



## **Centraal Bureau voor de Statistiek**

MSP

MIC/Bedrijvenconjunctuur

*Postbus 4000*

*2270 JM Voorburg*

---

## **Backcasting Nace-2: Business surveys, an example.**

**Gert Buiten, Arjan Neef, Sidney Vergouw**

*Summary:*

*Keywords:*

Kennisgeving:

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

---

*Projectnummer:*

*CBO-*

*BPA-nummer:*

*-CBO*

*Datum:*

*17 oktober 2007*

## **BACKCASTING NACE-2: BUSINESS SURVEYS, AN EXAMPLE.**

|     |                                                                                                                                |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------|----|
| 1.  | Introduction .....                                                                                                             | 2  |
| 2.  | Overview of Main Methods .....                                                                                                 | 3  |
| 2.1 | Use of a key .....                                                                                                             | 3  |
| 2.2 | Micro approach.....                                                                                                            | 3  |
| 2.3 | Macro approach .....                                                                                                           | 3  |
| 2.4 | Benchmark years/interpolation method.....                                                                                      | 4  |
| 3.  | Methodology of Business Surveys.....                                                                                           | 5  |
| 3.1 | Special features Business surveys .....                                                                                        | 5  |
| 3.2 | Methods for backcasting business survey indicators.....                                                                        | 5  |
| 4.  | Two macro methods: example 1: Aggregate3435.....                                                                               | 7  |
| 4.1 | Construction of “basic data”.....                                                                                              | 7  |
| 4.2 | Method I: Macro method, computing weighted balances based on the transition-matrix (table 3).....                              | 8  |
| 4.3 | Method II: Macro/micro method, convert balances of old series using a conversion-matrix of balances (based on micro data)..... | 9  |
| 4.4 | First conclusions about the methods and some additional remarks.....                                                           | 10 |
| 5.  | The “pure” micro method: example 2: Manufacturing Industry .....                                                               | 12 |
| 5.1 | Preliminary work and assumptions: unique weights, recoding and non-response                                                    | 12 |
| 5.2 | Calculations and summary of the results .....                                                                                  | 13 |
| 5.3 | Conclusions and remarks.....                                                                                                   | 15 |
| 6.  | Conclusions .....                                                                                                              | 16 |

## **1. Introduction**

An important part of the implementation of NACE Rev 2 is the conversion of the current historical time series from NACE Rev 1.1 to NACE Rev 2. This document debates possible procedures to backcast historical time series in the area Business Surveys.

The second chapter starts with the basic principles of reconstructing time series in general: the four main methods are discussed. Chapter three handles the special features of Business surveys. These features imply the main methods are not directly applicable for backcasting Business survey indicators. Therefore, there will be discussed three methods which are adapted to specific characteristics of Business surveys: two macro methods and one micro method. In chapter four the two macro methods are illustrated by using an example. The chapter ends with some first conclusions.

Subsequently, in chapter five an example is used to clarify the micro method. Based on the results some conclusions are drawn and remarks are made. Finally, this document ends with final conclusions, remarks and recommendations.

## **2. Overview of Main Methods**

There are several possible procedures to apply a revised classification to historical time series. They can be divided into four main methods.

1. Use of a recoding key on published series
2. Recalculating data by recoding units at the micro level (“reconstructing”)
3. Converting or reinterpolating published series using a conversion matrix (“backcasting”, “macro-approach”)
4. Combining the micro and macro approaches by estimating benchmark years with a micro method and interpolating with macro techniques (“benchmark years/interpolation method”)

### **2.1 Use of a key**

The method based on the use of a key is the most straightforward and simple of the four methods described in this report. The technique use a recoding “key” with which a classification at the lowest aggregation level is directly recoded to the revised classification E.g. the old code 4.3.2.1 is recoded to 1.2.3.4 and the historical data for 4.3.2.1. are assigned to 1.2.3.4. In its purest form, this method can only be applied if there are only 1-on-1 or n-on-1 changes from the old to the new classification. The key method assures a straightforward relationship between the old and the new results, since the old data are simply transferred or projected onto the new classification.

### **2.2 Micro approach**

The micro-approach means that the revised classification will be applied to the historical time series by assigning the revised classification to each statistical unit and for every period in the time series. That is, all statistical units used for calculating the old time series are coded again according to the new classification. After that, the statistical results (averages, totals etc.) are recalculated using the same calculation routines as for the old data. In fact, the entire production process is repeated starting from the micro level, but now using the new classification. Therefore technique is also known as the micro-approach.

The method is not dependent on the type of relationship between old and new codes and can also cope with for instance 1-on-n and n-on-m relations between the old and new classification. Because of the double coding of the units according to the old and new classification, there is an exact relationship between the old and new results.

### **2.3 Macro approach**

Opposite to the micro approach, the macro approach works at aggregate levels. The data based on the initial classification are redistributed according to the revised classification with the help of a set of conversion coefficients. These conversion coefficients are derived from a

conversion matrix. This means only one point in time is double coded according to the old and revised classification.

