# BACK CASTING HANDBOOK

# PART 1: INTRODUCTION

## 1.1 Background and scope of the handbook

The implementation of the revised NACE in EU statistics implies a disruption of all time series referring to NACE Rev. 1 or NACE Rev. 1.1. Such time series are available for many statistical domains, and to different kind of statistics (indexes, aggregates), produced and published according to different frequency (annual, quarterly or monthly) and at different levels of detail.
Long time series are of extreme importance for many users: typical examples of their use are the determination of growth rates, the identification of seasonal adjustment patterns or the application of forecasting models.

The provision of reconstructed time series in terms of NACE Rev. 2 is therefore a necessary target for statisticians dealing with the implementation of the revised NACE.

The reconstruction in terms of NACE Rev. 2 of existing statistical time series, currently expressed in terms of NACE Rev. 1 or NACE Rev. 1.1, is called "backcasting": this term is somehow derived from "forecasting".

This handbook aims at providing information to statisticians implementing NACE Rev. 2 in the European Statistical System. For each methodology, it presents the description, some examples and possible pros and cons.

The methodologies presented here are not intended to be exhaustive or prescriptive. In fact, there are no "best methods", as the choice among them depends on many factors:

- the kind of statistics to be back-casted (raw data, aggregates, indices, growth rates…)
- the availability of microdata
- the availability of microdata "double coded" according to both the old and the new classification
- the length of the "double coded" period
- the frequency of the existing time series
- the frequency and the level of detail of the requested back-cast series
- cost/effectiveness considerations
- etc.

Therefore, the choice on the method to apply for backcasting a specific time series should be done on the basis of many considerations. Not only the specific statistical domain, but also the national context will affect the decision on which method to apply.

## 1.2 Brief summary and presentation of content

All methods presented in this handbook assume that all the units recorded in the Business Register (BR) are double coded (according to the old and the new classification) for at least one point in time. The NACE Regulation does not impose a specific date for the double coding: it only requires that statistics referring to economic activities performed from 1 January 2008 onwards shall be produced according to NACE Rev. 2 (or a national classification derived from it). For most EU Member States, 2008 will be the year of "double coding".

Methods presented in section 2.1 are characterised by the so-called "Micro-data approach". The basic idea is the following: the time-series of interest is directly obtained from the microdata available in the BR, or from a database where individual observations are recorded. The "micro-data approach"

consists in substituting the old activity code (according to NACE Rev. 1.1) with the new activity code (NACE Rev. 2) and the re-compute the time series on the basis of the new code.
This method is the one providing most reliable reconstructed time series, but is very costly and coefficients of variation are high.

Methods presented in section 2.2 follow a "proportional approach" and make use of "conversion matrices", which allow the transformation of aggregated data, expressed in terms of NACE Rev. 1.1, into data expressed in terms of NACE Rev. 2 on the basis of proportions calculated according one or more statistics collected, in one point in time, according to the old and the new classification. These methods are less resource and time consuming than those based on micro-data, but they only approximate what the earlier observations may have been.

Methods presented in Section 2.3 combine the "micro-data approach" and the "proportional approach", as they require the double-coding of units for more than one year and interpolate between the two "double-coded" periods. These methods can be seen as an intermediate solution between those presented in chapters 2.1 and 2.2, in terms of both costs and quality of the reconstructed time series.

Part 3 of this publication presents the requirements for reconstructed time series, which are included in Community law, and are legally binding for all Member States.

Part 4 provides some additional information useful for those readers that intend to extend their knowledge of backcasting methodologies. More specifically:

- a list of references, with a short summary of the content or key points;
- examples or tables illustrating the methods or other topics presented in the previous parts of this handbook.

The handbook is the outcome of a collective work; Emmanuel Roulin (INSEE) drafted section 2.1, Ulrich Eidmann (Eurostat) drafted section 2.2.
This handbook, as well as the others of the series "Implementation of NACE Rev. 2", will be updated whenever there is reason for that. The latest version is available on the “Operation 2007” website http://forum.europa.eu.int/irc/dsis/nacecpacon/info/data/en/index.htm .

# PART 2: METHODS FOR BACK-RECALCULATION

## 2.1. Methods based on detailed re-working of individual data (micro-approaches)

### 2.1.1 What does it mean?

When changing a classification, "detailed reworking of micro-data" means to assign a new activity code (in terms of the new classification) to each statistical unit and for every period object of retropolation.

Once this assignment has been done, each series has to be re-worked, in order to have the series expressed in terms of the new classification.

Using this method, the only specific work is carried out at unit level, by assigning the new code corresponding to the principal activity; no other individual data or figure are modified in the database. The re-aggregation of the series simply consists in summing up the data corresponding to the various industries defined in terms of the new classification.

If the assignment of the activity code to each unit is made on the basis of detailed and reliable information, the approach based on micro-data will provide results more reliable than those obtained using methods based on macro-data.

### 2.1.2 The principal advantages of the "micro" approach

The main advantage of the micro approach with respect to macro-approaches consists in the fact that the micro-approach best retains the structural evolution of the economy.

Actually, in the various macro-approach, which work at aggregate levels, a unique conversion matrix for the sectors is used for each year target of the retropolation: for instance, if in year *t* section *S* corresponds to sectors *x* and *y* in terms of the new classification with the proportion of 30% and 70% respectively, then the same transformation will be applied to all the retropolated years. This method assumes therefore that the units classified in a section over the different years have the same intrinsic characteristics, and that the proportion 70/30 observed in year *t* has remained unchanged over the whole period. This is a very strong assumption, which in not required when applying the micro-approach, as each unit is re-classified according to its principal activity, for each retropolated year.

Another relevant advantage of the micro-approach is that it does not require the choice of a specific variable to work with: when applying macro-approaches, it is necessary to choose a variable of reference for the identification of the conversion factors to be used when retropolating, and it usually is the value added or the number of employees. As a consequence, it is only the structure observed on this variable that determines the conversion matrices, whereas the other possible variable of interests (e.g. turnover, investments etc.) may have a completely different structure. The following example shows this difference between the two approaches.

*Let's assume that sector S (old classification) is composed of units U1, U2 and U3 and the following figures have been observed:*

| | ***Turnover*** | ***Employees*** | ***Value added*** |
|---|---|---|---|
| ***U1*** | *100* | *1500* | *20* |
| ***U2*** | *200* | *2400* | *45* |
| ***U3*** | *150* | *2000* | *50* |
| ***Total sector S*** | ***450*** | ***5900*** | ***115*** |

*Let's suppose now that in terms of the new classification the new codes for the three units are respectively S1, S1 and S2*

*The macro-methodology based on the variable "value added" will provide therefore the following information in terms of the new classification:*

| | ***Turnover*** | ***Employees*** | ***Value added*** |
|---|---|---|---|
| ***Sector S1*** | *254* | *3335* | *65* |
| ***Sector S2*** | *196* | *2565* | *50* |

*The micro-method will provide the following information:*

| | ***Turnover*** | ***Employees*** | ***Value added*** |
|---|---|---|---|
| ***Sector S1*** | *300* | *3900* | *65* |
| ***Sector S2*** | *150* | *2000* | *50* |

*The macro-methodology[1] considers that sector S (old classification) splits into the two sectors S1 and S2 (new classification) according to the proportion 65/115and 50/115, where these proportion are derived on the basis of the value added. The same proportion applies to the other variables, even if this does not correspond to reality. The micro-methodology works on the basis of units and therefore it is not necessary to make this assumption.*

A third advantage of the micro-methodology concerns the variables which are directly linked to the assignment of the principal activity (value added and possible proxies for each of the activities carried out by the unit). As the micro-approach works at individual level, it ensures the consistency between these variables determining the principal activity and the principal activity itself.

Finally, it should be stressed, as a further advantage of the micro-approach, that the different retropolated series are consistent after the retropolation, as the same statistical unit considered in the different series will be accounted in the same way in the retropolation framework: the principal activity assigned to this unit is the same for each series referring to this unit.

### 2.1.3 Some drawbacks of the micro approach

The major drawback of the micro-approach is its cost: the work has to be done for each unit and for each year included in the series: therefore the cost is greater than it would be if working with aggregated data.
However, this cost should not be overestimated: the starting point of the two approaches (micro and macro) is always the double-coding of the activities carried out by each unit. The initial cost is therefore the same for the two approaches. With the micro-approach, the individual information (economic activities of each unit) is retained and considered in all retropolated series. Eventually, the main cost with the micro-approach, when compared to the macro-approach, is mainly due to the fact that each series (one for each variable) has to be recalculated individually after recoding the units. The additional cost for the micro-approach depends on the costs of re-elaborating the whole sets of data.

