

# Use of Proximity Data for the Assessment of Geographical Comparability

Michalis Petrakos<sup>1</sup>, Håkan Lindén<sup>2</sup>, Photis Stavropoulos<sup>1</sup>, George Petrakos<sup>1</sup>

## Introduction

One of the major uses of statistical figures reported by National Competent Authorities (NCAs) is to permit comparisons between countries. Quite often statistics from a country or a group of countries is presented along with those of other ones so as to make the reported numbers more meaningful. Another important use is the compilation of figures referring to groups of countries; an overall unemployment rate for the EU, for example, is compiled by member states' unemployment rates. In all such cases the users assume that these numbers are comparable with each other but this is not always or completely true. In both occasions it is quite useful to be able to assess the level of comparability between pairs of statistics coming from different countries.

The issue of geographical comparability (like comparability in general) arises whenever two or more statistical figures, describing, at least in intention, the same statistical characteristic in different geographical areas, are examined in relation to the others. A reasonable definition is to say it is the extent to which differences between figures are attributed to differences between the true values of the statistical characteristic. For example, is unemployment in, say, Greece really two percentage points greater than Spain's or some of this difference is caused by the respective rates not being fully comparable? If the latter is true, what is the true difference in unemployment between the two countries?

The discussion above makes clear that comparability is a relative quality of a statistic. It cannot be defined if not in relation to another statistic (a standard or a similar quantity from another country). Moreover, comparability, or the lack of it, may be the result of a multitude of factors. An NCA cannot know in advance in what comparisons a statistical figure it produces is going to be used, and which factors will affect each comparison.

The factors that may cause two statistical figures to lose comparability are attributes of the surveys that produce them. These attributes may be grouped into two major categories: concepts of the survey and measurement methodology.

Each survey has a strictly defined objective. To measure or estimate, as accurately as possible, a statistical characteristic of a population. In order to plan the survey a lot of entities must be defined in advance. Such entities include the reference population, the characteristic itself, the classes of a classification of the population, etc. There are more entities which are presented in a list further below. If two surveys of the same characteristic, in different countries, do not use exactly the same definitions, their

---

<sup>1</sup> Agilis SA, Greece. Corresponding author's e-mail [michalis.petrakos@agilis-sa.gr](mailto:michalis.petrakos@agilis-sa.gr)

<sup>2</sup> Eurostat.

statistical products will lack comparability. The extent of the lack will be directly related to the difference in definitions.

Errors also creep in statistical figures because of imperfections in the conduct of surveys. Sampling is an additional source of errors (sampling variation) in sample surveys. These errors affect the accuracy of figures, which in turn may render figures not comparable to similar ones from other countries, since one cannot know what proportion of a difference is due to the sampling variation of estimates. There is a multitude of aspects of a survey that can affect sampling and non-sampling error. Some of them include the sample design, the condition of the sampling frame or master list, the training of the interviewers, the handling of non-response, the quality of data entry, the procedures used for data editing, etc.

In order to measure the effect of differences several studies that examine specific statistics from small groups of countries (usually two) have been conducted in the past decade. They have given us a lot of insight about how comparable international statistics are, what affects their comparability and, quite importantly, by how much. Dalen (1998) used simulations, to examine the figures for the CPI of Sweden and Finland under. This study showed how changes of conceptual aspects affect the CPI and since it was carried out for a number of years (albeit only for Sweden and Finland) it allows for some cautious generalisation. Sorrentino (2000) examined if the unemployment rates of USA, Canada, EU aggregates, and six European countries differ, even after adjustments for some conceptual differences, because of differences in the definition of unemployment and the implementation of the ILO guidelines. The adjusted figures were compared with reported statistics (for Europe this was done for the spring 1998 LFS). When US definitions were used, EU unemployment was found lower by 0.4 percentage points compared to that estimated using EU concepts. Interestingly enough, the US unemployment rate, did not change when measured with EU concepts. US-Canadian difference due to concepts was close to 0.8 of a percentage point. It should be noted, however, that the findings apply to the specific indicator only and also are valid for only a year and a particular pair of geographical regions. A review of similar studies up until 1999 is contained in Decoster (2000).

These studies directly measure the lack of comparability that can be attributed to specific conceptual differences. Unlike other measures of data quality they not only quantify the cause of the problem but the problem itself. They are therefore quite valuable for any user of international statistics. They have a few limitations however. They normally require access to microdata in order to perform the adjustments and therefore can be done by NSIs or Eurostat only and the results are valid for the specific pairs of countries and reference periods that were studied. They also take into account conceptual differences only and do not consider differences in data collection and processing methodologies. We may therefore say that although they provide high quality information this is bound to be sparse since the number of possible combinations in statistics, countries and years is huge.

In this work we present a conceptually simpler approach that uses statistical metadata and is based on a much lower level of measurement and more significantly measures the causes of lack of comparability and not the lack itself. It can moreover be performed by third parties since it relies on public information and therefore can be

cheaply implemented in any type of statistics from any country provided that they are well documented.

## ***Outline of using metadata scoring for the assessment of geographical comparability***

Statistical figures may be compared by comparing the respective metadata of the surveys that produced them. If the associated metadata are similar, the figures will be comparable. Any differences in the metadata must be quantified. The first step therefore is to determine a comprehensive list of aspects that can cause incomparability.

### **Determining a List of Sources of Lack of Geographical Comparability**

The basis of the proposed analysis is a list of sources of lack of comparability. The two principal characteristics of the list are to contain metadata items which may affect the comparability of the results and which are available for all countries under consideration in a harmonised manner. We have divided them in two categories, items that are related to Concepts (mainly definitions) and items that are related to Measurement. Following is a rather generic list of such items that can be adapted to many domains of application by adding aspects that may affect comparability and are pertinent to the domain (items 1.8 and 2.6).

- 1 Concepts
  - 1.1 Statistical characteristics
  - 1.2 Statistical measure (indicator)
  - 1.3 Statistical unit
  - 1.4 Reference population
  - 1.5 Reference period
  - 1.6 Study domains
  - 1.7 International Standards
  - 1.8 Conceptual aspects specific for a Domain under study
- 2 Measurement
  - 2.1 Sampling design
  - 2.2 Sampling errors
  - 2.3 Data collection
  - 2.4 Data processing
  - 2.5 Estimation
  - 2.6 Measurement Domain Specific (a list of characteristics pertinent to any specific domain, e.g. thresholds for Foreign Trade Statistics)

This list very much follows the structure of sources for lack of comparability in Linden (1999) with a second level adding to a total of 40 entries (the full taxonomy is provided as Annex 1). Some of the information for these entries are contained in quality reports or methodological notes provided along with European Statistics. It should be noted that this is just an indicative list and it should be modified in order to address comparability issues for a specific domain (some entries may be irrelevant - like sampling errors for censuses- and some may require more detailed analysis).

Different lists can be developed for international comparisons based on IMF's SDDS or other metadata standards.

## The scoring Scheme

A scoring scheme should be introduced which assigns a score to each metadata difference between statistics from two countries according to its effect on comparability. In practise, since most of the methodological information is textual (i.e. not machine understandable) a person will assign the weights after assessing the information provided. Here bilateral studies can provide a lot of insight and help to correctly appraise the weights. When little or no information on the severity of the effects of methodological differences on comparability exists then one can just record if there is a methodological difference that can potentially affect comparability or not, opting for a weighting scheme like:

$$w_{i,j}^k = \begin{cases} 0 & \text{no difference} \\ 1 & \text{a difference exists} \end{cases}$$

Where  $w$  is weighting factor (score) for every methodological aspect  $k$  reported from countries  $i, j$ .

Evidently, the effects of methodological differences can be wide ranging. Some might not affect comparability while some others (like differences in definitions) might have severe consequences regarding data comparability. To measure this a more elaborate scoring scheme may be introduced like with a weighting factor like:

$$w_{i,j}^k = \begin{cases} 0 & \text{small or no effect} \\ 0.5 & \text{medium effect} \\ 1.0 & \text{large effect} \end{cases}$$

Subjectivity exists in both schemes but is more evident in the latter scheme since someone must decide what is a small, a medium and a large difference (instead of whether there is a difference that affects comparability or not). In any case the resulting score can be treated as the methodological distance between the respective countries.