When applying a conversion matrix from one period also for other periods, assumptions have to be made. Basically, one has to assume that a specific aspect of the structure from the conversion matrix remains constant over time. E.g. one may assume that for the entire time series, 60% of the value of the old group 4.3.2.1 is assigned to the new group 1.2.3.4.

The use of conversion matrices enables this method to cope with all types of relationships between old and new groups. That means that also 1-on-n and n-on-m relationships can be handled. In the case of 1-on-1 changes, this method will act the same as the key method.

## **2.4 Benchmark years/interpolation method**

This technique is a combination of the micro and the macro approach. According to this method, two periods have got to be double coded. These periods are called the benchmark periods. The optimal benchmark periods are to be determined by subject matter experts. This method can deal with all types of relationships between old and new Nace groups, including 1-on-n and n-on-m relations

First of all, the micro data for the benchmarking periods are recoded to the revised classification. After that, two sets of conversion coefficients are obtained to convert the aggregated estimates from the initial to the revised classification. For the periods between the two benchmark periods, the coefficients are interpolated. For some subclasses, the evolution between the two benchmark periods might not have been linear. Therefore, a non-linear interpolating method could be used. As mentioned in the macro-approach, one could in such cases make use of experts' opinions.

A possible variation of the method described consists in combining the coefficients determined for the two benchmark periods into a single set and then apply these conversion coefficients to all the periods of the time series.

### **3. Methodology of Business Surveys**

The macro approach discussed above is mainly meant to backcast index levels. This implies these techniques are not directly applicable to business survey indicators, since there's a conceptual difference. Index levels are based on e.g. turnover, value added or number of employees, while business survey indicators are based on the percentage of optimists, pessimists and neutrals. This makes it impossible to divide or aggregate these results in the same way as e.g. turnover. Besides that, business survey indicators don't have a long term growth or decline, since they fluctuate around a long term trend. This makes it impossible to make assumptions about differences in growth within one subclass. Last, time series related to business indicators can never result in scores less than zero or more than hundred percent.

Therefore, we've developed and tested three methods which are basically based on the same principles of the discussed methods above, but are adapted to the specific characteristics of business surveys. However, before these techniques can be explained and a practical example can be shown, it's necessary to know the main characteristics and procedures related to the Dutch business survey. This will be done in the next section.

#### **3.1 Special features Business surveys**

Business surveys are designed to give a fast first impression of economic tendencies and sentiment of businesses and consumers. To create a fast first impression, the questions asked are very simple qualitative indications of developments. For example: how has the production evolved last month: it is improved, it is unchanged, it is decreased? With the individual answers from the units, weighted balances (in percentage of the businesses) of the positive and negative answers can be computed.

The units are classified according to the Nace-1 and put in groups. Within each group, each individual unit has its own weight. This "inner weights" are used to calculate the results of the group. For aggregating groups into division or aggregates "outer weights" are used. Based on turn-over data from SBS, each group is assigned a coefficient based on absolute turn-over. Different aggregates are composed with different groups, and due the outer weights aggregation is possible.

#### **3.2 Methods for backcasting business survey indicators**

The first two methods tested are based on the principles of the macro approach. The first macro approach converts aggregated confidence indicators data according to revised classification by using a transition matrix of the weighting scheme to calculate weighted averages of the "old" data for every new publication aggregate.

First, the "outer weights" of the old Nace-1 groups have to be distributed over the elements of these groups that go to different new Nace-2 groups. This will be done with the use of a conversion matrix. Subsequently, these distributed weights are used to calculate a weighted average percentage for the answer from the business survey. This approach implicitly assumes that confidence is evenly distributed within all components of the old Nace-1

aggregate. In other words: every element of the old NACE aggregate that corresponds with a different new Nace-2 aggregate has the same confidence as the other components.

The second macro accepts that confidence levels (e.g. the percentage of positive answers) may differ within the components of an old Nace-1 group. It calculates the ratio between confidence of every cell in the transition matrix and the confidence of the corresponding entire old Nace-1 group. This method implicitly assumes that this ratio is stable over time. The confidence levels of the cells composing a given new Nace aggregate thus differ, but all cells coming from the same old aggregate still follow the same pattern over time.

In the second part of the analyses, the micro method is tested. Basic idea is to recode every unit and use the original data to repeat the production for the new Nace-2 classification, for each period available. To make use of the original inner and outer weights, new unique weights for each statistical unit (business) are assigned. New aggregation under the Nace-2 classification takes the historical weights into account and is therefore comparable with the aggregates under the Nace-1 classification.