[1] This example is based on a simple macro-approach (based on one variable only): other macro-approaches make use of combined variables simultaneously). Other drawbacks could be then observed.

### 2.1.4 When to use the micro approach?

The micro-approach is specifically indicated for the retropolation of series where the statistical unit is the enterprise. Therefore, the business statistics series are specifically suitable for this method.

Actually, for this type of series, the impact of the change of classification concerns almost exclusively the code of the principal activity associated to each unit. However, it should be considered that some variables of the series may be impacted by the change of the classification. For instance, the part of the turnover originated by the principal activity or the shares of turnover produced in great economic sectors[2].

On the other side, those series where the statistical unit is not the enterprise cannot be retropolated via the micro-approach being described here: these are price indexes and, more generally, those series where the observation unit is the product.

Paragraphs below describe in more detail how the micro-approach can be applied:

- a sophisticated way (and more precise in terms of retropolation), dealing with the most detailed level of reporting unit (the local kind of activity unit);
- another way, which refers to the principal activity only.

The micro-approach is very well suitable when the National Statistical Institute can recode the principal activity of the unit on the basis of the detailed observation of the activities carried out by the units[3]. However, the micro-approach is applicable only if the complete information on the economic activities of the units observed in the series is still available.

### 2.1.5 Census and sampling

When retropolating series obtained on the basis of sample surveys, the loss of information does not depend on the method applied for the retropolation (micro or macro approach): from this point of view, the micro-approach does not provide advantages if compared with the macro-approach.

The method to be applied is the same: the micro-approach can be assimilated to a reclassification of units analogous to the one applied when the units are misclassified in the reference population. Such a misclassification does not produce bias (it is assumed that all domains are correctly covered by the sample), but increases the variance. This variance is essentially due to the fact that for a sector expressed in terms of the new classification, the available sample might be very small.

In case of census, and analogously to the case of misclassification in case of sample survey, phenomena like those described above cannot verify: by construction, all the reference population is considered in the statistical results.

[2] For the series referred to in the Structural Business Statistics regulations, they are variable "18 xx".

[3] This is the case of France, where all enterprises with more than 20 employees are surveyed every year within the annual structural business survey on their turnover broken down according to economic activity for identifying the principal activity (a sample is surveyed for enterprises with less than 20 employees).

### 2.1.6 Double-coding at least one year

The micro-approach requires at least one year (or one reference period) of double-coding, according to the principal activity of the units. This double coding, for a given year, provides for each unit the conversion between the principal activity expressed in terms of the old and the new classifications. This correspondence is applied to all the years (periods) of the series where the unit appears.

Several methods can be applied for the double coding: the simplest one simply consists in asking the unit itself to describe its principal activity in terms of the old and the new classification. In this case, detailed explanatory notes need to be provided to the unit.

However, this method does not allow determining the principal activity, as the top-down method should be applied. The top-down method consists in determining first the main NACE Section, then the main NACE Division and further down until the main Class according to the share of the value added produced by the unit for each elementary activity.

The top-down method should be applied in case the unit carries out multiple activities. In order to apply the top-down method, the shares of value added corresponding to each activity carried out by the unit must be known, or at least a variable which can work as a *proxy* of the value added.

### 2.1.7 Re-working at the level of elementary activities

When changing the activity classification, it is extremely burdensome asking the unit to provide the figures corresponding to the value added generated by each elementary activity according to both the old and new classifications. Therefore it is preferable to use an intermediary classification in order to obtain the necessary information.

The intermediary classification which is provided to the unit when asking the shares of value added is developed in such a way that allows the unambiguous identification of its principal activity according to both the old and new classifications. This intermediary classification is the cartesian product of the two classifications, as described in detail below and in the annex to this chapter.

*Let's assume that the old classification:*

- *is composed of two classes only, A and B;*
- *these two classes can be split into A1, A2 and A3, on one side, and B1 et B2, on the other side*[4].

*Let's assume also that the new classification consists in a reorganisation of the sub-classes mentioned above as it follows:*

*The new class X consists of the sub-class A1 ;*
*The new class Y consists of the sub-classes A2, A3 and B1 ;*
*The new class Z consists of the sub-class B2.*

*The following classification will be an intermediary between the two old and new classifications*[5] *:*

*M = {A1} ; N = {A2, A3}, O = {B1}, P = {B2}*

*It is therefore possible to directly define both the old and the new classifications using the intermediary classification:*

[4] This decomposition in sub-classes may correspond to both sub-classes in strict sense, or to products which define the classes; the sub-classes may already exist, as a consequence of previous needs, and in that case the codification would be easier to use.

[5] "An" and not "The", as it is always possible to build up several intermediary classifications, at higher levels of detail: however, all these intermediary classifications are derived from the "minimal" (most detailed) intermediary classification.

*A = {M, N}, B = {O, P}*
*X = {M}, Y = {N, O}, Z = {P}*

The observation of the elementary activities of a unit in terms of the old and the new classification (for instance using an intermediary classification), is of great advantage for the organisation of the retropolation activities. If, for instance, one is interested not only in statistics at sector level, but also at branch level [6], the observation of all the elementary activities in terms of the two classifications allows a retropolation which is simultaneous (based on the same information) and consistent among sectors and branches: this is possible because at individual level, the principal activity is defined on the basis of the observation of all the elementary activities expressed in terms of the new classification.

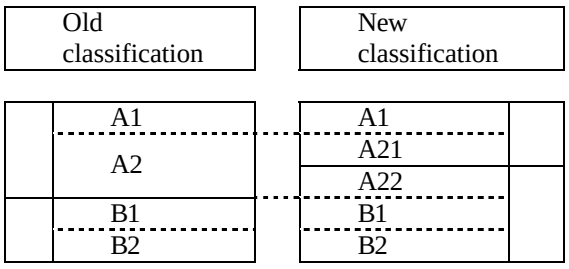
Another advantage, more important than the previous one, is that the observation in terms of the new nomenclature of all the elementary activities allows the weakening of the assumption of "constant structure" on which all methods of retropolation are based. In the macro-approach, the conversion structure between the old and the new classification is applied to all years of the series, as defined according to the observation of the "double coding year". In the micro-approach, this assumption of "constant structure" is less strong, as it is made at level of the enterprise and not at the aggregated level. When, in the micro-approach, the work is done at the elementary activities level, the assumption of "constant structure" is made at the level of elementary activity.

Actually, the more the "constant structure assumption" (to be assumed in any retropolation methodology) is made at a lower level, the less it is strong and therefore the structural developments of the economy are maintained in the retropolated series.

When the elementary activities of the enterprises have been the target of the reclassification, it is at that level that the conversion matrices are defined: they are then applied to each elementary activity of the years to be retropolated. For each year to be retropolated, the transition matrix will be determined on the basis of the principal activity assigned to each enterprise.

The procedure described above will best respect the possible changes of the composition of activities of each unit during the years. The following example illustrates both the principle of this methodology and its advantages.

*Let's assume that the old classification consists of the two groups A et B, each one broken down into two classes: A1, A2 and B1, B2.*
*Assume that the new classification is different from the old one as follows: a split of class A2 into two classes A21 and A22, a new group U composed of A1 and A21, and a new group V composed of A22, B1 and B2. The correspondence transition from one classification to the new one is shown below:*




*Consider the unit E, whose share of the value added in the double-coded year is as follows[7]:*

[6] A sector (also called "industry") is constituted of all the units which have the same principal activity, and includes also their possible secondary activities. A branch is constituted of all secondary activities (having the same NACE code) of all the units, independently of their principal activity.
[7] In this example, the double coding is equivalent to the new coding, as the new classification consists of a simple split of class A2 into classes A21 and A22.

| Year T | |
|---|---|
| A1 | 10 |
| A21 | 10 |
| A22 | 20 |
| B1 | 20 |
| B2 | 10 |

*In year T, applying the top-down method the principal activity is A (40 versus 30) according to the old classification, and V according to the new classification (50 versus 20). Moreover, according to this double-coded observation, for enterprise E in year T its elementary activity A2 is split into activities A21 and A22 in the proportion 1/3 and 2/3 respectively.*

*Let's assume now that in year R, the year to be retropolated, the share of the value added according to the old classification was as follows:*

| Year R | |
|---|---|
| A1 | 30 |
| A2 | 12 |
| B1 | 10 |
| B2 | 10 |

*Therefore, in terms of the old classification, the principal activity of unit E was A (42 against 20).*

*Applying the micro-approach directly to the principal activity, the code A (old classification) associated to enterprise E in year R would have been converted in the code V (new classification), on the basis of the transition matrix observed in year T.*
*Conversely, working at the more detailed level of the elementary activities, the observation in year T would have determined the split of the activity A into the two activities A21 and A22 in the proportion 1/3, 2/3, and therefore to estimate the following share of the value added, in terms of the new classification:*

| Year R (estimates) | |
|---|---|
| A1 | 30 |
| A21 | 4 |
| A22 | 8 |
| B1 | 10 |
| B2 | 10 |

*We can deduce therefore that a "retropolated principal activity" is U and not V (34 against 28). The retropolation worked out at the most detailed level (the elementary activity of the unit) keeps into account the modification of the structure of the unit between years R and T: it is more reliable and reflects the reality.*

### 2.1.8 Two specific cases: an elementary activity or a unit that have not been double coded; the necessity to use transition matrix

The procedure described above (working at the level of elementary activities) can be applied only to elementary activities observed and recorded in the year of double coding (year T). Two conditions must be fulfilled:

- first, the enterprise E for which we need the retropolated data in year R must be observed in year T,
- second the elementary activity S to be retropolated in year R must be observed in year T.