## Producing Comparability Indicators

Several indicators can be computed from this semi-quantitative scheme. First a coefficient, that indicates how different the methodologies used by two countries  $i, j$  are can be computed by summing the weights for all methodological entries  $M$

$$P_{ij} = \frac{\sum_{m=1}^M w_{ijm}}{M} \quad (1)$$

If a number  $g$  of metadata are missing (not provided) from one of the two countries then, obviously, a comparison cannot be made; for these aspects the corresponding

weights will be zero and we will evaluate M-g cases and change the denominator accordingly.

An indicator of how much geographically comparable a country j is, can be derived by averaging its methodological distances  $p_{ij}$  with respect to the group of countries included in the study.

$$p_j = \frac{\sum_{i \neq j} p_{ij}}{N-1} \quad (2)$$

An overall coefficient that reveals how geographically comparable is the whole set of country statistics is,

$$p = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N p_{ij}}{\binom{N}{2}} \quad (3)$$

### **Analysis of the Dissimilarity Matrix to assess Comparability**

For pair wise studies of geographical comparability among a group of countries, the scoring scheme will lead to a two-way symmetrical matrix with the countries defining its rows and columns. The entry of each cell will be the comparison score of the two countries to which the cell corresponds. This matrix can also be used as input for further analysis. Applying Multidimensional scaling methods for example, will produce a two or three dimensional scatter diagram of countries where the relative magnitude of the distances between them is proportional to their methodological differences. The closest two countries are on the map the more comparable their statistical figures are.

Tree clustering methods can also be used with the same dissimilarity matrix as an input to create a tree diagram where the branches include countries with similar methodologies. The diagram should reveal clusters of countries that follow similar methodologies, if they exist.

This approach may be of great interest for international organisations by, for example, Eurostat or OECD. They can for example compare the Labour Force surveys of the member states and produce a score matrix or a scatter diagram which will show how harmonized the surveys are and therefore how comparable are their statistics. The same approach can be used for assessing the comparability of statistical figures from candidate EU member states. The methodology can be also quite useful in relatively new series when methodologies are not yet mature and different approaches are implemented at the same time. By identifying methodological clusters Working Groups or other parties with an interest in developing a harmonised methodology for the statistics under consideration can decide which methodological choices should be adopted for the EU area.

## ***Proximity data derived from official statistics (Mirror Statistics)***

Mirror statistics can be obtained when similar statistical characteristics are measured by different NCAs. They are common practice in the measurement of flows (like Foreign trade, tourism and transportation statistics). Although they are mainly used for the assessment of data consistency, which is associated with coherence, the geographical comparability of a source can be indicated by looking at the size of discrepancies.

While in the previous case, we studied methodological differences in order to assess the comparability of data, in this case we study differences in the data in order to assess the comparability of methodologies.

Data typically consists of two matrices with inbound flows in one and outbound flows in the other. Matrices on tourist movement between EU member states will show the numbers of tourists leaving each state for the other states and the number of tourists each state receives from other states. One can then look, for example, at how many tourists does Greece report as leaving it for Austria and at how many tourists does Austria report as receiving from Greece.

Absolute differences of the two reported figures for each direction of flows, for a pair of countries, can be normalized by dividing with the average reported quantity. This results into two dissimilarity matrices one for each direction of flows. These can then be averaged up into a single figure thus yielding an indicator of discrepancies between the two countries. Since only the size is of importance at this stage the absolute values are retained. This way a dissimilarity matrix is constructed that can provide indicators and input for analysis in the same way as in the previous paragraphs. By identifying what is common in the countries with small discrepancies and what differs in countries that are far apart in the multidimensional scaling map we can make some suggestions on what might contribute to the discrepancies. This can be beneficial since by establishing statistics that are consistent we can confidently move one step further and abolish the double collection scheme (and have each country reporting one flow only) thus offering big savings to the European taxpayer.

## **Case studies**

### ***A case study of Labour force statistics.***

As a case study we analysed metadata reported to the OECD and included in the Standards and methods publication for indicators about employment in industry. For the purposes of the study we devised a list of causes for lack of comparability that follows OECD's standard list of metadata items with two levels, the first of which is:

Table 1. First level of the taxonomy of causes of lack of comparability based on OECD's standard list of Metadata items.

Code	Name	Description
1	Source	Characteristics of the source like available breakdown, periodicity.

2	Concepts & Coverage	Definitions, reference period, statistical population/exclusions. Classification and Geographical coverage.
3	Standards	Followed standards, Classifications, Departures from International standards.
4	Data collection	Reporting units & method, Master list.
5	Data Manipulation	Aggregation, Grossing-up, seasonal adjustment.
6	Quality	Sampling error, measurement error, and non-response rate.
7	Other	Not elsewhere mentioned

The full taxonomy is shown in Annex 2.

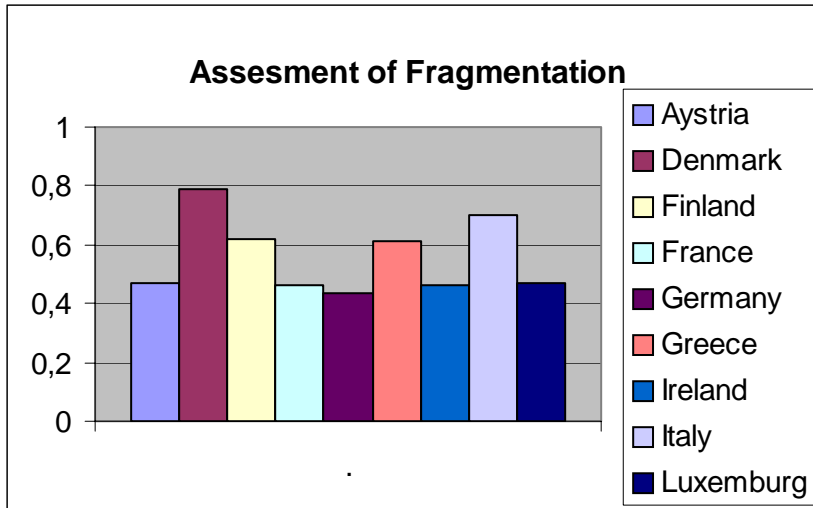
We also used a binary scoring scheme to judge if there is a methodological difference between two countries with respect to a specific entry (at the second level of the taxonomy).

The symmetrical table of the  $p_{ij}$  measure of dissimilarity of Eq. 1 between 9 European countries reporting to the OECD is shown in Table 3.

Table 2. Table of methodological distances.

	Austria	Denmark	Finland	France	Germany	Greece	Ireland	Italy	Lux
Austria	0,000	0,733	0,615	0,400	0,167	0,364	0,429	0,769	0,273
Denmark	0,733	0,000	0,929	0,667	0,750	0,917	0,786	0,714	0,833
Finland	0,615	0,929	0,000	0,615	0,667	0,800	0,538	0,167	0,600
France	0,400	0,667	0,615	0,000	0,250	0,500	0,200	0,714	0,333
Germany	0,167	0,750	0,667	0,250	0,000	0,500	0,300	0,727	0,125
Greece	0,364	0,917	0,800	0,500	0,500	0,000	0,500	0,900	0,400
Ireland	0,429	0,786	0,538	0,200	0,300	0,500	0,000	0,692	0,273
Italy	0,769	0,714	0,167	0,714	0,727	0,900	0,692	0,000	0,900
Lux	0,273	0,833	0,600	0,333	0,125	0,400	0,273	0,900	0,000
Average	0,469	0,791	0,616	0,460	0,436	0,610	0,465	0,698	0,467

Apparently Denmark is quite distant from all other countries and Italy and Finland are located close to each other and distant from everybody else. This in fact reflects reality quite well since the rest of the countries report results from Industrial surveys where companies are required to report the personnel they were employing in the reference period. Italy and Finland report results from their Labour Force Surveys covering households, while Denmark in a remarkably cost efficient way reports social security contributions from companies divided by the required contribution for a single employee. Average dissimilarities reflected by the coefficient of Eq. 2 are shown in Figure 1.

