, *Direction A, Eurostat, Luxembourg.*
- Defays, D., Kokolakis, G. and Nanopoulos, P. (2000).** Clustering analysis under a cardinality constraint. *Journal of Classification* (to appear).
- Defays, D. and Nanopoulos, P. (1992).** Panels of enterprises and confidentiality-the small aggregates method. *Proceedings of 1992 symposium on design and analysis of longitudinal surveys, Ottawa, Statistics Canada.*
- Den Bakker, P. G. and Van Rooijen, A.M.R. (1997).** Backward calculation of Dutch national accounting data. Lessons from the past towards a new approach. Department of National Accounts, *Statistics Netherlands.*
- Direction des etudes et syntheses economiques. (1997).** Backward calculation of national accounts. Department des Comptes Nationaux, *INSEE.*
- Eurostat, Direction E. (1999).** Labour force survey, methods and definitions. Internal Publication, *Eurostat, Luxembourg.*
- Eurostat, Direction E. (1999).** The European Community household panel, survey methodology and implementation. Internal Publication, *Eurostat, Luxembourg.*
- Eurostat Direction E. (1999).** Household budget surveys in the European Union, methodology and recommendations for harmonisation. Internal Publication, *Eurostat, Luxembourg.*

- First Community Innovation Survey (1992).** The micro-aggregation process used in the first Community innovation survey. Internal Report, *Direction A, Eurostat, Luxembourg.*
- Kazemier, B. (1997).** Time-series revision. Department of National Accounts, *Statistics Netherlands.*
- 36th Meeting of the Statistical Programme Committee (2000).** Progress report of the internal task force on the procedures for producing internal publication and publishing estimations at Eurostat. *Eurostat, Item 6b of the agenda, Luxembourg.*
- Strudler, M., Lock Oh, H. and Scheuren, F. (1986).** Protection of taxpayer confidentiality with respect to the tax model. *American Statistical Association, Proceedings of the Section on Survey Research Methods, Alexandria.*
- Tzavidis, T., N., (2000).** Aspects of Estimation Procedures at Eurostat with Some Emphasis on the Over-Space Harmonisation, M.Sc. Thesis, Athens University of Economics and Business, ISBN 960-7929-49-7