#### 4. Two macro methods: example 1: Aggregate 3435

For analytical purposes, we have used a simplified example from our business survey Industry. The choice for the aggregate 3435 (transport industry) was made because of the relatively simple transition of the Nace-1 groups over a limited number of Nace-2 groups (3 digit). The original five Nace-1 groups which made up this aggregate are converted into six Nace-2 groups. Within the five groups there are 98 units. The first step was to make sure every unit was recoded into a Nace-2 classification.

##### 4.1 Construction of “basic data”.

Both methods need the same basic data.

The first table is a conversion matrix based on turnover. This matrix is a provisional result of the transition of turnover figures (monthly statistic April 2007) from Nace-1 to Nace-2. Within this example, it shows that the groups 341, 342 and 343 are converted one on one to the new group. Groups 351 and 353 each split into two new groups. In total, six nace-2 groups are constructed. The Nace-2 group 331 consist of parts from two Nace-1 groups. In the Nace-2 group 293 about one percent of the turnover comes from a group outside the 3435, but for simplification this is taken out of the example.

Table 1: Conversion matrix based on turnover (preliminary figures, monthly statistic April 2007)

| Nace-1 1993 | Nace-2 2008 |     |     |     |     |     | total<br>Nace-1 1993 |
|-------------|-------------|-----|-----|-----|-----|-----|----------------------|
|             | 291         | 292 | 293 | 301 | 303 | 331 |                      |
| 341         | 430         |     |     |     |     |     | 430                  |
| 342         |             | 125 |     |     |     |     | 125                  |
| 343         |             |     | 103 |     |     |     | 103                  |
| 351         |             |     |     | 274 |     | 79  | 353                  |
| 353         |             |     |     |     | 30  | 34  | 64                   |

The 2<sup>nd</sup> table shows the weighted balances of positive and negative answer for the question how the production has evolved the last month. Furthermore the “outer weights”, which are used to calculate the aggregate, are included.

Table 2: Historical series

| Evolution of production |             | Balance of positive and negative answers |                |               |                |
|-------------------------|-------------|------------------------------------------|----------------|---------------|----------------|
|                         | Weights     | May 07                                   | June 07        | July 07       | August 07      |
| 341                     | 682         | 0                                        | 0              | 45            | 45             |
| 342                     | 1324        | 8                                        | 0              | 3             | 9              |
| 343                     | 1240        | -3                                       | 15             | 26            | 17             |
| 351                     | 2106        | 8                                        | 12             | 16            | 6              |
| 353                     | 442         | 0                                        | 7              | 0             | 1              |
| <b>Aggregate 3435</b>   | <b>5794</b> | <b>4,0939</b>                            | <b>8,10597</b> | <b>17,362</b> | <b>13,2489</b> |

The 3<sup>rd</sup> table is the calculated transition-matrix, based on turnover. The outer weights are divided over the new groups, based on the division of turnover. The six new Nace-2 groups get weights to calculate the “new” 3435 aggregate.



Table 3: Calculated transition-matrix, weights divided over the Nace-2 groups, based on turnover

| Nace-1 1993                 | Nace-2 2008 |      |      |      |     |     | total       |
|-----------------------------|-------------|------|------|------|-----|-----|-------------|
|                             | 291         | 292  | 293  | 301  | 303 | 331 | Nace-1 1993 |
| 341                         | 682         | 0    | 0    | 0    | 0   | 0   | 682         |
| 342                         | 0           | 1324 | 0    | 0    | 0   | 0   | 1324        |
| 343                         | 0           | 0    | 1240 | 0    | 0   | 0   | 1240        |
| 351                         | 0           | 0    | 0    | 1635 | 0   | 471 | 2106        |
| 353                         | 0           | 0    | 0    | 0    | 207 | 235 | 442         |
| <b>Total weights Nace-2</b> | 682         | 1324 | 1240 | 1635 | 207 | 706 | 5794        |

The 4<sup>th</sup> table is the transition-matrix of the weighted balances. For the Nace-2 groups 301 and 331 the original answers of the units in the Nace-1 351 are used to recalculate new balances. The old weighted balances of the 351 was 8, in the new classification the units which move to the 301 produce a weighted balance of -4, and the units which moves to the 331 produce a balance of 34. This example clearly shows there's no evenly distribution of the balances within components of one Nace-1 group.

Table 4: transition-matrix of the weighted balances

Balance of positive and negative answers, recalculated from the historical microdata

| Nace-1 1993 | May-07 | Nace-2 2008 |     |     |     |     |     | Aggregate | Original |
|-------------|--------|-------------|-----|-----|-----|-----|-----|-----------|----------|
|             |        | 291         | 292 | 293 | 301 | 303 | 331 |           |          |
| 341         |        | 0           |     |     |     |     |     | 0,0       | 0        |
| 342         |        |             | 8   |     |     |     |     | 8,0       | 8        |
| 343         |        |             |     | -3  |     |     |     | -3,0      | -3       |
| 351         |        |             |     |     | -4  |     | 34  | 4,5       | 8        |
| 353         |        |             |     |     |     | 0   | 0   | 0,0       | 0        |

\*1: The newly calculated Nace-1 aggregate for group 351 differs from the historical value because of the division of the weights based on turnover. The individually weights of the underlying units split in a different proportion. For the exact calculations, see appendix 2.