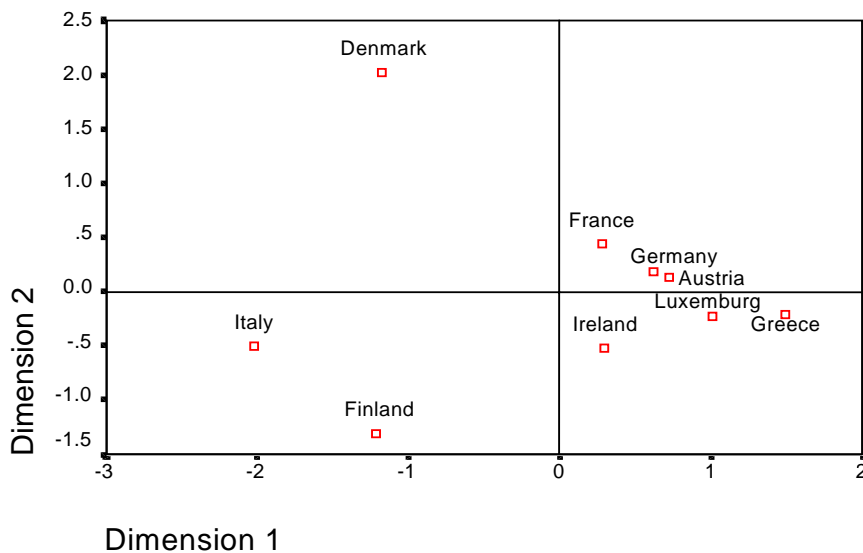


**Figure 1. Average dissimilarities for all countries under consideration**

Since a matrix of dissimilarities is at hand a two dimensional map of practices can be drawn up using multidimensional scaling. The resulting map is shown in Figure 2. A three-dimensional multidimensional scaling reveals the differences between the three different main methodological choices mentioned before but also indicates that Greece differs in the way it implements the industrial survey from the rest (owing to the classification used, ISIC 68 and population exclusions).

### A Map of statistical methodologies

#### Euclidean distance



**Figure 2. A Map of concepts and practices drawn using Multidimensional scaling.**

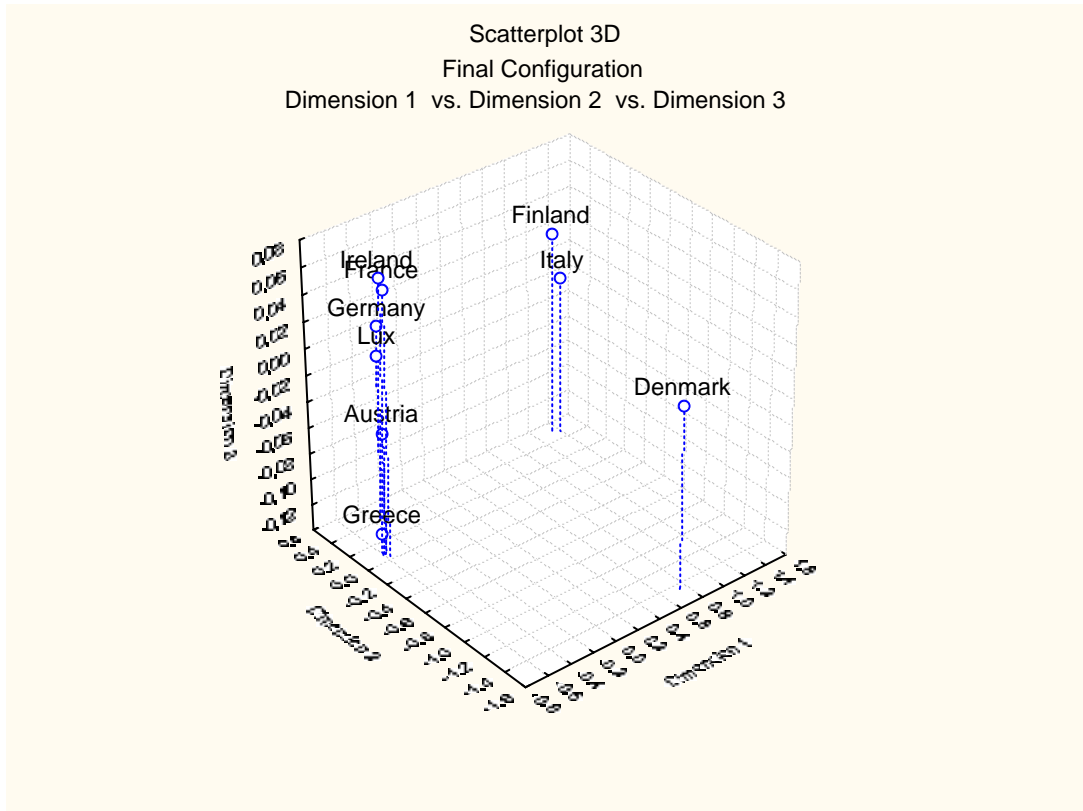


Figure 3. 3-D multidimensional scaling

A tree clustering method using the same dissimilarity matrix gives the following tree:

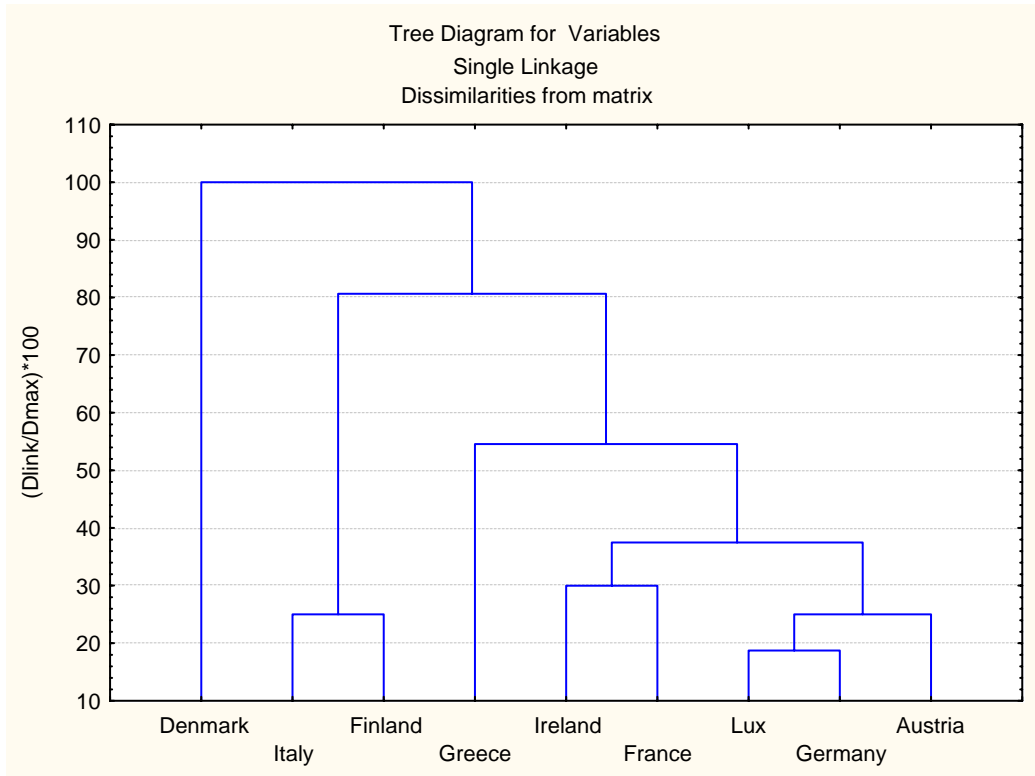


Figure 4. Result of tree clustering

## **ICT survey on enterprises**

The survey on ICT usage on enterprises is conducted every year in all EU-15 MSs and Norway. It is a highly harmonised survey following a model questionnaire with a number of mandatory questions plus some optional ones. All differences between statistical methodologies followed by the 15 Member States and Norway are combined to give an overall assessment of the level of comparability of the data

produced. In total,  $\binom{16}{2} = 120$  pair-wise comparisons between countries have been

performed based on a list containing the most important elements regarding the methodology for the execution of the ICT survey. This list was specially created for this purpose and is by no means uniquely defined; the most important indicators that could provide a clear-cut view of the methodological differences and thus cause lack of comparability have been defined.

The ICT specific list contains:

### ◆ Differences in implementing the Eurostat model questionnaire

This refers to the differences between national surveys regarding the use of the mandatory and optional questions as they were proposed by Eurostat. These differences were treated separately in the analysis. For the case of the use of mandatory questions, the mathematical expression for the differences is:

$$w_{i,j} = \text{number of breaks between countries } i, j \quad (4)$$

The number of breaks between the two countries (say,  $i$  and  $j$ ) signifies the total number of questions that are used by only one of the two countries. If a mandatory question is used by both countries or by none of them, it is not counted in this sum. The value that this coefficient can obtain is an integer lying between 0 (when the two countries have everything in common) and 22, which is the total number of mandatory questions in the Eurostat model questionnaire. Thus, the smaller the value of the coefficient, the more harmonised the surveys are (with respect to this methodological element).