Two specific situations could verify, corresponding to these two conditions. Actually, these two situations could be combined: it is in fact the treatment of an elementary activity (its re-codification in terms of the new classification) which has not been observed (in the same unit) in year T (double coding). The case of a unit which has not been observed in year T (either it did not exist anymore in year T, or it has not been included in the sample, or did not reply) corresponds to the situation where no elementary activity has been observed for the unit.

A possible solution for solving these situations consists in using transition matrices[8]; they are determined on the basis of a great number of observations, and therefore correspond to average (see below). The principle is the same as applied in the macro-approach, but some specificity is described here.

Transition matrices are then applied directly to elementary activities to be retropolated, in order to determine their developments in terms of the new classification. Once the developments are known, the micro-approach described in the previous paragraphs can be applied: it is therefore a combined use of the micro and the macro approach.

### 2.1.9 How to treat a unit with unknown elementary units (never observed)?

It may happen that the elementary activities carried out by a unit which needs to be retropolated for year R are not known (e.g. not collected, unit not surveyed or not responded).

A possible solution consists in retropolating directly the principal activity of the enterprise: two possible situations may present, namely:

- The unit has been observed in the year T of double coding, but not in year R. In this case, the safest solution consists in assuming that the principal activity in year R was the same as in year T.
- The unit has not been observed in year T. In this case, the only possible solution consists in applying a transition matrix for identifying the principal activity of year R in terms of the new classification. Both a micro (work directly on the unit) and a macro-approach (use of a conversion matrix) are combined.

### 2.1.10 Conversion matrix or hot-deck procedure?

A conversion matrix presents the probability that an element (elementary activity, principal activity, etc.), coded as *i* in the old classification is coded *j* according to the new classification. These probabilities are determined on the basis of the empirical frequencies observed on the reference population in the year T of double-coding.

Two main kinds of "conversion" may affect a class *i* of the old classification[9]:

- either there is a one-to-one correspondence to class *j* of the new classification (with or without a change of the code)
- or class *i* is split in two or more classes in the new classification (one-to many correspondences).

In the first case, there will be one and only one conversion coefficient $c_{i,j} = 1$ (all the other elements of the *i*-th row are 0). In the second case, there will be several coefficients different from 0 and with value between 0 and 1, whose sum is 1.

[8] See below for the definition of conversion matrices.

[9] It must be stresses that even if there is equivalence between class *i* and class *j*, there is no reason to assume that a unit with principal activity having class *i* in the old classification has class *j* as principal activity in the new classification. This can be assumed if and only if the unit carries out only one activity. Otherwise, the application of the top-down method may affect the identification of the principal activity in terms of the new classification.

In the second of the cases just mentioned, the use of such a conversion matrix for recoding the elementary activity may artificially modify the structure of the activities carried out by the enterprises.

For instance, let's assume that the class A (old classification) is split into two classes U and V in the new classification.

*Assume, for simplicity reasons, that all units carrying out the activity A have A as a unique activity (and therefore the principal activity i terms of the old classification is A).*
*Assume that for 70% of these enterprises the old activity A corresponds to new activity U, that for 20% the old activity A corresponds to both activities U and V (according a proportion of 60% and 40%), and that for 10% the activity A corresponds to activity V.*
*Then, the determination of the conversion coefficients will provide the following results[10]:*

$c_{A,U} = 82\%$

$c_{A,V} = 18\%$

*Therefore, the application of these conversion coefficients to units to be retropolated for year Y-1 and carrying out only activity A (according to old classification), will associate for each unit the amounts corresponding to 82% and 18% for the two activities U and V (value added or number of employees). So, the identification of the principal activity in terms of the new classification will lead to activity U for all these units, and none of them will have V as principal activity, even if 10% of them have been observed as such in the year of double coding.*

It is therefore necessary to prevent the risk described in the previous paragraph. Different solutions may be considered:

- The first one consists in applying a retropolation of elementary activities in two steps:
  - the first step consists in randomly determining in how many new activities the old elementary activity should be retropolated[11].
  - The second step will then consist, on the basis of the outcome of the previous step, in establishing the one or more activities in terms of the new classification.

- The other procedure, more simple from the application point of view, may be used in order to control the risk mentioned above: to apply a "hot deck procedure" instead of the conversion matrices. The hot deck procedure consists in finding the "closest" unit to the one which has problem for retropolation[12]. The retropolation of the elementary unit will be made in the same way as made for the "closest unit"[13].

The risk previously mentioned is limited to the use of conversion matrices for retropolating elementary activities. It does not exist when these matrices are applied for retropolating directly the principal activity. Actually, in this last case, the only target is the identification of the new code of the unique principal activity. The different coefficients $c_{i,j}$ will be applied without any risk.

A supplementary caution should be considered when applying the retropolation procedure directly to the principal activity.
*Let's suppose that a unit has the same principal activity A in the three years R-2, R-1 et R to be reptropolated. Let's suppose that no information is available on its elementary activities for each of these three years. Moreover, let's suppose that the retropolation procedure applying the conversion matrix for year R transforms the principal activity A into the principal activity X.*

[10] We assume here the hypothesis that the coefficients are calculated on the basis of non-weighted conversions: a weighting made on the basis of the value added or the number of employees, might marginally modify the coefficients.
[11] This will be done on the basis of the observation made in the year T of "double –coding".
[12] That because this activity has not been observed, in this same unit, in the year T of double coding.
[13] This "closes neighbour" should carry out the same elementary activity as observed in the year T of double coding.

*Then, it is preferable to convert this activity A into X also for years R-1 and R-2, than using again the random procedure.*

## 2.2 Methods based on conversion coefficients (macro-approaches)

*…including application (real or examples) of approaches to the various statistical domains; pros and cons of the various approaches*

*The following chapter draws upon a number of very valuable articles and documents provided by Destatis and by Statistics Canada. For further consultation, the exact references are listed at the end of this handbook..*

### 2.2.1 What are "proportional methods"?

The "proportional method" offers a simple technique to carry out backward calculation, especially in a first attempt to determine the new path of the involved time series. A transitory period is expressed both under the new and the old classification system. Then in order to reconstruct the historical series under the new classification, a proportional rule – meaning a set of so-called "conversion coefficients" – is applied to the historical part of the time series under the old classification.

The proportional method is applied at "macro" level – in its most simple form, when conversion coefficients are estimated on the basis of the number of units only, it does not require going back to the micro data of the individual units at all. It is thus a low resource and time consumption approach to the backward calculation, but it only approximates what the earlier observations may have been without analyzing in a deep way the revision effects on time-series.

*Simple vs. sophisticated*

The proportional method is equivalent to applying the growth rate of the former time series to the revised level established under the new classification. In its most simple form, the procedure follows thus the rule of three throughout the whole historical series. But there exist also more sophisticated methods where coefficients are adjusted for particular years – which in turn can be done on the basis of experts' opinion or on the basis of more or less sophisticated estimation techniques.

Specific measures will have to be taken in order to deal with the breaks which can be expected to appear between the different parts of the time series.

*Assumptions underlying the proportional methods*

The proportional method modifies only the estimates and does not consider or modify the micro-data used for the construction of these estimates. There is thus no longer a link between historic micro and macro data.

The use of the same set of coefficients through time is based on the assumption that the distribution of the variables of interest between the old and the new classification does not change. For example, for a given NACE Rev. 1.1 industry, the proportion of turnover going to a specific NACE Rev. 2 industry might change over time.

***

In the remainder of this chapter, we will first go through the individual steps of applying the proportional method. This brief – theoretical – introduction will be followed by examples from Destatis (the German Federal Statistical Office) and Statistics Canada. The chapter will be concluded by a discussion of the pros and cons of the proportional method pointing out measures that might be taken in order to deal with one or the other of the shortcomings of this method.