#### 4.2 Method I: Macro method, computing weighted balances based on the transition-matrix (table 3).

The first method considered, is a macro method, which uses a transition-matrix to compute new balances out of the historic series. There is one assumption, there is supposed to be equal confidence within each cell of a row. In other words, each part (or component) of a Nace-1 group which split in more Nace-2 groups, has an equal balance.

The first step is to create a matrix with the historical balances under Nace-1, assigned to the Nace-2 groups (table 5).

Table 5: create a matrix with the historic balances.

| Nace-1 1993 | Nace-2 2008 |     |     |     |     |     | total       |
|-------------|-------------|-----|-----|-----|-----|-----|-------------|
|             | 291         | 292 | 293 | 301 | 303 | 331 | Nace-1 1993 |
| 341         | 0           | 0   | 0   | 0   | 0   | 0   | 0           |
| 342         | 8           | 8   | 8   | 8   | 8   | 8   | 8           |
| 343         | -3          | -3  | -3  | -3  | -3  | -3  | -3          |
| 351         | 8           | 8   | 8   | 8   | 8   | 8   | 8           |
| 353         | 0           | 0   | 0   | 0   | 0   | 0   | 0           |

Next, the transition-matrix (table 3) and the matrix with historical balances (table 5) are multiplied. The historical balances are weighted and divided over the Nace-2 groups. This new balances can be weighted with the new weights of the Nace-2 groups (last row table 3) and the "new" aggregate 3435 under Nace-2 is calculated. This is (as expected) the same aggregate calculated under Nace-1. With the same method, balances for June, July and August are computed and the aggregate 3435 (table 6).

Table 6: calculate weighted balances.

| May 07            |  | Nace-2 2008 |     |      |      |     |      |        |
|-------------------|--|-------------|-----|------|------|-----|------|--------|
| Nace-1 1993       |  | 291         | 292 | 293  | 301  | 303 | 331  |        |
| 341               |  | 0           | 0   | 0    | 0    | 0   | 0    |        |
| 342               |  | 0           | 8   | 0    | 0    | 0   | 0    |        |
| 343               |  | 0           | 0   | -3   | 0    | 0   | 0    |        |
| 351               |  | 0           | 0   | 0    | 8    | 0   | 8    |        |
| 353               |  | 0           | 0   | 0    | 0    | 0   | 0    |        |
| Weighted balances |  | 0           | 8   | -3   | 8    | 0   | 5,3  | 4,0939 |
| June 07           |  | 0,0         | 0,0 | 15,0 | 12,0 | 7,0 | 10,3 | 8,1    |
| July 07           |  | 45,0        | 3,0 | 26,0 | 16,0 | 0,0 | 10,7 | 17,4   |
| August 07         |  | 45,0        | 9,0 | 17,0 | 6,0  | 1,0 | 4,3  | 13,2   |

Example Nace-2 group 331 = (weight351\*balance351 + weight 353 \* balance353) / weight (351+353) = (471\* 8 + 235\* 0) / 706 = 5,3.

#### 4.3 Method II: Macro/micro method, convert balances of old series using a conversion-matrix of balances (based on micro data).

The second method consists of a combination of a macro and micro approach. The old series of weighted balances are converted to the Nace-2 groups using a conversion-matrix which is a combination of the transition-matrix (table 3) and the transition-matrix of weighted balances (table 4). There is the assumption of a fixed ratio between the balance in a cell (Nace-2) and the total of the line. In other words, where the Nace-1 group splits in components for more groups, each component gets its own balance. The ratio between these balances is used in a conversion matrix which is used to convert the old series and is presumed fixed in time.

The first step is the calculation of the fixed ratios (table 7). The transition-matrix of weighted balances (table 4) is used to compute this. This matrix is based on the provisional conversion matrix based on turnover (table 1). Therefore, the “new” calculated balance for the Nace-1 group has to be used.

Table 7: Calculation of the ratio cell / line, out of table 4.

| May 07      |  | Nace-2 2008 |     |     |         |     |        |       |
|-------------|--|-------------|-----|-----|---------|-----|--------|-------|
| Nace-1 1993 |  | 291         | 292 | 293 | 301     | 303 | 331    | Total |
| 341         |  | 1           | 0   | 0   | 0       | 0   | 0      | 1     |
| 342         |  | 0           | 1   | 0   | 0       | 0   | 0      | 1     |
| 343         |  | 0           | 0   | 1   | 0       | 0   | 0      | 1     |
| 351         |  | 0           | 0   | 0   | -0,8881 | 0   | 7,5484 | 1     |
| 353         |  | 0           | 0   | 0   | 0       | 1   | 1      | 1     |

Example Nace-1 group 351: The two computed new balances for the Nace-2 groups 301 and 331 are divided by the aggregate 4.5. (-4/4.5 and 34/4.5).