Similar to this, the coefficient indicating the differences in the use of optional questions is now expressed as:

$$w_{i,j} = \frac{\text{number of differences between countries } i, j}{2} \quad (5)$$

The value of the coefficient lies between 0 and 3.5 ( $=7/2$ , where 7 is the total number of optional questions in the model questionnaire). As in the previous case, the greater the value of the coefficient, the more different the use of optional variables is. It should also be noted that the reason for applying a smaller weight to this coefficient is due to the greater importance of the differences in the use of mandatory variables than the optional ones for the harmonisation analysis.

### ◆ Differences in classification coverage of NACE sections

This aspect deals with the coverage of the sample size, in what regards economic activity. Four coefficients relating to economic activity (according to NACE Rev. 1) have been considered separately: differences in NACE Sections D, G, H, I, K, in NACE Division J 67, in NACE Division F 45 and in NACE Divisions O 92, 93. A scoring factor was introduced for every of these coefficients  $k$  reported from countries  $i, j$ , which takes values

$$w_{i,j}^k = \begin{cases} 0 & \text{no difference} \\ 1 & \text{a difference exists} \end{cases} \quad (6)$$

#### ◆ The size threshold for inclusion in the survey

In this part, differences in the enterprise size threshold that was included in each survey are analysed. The threshold values of the number of employees are split into three groups: A for none or at least one employee, B for at least five employees and C for at least 10 employees. The scoring factor in this case becomes:

*If there is no difference ( $i, j$  both belong to A, B or C) :  $w_{i,j} = 0$*

*If  $i \in A, j \in C$  and vice versa :  $w_{i,j} = 1$*  (7)

*Otherwise:  $w_{i,j} = 0.5$*

#### ◆ The size threshold for complete enumeration

Another aspect, which is relevant to the survey design and more specifically, to the sampling methodology, is the enterprise size threshold for complete enumeration. This coefficient indicates the differences between national methodologies in what regards the threshold values of the enterprise size for applying a census survey. Similar to the previous case, the size is split into three groups: A for up to 100 employees, B for between 100 and 250 employees and C for at least 250 employees. The scoring factor now obtains exactly the same form as the one appeared in (4).

#### ◆ The sampling fraction

The sampling fraction is defined as the ratio of the sample size to the population size, i.e. the percentage of the population which is under survey. This methodological aspect is also included in the list, since it is an absolute measure that enables comparability between countries, in contrast with the sample size, which is a relative measure (for example, a sample size of 6 000 enterprises is normal for Ireland but small for Germany). The coefficient relating to the sampling fraction is mathematically expressed by (4), but this time A, B and C are the groups in which the sampling fraction is below 10%, between 10 % and 20% and above 20% respectively.

#### ◆ The mode of data collection

In all Member States and Norway, post remains the most popular mode of data collection. However, other methods are sparsely used, such as interview, online questionnaires, etc. Thus, in order to assess any similarities in the use of these

additional means of data collection, this aspect was also included in the list. The scoring factor in this case takes the same values as in (6).

#### ◆ The data collection period

The period of data collection relates to the timeliness of data and whether Eurostat guidelines were followed and therefore, it could not be absent from this list. The corresponding scoring factor addresses the differences in data collection periods (in the first quarter of 2002 –as Eurostat proposed– or otherwise) and consequently, identifies the groups of countries that succeeded or not in conducting the survey on time, in accordance with Eurostat’s recommendations. It is defined as in (6).

#### ◆ The system of reminders

Since the use of reminders has been proved to have an effect on the response rate of the ICT survey, it is treated in this section. Similar to the previous case, the corresponding scoring factor is expressed by (6), and it exhibits the differences in the use (or not) of reminders.

#### ◆ The response rate

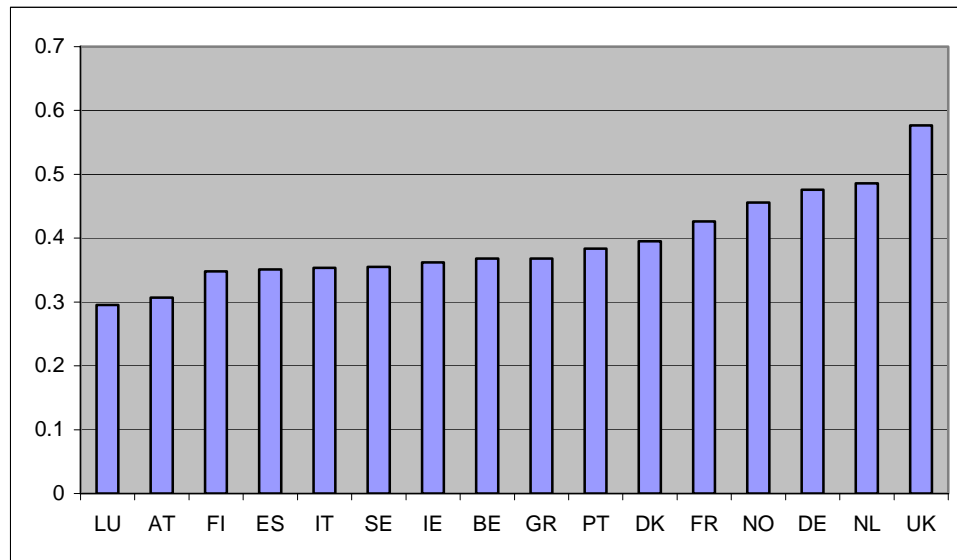
Along with the system of reminders, the differences in the response rates are also included in the harmonisation analysis. The scoring factor appears as in (7), with the groups A, B and C being defined as those containing national surveys that experienced response rates up to 50%, between 50% and 75% and at least 75% respectively.

Based on the methodological information obtained from all countries, this symmetric table of methodological differences was constructed