### 2.2.2. Step-by-step in theory

*Starting point – concordance tables*

Concordance tables are the starting point for establishing the link between old and new classification systems. These tables depict the relations from old to new and from new to old, and provide thus (mostly qualitative) information on the transition between the two systems.

To the users these tables are helpful in understanding the relationship between the old and new codes, in discerning the industrial scope of changes, and in understanding how the revision affects the historical continuity of estimates. For the producers of statistics, concordance tables are the basis for the calculation of conversion coefficients.

Concordance tables can be more or less detailed. Major concordance tables can have exhaustive explanatory notes with detailed comparisons between the old and the new system. For the purpose of converting data from one system to the other, it is however sufficient if the concordance table provides (1) lists of all industries within each category, and (2) changes in industrial scope (additions and subtractions) from the old to the new system and vice versa.

*Step 1 – Estimation of conversion coefficients*

Conversion coefficients are factors based on a measured reallocation of data at aggregate industry levels that reflect the changes between the old and new classification systems ([14]). They should be calculated at the most detailed level possible.

On the basis of concordance tables, the conversion coefficients can be calculated for each classification based on the number of units. Alternatively, the conversion coefficients can also be calculated on the basis of variables such as turnover, employment, earnings, sales, etc. This will require the availability of micro level data. It is possible that different sets of coefficients are used according to the variables of interest.

Conversion coefficients show how much each industry has changed (either in terms of units or in terms of a variable), where the movements took place, and between which industries the movements occurred and in which direction. In a way, conversion coefficients are a quantitative representation of the concordance tables.

The coefficients can be computed at a single time point or at several time points. The advantage of measuring them at several time points is that one can determine whether the conversion coefficients at a single point in time are appropriate. From the theoretical point of view, it might appear ideal to have conversion coefficients calculated for every point of the time series but in practice this will be too demanding in resources. A possible compromise could be to calculate coefficients for two different points in time (such as at the beginning and the end of the historical series to be converted) and to obtain the coefficients for the time points between these two by interpolation.

Ideally, the conversion coefficients would be calculated on the basis of data for at least one year, which would be the changeover year between one classification and another. For improving the quality of the conversion coefficients it is recommended to extend the period of double coding, for instance by another year, in order to give the new classification time to settle down, and to have the coefficients calculated on the basis of data which has already undergone some corrections.

*Step 2 – Combination of estimates from the old classification with conversion coefficients*

([14]) See Handbook on Sampling and estimation in the context of implementing NACE Rev. 2 for a discussion of the options for calculating calibration factors used for the calculation of industry aggregates according to the old and the new classifications.

In a second step, industry estimates according to the new classification are obtained as a weighted sum of industry estimates from the old classification, the conversion coefficients being used as weights. As an example, a given industry A according to the new classification might be composed of two parts coming from two different industries $A_1$ and $A_2$ according to the old classification. The conversion coefficients are a measure of the relative importance of $A_1$ respectively $A_2$ in the new industry A. This is shown in more detail in the practical examples in Sections 2.2.3 and 2.2.4.

Sometimes (when only one set of conversion coefficients is applied to the whole time series) the calculation is referred to as "weighted linear combination".

*Step 3 – Linkage of the estimates from the three time-segments*

The overall purpose of the back-casting exercise is to constitute historical series according to the new classification, from the existing series with the previous classification.

These "historical" series will consist of three segments:

1. The historical time segment where only the old classification existed. This is the segment for which the conversion coefficients have been estimated.

2. The transitory time segment where the old and new classifications are present. For this segment, conversion coefficients can be "observed".

3. The final time segment where only the new classification will be used.

Regardless of the method used to obtain estimates over the historical segment, a break will typically occur between the first (historical) and the second (transitory) segment. This break, or jump, will be caused mainly by the change in the field of observation which in turn will be the result of the change in the classification.

The purpose of linking, in the present step, is to alleviate the jump. One approach is to raise the converted historical segment to the level of the transitory segment, which eliminates the jump; another is to “wedge” the jump, i.e. to spread it over a number of months or years. Other variants exist.

The expert knowledge of subject-matter analysts will be required at this stage to review the series and adjust them to agree with their prior knowledge.

*Step 4 – Final adjustments for consistency*

Once the new table of continuous time series is produced, it may be necessary to restore contemporaneous additivity.

*Step 5 – Seasonal adjustment*

One of the principal objectives of backcasting is to establish a historical time series which subsequently serves as basis for seasonal adjustment. The procedure described so far is not used to produce the seasonally adjusted series directly.

### 2.2.3. Example 1 (Destatis): Rebasing the indices of production industries on 1991

*Background*

In January 1995, the Industrial Classification of Economic Activities 1979 Edition (SYPRO) ([15]) was replaced by a new edition, the WZ 93 ([16]), a classification corresponding, at the four-digit level (classes), to the NACE Rev. 1.1 but with a further break-down of the classes into branches (five-digit level).

The change in the classifications made it necessary to recalculate the data obtained on the basis of the SYPRO and the 1989 Product Classification for Production Statistics (GP 89) up to and including December 1994, including all months from 1991 to 1994, in line with the new classification. The GP 89 was used as the basis for a reporting nomenclature according to which data on quantities and values was collected for the update of the monthly production indices for approximately 1 000 products.

The macro approach was chosen because of its simplicity and the short time required for its implementation. Furthermore, access to micro data, in the framework of the back-casting, would have been limited or impossible.

*Calculation of allocation factors*

For the purpose of converting the monthly data structured in line with the SYPRO to the WZ 93, it was assumed that each product covered under the GP 89 could be assigned completely to a new WZ 93 heading. This way, local kind-of-activity units were formed on the basis of WZ 93 and defined by the products in accordance with GP 89. With the SYPRO being defined through the GP 89 as well, factors could be calculated – from the gross production values according to the GP 89 – for the conversion of the SYPRO data to the WZ 93.

Table 1 shows the above approach in a schematic way, taking as an example the allocation of two SYPRO classes of economic activity to three four-digit headings of WZ 93.

**Table 1: Calculation of allocation factors for SYPRO**

| **$S_j$** | **$GP_{ij}$** | **$BPW_{ij}$ (€)** | **$BPW_j$ (€)** | **$A_{ij}$** | **$W_k$ pro rata** | **$W_k$ total** | **$BPW_k$ (€)** |
|---|---|---|---|---|---|---|---|
| | $GP_{11}$ | 150 | | 0.3 | $W_1$ | | |
| $S_1$ | $GP_{21}$ | 100 | | 0.2 | $W_3$ | $W_1$ | 650 |
| | $GP_{31}$ | 250 | | 0.5 | $W_2$ | ---------- | ---------- |
| *Total $S_1$* | *$GP_{.1}$* | *500* | *500* | *1.0* | | $W_2$ | 350 |
| | $GP_{12}$ | 100 | | 0.1 | $W_2$ | ---------- | ---------- |
| $S_2$ | $GP_{22}$ | 400 | | 0.4 | $W_3$ | $W_3$ | 500 |
| | $GP_{32}$ | 500 | | 0.5 | $W_1$ | | |
| *Total $S_2$* | *$GP_{.2}$* | *1000* | *1000* | *1.0* | | | |

where

$S_j$ = SYPRO class j of economic activity (four-digit heading)

$W_k$ = WZ 93 class k of economic activity (four-digit heading)

([15]) *Systematik der Wirtschaftszweige (Ausgabe 1979), Fassung für die Statistik im Produzierenden Gewerbe* – Industrial Classification of Economic Activities (1979 Edition), Version for Statistics of Production Industries.

([16]) *Klassifikation der Wirtschaftszweige* – Industrial Classification of Economic Activities (1993 Edition).

$GP_{ij}$ = product i allocated to class of economic activity j (corresponding to GP 89)

$BPW_{ij}$ = gross production values of GP headings allocated to class j (in DM)

$BPW_j$ = gross production values of class j (SYPRO)

$BPW_k$ = gross production values of class of economic activity k (WZ 93)

$A_{ij}$ = factors for the allocation of SYPRO gross production values to WZ 93

Factors $A_{ij}$ were used for recomputing all absolute values included in the index calculation. In the case of the production indices these were data on value added required for weighting purposes.

*Calculation of conversion factors*

For the conversion of the SYPRO-based indices for the classes of economic activity j to WZ 93 indices for the classes of economic activity k, suitable conversion factors $U_{jk}$ were required. The following Table 2 shows how these factors $U_{jk}$ were calculated.