Remark: When there are zeros in the balances, this step won't work properly. Probably, this potential problem could be solved by for example raising the balances with 100. We're still thinking about a proper solution.

The next step is to create a matrix with the historical balances under Nace-1, assigned to the Nace-2 groups (table 8), of the next (or previous) period.

Table 8: create a matrix with historic balance of the next period (June 2007).

| June 07     | Nace-2 2008 |     |     |     |     |     | Total |
|-------------|-------------|-----|-----|-----|-----|-----|-------|
| Nace-1 1993 | 291         | 292 | 293 | 301 | 303 | 331 |       |
| 341         | 0           | 0   | 0   | 0   | 0   | 0   | 0     |
| 342         | 0           | 0   | 0   | 0   | 0   | 0   | 0     |
| 343         | 15          | 15  | 15  | 15  | 15  | 15  | 15    |
| 351         | 12          | 12  | 12  | 12  | 12  | 12  | 12    |
| 353         | 7           | 7   | 7   | 7   | 7   | 7   | 7     |

Next, the ratio matrix (table 7) and the matrix with historical balances (table 8) are multiplied. The historical balances are divided over the Nace-2 groups, based on the ratio of dividing.

Table 9: conversion of the historic balances by using ratios.

| June 07     | Nace-2 2008 |     |     |         |     |       | Total |
|-------------|-------------|-----|-----|---------|-----|-------|-------|
| Nace-1 1993 | 291         | 292 | 293 | 301     | 303 | 331   |       |
| 341         | 0           | 0   | 0   | 0       | 0   | 0     |       |
| 342         | 0           | 0   | 0   | 0       | 0   | 0     |       |
| 343         | 0           | 0   | 15  | 0       | 0   | 0     |       |
| 351         | 0           | 0   | 0   | -10,657 | 0   | 90,58 |       |
| 353         | 0           | 0   | 0   | 0       | 7   | 7     |       |

The final step is the calculation of weighted balances and the aggregate for the Nace-2 groups (table 10). These new balances can be weighted with the new weights of the Nace-2 groups (last line table 3) and the “new” aggregate 3435 under Nace-2 is calculated. This is (as expected) the same aggregate calculated under Nace-1 and with the first method, as in table 6.

Table 10: calculate weighted balances.

| June 07                  | Nace-2 2008 |     |     |        |     |        | Total |
|--------------------------|-------------|-----|-----|--------|-----|--------|-------|
| Nace-1 1993              | 291         | 292 | 293 | 301    | 303 | 331    |       |
| 341                      | 0           | 0   | 0   | 0      | 0   | 0      |       |
| 342                      | 0           | 0   | 0   | 0      | 0   | 0      |       |
| 343                      | 0           | 0   | 15  | 0      | 0   | 0      |       |
| 351                      | 0           | 0   | 0   | -10,66 | 0   | 90,581 |       |
| 353                      | 0           | 0   | 0   | 0      | 7   | 7      |       |
| <b>Weighted balances</b> | 0           | 0   | 15  | -10,66 | 7   | 62,787 | 8,106 |

Example Nace-2 group 331 = (weight351\*balance351/ weight (351+353)) + (weight 353 \* balance353 / weight(351+353))= (471\* 90,58/706) +( 235\*7 / 706) = 62,79.

#### 4.4 First conclusions about the methods and some additional remarks.

Both methods are working technically as expected. When comparing the two methods, the different new weighted balances for Nace-2 groups 301 and 331 are explainable by the different assumptions for the methods. The transition-matrix for weighted balances (table 4) shows clearly that for the group 351 the assumption of supposedly equal confidence within each cell of a row will not hold. To underpin this conclusion it's desirable to perform this analysis for every division (2 digits) and groups (3 digits) that split into more Nace-2 divisions/groups, preferably for more than one period. If the majority of these divisions and groups do have a different weighted balance for the different components, then method I is not suitable for backcasting.

Furthermore, the assumption of a fixed ratio between the different cells in a row will probably not hold as well. The dividing of the confidence on one selected point in time is not satisfactory. However, it turned out that for every other period we chose to make table 4, the ratios differ significantly.

*Table 11: Comparison of the results for June 2007.*

|                 | <b>Nace-2 2008</b> |            |            |            |            |            | <b>Total</b> |
|-----------------|--------------------|------------|------------|------------|------------|------------|--------------|
|                 | <b>291</b>         | <b>292</b> | <b>293</b> | <b>301</b> | <b>303</b> | <b>331</b> |              |
| <b>Method 1</b> | 0,00               | 0,00       | 15,00      | 12,00      | 7,00       | 10,34      | <b>8,11</b>  |
| <b>Method 2</b> | 0,00               | 0,00       | 15,00      | -10,66     | 7,00       | 62,79      | <b>8,11</b>  |

Some final remarks:

- The use of the conversion matrix based on turnover (table 1, in example 2007) results in distortion of the analyses. The weighting scheme in Nace-1 groups shows, when split into two groups, another proportion than the dividing based on turnover figures from a completely different year. In future analysis with these methods, it will be best to create a conversion matrix which takes the weighting scheme into account.
- When using the historic weighted balances, be aware that zeros in the matrix will create some problems. Probably, this potential problem could be solved by for example raising the balances with 100. We're still thinking about a proper solution.