Table 3: Table of methodological dissimilarities

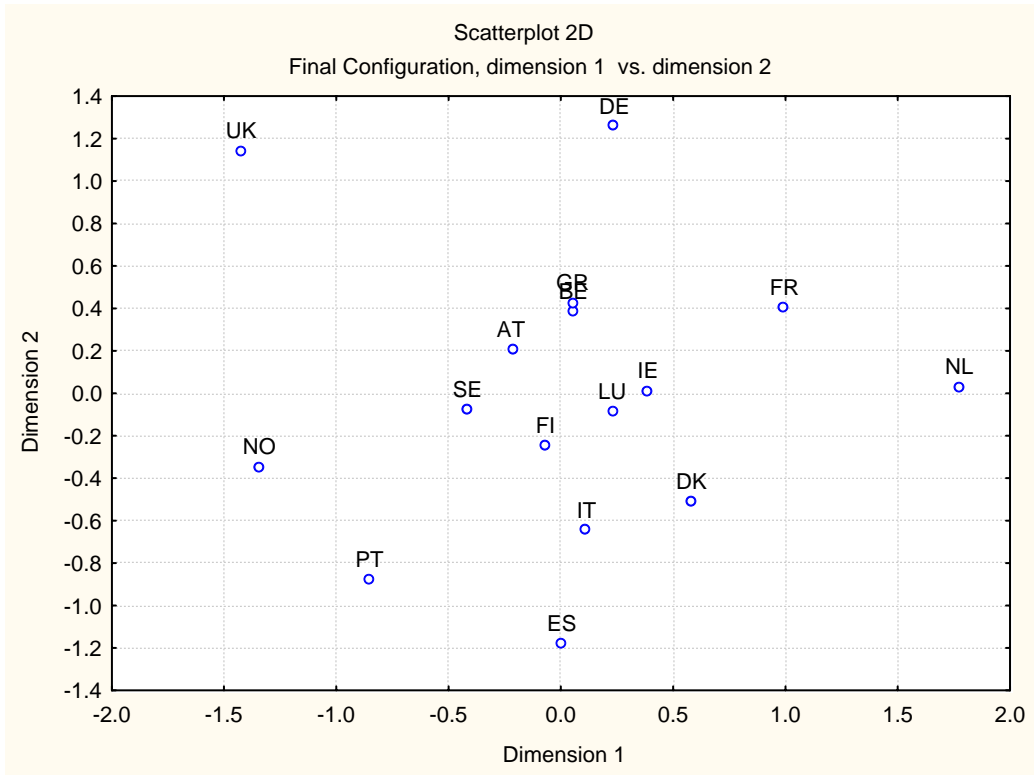
Total metadata	AT	BE	DE	DK	ES	FI	FR	GR	IE	IT	LU	NL	PT	SE	UK	NO
AT	0	0.23	0.36	0.45	0.27	0.27	0.43	0.23	0.32	0.3	0.25	0.523	0.18	0.16	0.5	0.43
BE	0.23	0	0.55	0.48	0.32	0.45	0.52	0.07	0.16	0.27	0.25	0.523	0.41	0.39	0.659	0.61
DE	0.36	0.55	0	0.5	0.61	0.43	0.43	0.5	0.59	0.55	0.52	0.432	0.5	0.43	0.591	0.61
DK	0.45	0.48	0.5	0	0.45	0.23	0.43	0.5	0.32	0.34	0.27	0.364	0.55	0.43	0.591	0.41
ES	0.27	0.32	0.61	0.45	0	0.41	0.41	0.39	0.34	0.25	0.18	0.477	0.2	0.39	0.5	0.41
FI	0.27	0.45	0.43	0.23	0.41	0	0.32	0.39	0.3	0.3	0.23	0.568	0.43	0.3	0.636	0.32
FR	0.43	0.52	0.43	0.43	0.41	0.32	0	0.57	0.48	0.34	0.36	0.523	0.45	0.36	0.705	0.48
GR	0.23	0.07	0.5	0.5	0.39	0.39	0.57	0	0.23	0.32	0.27	0.545	0.36	0.34	0.636	0.55
IE	0.32	0.16	0.59	0.32	0.34	0.3	0.48	0.23	0	0.27	0.14	0.455	0.5	0.48	0.773	0.45
IT	0.3	0.27	0.55	0.34	0.25	0.3	0.34	0.32	0.27	0	0.18	0.523	0.41	0.36	0.705	0.55
LU	0.25	0.25	0.52	0.27	0.18	0.23	0.36	0.27	0.14	0.18	0	0.409	0.34	0.32	0.636	0.36
NL	0.52	0.52	0.43	0.36	0.48	0.57	0.52	0.55	0.45	0.52	0.41	0	0.61	0.41	0.727	0.68
PT	0.18	0.41	0.5	0.55	0.2	0.43	0.45	0.36	0.5	0.41	0.34	0.614	0	0.3	0.455	0.43
SE	0.16	0.39	0.43	0.43	0.39	0.3	0.36	0.34	0.48	0.36	0.32	0.409	0.3	0	0.568	0.45
UK	0.5	0.66	0.59	0.59	0.5	0.64	0.7	0.64	0.77	0.7	0.64	0.727	0.45	0.57	0	0.55
NO	0.43	0.61	0.61	0.41	0.41	0.32	0.48	0.55	0.45	0.55	0.36	0.682	0.43	0.45	0.545	0
Average	<b>0.31</b>	<b>0.37</b>	<b>0.48</b>	<b>0.39</b>	<b>0.35</b>	<b>0.35</b>	<b>0.43</b>	<b>0.37</b>	<b>0.36</b>	<b>0.35</b>	<b>0.3</b>	<b>0.486</b>	<b>0.38</b>	<b>0.36</b>	<b>0.577</b>	<b>0.46</b>

The average dissimilarity which is also shown in Figure 5, indicates the level of incomparability of each country with the rest of the countries. Apparently, the UK is the most dissimilar, followed by Netherlands, Germany, Norway and France, while Luxembourg and Austria are the countries that are more comparable.



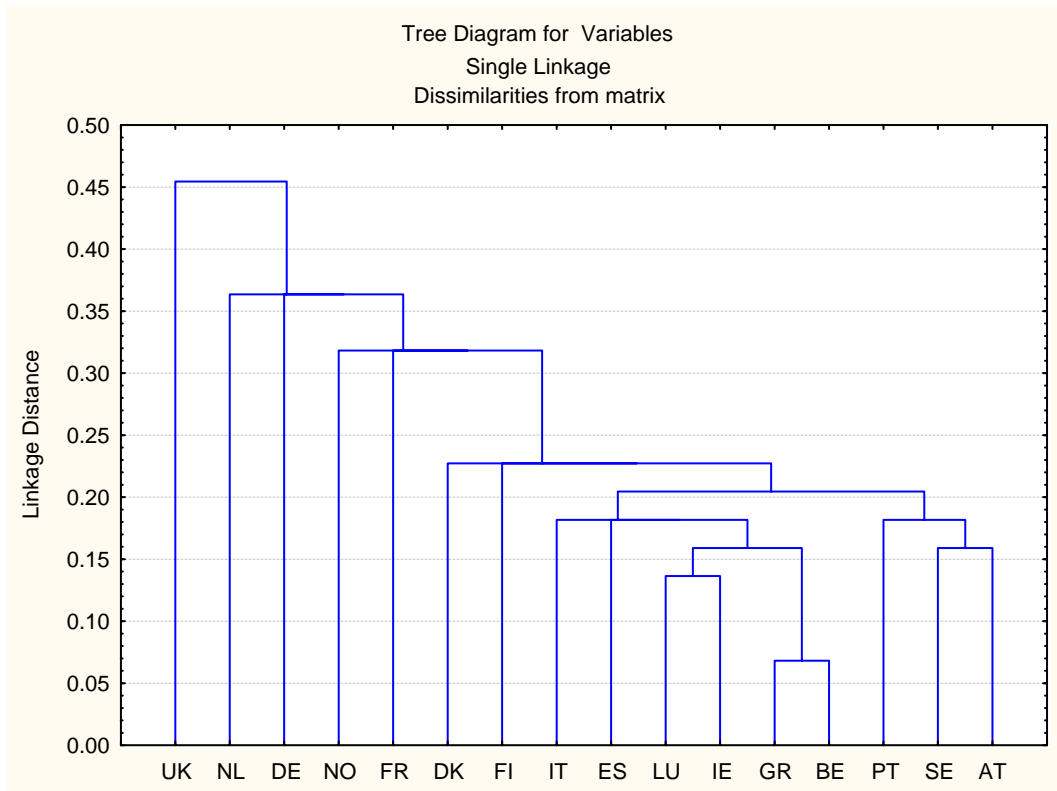
**Figure 5. Average dissimilarities for all countries under consideration**

Since a matrix of dissimilarities is at hand, a two dimensional map of practices can be drawn up using multidimensional scaling. The resulting map, shown in Figure 6, reveals the existence of a cluster which includes AT, SE, GR, BE, LU, FI, IE and perhaps IT and DK with the rest of the countries falling in some distance.



**Figure 6. A map of concepts and practices drawn using Multidimensional scaling**

A tree clustering method using the same dissimilarity matrix and a simple linkage average amalgamation rule gives the following tree. The same cluster identified by multidimensional scaling seems to form again, but this time it includes ES and PT as well.



**Figure 7. Result of tree clustering**

## Case study with mirror statistics from Foreign trade

One application of mirror statistics as a tool for assessing Geographical comparability and Coherence is in Foreign Trade Statistics. Twice a year, a Mirror Leaflet is published by Eurostat providing a general overview of Intra-EU asymmetries within external trade statistics. The total flows of goods for the year 2001 between the 15 MS of the EU (see Annex 3) were used to compute the following dissimilarity matrix:

FR	NL	DE	IT	UK	IE	DK	GR	PO	ES	BE	LU	SE	FI	AU
0.00	8.87	3.39	4.52	9.92	9.45	0.16	10.03	3.20	6.09	1.88	12.70	2.90	6.37	8.00
8.87	0.00	6.24	6.65	24.10	15.63	19.73	1.74	3.61	1.14	5.33	6.39	6.89	29.83	15.79
3.39	6.24	0.00	6.85	6.98	8.08	8.57	5.69	2.01	1.46	8.15	7.13	3.93	19.58	4.76
4.52	6.65	6.85	0.00	5.87	14.88	4.02	1.11	9.01	5.93	1.79	32.43	1.00	0.76	6.07
9.92	24.10	6.98	5.87	0.00	18.40	11.36	20.43	11.11	6.87	6.86	25.82	2.82	3.26	9.07
9.45	15.63	8.08	14.88	18.40	0.00	8.89	6.71	11.65	11.76	11.49	103.42	5.03	13.59	19.08
0.16	19.73	8.57	4.02	11.36	8.89	0.00	0.24	6.03	10.62	5.03	68.16	0.17	7.42	7.62
10.03	1.74	5.69	1.11	20.43	6.71	0.24	0.00	4.79	9.92	29.50	55.30	10.54	11.58	8.57
3.20	3.61	2.01	9.01	11.11	11.65	6.03	4.79	0.00	8.18	5.43	14.23	0.64	7.55	19.17
6.09	1.14	1.46	5.93	6.87	11.76	10.62	9.92	8.18	0.00	13.04	23.84	7.59	20.11	18.27
1.88	5.33	8.15	1.79	6.86	11.49	5.03	29.50	5.43	13.04	0.00	14.71	1.37	14.13	5.11
12.70	6.39	7.13	32.43	25.82	103.42	68.16	55.30	14.23	23.84	14.71	0.00	41.02	88.18	1.19
2.90	6.89	3.93	1.00	2.82	5.03	0.17	10.54	0.64	7.59	1.37	41.02	0.00	15.42	0.49
6.37	29.83	19.58	0.76	3.26	13.59	7.42	11.58	7.55	20.11	14.13	88.18	15.42	0.00	12.92
8.00	15.79	4.76	6.07	9.07	19.08	7.62	8.57	19.17	18.27	5.11	1.19	0.49	12.92	0.00