**Table 2: Construction of SYPRO conversion factors**

| **$W_k$** | **$GP_{ij}$** | **$BPW_{ij}$ (€)** | **$S_j$ pro rata** | **$U_{jk}$** |
|---|---|---|---|---|
| $W_1$ | $GP_{11}$ | 150 | $S_1$ | 0.231 |
| | $GP_{32}$ | 500 | $S_2$ | 0.769 |
| *Total $W_1$* | | *650* | | *1.000* |
| $W_2$ | $GP_{31}$ | 250 | $S_1$ | 0.714 |
| | $GP_{12}$ | 100 | $S_2$ | 0.286 |
| *Total $W_2$* | | *350* | | 1.000 |
| $W_3$ | $GP_{21}$ | 100 | $S_1$ | 0.200 |
| | $GP_{22}$ | 400 | $S_2$ | 0.800 |
| *Total $W_3$* | | *500* | | *1.000* |

The total production value of a WZ 93 class of economic activity is composed of the production values of various SYPRO class. Thus in each case, a "weighting structure" can be computed for the aggregation of the SYPRO classes concerned to form a class of the new classification.

Taking the new class of economic activity 35.42 (Manufacture of bicycles) as an example, Table 3 shows the calculation of SYPRO conversion factors. Both SYPRO class 3324 (Manufacture of bicycles) and part of SYPRO class 3327 (Manufacture of parts for motor-cycles and_bicycles) were assigned to that WZ 93 class.

**Table 3: Example for computing conversion factors**

| **Class of economic activity** | | **Production value assigned to class of economic activity 35.42** | **Conversion factors** |
|---|---|---|---|
| **WZ 93** | **SYPRO** | **(1 000 €)** | **$U_{jk}$** |

| | | | |
|---|---|---|---|
| 3542 Manufacture of bicycles | 3324 Manufacture of bicycles | 1 059 322 | 0.669 |
| | 3327 Manufacture of parts for motor-cycles and bicycles | 524 115 | 0.331 |
| *Total* | | *1 583 437* | *1.000* |

All in all, DM 1.06 billion of the total production value of SYPRO class 3324 were allocated to WZ 93 class 35.42, while a total of DM 0.5 billion of the total production value of SYPRO class 3327 were assigned to that class. The trend of the production index for the WZ 93 four-digit heading 35.42 was then represented by the two indices for kind-of-activity units of SYPRO classes 3324 and 3327 combined by conversion factors $U_{jk}$.

*Application of conversion factors*

Before constructing long-term index series, the SYPRO indices had to be rebased on 1991 = 100.

**Table 4: Application of conversion factors – (1) Rebasing of SYPRO indices**

| **SYPRO branches of economic activity** | **Period** | **Production indices (calendar month)** | |
|---|---|---|---|
| | | **base 1985 = 100** | **1991 = 100 (rebased)** |
| 3324 Manufacture of bicycles | 1988 | 110.3 | 70.1 |
| | 1989 | 136.0 | 86.4 |
| | 1990 | 161.2 | 102.4 |
| | 1991 | 157.4 | 100.0 |
| 3327 Manufacture of parts for motor-cycles and bicycles | 1988 | 77.8 | 95.5 |
| | 1989 | 123.0 | 150.9 |
| | 1990 | 98.6 | 121.0 |
| | 1991 | 81.5 | 100.0 |

The purpose of the next step is to aggregate the rebased SYPRO indices by means of the conversion coefficients $U_{jk}$ to obtain the WZ 93 index of class 35.42. With the conversion factors Ujk any class of economic activity of WZ 93 can be constructed this way.

**Table 5: Application of conversion factors – (2) Aggregation of rebased SYPRO indices**

| **Branches of economic activity** | | **Conversion factors $U_{jk}$** | **Production index (calendar month) 1991=100** | | |
|---|---|---|---|---|---|
| **SYPRO** | **WZ 93** | | **1988** | **1989** | **1990** |
| 3324 | | 66.9 | 70.1 | 86.4 | 102.4 |
| 3327 | | 33.1 | 95.5 | 150.9 | 121.0 |

| | | | | | |
|---|---|---|---|---|---|
| | **3542** | | **78.5** | **107.7** | **108.6** |

### 2.2.4. Example 2 (Statistics Canada): Monthly Wholesale and Retail Trade Survey (MWRTS)

Throughout the years, Statistics Canada has used different versions of the Standard Industrial Classification (SIC) system and the North American Industry Classification System (NAICS) for industrial classification. The Monthly Wholesale and Retail Trade Survey (MWRTS), a major survey conducted by Statistics Canada, was developed in the late 1980's to produce sales and inventories estimates for SIC-based industrial sectors.

The MWRTS had to be redesigned to permit conversion to NAICS because the existing survey system did not permit the sample to be redrawn. The plan for conversion and back-casting included a parallel run for reference year 2003 and the release of NAICS based estimates toward the end of 2003. The stratification and sampling of the MWRTS was updated in 1998. Hence two different procedures were applied, one to the years prior to 1998 (to bring them into line with the 1998 results) and another to the years from 1998 onwards. The approach adopted for this work was the "macro" approach.

The following paragraphs are based on a paper presented by S. Fortier to the Statistical Society of Canada at its annual meeting in June 2003.

*Estimation of conversion coefficients*

The conversion coefficients $\alpha_{ij}(a,m)$ represent the percentage of the total of the group $i$ (old classification) allotted to group $j$ (new classification).

For the MWRTS, the values of the conversion coefficients have been derived from the data sampled for 48 months between January 1998 and December 2001. The coefficients considered by experts to be invalid or lower than 0.3% in absolute value were eliminated and reallocated. This reduced the number of series from 1000 to 230. The remaining series were analyzed graphically to detect the presence of regional differences, seasonality, or outliers. Finally it was decided to estimate the monthly coefficients for the years 1991 to 1997 on the basis of the average of the coefficients calculated for the corresponding months of the years 1998 to 2001, for each region.

The conversion coefficients can thus be written in the form

$$\hat{\alpha}_{ij}^{r}(1991,m) = \cdots = \hat{\alpha}_{ij}^{r}(1997,m) = \frac{1}{k}\sum_{a=1998}^{2001}\delta_{ij}^{r}(a,m)\alpha_{ij}^{r}(a,m) \qquad \text{for} \quad m = 1,\ldots,12$$

where the variable $k$ equals the sum (over the four years) of the indicatrix $\delta_{ij}^{r}(a,m)$ defined as

$$\delta_{ij}^{r}(a,m) = \begin{cases} 0 & \text{if } \alpha_{ij}^{r}(a,m) \text{ is considered invalid;} \\ 1 & \text{otherwise.} \end{cases}$$

This allows to withdraw outliers from the calculation of the average. The coefficients obtained are readjusted to sum up to 100% for each combination of year $a$, month $m$, region $r$ and trade group $i$ of the old classification (SIC).

*Combination of estimates from the old classification with conversion coefficients*

For each trade group under the new classification, a weighted linear combination of the total of each group according to the old classification was used. The total $X_j(a,m)$ of the trade group $j$ according to the new classification for the year $a$ and the month $m$ is given by

$$X_j(a,m) = \sum_i \alpha_{ij}(a,m) X_i(a,m)$$

where $X_i(a,m)$ is the sum of the trade group $i$ according to the old classification. The weights of the linear combination are the conversion coefficients $\alpha_{ij}(a,m)$.

*Continuity of the series under the NAICS*

The series under the new classification (NAICS) are divided into three segments. A first from January 1991 to December 1997 where the estimates were obtained using estimated conversion coefficients. The second segment starts in January 1998 and finishes at the time when the survey based on the old sample was discontinued. In this second part, the series under the NAICS were obtained by domain estimates, on the basis of observed conversion coefficients.

The third segment starts with the new survey. There is an overlap of a few months where both the old and the new survey were in production (tested in parallel). There was expected to be a break in the series at the time of the switch to the new survey this break being explained by the change of classification but also by other methodological changes. It had been envisaged to use the results of the test period to adjust the level of the retrofitted series applying a constant multiplicative adjustment over time to adjust the historical series to the levels resulting under the new survey.

There was also a break observed in January 1998 when switching from the estimated to observed coefficients. To reduce this effect, all the retrofitted data for 1998 were recomputed by using the estimated coefficients (i.e. the coefficients calculated on the basis of the four years average, including 1998). In fact, the 1998 coefficients were differing from the average more strongly than the other three years. Outliers were removed from the average calculation. By extending the first segment of the series until December 1998, the break between the first two parts was cancelled out.

*Sources of errors*

A first source of potential error is the sample frame itself. A classification error for a given month between 1998 and 2001 affects not only that particular month in question but also the corresponding month each year between 1991 and 1997. In order to reduce the impact of the wrongly classified units, the large contributors were manually checked and re-coded where needed. In the case of corrections of the sample frame since 1998 the series estimated under the new classification (NAICS) were adjusted.