## 5. The “pure” micro method: example 2: Manufacturing Industry

The macro methods tested in the previous chapter both have their limitations. The fact that both the assumptions made in the methods probably won't hold in practice is a serious reason to consider another approach. Therefore a “pure” micro method is tested. The basic idea is to recode every unit and use the original data to repeat the production for the new Nace-2 classification, for each period available.

The methodology in brief:

- For all units in the panel the “unique weights” are calculated.
- All units in the panel are recoded to a Nace-2 classification.
- The preliminary publication / delivery scheme under Nace-2 is constructed, consisting of aggregates, divisions (2-dig) and groups (3-dig).
- Every unit is assigned to a group in Nace-2.
- The historical answers of each unit and the unique weights are used to calculate new balances for groups, divisions and aggregates. In other words, the production is done under the new conditions.
- Analyses for the transition schemes Nace-1 → Nace-2 makes it possible to “predict” effects on aggregates and divisions. For the total aggregate for the Manufacturing Industry, for example, there will be no/very limited effect. Most of the “shifts” are between (some) divisions and groups.
- The macro methods can in this case be used as a tool to create “expected” balances under Nace-2 and use these as a check on the produced balances.

### 5.1 Preliminary work and assumptions: unique weights, recoding and non-response

The previous examples did not take into account the special system of weights in the business surveys. The weighting within groups (inner weight) and weighting between groups (outer weight) are hard to catch in a transition-matrix. Within the pure micro method, for every unit its own unique weight in the Nace-1 aggregates can be computed, by multiplying the inner weight (of the unit) and the outer weight (of the original group the unit belongs to). After the recoding, these unique weights can be used to compute the balances under Nace-2.

The second condition for this method is the availability of the completely recoded sample, with even the expired units (i.e. when a company is bankrupt it will be replaced in the sample). The further the backcast period will be extended, the more expired units will appear. To use the same method for older periods these units must be recoded in Nace-2 as well.

For this example (period April 2007), all the records are recoded in a straightforward manner. When a Nace-1 group splits into more Nace-2 groups, the units are recoded evenly to the new groups. This method is not sophisticated enough for the real backcasting, but for this example it will do.

The third assumption that is made, concerns the treatment of non-response. Within the business surveys, the non-responding units are filled with the average of the group they're part of. When the sample is transferred into Nace-2, there are two options. The first option is to assign the original used average to the non responding units. This is the balance they got in the historical series, and therefore what is used to compute the old aggregates. The second option is to compute the new averages for the Nace-2 group and assign this average to the non-responder. When this option is chosen, it is possible that the computed aggregates will differ significantly of the old series. Therefore, in this example the first option is chosen, the use of the original assigned values for non-response.

## 5.2 Calculations and summary of the results

For the period April 2007 the answers for the question "how the production has evolved last month" per unit are collected. After assigning unique weights and recoding the units into Nace-2 the calculations have been performed. In this example the aggregates are calculated for the total Industry, six publication sub-aggregates and the divisions (2 digits). For all calculations, see the appendix (the Excel sheet).

As a first step, for the Nace-1 aggregates and divisions balances are recalculated, just as an extra check that calculating with the unique weights per unit produces the same results as the "normally" calculation in the production process. As expected all the results are exactly the same.

Next, the Nace-2 aggregates are computed by the following method:

Figure 5.1: Method of calculation balances Nace-2

| Unit  | Unique weight | Answer possibility: |       |      |                         |
|-------|---------------|---------------------|-------|------|-------------------------|
|       |               | +                   | =     | -    |                         |
| a     | 100           | 1                   | 0     | 0    |                         |
| b     | 150           | 1                   | 0     | 0    |                         |
| c     | 200           | 0                   | 1     | 0    |                         |
| d     | 100           | 0                   | 0,5   | 0,5  |                         |
| Total | 550           | 250                 | 250   | 50   | <b>Weighted balance</b> |
|       |               | 45,45               | 45,45 | 9,09 | <b>36,36</b>            |

*Step 1: Sum of product weight \* answer possibility.*

*Step 2: Divide the sum by the total weight, multiplies with 100*

*Step 3: balance + and -: weighted balance of the Nace-2 group/division/aggregate.*

The results of the calculation for the six chosen aggregates are shown in figure 5.2. There are only minor differences, mainly due to rounding, three units which are classified outside the Manufacturing, and some wrongly classified units under Nace-1, which are now moved to the correct group. Bear in mind that the official publication aggregates under Nace-2 are unknown for the moment, but these are most probably the publication sub-aggregates which Statistic Netherlands will use.

