Conversion of short-term indicators' series on production, turnover and prices for industrial activity according to NACE Rev. 2. at the Statistical Office of the Republic of Slovenia

Iris Rošker and Laura Šuštar Kožuh, Statistical Office of the Republic of Slovenia

Key words: Short-term statistics - NACE Rev. 2 – new activity classification – backcasting series – new base year 2005=100 – weights – indices – industry – industrial production – industrial turnover indices – industrial output price indices - import price indices

The STS regulation (Council Regulation (EC) No 1165/2005) with the accompanying amendments and other documents required that the beginning of transition to the newly implemented NACE Rev. 2 activity classification started with the data for January 2009, simultaneously with the switch to the new base year 2005 (or 2006 for some prices). According to the demands of the regulation, the whole series of data, back to at least 2000 or further if possible, had to be provided. There were several possibilities for backcasting of data, e.g. on micro level, on macro level, or a combination of both. For data on production and turnover the model of macro transference was chosen with the application of the translation tables, in which the shares of 'old activities' (according to NACE Rev.1) that appear in 'new activities' (according to NACE Rev. 2) are stated. For backcasting of the industrial production index (IPI) also new weights had to be prepared, while for turnover and other non-weighted variables there was no need for that.

In the fields of output price indices for the domestic and non-domestic markets and import price indices, the data were recoded to the new classification - NACE Rev. 2 at the lowest level, i.e. the product level. On the basis of the Prodcom correspondence table all data for all months from January 1998 on (for import price indices from January 2006 on) were recoded to new codes from Prodcom 2008, which was adapted to the new NACE Rev. 2.

With recoded data the new indices were calculated according to the new classification. All data were backcast at the same time. Because all data were recoded at the lowest level, backcasting was much easier. We used the same program to calculate indices. We created two databases, one with data according to the old classification and one with data according to the new classification. For output price indices on the domestic market and the non-domestic market and for import price indices the switch to a new base year 2005 was already made in 2007.

The entire process of the new statistical classification of activities of business entities NACE Rev. 2 coming into force and the notifying of users and public in general was centrally led. A special group of experts from different statistical fields was established with the task to introduce the new classification - NACE Rev. 2 in statistical surveys.

1. First step: preparation of weights according to new classification NACE Rev. 2 and the new base year = 2005

Because of the introduction of the new activity classification and the new base year it was neccessary to prepare the new weights, not only for the calculation of the European aggregates (e.g. for Eurostat) but also for our own national calculations of aggregated activity levels of industrial production indices. The main source for the calculation of new weights for the area of industrial short-term indicators were the double-coded micro data from the structural business statistics (SBS). But because of the lack of some (non-domestic) variables within the SBS statistics, we had to find also additional sources for the remaining variables within the final accounts of the enterprises and within the international trade statistics data. Thus, the sources of the newly calculated industrial STS indicators' weights were as follows:

Table 1: The sources of the newly calculated industrial STS indicators' weights

Requested variable	Purpose (weight for)	Dissemina-tion code	Sources
Value added at factor cost	PROD (production)	V12150	SBS
Turnover total	TOVT, TOVV, ORDT, PRON (total turnover non-deflated, turnover deflated or volume of sales, new orders received (total), output prices for industry and services -total)	V12110T	SBS
Turnover domestic	TOVD, ORDD, PRIN (turnover domestic market non-deflated, new orders received – domestic market, output prices - domestic market)	V12110D	final accounts
Turnover non-domestic	TOVE, ORDE, PREN (turnover non-domestic market non-deflated, new orders received non-domestic market, output prices non-domestic market)	V12110E	final accounts
Turnover non-domestic € area	TOVZ, ORDZ, PREZ (turnover non-domestic market non-deflated € area, new orders received non-domestic market € area, output prices non-domestic market € area)	V12110Z	international trade statistics (exports)
Turnover non-domestic non-€ area	TOVX, ORDX, PREX (turnover non-domestic market non-deflated non-€ area, new orders received non-domestic market non-€ area, output prices non-domestic market non-€ area)	V12110X	international trade statistics (exports)
Imports total	IMPR (import prices total)	IMP-T	international trade statistics (imports)
Imports € area	IMPZ (import prices €-area)	IMP-Z	international trade statistics (imports)
Imports non-€ area	IMPX (import prices non-€ area)	IMP-X	international trade statistics (imports)

2. The course of transference of SBS data

Because the data from the above listed sources (for 2005, some of them for 2006) were gathered still according to the old activity classification, it was neccessary that at the first stage they were transferred according to the new demands. That was achieved so that the enterprises were provided with the new classification codes from the already translated statistical business register (SBR). Namely, at the Statistical Office of the Republic of Slovenia (SORS) we decided to translate all of the yearly states of SBR back to 2000. Thus, when we had all the enterprises translated according to the new classification, their accompanying data were aggregated according to the NACE Rev. 2.