A natural way to analyse the matrix is to compute average dissimilarities for each country. These are presented in the next figure:

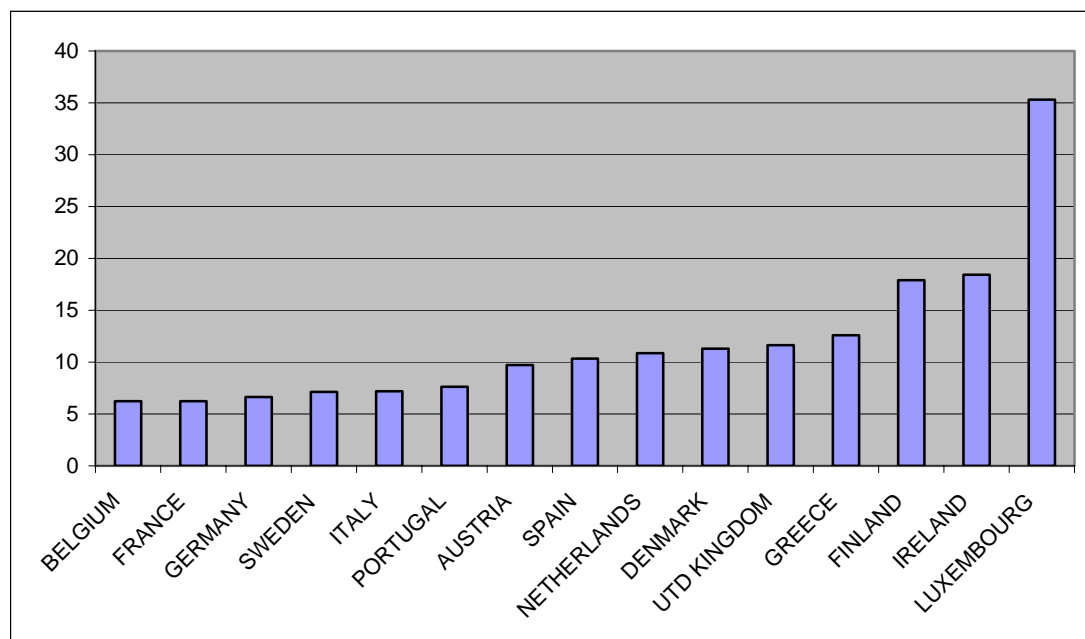


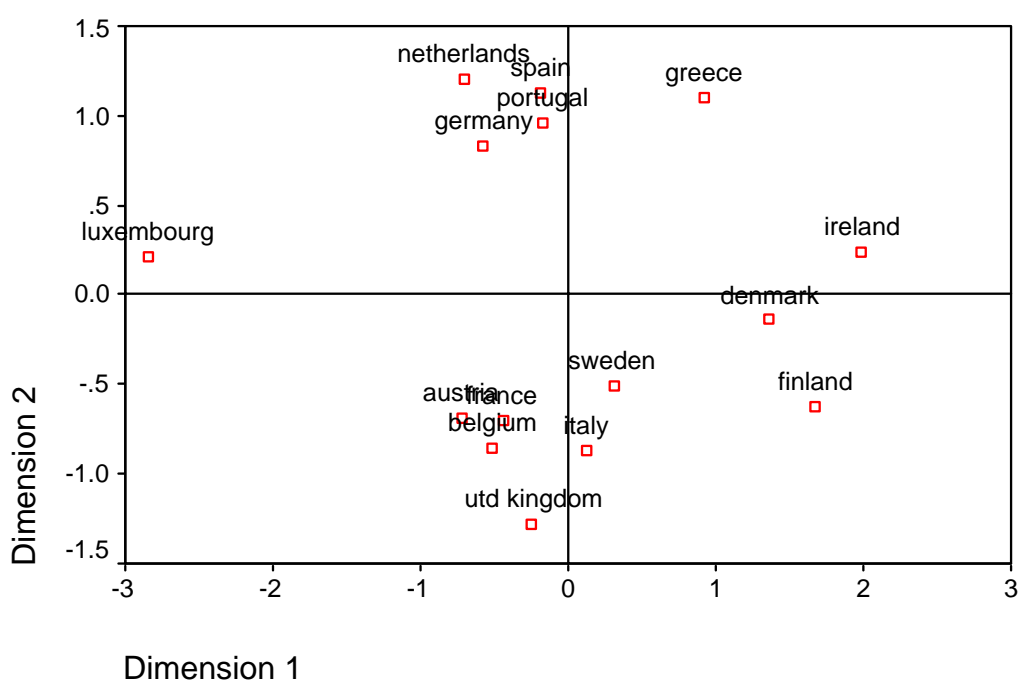
Figure 8. Average dissimilarities of external trade flows in 2001 for the 15 MS of the EU.

An observation that can be made based on this figure is that small countries have the larger dissimilarities, something obvious for Luxemburg, with average 35.3% and Ireland and Finland with average dissimilarities 18.4% and 17.9% respectively. This is a reasonable result since large countries report much more transactions and discrepancies tend to cancel out, thus lowering the dissimilarity indicator.

A multidimensional scaling routine can be very useful at visualising the dissimilarities of countries.

## Derived Stimulus Configuration

### Euclidean distance model



**Figure 9. A map of dissimilarities between countries for foreign trade statistics (2001)**

Two clusters are clearly formed, the first includes (Netherlands, Spain, Portugal and Germany) the second includes Austria, France, Sweden, Italy, Belgium and the UK. Greece is close to the first cluster, Denmark and Finland to the second. Luxemburg cannot be associated with any of the two clusters and Ireland is almost equidistant from the two clusters (although closer to them than Luxemburg).

This map can be really useful to Foreign trade statistics experts. By finding what aspects of methodology are shared among countries belonging in the same cluster (and different from those of countries in the other cluster) they can potentially identify the methodological differences that cause the discrepancies in mirror statistics.

Although this is just an example and it is rather difficult and perhaps premature to draw conclusions from this clustering, one observation is worth mentioning. Most

countries that stand out of the two clusters share a common characteristic which might be associated with their position in the MDS map. They have different thresholds for dispatches and arrivals.

In the following table assimilation thresholds are shown for each country in 2001 (source Foreign Trade Statistics Quality report (Eurostat, 2003))

Country	Assimilation Threshold		
	Arrivals	Dispatches	Percentage difference
Belgium	250,000	250,000	0.00%
Denmark	202,000	336,000	66.34%
Germany	200,000	200,000	0.00%
Greece	29,000	44,000	51.72%
Spain	96,162	96,162	0.00%
France	100,000	100,000	0.00%
Ireland	190,461	634,869	233.33%
Italy	102,500	154,000	50.24%
Luxembourg	103,000	103,000	0.00%
Netherlands	225,000	225,000	0.00%
Austria	145,000	145,000	0.00%
Portugal	60,000	85,000	41.67%
Finland	100,000	100,000	0.00%
Sweden	172,500	172,500	0.00%
United Kingdom	374,674	374,674	0.00%

We can see that Greece, Ireland and Denmark have assimilation thresholds for dispatches much higher than those for arrivals, although this is not true for Finland which also is distant from the clusters. Furthermore Portugal and Italy fit very well in the clusters and do have substantial differences between thresholds.

## **Conclusions**

The analysis of proximity data can be a useful tool for assessing geographical comparability. When different countries release similar data on different populations then, proximity data can be based on metadata and the results may indicate the extend to which the data are comparable. When different countries release data on the same population as is the case with flows and mirror statistics then the proximity data can be based on mirror statistics and the results can help identify the parts of methodology that cause the discrepancies.