The second type of error comes from the use of the conversion coefficients calculated over recent years (1998-2001) to estimate the coefficients of conversion of the former years. This method is appropriate if the distribution according to the old classification is stable from one year to another. If not we can nevertheless assume the risk of error to be minor for 1997 than for 1991. In most of the cases, the assumption of stability was accepted.

The use of conversion coefficients calculated as described above is however not applicable (or should rather not be applied) where an industry underwent an important change. Where a specific industry had little importance at the beginning of the observation period, the conversion coefficients for those years have to be revised downwards while the coefficients of the other sectors have to be revised upwards.

The value of the adjustment is based on experts' opinion and the results of a partial classification under the new classification at micro level. This type of adjustment makes it possible to model the variations over time of the coefficients. Due to the low number of observations at the moment of the analysis, no adjustments were made for calendar and working / trading day effects.

An additional source of errors arises from the use of coefficients based on one variable and used on another. There are two variables of interest in the MWRTS, the sales and stocks. The whole work on the coefficients was carried out with reference to the sales. The series of stocks was calculated by applying ratios to the retrofitted series of sales.

### 2.2.5. Example 3 (Statistics Canada): Survey of Employment Payroll and Hours (SEPH)

*Conversion of historical series from the 1980 Standard Industrial Classification (SIC80) to the North American Industrial Classification (NAICS)*

In order to convert the SIC80 series to NAICS, data from two periods of three consecutive months of 1998 and an additional period of three months in 1999 were used. For these periods, information obtained from the Statistics Canada Business Register was used to re-code the micro-data (at the establishment level) from SIC80 to NAICS. Estimates were then computed for the combinations of each detailed SIC industry by each detailed NAICS industry for each province and variable ([17]). Conversion ratios were created by dividing these estimates by the corresponding estimate at the detailed 1980 SIC level (by province and variable). These ratios were then used to convert 1980 SIC estimates into NAICS estimates for the full period from January 1991 to December 2000.

Once converted, the new series were analyzed for consistency. Historical corrections (accumulated since the beginning of Phase II in May of 1996) were also incorporated in the data series. Because of the number of series involved, the analysis was concentrated on the most significant variables for each province (such as average weekly earnings, employment, etc.).

Users should note that the conversion method used has some limitations. The choice of the method was constrained by the non-availability of NAICS establishment information for the business population for earlier periods and also the inherent changes in the target population (establishments with employees) through time (births and deaths of establishments). Had the information to re-code each monthly data file been available, the resulting series would have been somewhat different. For example, new industries came into existence within the decade while others might have disappeared in some provinces. This has a negative impact on the quality of the converted series especially as one moves further back from the time the conversion ratios were estimated.

In addition, because a conversion ratio method was used, the patterns observed for a closely related set of 4 digit NAICS series may in some cases, be very similar over the 1991-2000 period since the conversion may have been based on a higher level of aggregation. In these cases, the close relationship shown by the 4 digit NAICS series may end in January 2001 as each of these more detailed series will now be analyzed separately.

### 2.2.6. Advantages and shortcomings of the Proportional Methods

*Advantages*

([17]) SEPH produces estimates for eleven base variables from which all other derived variables are calculated.

** The proportional method being applied at "macro" level one does not need to go back to the micro data of the individual units. It is thus a low resource and time consumption approach to the backward calculation, but it offers only an approximate solution that does not analyze in a very deep way the revision effects on time-series.

*Shortcomings*

** The application of coefficients to data classified under the old system, in order to convert these data to the new standard, is just an approximation of what the earlier observations may have been.

-- It would be ideal to have conversion coefficients calculated for every point in time within the historical and transitory segments, and for every variable of interest. However, for reasons of limited resources, there will often be just one set of conversion coefficients, for a single year (or whatever reference period) and calculated based on one variable (e.g. employment) and applied to another variable (e.g. earnings). Combining these coefficients with the estimates from the old estimation is working with fixed weights which are entirely driven by the chosen reference period.

This might work well for short periods of time. But the assumptions underlying the coefficients will become invalid over longer periods where the new system's economic structure differs substantially from that of the old.

-- HOWEVER, conversion coefficients could be established at least for a number of ("benchmark") years. One could on this basis determine whether the conversion coefficients at a single point in time are appropriate.

-- ADDITIONALLY, evidence from such a benchmarking exercise, or simply experts' opinion, might be used to carry out adjustments of conversion coefficients for certain years (e.g. giving DVD salesmen lower coefficients at the beginning of the nineties). The conversion coefficients of particularly important industries could be fine-tuned by means of micro-level techniques.

-- FURTHERMORE, complications arising from important shifts in the composition of industry groups over time, and especially the problem of "previously-out-of-scope" units, are not specific to macro approaches.

-- If the same concordances are used for every month in a year, the seasonal pattern of the reconstructed historical series will be distorted.

-- A revision of the economic classification system usually encompasses broad changes to reflect innovations in industrial composition. The new classification principles are likely not to reflect the economic reality of historical data.

*Provisional conclusion*

** In practice, nobody will probably rely on the exclusive use of either macro or micro techniques. The macro approach has the principal advantage that it is cheap and fast. Its main disadvantage is that, when applied in its most simple form, the results will be meaningful just for a short period of time. However, there are ways, including micro methods, to overcome some of these shortcomings.

### 2.3. Methods applying interpolation between benchmarks (combined micro and macro-approaches)

The reconstruction of disrupted time series can be done at micro-level, at macro level or a combination of the two. This section presents the methods which require the recoding of units at micro-level for two periods (months, quarters or years depending on the periodicity of the statistics). The essential scope is to derive as many conversion coefficients as the periods included in the time series to be reconstructed.
The two periods which are "double coded" are indicated by A and B. and are also called "benchmark periods"**:** the optimal benchmarking periods are to be determined by subject matter experts.

According to this method, the micro-data for period A and B are recoded to the new classification. Then, two sets of conversion coefficients are obtained to convert the aggregated estimates from NACE Rev. 1.1 to NACE Rev. 2. For the periods between A and B, the coefficients are interpolated. Finally, these interpolated coefficients are applied to convert the estimates and revise the series.
The interpolation of the coefficients between periods A and B allows taking into account the evolution that might have occurred in the NACE distribution. For example, the proportion of units classified under "retail sale of computer" has certainly increased since 1982. This evolution might have been not linear, and therefore a non-linear interpolating method could be used.

A single set of ratios could be used for all the variables of interest (e.g. turnover, value added, employment, etc.) or one set of ratio for variable of interest (i.e. one set for turnover, another set for value added, etc.). Using one set of ratio keeps the consistency between the variable of interest; for example, the ratio of value added to the turnover would not be affected. But using one set of ratios per variable of interest reflects more the different splits that can occur by converting from NACE Rev. 1.1 to NACE Rev. 2. For examples, consider the split of a NACE Rev. 1.1 class into two NACE Rev. 2 classes where one NACE Rev. 2 class contains the high share of value added; then, using the same ratio for value added and turnover would not reflect the movement of high value added to a particular NACE Rev. 2 class.

A possible variation of the method just described consists in combining the coefficients determined on the basis of A and B into a single set (mean of the two's) and then apply these conversion coefficients to all the periods of the time series. This is quite a crude assumption, but less crude than the one made when applying simple proportional methods.

**PART 3: REQUIREMENTS FOR TIME SERIES RECONSTRUCTED ACCORDING to NACE Rev. 2**

## 3.1 Introduction

According to the regulation establishing NACE Rev. 2, the European Commission will apply NACE Rev. 2 to all statistics classified according to economic activities. As a consequence, existing statistical time series referring to NACE Rev. 1.1 will be disrupted and this will create huge problems for users of economic statistics. Therefore, the provision of time series reconstructed according to NACE Rev. 2 is a crucial element of the activities related to the implementation of NACE Rev. 2.

The European Statistical System (ESS) is undertaking all efforts to implement NACE Rev2 in a strictly co-ordinated manner in order to fulfil users' request. However, there is a trade-off between the conversion of "old" NACE data (which, in many cases, did not provide the breakdown and details as included in the "new" NACE) and the provision of economically meaningful time series. For example, the use of NACE Rev. 2 items in statistical time series covering historical periods of 30, 40 or more years might not always be possible, given that some economic activities did not exist by that time. A careful assessment on the time span for which back-data will be made available by the ESS is therefore necessary.

The next section lists the requirements for reconstructed time series, which are included in Community law, and are legally binding for all Member States. The chapter will be continuously updated according to upcoming information.