Figure 5.2: comparison of aggregates under Nace-1 and Nace-2.

| Publication aggregates                  | Nace-1AGG | Balances (+/ =/ -/ B) | Nace-2AGG   | Balances (+/ =/ -/ B) |
|-----------------------------------------|-----------|-----------------------|-------------|-----------------------|
| Food, beverages and tobacco             | 15-16     | 23/69/8/15            | 10,11,12    | 23/69/8/15            |
| Textiles, clothing and leather          | 17-19     | 20/80/1/19            | 13-14-15    | 20/79/1/19            |
| Wood, furniture, building materials     | 20-26-36  | 28/71/2/26            | 16-23-31    | 28/70/2/26            |
| Paper products, printing and publishing | 21-22     | 51/35/15/36           | 17-18       | 51/35/15/36           |
| Petroleum, chemical, plastic products   | 23-24-25  | 28/63/9/18            | 19-20-21-22 | 27/63/9/18            |
| Metal and electro technical products    | 27/35     | 26/70/4/23            | 24/30-33    | 27/70/3/23            |

The results of the calculation for the divisions (2 digits) are shown in figure 5.3. Besides the new balances, the transition from Nace-1 to Nace-2 is visualised by colours. Comparison of the results is more difficult because of the dividing of divisions and the building of several groups into Nace-2 divisions. For the divisions of Nace-1 and Nace-2 which have a one on one relation (green, see 2<sup>nd</sup> column Nace-1), the results must be equal and this is the case (besides the minor differences explained in the previous paragraph). In case of Nace-1 divisions which split in more divisions, the calculations shows again that the assumption of equal balances for each of the components (macro method I) will not hold. See for example the Nace-1 division 24, which split into the Nace-2 division 20 and 21. Both components differ in balances, after the recalculation.

Figure 5.3: Summary divisions ( 2 digits)

| Nace-1 | (+ = - S)   | nace-2 | (+ = - S)   | Nace-1         | Nace-2                         |
|--------|-------------|--------|-------------|----------------|--------------------------------|
| 1516*  | 23/69/8/15  | 10     | 25/67/8/17  | Divides 1-on-3 | From Nace-1 1516               |
|        |             | 11     | 33/67/0/33  |                | From Nace-1 1516               |
|        |             | 12     | 3/81/15/-13 |                | From Nace-1 1516               |
| 17     | 20/80/0/20  | 13     | 19/80/0/19  | Divides 1-on-2 | From Nace-1 17                 |
| 18     | 7/92/1/6    | 14     | 13/86/1/13  | 1 on 1         | From Nace-1 17 and 18          |
| 19     | 44/55/1/43  | 15     | 44/55/1/43  | 1 on 1         | From Nace-1 19                 |
| 20     | 49/47/4/46  | 16     | 49/47/4/46  | 1 on 1         | From Nace-1 20                 |
| 21     | 41/52/8/33  | 17     | 40/52/8/32  | 1 on 1         | From Nace-1 21                 |
| 22     | 54/30/17/37 | 18     | 54/30/17/37 | 1 on 1         | From Nace-1 22                 |
| 23     | 9/91/0/9    | 19     | 9/91/0/9    | 1 on 1         | From Nace-1 23                 |
| 24     | 30/60/10/20 | 20     | 37/51/13/24 | Divides 1 on 2 | From Nace-1 24                 |
|        |             | 21     | 0/99/1/0    |                | From Nace-1 24                 |
| 25     | 30/60/10/19 | 22     | 30/60/10/19 | 1 on 1         | From Nace-1 25                 |
| 26     | 21/77/2/18  | 23     | 20/77/2/18  | 1 on 1         | From Nace-1 26                 |
| 27     | 24/74/2/22  | 24     | 25/74/2/23  | 1 on 1         | From Nace-1 27                 |
| 28     | 27/68/5/22  | 25     | 26/68/6/21  | Divides 1 on 2 | From Nace-1 28                 |
| 3132** | 15/80/6/9   | 26     | 15/82/3/12  | Divides 1 on 3 | From Nace-1 30, 3132 and 33    |
| 30     | 8/91/1/7    | 27     | 25/67/7/18  | Divides 1 on 2 | From Nace-1 3132               |
| 33     | 25/75/24    | 28     | 39/58/3/36  | Divides 1 on 2 | From Nace-1 29                 |
| 29     | 43/54/3/40  | 29     | 18/80/2/16  | Divides 1 on 2 | From Nace-1 34                 |
| 34     | 24/73/3/22  | 30     | 7/85/8/0    | Divides 1 on 2 | From Nace-1 35                 |
| 35     | 16/78/6/11  |        |             | Divides 1 on 2 |                                |
| 36     | 28/71/1/26  | 31     | 28/71/1/26  | 1 on 1         | From Nace-1 36                 |
|        |             | 32**   | 53/47/1/52  |                | From Nace-1 33                 |
|        |             | 33***  | 43/57/1/42  |                | New class, from 28, 29, 33, 35 |