## **Acknowledgments**

This study was partly financed by the European Commission, Eurostat. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Commission.

## References

**Colledge, M.J., (2000)**, Metadata Standards to Ensure International Comparability of Short Term Indicator Data. Bulletin of the International Statistical Institute, Proceeding of the 52<sup>nd</sup> session.

Available at: [www.stat.fi/isi99/proceedings/arkisto/varasto/coll0429.pdf](http://www.stat.fi/isi99/proceedings/arkisto/varasto/coll0429.pdf)

**Dalén, J. (1998)**, "*Studies on the Comparability of Consumer Price Indices*" *International Statistical Review*, Vol. 66, No.1, pp.83-113

**Elvers, E. (1998)** Other Aspects of Quality:Comparability and Coherence, in Davies, P. and Smith, P. (eds), *Model quality report in business statistics*, Eurostat, Final Deliverable of SUPCOM97 Lot.

**Everitt, B.S., Rabe-Hesketh, S., (1997)** The Analysis of Proximity Data, Kendall's Library of Statistics No 4, Arnold, pp.178.

**Eurostat (1998)**, An approach for improved quality of statistics for international comparisons, Doc. Eurostat/A4/Quality/98/General/Comparability.

**Eurostat (1999)**, "*Comparability and Coherence*", Doc. Eurostat/A4 /Quality/99/General/ Comparability and Coherence / State of play

**Eurostat, (2000)**. "*Standard quality report.*" Available in electronic form at: <http://www.forum.europa.eu.int/Public/irc/dsis/qjs/library?l=/public&vm=detailed&sb=Title>.

**Eurostat (2001a)**."*Comparability of Statistics*", Item 8, Doc. Eurostat/A4/Quality/01/Comparability.

**Eurostat (2001b)**."*How to make a quality report*", version 2.

**Eurostat (2002)**, "*Mirror Leaflet*", Eurostat Unit C4.

**Eurostat (2003)**, "*Foreign Trade Statistics-Quality Report*", Office for Official Publications of the European Communities, Cat. No KS-AS-03-001-EN-N.

**Linden, Hakan (1999)** "Comparability - Examples and Indicators".

**Papageorgiou H., Petrakos M, Vardaki M., Theodorou I., and Pentaris F., (2001)**, Assessing the level of Fragmentation of Data Series and Multisource Statistical Tables based on Metadata, in proceedings, NTTS 2001, Crete. Also Available at:

[http://webfarm.jrc.cec.eu.int/etk-ntts/Papers/final\\_papers](http://webfarm.jrc.cec.eu.int/etk-ntts/Papers/final_papers).

**Sorrentino, C. (2000)**, "*International Unemployment Rates: How Comparable Are They?*" *Monthly Labor Review*, Vol.123, No. 6, June 2000, pp.3-20. Also Available in electronic form in <http://stats.bls.gov./opub/mlr/2000/06/art1full.pdf>.

## ***ANNEX1 The full general taxonomy for sources of lack of Comparability***

- 1 Concepts
  - 1.1 Statistical characteristics
    - 1.1.1 Definition of the characteristic.
    - 1.1.2 Definition of variables
    - 1.1.3 Classifications used (type and version)
    - 1.1.4 Application of the Classification
    - 1.1.5 Breakdown available
  - 1.2 Statistical measure (indicator)
    - 1.2.1 Definition
    - 1.2.2 Unit of measurement
    - 1.2.3 Summary statistic used
    - 1.2.4 Completeness of the statistics (SQR:7)<sup>3</sup>
  - 1.3 Statistical unit
    - 1.3.1 Delineation of Statistical Unit
  - 1.4 Reference population
    - 1.4.1 Definition
    - 1.4.2 Exclusions
  - 1.5 Reference period
    - 1.5.1 Date of the beginning of the reference period
    - 1.5.2 Date of the end of the reference period.
  - 1.6 Study domains
    - 1.6.1 Classification Coverage for main characteristic(s)
  - 1.7 Standards
    - 1.7.1 International or European Standards the Survey is following
    - 1.7.2 Departures from the above standards
  - 1.8 Structure effects
  - 1.9 Conceptual aspects specific for a Domain under study
    - 1.9.1 Administrative rules
    - 1.9.2 Legal aspects that affect the comparability of estimates
- 2 Measurement
  - 2.1 Sampling design
    - 2.1.1 Type of survey
    - 2.1.2 Frame or Master List type (whether it is constructed from a census or an administrative register)
    - 2.1.3 Frame or Master List problems (frame's undercoverage, overcoverage, misclassification, duplication as well as frame's reference period) (SQR:2.2.1)
    - 2.1.4 Sampling Method
    - 2.1.5 Sample size or sampling ratio
  - 2.2 Sampling errors
    - 2.2.1 Coefficients of variation for major estimates (SQR:2.1)

---

<sup>3</sup> Indicates the paragraph number of the Standard quality Report from where this information can be collected.

- 2.2.2 Method used for sampling error calculation (SQR:2.1)
- 2.3 Data collection
  - 2.3.1 Reporting Method
  - 2.3.2 Response rate (SQR:2.2.4)
  - 2.3.3 Method used for reducing non-response (SQR:2.2.4)
  - 2.3.4 Item non-response (missing data)
- 2.4 Data processing
  - 2.4.1 Data editing (SQR: 2.2.3)
  - 2.4.2 Rates of failed edits (SQR: 2.2.3)
  - 2.4.3 Imputation rules-method.
- 2.5 Estimation
  - 2.5.1 Aggregation method
  - 2.5.2 Aggregation weights
  - 2.5.3 Grossing-up method
  - 2.5.4 Seasonal adjustment method
- 2.6 Measurement aspects specific for a Domain under study (these include characteristics pertinent to any specific domain, e.g. thresholds for Foreign Trade Statistics)

**ANNEX 2. The full taxonomy used in the Example on Labour Force Statistics**

C o d e	LEVEL I name	C o d e	T i m e	S p a c e	LEVEL II name
1	Source	11	+	+	Type of source
		12	+	-	Periodicity
		13	-	+	Reference period
		14	+	+	Unit of measurement
2	Concepts & coverage	21	+	+	Indicator Definition
		22	+	-	Geographical Coverage
		23	+	+	Classification Coverage
		24	+	+	Statistical population
		25	+	+	Population Exclusions
3	Standards	31	+	+	Standards framework
		32	+	+	Departure from framework
		33	+	+	Classifications
4	Data collection	41	+	+	Reporting unit
		42	+	+	Reporting method
		43	+	-	Master list
		44	+	+	Sample size
5	Data Manipulation	51	+	+	Aggregation method
		52	+	+	Grossing-up method
		53	+	+	Seasonal adjustment
6	Quality	61	+	+	Sampling error
		62	+	+	Non-response rate
7	Other	70	+	+	Other