### 3.2 Legal requirements

The establishment of NACE Rev. 2 affects statistical domains, which are regulated by EU legal acts and which present statistics according to economic activities. Such legal acts (e.g. Council Regulations, Commission Regulations) *inter alia* specify the reporting obligations of Member States with regard to the level of detail, the frequency and the starting period of data.

The table below lists the statistical domain, the starting year of application (i.e. the availability of data), and the current provisions related to the transmission of reconstructed time series or double reporting of data.

| Domain | Ref. Year | Delivery first data according to NACE Rev. 2 | Back-cast time series: length | Delivery of back-cast time series | Reference Year for dual coding |
|---|---|---|---|---|---|
| | | | | | |
| Energy | 2008 | November 2009 | - | - | - |
| Labour Force Survey | 2008 | June 2008 | Voluntary basis | Voluntary basis | Voluntary basis |
| Structural Business Surveys | 2008 | October 2009 | Voluntary basis | Voluntary basis | 2008 |
| EU-Survey on Income and Living Conditions | 2008 | December 2009 | - | - | 2008 |
| Science & Technology | 2008 | October 2009 | 2003-2007 | October 2009 | 2009-2010 |
| FATS-inward | 2008 | August 2010 | Voluntary basis | Voluntary basis | 2008 |
| Labour Cost Survey | 2008 | June 2010 | - | - | 2008 |
| Short Term Statistics | 2009 | March 2009 | 1998-2008 | March 2009 | - |
| Labour Cost Index | 2009 | June 2009 | 2000-2008 | June 2009 | |

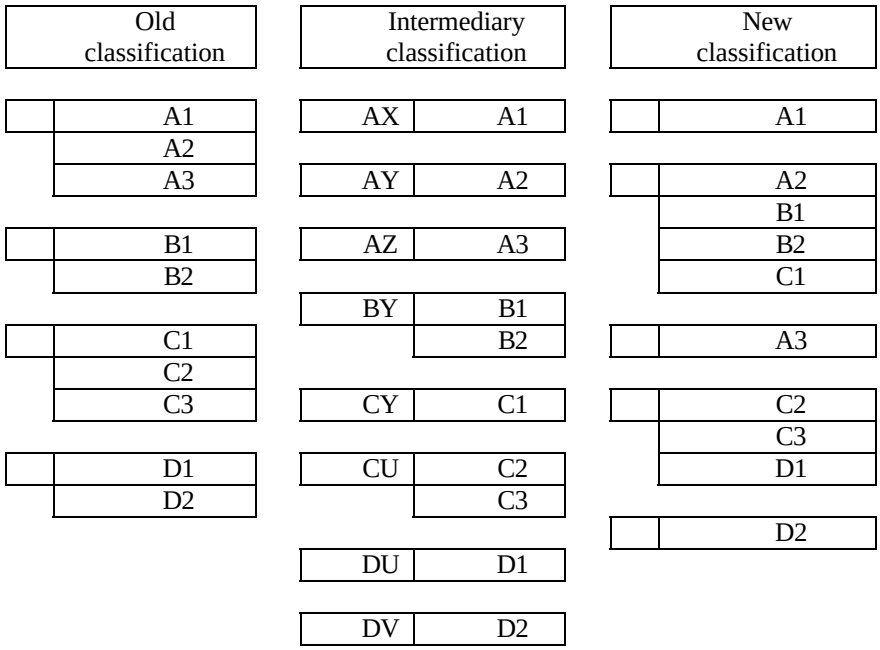
| | | | | | |
|---|---|---|---|---|---|
| Inform. Soc. | 2009 | October 2009 | Under discussion: 2003-2008 for a list of core indicators | Under discussion: June 2009 | 2009 |
| Balance of Payments | 2010 | September 2011 | 2008-2009 | Sept. 2011 | 2009 |
| FATS-outward | 2010 | September 2012 | 2008-2009 | Sept. 2012 | - |
| National Accounts | 2010 (annual data) 2011Q2 (quarterly data) | September 2011 | Under discussion: 1990-2010 (specific variables, new MS: 1995-2010) | Under discussion: In two batches: Sept. 2011 and Sept. 2012 | - |
| Structure of Earnings Survey | 2010 | June 2012 | - | - | - |
| European Agriculture Accounts | 2010 | September 2011 | 1995-2010 | Sept. 2011 | - |
| Waste Statistics Regulation | 2008 | June 2010 | Eurostat will reconstruct 2004 and 2006 on the basis of the 2008 information | - | 2008 limited to major changes on a voluntary basis |
| Business register | 2008 | May 2008 | Voluntary basis | - | Length optional |
| Community Vocational Training Survey | 2010 | 2012 | - | - | - |
| Job Vacancy survey | 2009 | June 2009 | Voluntary basis | Voluntary basis | 2008 |

- means that no back cast or double reporting is foreseen

## PART 4: REFERENCES AND TABLES

### 4.1 Annex to Section 2.1

Example of building up of an intermediary classification




The intermediary classification is the result of the cartesian product of the old and the new classifications: the codes AU, AV, BX, BZ, BU, BV, CX, CZ, CV, DX, DY and DZ do not appear, as they corresponds to empty sets (there are no intersection between the two codes).

We note that several codes of the intermediary classification corresponds to codes either to codes of the old classification, or to codes of the new classification: AX=X, AZ=Z, BY=B et DV=V.

The elements A1, A2, etc., D1, D2 are the "breakdowns" of codes the old or of the new classification; these breakdowns could be motivated by breakdowns of activities, or of products, or of groups of products.
The detail of breakdowns provided here is greater than strictly necessary for this intermediary classification: the detail C2, C3 is not more informative than the set {C2, C3}; the same holds for B1, B2. On the other side, the number of codes of the intermediary classification (8) is the smallest possible in this example, and much smaller than the theoretical one (20, from 4x5).

Conversion matrices

A conversion matrix is a matrix of dimension IxJ, where I is the number of classes of the old classification and J the number of classes of the new classification. The generic element of the matrix $c_{i,j}$ is the probability that an activity codified as "i" in the old classifications codified as "j" in the new classification. Therefore, $c_{.,j} = 1$.

These probabilities are determined by the empirical conversions, observed in the double coding year on the observed units. The necessary information for the identification of the conversion matrix is the elementary and the principal activity of units.

Major precision could be obtained by constructing conditional conversion matrices, keeping into account the size of the statistical units, or its principal activity, or the number of its elementary activities carried out by the unit.

### 4.2 Commented references

| | Author | Title | Short summary or key points |
|---|---|---|---|
| 1 | ESTAT Classification and Euro-indicators sections | Back-casting time series broken by new classification coding | • Methodologies based on aggregated (macro-data) series<br>• Reference list |
| 2 | INSEE | The role of SBS surveys within classification changes | *The paper presents some options for backcasting the estimates of the business surveys. Both micro and macro methods are briefly mentioned.*<br>*Micro method is preferred to handle classification change because it overcomes the constant structure constraints of the macro method.*<br>*For the implementation of the micro approach, the same conversion factors as considered in the double coding span in the activities of a single enterprise (for ex. referring to the outcome) are applied to the same elementary activities in the previous years.*<br>*Only activities carried out previously by the same enterprise and not present in the "double code year" are recoded using an average transition matrix drawn up from all the businesses (or a more uniform sub-population).* |
| 3 | INSEE | Long-period series in base 1995: manual recalculation and econometric retropolation of the IPI | The retropolation program estimates linear models between series of two successive bases over a common period which should be as long as possible (generally 7 years). The retropolation program draws linear approximations of the retropolated series in order to produce estimations of the missing values from the past. The dynamics of the series to be retroplated and the turning points of the retropolating series are generally well taken in account. The models are constructed by maximizing the log likelihood ratio calculated using the Kalman filter method. Two types of dynamic models are tested. |