1516\*: Combination 15 and 16. 32\*\*: not in sample

3132\*\* Combination 31 and 32 33\*\*\*: New class

The final experiment was an extra test to check to which extent the assumption of a fixed ratio in time between the balance in a cell (Nace-2) and the total of the row (macro method II) holds. Therefore, for four divisions which split the balances, the “new” components are calculated for several periods. At a first glance it appears that there is no fixed ratio between the balances. When fixed ratios are used in the macro method II, the results will probably deviate from those of the micro method.

Figure 5.4: Summary division which split, balances of more periods

| Nace-1    | April    |          | May      |          | June     |          |
|-----------|----------|----------|----------|----------|----------|----------|
|           | Nace-2 A | Nace-2 B | Nace-2 A | Nace-2 B | Nace-2 A | Nace-2 B |
| <b>29</b> | 38       | 51       | 4        | 18       | 19       | 41       |
| <b>30</b> | 10       | 5        | -6       | -4       | 7        | 4        |
| <b>31</b> | 0        | 18       | 5        | 11       | 2        | 0        |
| <b>35</b> | 6        | 26       | -3       | 16       | 5        | 11       |

### 5.3 Conclusions and remarks

It seems possible to produce new weighted balances under the new Nace-2 classification based on historical micro data. For aggregates the results match with the “expected” results. For divisions it creates balances based on historical input from recoded units. These results are contrary to the assumptions for the macro methods and thereby confirm the conclusion that the assumptions related to the macro methods do not hold.

Some remarks:

- When a macro method is used to check the computed balances, or to create “expected” balances, the results will differ, when divisions or groups split in the new classification. In example: the Nace-1 division 1516 would by using a macro method for the Nace-2 divisions 10, 11, 12 produce the same balances. The micro approach show different balances.
- When using the micro method, it is possible that the new groups or divisions don’t have enough units in it to be representative. The old sample is designed with sample-fractions in order to get a representative and publishable sample. When the units are allocated according to the Nace-2 classification, the new fractions have to be calculated and analysed, to make sure the impact on representative is limited.
- The old data must be available, in order to produce the balances according the new classification for a substantial number of periods. Furthermore the expired units have to be recoded. The Statistic Netherlands is researching the possibility to reproduce the balances in Nace-2 for 84 periods (January 2001- December 2007).



## 6. Conclusions

The three proposed methods seem to be working as expected for the simplified examples. The differences between the resulting backcasted balances are explainable. Against the background of methodologically most preferable and practically best workable, the three methods described each have their strength and weaknesses.

The first macro-method is relatively simple and easy to perform. Its major drawback is the fact that it doesn't consider possible different developments within components of groups that divide under the Nace-2 classification. The choice of turn-over as an indicator to use for the creation of transition matrices (chapter 4, table 1 and 3) gives new balances which are less comparable with Nace-1. This can be solved to use old inner-weights for this step. This method can be applied on the historical series and without the availability of micro-data.

The second macro-method, combined with some micro data, has an extra step in processing. For one period double coded micro data is required. This micro data is used to find the different balances for components of groups which divide under the Nace-2 classification. The assumption that this ratio is fixed over time is of course questionable, but on the other hand this method can be applied on the historical series, with only the need for one period of double coded micro-data. The earlier mentioned issue of the use of turn-over in the transition matrices applies for this method as well.

The micro-method produces balances for Nace-2 aggregates which are the result of reproducing the production process with recoded historical data. This requires some assumption and conditions like the availability of the micro data, the possibility to recode every statistical unit (including expired units) into Nace-2 and the treatment of non-response. When these conditions are met, the repetition of the production process for a number of periods is rather easy for the institutes involved. One possible disadvantage is the representativeness of the newly calculated balances, for classes and groups. The original sample designs were created around sample fractions for the old classes and groups. The reclassification of the units can lead to different and insufficient sample fractions.

Which method or combination of methods is best suitable for the Business surveys, is hard to tell in general. All methods have their advantages and disadvantages. Perhaps is a combination of micro and macro methods the best working in practise: for two or three years reproducing balances for the Nace-2 classification and use these results to create benchmark/transition matrices. The older series can then be recalculated. All depends on the feasibility of these methods within the partner institutes.

So far, this analysis contains only examples for the Manufacturing Industry. Some special attention is needed for the MIG's, since research by Eurostat suggests some difficulty in backcasting these series. Furthermore, this kind of analysis must be repeated for services, retail and construction as well. Especially for services some tough problems are expected. This preliminary report can be classified as work in progress and will be updated after further research.