## ANNEX 3. The Data used for the Mirror Statistics example

YEAR: 2001		Importing Countries																		Stores & Provisions	NGA (**)	EU15
Exporting Countries		France	Netherlands	Germany	Italy	Utd. Kingdom	Ireland	Denmark	Greece	Portugal	Spain	Belgium	Luxembourg	Sweden	Finland	Austria	EU15 (*)					
France	1		14,226	53,708	30,476	35,039	3,163	2,694	2,717	7,478	31,912	24,431	2,638	4,938	1,895	3,481	218,795	0	769	219,565		
	2		12,666	50,891	29,019	30,310	2,679	2,890	1,991	4,500	30,443	26,867	1,567	4,353	1,632	3,175	202,984					
	3		1,580	2,816	1,457	4,729	484	-197	726	2,978	1,469	-2,436	1,071	585	264	306	15,811					
	4		12.3%	5.5%	5.0%	15.6%	18.1%	-6.8%	36.5%	66.2%	4.8%	-9.1%	68.3%	13.4%	16.2%	9.6%	7.8%					
Netherlands	1	27,500		68,899	15,970	28,299	2,542	3,626	2,107	2,304	9,040	29,653	806	5,251	2,640	4,163	202,800	0	78	202,878		
	2	25,864		60,566	16,047	23,704	2,075	3,468	1,763	2,123	8,387	34,139	629	4,301	1,324	3,801	188,192					
	3	1,636		8,332	-77	4,595	466	158	344	181	653	-4,486	177	949	1,317	362	14,608					
	4	6.3%		13.8%	-0.5%	19.4%	22.5%	4.6%	19.5%	8.5%	7.8%	-13.1%	28.1%	22.1%	99.5%	9.5%	7.8%					
Germany	1	69,601	40,011		47,119	52,764	3,946	10,485	5,144	6,369	27,841	32,270	2,918	12,978	6,682	33,486	351,612	0	0	351,612		
	2	68,635	40,170		46,171	45,226	3,591	10,868	4,310	6,072	28,266	32,427	3,099	11,786	5,230	36,189	342,030					
	3	965	-169		948	7,537	365	-373	834	296	-425	-158	-181	1,192	1,452	-2,702	9,582					
	4	1.4%	-0.4%		2.1%	16.7%	9.9%	-3.4%	19.4%	4.9%	-1.5%	-0.5%	-5.8%	10.1%	27.8%	-7.5%	2.8%					
Italy	1	33,007	7,143	39,220		18,065	1,563	2,108	5,240	3,568	16,549	8,163	625	2,484	1,267	5,795	144,706	25	0	144,731		
	2	31,667	6,222	34,898		16,038	1,085	2,165	3,785	2,994	15,752	8,520	211	2,154	1,277	5,596	132,364					
	3	1,340	921	4,322		2,047	478	-57	1,455	564	798	-357	313	330	-11	199	12,342					
	4	4.2%	14.8%	12.4%		12.8%	44.0%	-2.6%	38.4%	18.8%	5.1%	-4.2%	148.2%	15.3%	-0.8%	3.6%	9.3%					
Utd. Kingdom	1	29,953	22,830	36,873	13,172		22,589	3,620	1,767	2,494	13,024	14,888	443	6,236	2,493	1,907	172,290	0	2,755	175,045		
	2	26,386	16,785	37,405	13,205		21,414	3,778	1,409	2,212	12,320	15,497	514	5,839	2,332	2,009	163,104					
	3	1,567	6,046	-532	-34		1,175	-158	359	262	703	-608	-71	398	161	-1	9,187					
	4	5.5%	36.0%	-1.4%	-0.3%		5.5%	-4.2%	25.5%	12.7%	5.7%	-3.9%	-13.8%	6.8%	6.9%	-0.0%	5.6%					
Ireland	1	5,527	4,267	11,637	3,295	20,654		598	326	302	2,276	4,430	91	1,350	406	408	55,566	0	2,627	58,193		
	2	5,400	3,821	15,051	3,511	15,041		595	278	267	2,345	5,592	57	1,240	392	429	54,019					
	3	128	445	-3,415	-216	5,613		3	48	35	-70	-1,162	34	110	14	-21	1,546					
	4	2.4%	11.7%	-22.7%	-6.1%	37.3%		0.5%	17.1%	13.2%	-3.0%	-20.8%	60.1%	8.9%	3.5%	-0.0%	2.9%					
Denmark	1	3,135	2,662	11,531	2,057	5,529	799		463	324	1,399	967	150	6,671	1,715	512	37,913	0	3	37,915		
	2	2,932	1,869	9,375	1,848	4,214	672		378	257	1,250	901	33	5,409	1,306	459	30,904					
	3	203	792	2,156	209	1,315	127		85	67	148	67	118	1,262	409	53	7,009					
	4	6.9%	42.4%	23.0%	11.3%	31.2%	18.9%		22.5%	26.2%	11.9%	7.4%	359.5%	23.3%	31.3%	11.4%	22.7%					
Greece	1	416	269	1,246	934	900	40	91		76	360	109	5	106	55	73	4,680	0	0	4,680		
	2	700	333	1,669	1,323	749	41	112		98	399	337	3	120	81	105	6,070					
	3	-285	-64	-423	-389	151	-1	-21		-22	-39	-228	2	-14	-26	-32	-1,391					
	4	-40.6%	-19.2%	-25.3%	-29.4%	20.1%	-3.6%	-18.9%		-22.5%	-9.8%	-67.5%	77.2%	-11.8%	-32.4%	-30.1%	-22.9%					
Portugal	1	3,450	1,141	5,232	1,253	2,812	144	292	103		5,202	1,463	30	411	131	224	21,887	6	1	21,893		
	2	6,142	1,152	5,272	1,243	2,538	129	326	88		4,842	1,974	23	366	169	197	24,452					
	3	-2,692	-11	-40	9	274	15	-34	14		360	-511	7	55	-38	27	-2,566					
	4	-43.8%	-1.0%	-0.8%	0.7%	10.8%	11.3%	-10.6%	16.4%		7.4%	-25.9%	30.2%	15.5%	-22.5%	13.7%	-10.5%					
Spain	1	25,330	4,692	15,387	11,697	11,719	799	861	1,421	13,226		3,719	168	1,308	476	1,165	91,866	291	819	92,976		
	2	23,508	4,838	14,719	10,914	10,798	612	779	1,049	12,064		3,895	117	1,054	666	895	85,909					
	3	1,821	-245	668	782	921	188	81	371	1,161		-176	52	254	-190	270	5,958					
	4	7.7%	-5.1%	4.5%	7.2%	8.5%	30.7%	10.4%	35.4%	9.6%		-4.5%	44.1%	24.1%	-28.6%	30.1%	6.9%					
Belgium	1	36,966	25,893	38,526	12,266	12,376	1,061	1,730	1,412	1,617	8,204		4,176	2,959	1,179	2,307	160,262	0	0	160,262		
	2	32,363	20,194	32,560	11,338	17,894	1,038	1,679	910	1,338	6,020		4,213	2,473	828	2,063	134,910					
	3	4,593	5,698	5,966	928	3,482	623	51	502	279	2,184		-37	486	390	244	25,351					
	4	14.2%	28.2%	18.3%	8.2%	19.5%	60.1%	3.0%	56.2%	20.9%	36.3%		-0.9%	19.7%	42.3%	11.8%	18.8%					
Luxembourg	1	2,111	467	2,661	706	940	120	74	43	99	374	1,600	137	123	124	124	9,579	0	527	10,105		
	2	2,729	526	2,171	867	471	13	68	23	97	332	1,179		119	41	170	8,806					
	3	-619	-59	490	-160	469	107	6	21	2	42	421		18	81	-46	773					
	4	-22.7%	-11.3%	22.6%	-18.5%	99.5%	810.8%	9.2%	91.7%	2.0%	12.6%	35.7%		15.2%	197.7%	-27.0%	8.8%					
Sweden	1	4,227	4,128	8,758	2,933	7,318	452	4,815	418	419	2,060	3,785	20		4,395	811	44,528	0	1,600	46,128		
	2	4,524	4,388	8,915	3,451	7,387	445	5,918	456	490	2,184	4,656	57		3,707	996	47,573					
	3	-297	-260	-156	-518	-69	7	-1,103	-38	-71	-134	-871	-37		688	-188	-3,045					
	4	-6.6%	-5.9%	-1.8%	-15.0%	-0.9%	1.5%	-18.6%	-8.4%	-14.5%	-6.1%	-18.7%	-65.2%		18.6%	-18.6%	6.4%					
Finland	1	2,220	1,880	5,958	1,751	4,637	378	1,173	364	267	1,197	1,181	38	4,091		509	25,645	0	280	25,925		
	2	2,269	2,011	5,138	1,711	4,644	298	1,326	313	241	1,283	1,263	17	3,551		709	24,774					
	3	-49	-131	820	41	-6	79	-153	51	26	86	-82	21	540		-200	871					
	4	-2.2%	-6.5%	16.0%	2.4%	-0.1%	26.6%	-11.6%	16.4%	11.0%	-6.7%	-6.5%	120.1%	15.2%		-28.2%	3.5%					
Austria	1	3,906	1,934	25,528	6,866	3,782	245	588	429	411	2,038	1,366	170	903	443		48,588	11	9	48,609		
	2	3,649	1,543	21,469	6,297	2,991	158	544	249	318	1,638	1,379	121	728	413		41,695					
	3	257	391	4,059	569	791	98	24	180	94	200	-13	49	175	30		6,894					
	4	7.0%	25.4%	18.9%	9.0%	26.4%	55.7%	4.4%	72.1%	29.5%	10.9%	-0.9%	41.0%	24.0%	7.3%		16.5%					
EU15 (*)	1	247,337	131,442	325,163	150,495	213,854	38,439	32,733	21,953	38,944	121,466	128,026	12,178	49,822	23,899	54,964	1,590,716	332	9,467	1,600,515		
	2	238,768	116,519	300,100	146,945	182,006	34,250	34,506	17,001	33,872	115,862	138,626	10,860	43,482	19,397	56,792	1,487,786					