| | | | |
|---|---|---|---|
| 4 | ISTAT – Moauro | Modelling a Change of Classification by a structural Time Series Approach | *In this paper a new approach of backward calculation is suggested. The change of economic sectoral time series data is examined by a conversion matrix approach. A state space form is set up considering the new sectoral standards figures to be reconstructed as unobserved and the few available observations as time-varying restrictions. The Doran methodology of constraining the Kalman filter to satisfy time varying restrictions is applied to increase efficiency of the estimates.* |
| 5 | University of Ca`Foscari Venezia | Methodological aspects of time series back-calculation | *This paper provides theoretical and operational framework for back-casting. The authors used an ARIMA model to produce estimates. To get a reliable estimation we need a relative long time span. In our case we can't use this method.* |
| 6 | OECD - ISTAT | Retrapolating Italian annual national accounts data according to ESA95 | |
| 7 | UK – N.I.E.S.R. | Backward calculation of national accounts data (Retrapolation) | |
| 8 | ISTAT | Time series reconstruction by the Kalman Filter | *In this paper the main instrument for time series reconstruction by state space models are provided. The Kalman filter provides a well-established procedure to obtain optimal parameter estimation of a state space model. This work gives a general description of the Kalman filter, the Doran and Doran and Rambaldi methodology. Basic tools on initial conditions, missing observations are provided. This report is a good theoretical summary of the state space* models. |
| 9 | Eurostat | Backward calculation techniques - 1 | The introduction of EURO is an economic event that has a big impact on the national accounting system. Member States have to convert their historical time series, expressed in national currency, in Euro series. Two methods of backward calculation are distinguished: the annual backward calculation and the benchmark years and interpolation. The latter one is based on a two step procedure. In the first step detailed estimates for one or more benchmark years are calculated. In the second step, figures for the |
| 10 | Eurostat | Backward calculation techniques - 3 | |
| 11 | Eurostat | Backward calculation techniques - Bibliography | |

| | | | |
|---|---|---|---|
| | | | remaining years are determined by interpolation. Two methodological approaches are described: the Netherlands and the France case. The former is a variant of the layer correction method belonging to the benchmark years and interpolation category, the latter focuses more on theoretical aspects of the Kalman filter. |
| 12 | Università di Padova | Constrained retropolation of high-frequency data using related series | |
| 13 | OECD - Voorburg | Compilation manual for an index of service production | |
| 14 | U.S. Bureau of Economic Analysis | The impact of classification revisions on time series | *This paper describes in general terms how pros and cons could be balanced when a revision of a classification is considered. The reconstruction of the broken time series is performed by creating linkages where the series break. A concordance between the new and the old series can be developed via "dual classification". However, because of the new classification principles of the revised classification, the series do not necessarily reflect the economic reality of the historical data. The author recommends the use of microdata.*<br>*In a separate paragraph the paper lists all the kinds of activities and costs requested by the revision project..*<br>*In the last paragraphs the rationale to revise a classification is considered. When is the analytical gain from improving the classification high enough to justify the costs of broken historical continuity? The paper doesn't provide a definitive answer, but shows the relevant considerations.* |
| 15 | U.S. Bureau of the Census | Methods used to develop retail and wholesale time series under the north American industry classification system | |
| 16 | UN – ADB - ESCAP | Basic principles and practices in rebasing and linking national accounts series | |
| 17 | Statistics Canada | Introduction to concordances | |

| | | | |
|---|---|---|---|
| 18 | Statistics Canada | Implementing a NAICS-based time series into Canadian System of NA | *Focus on Input-Output tables.*<br>*The paper describes the changeover method used by national accounts in Statistics Canada when the new NAICS classification was implemented. First the new classification was implemented in the input-output table 1997. This work relied on a very good concordance old-new classification for 1997, based on some preliminary work. Afterwards, the series 1961-1996 was reconstructed with this correspondence. Basically, the correspondence in 1997 was applied to all the years 1961-1996, although some adjustments were done for products that disappeared from the markets in 1997. The correspondence was applied separately to outputs and inputs in each industry (=group of firms in the same economic activity), keeping very tight control on some accounting constraints and allowing some others to "float". Finally, an automatic balancing algorithm was used to "fix" the accounting rules that had "floated". The purpose of this was to keep value added by industry under control and avoid GDP and growth rates to change as a result of the new classification.*<br>*The paper underlines the importance of preliminary work, before national accounts changed over. It is said that the business register had double coding for several years. They also say that a good concordance old-new classification was established for year 1997, combining the double-coded business registers with administrative information (from firms' tax registers).*<br>*The Canadian national accountants work with 4 aggregation levels (i.e. working detail of the classification). The second most detailed aggregation level was created explicitly to be used for the backcasted series 1961-1996. Actually, it was defined with a view to ensure a smooth transition between the old and new classification. "Smooth transition" means here that the value added of old and new industries is approximately the same. It is said that this aggregation level was not analytically useful; it looks like a mere tool.*<br>*The backcasting was macro-data based. A key message is that more importance was given to consistency induced by national accounts accounting rules than to a very sophisticated correspondence old-new classification for a long period of time (only 1-year long correspondence* |

| | | | |
|---|---|---|---|
| | | | *was used, for 1997). These constraints imposed by accounting rules (= relations between variables that must perfectly match) have no equivalent out of national accounts domain (e.g. in business surveys), or do not impose a comparable level of rigidity.*<br>*Although the backcasting was macro-data based, there was some limited complementary use of micro-data based calculations, only for few activities. The paper mentions (pg 7) that the manufacture and mining survey was re-processed (i.e. micro data re-classified and grossed up again with new weights), but only for year 1997. This exercise "greatly helped to produce consistent results". Pg. 4 says that it would have been very expensive, if possible, a micro-data based approach back to 1961, and therefore other options were developed to back-cast the national accounts series.* |
| 19 | Statistics Canada | Annex to "Implementing a NAICS-based time series into Canadian System of NA" | |
| 20 | Statistics Canada | Press release on historical national input-output tables | |
| 21 | Statistics Canada – Mike Hidiroglou, Benoit Quenneville, Guy Huot | Methodological problems and options for SIC-NAICS conversion<br>(October 2001) | • Methodologies based on aggregated (macro-data) series<br>• Reference list<br>*The paper refers to conversion tables and concordance coefficients for assuring historical continuity of time series. The conversion tables provide a comparison of codes in the old and new systems; the concordance coefficients are conversion factors, showing how much each industry has changed. Concordance coefficients can be computed at a single time point or at several points in time.*<br>*Both micro and macro methods are mentioned for the re-construction of the series. Regarding micro approach, domain estimation can be carried out using re-coded records. This requires first assigning new codes to all sampled units in the historical span, then the production re-run with domain estimation for all time points (using survey weights from the old classification).* |

| | | | |
|---|---|---|---|
| 22 | Statistics Canada -Canadian Statistical Society | The conversion of historical time series according to a revised classification in the wholesale and retail sale monthly survey (June 2003) | *Both micro and macro approaches are considered and their pros and cons presented. Despite the higher level of precision, the micro approach is considered complex and the macro approach is proposed for implementation.*<br>*Regarding the micro approach, the following steps are envisaged:*<br>*- the double code is assigned to the units included in the sample of the survey (assuming that no change of activities has occurred in that year);*<br>*- To the other units, a code is assigned according to the probability of assignment established empirically by the frequency of each relationship old-new observed in the" double code year".*<br>*- In case of one-to-many relationship, the reclassification can be done using a "division method"(a certain percentage of the variable of interest is recoded according to a factor of division derived from the data for which classification is known under the two systems).*<br>• Reference list |
| 23 | Statistics Canada | Statistics Canada's Experience with NAICS 1997 Implementation and Back-casting | Very detailed report on Statistics Canada's experience in implementing a new classification. Covered topics include: management, dissemination and implementation in specific statistical domains. |
| 24 | UN? | Review of country practices on rebasing and linking National Account series | |
| 25 | Caporin- Sartore for Eurostat | Methodological aspect of time series back-calculation for selected PEEI | |
| 26 | ECB | Technical note on the derivation of historical time series of monetary aggregates | |
| 27 | ECB | Interpolation and backdating with a large information set | |

| | | | |
|---|---|---|---|
| 28 | Wallgren & Wallgren – Statistics Sweden | Register statistics – administrative data for statistical purposes<br>Chapters 8 and 9 | Chapter 8 "Calibration and imputation" and Chapter 9 "Estimation with combination objects", presents two methods for linking time series (backcasting). Both methods use micro data and all time series based on these microdata are backcasted and completely consistent with each other. One method uses calibration of weights and the other method combines the detailed information in the Business Register. Numerical examples illustrate the methods, based on real cases observed at Statistics Sweden. |
| 29 | Statistics Canada – R. Laflotte, S.Lavallée, P.Lavallée | Converting the SEPH historical series to NAICS | SEPH is the Survey of Employment, Payrolls and Hours. The paper is very clear and presents several methods used at Statistics Canada for reconstructing time series broken by the change of the industrial classification**.** The methods presented combine micro and macro approaches, together with pros and cons and possible drawbacks and advantages. |
| 30 | Statistics Canada M. Morry | Backcasting time series at Statistics Canada under NAICS | Power-point presentation with an excellent review of micro and macro methods for backcasting, including examples. |
| 31 | Destatis – C. Bald-Herbel, N. Herbel | Rebasing the indices of production industries on 1991 (1996) | |
| 32 | Statistics Canada - J. Leduc | SEPH estimates are now based on North American Industrial Classification System (NAICS) (2001) | |