For the calculation of weights for the turnover split according to the domestic and the non-domestic markets, the final accounts data of the enterprises were used. By means of the final accounts items 'turnover from the domestic market sales of production and services' (without rentals) and 'turnover from the foreign market sales of production and services' we calculated the shares for each of these two items in the total turnover from final accounts. These shares were finally applied to the already

(namely in the first step) calculated values of the total turnover (source: SBS) in order to obtain the three comparable values:

- total turnover,
- turnover domestic market,
- and turnover non-domestic market,

of which the final two were estimated on the basis of the abovementioned shares from the final accounts items and the value of the SBS total turnover.

A similar method was further applied in order to calculate the euro and non-euro indicators. Here the export data from the international trade statistics were used to divide the shares of turnover among the euro and non-euro zone by having applied the EU-15 list of countries.

Consequently the calculation then demonstrated as follows:
- turnover total = turnover (domestic market) + exports
- exports = exports (€ zone) + exports (non-€ zone)

We used such method with shares so as to gain comparable results from the three different sources (SBS, final accounts and exports statistics), that all have different methodology and thus more or less different results. Otherwise, should we have used the absolute values instead of shares of final accounts and exports statistics, we would never have got in accordance with the SBS results.

3. Conversion of data series according to NACE Rev. 2

The expert group in charge for the coordinated switch of the statistical surveys to the new classification in SORS prepared a presentation for statistitians with the aim to smooth the transference, and thus prepared the plan of actions with the schedule. The group also prepared the list of possible methods that could be used with the guidelines when and where each of these methods should be used.

Methods of recalculation of the statistical series were to be used in these cases:
- translation key – in situation of relation 1:1 or n:1 between the old an new classifications,
- micro level recalculation with the transference of activity on the enterprise (or LKAU) level and the new calculation of indices,
- macro level recalculation with the use of translation tables or matrices,
- combination of micro and macro level methods, e. g. recalculation using micro level method for the recent years and matrices (macro level) for the preceeding years.

Every method has its advantages and disadvantages. While the main advantage of the recalculation of data according to micro method is its accuracy, it has several disadvantages, mainly connected with its exhaustive course from stage to stage:
- firstly the new activity code for all of the enterprises and their LKAU's for several past years has to be defined,
- then all the data processing for all past periods has to be repeated,
- then the data imputation problem arises (preparation of the new imputation model according to the new classification),
- in case of sampling surveys, the design weights have to be re-calculated,
- finally we are confronted with the problem concerning the missing data for certain new activities that were out of scope in case of the previous classification, but they are included in the observation within the new classification.

The advantages of the macro method or the recalculation of the backcast series using the translation matrices are:
- it is not neccesary to provide the translation tables for all of the reference periods (i.e. months), but only for the yearly situations,

3

- recalculation is done on indices and not on the micro data of enterprises,
- and thus the whole process is less extensive and exhaustive.

The disadvantage of this method is that the direct connection between the indices on the aggregated level and the micro data is threatened.

4. Transference of the industrial production and turnover

The macro method was selected for the industrial production and turnover series backcasting. For the periods 2004, 2005, 2006 and 2007 we applied the data on newly recoded industrial groups and new weights to aggregate the results on higher activity levels, while for the years from 2003 back to 1998 we used the translating matrices directly. The linking of both series was done with the use of the chain indices. The reason for such an act was based on the fact that the methodology of the industrial production index calculation was changed in 2004, so that from 2004 on its calculation was based on the value of turnover considering the changes of stocks. In the previous period the index of industrial production was calculated from the volume data and that was the reason why the index of turnover for the period before 2004 was estimated. This estimation was performed by using the data on value of the sold products on the NIP level, which were available in the „old methodology data". Other motives for the macro method pick up are: the easier acquirement of data for the activities that were previously out of scope, great extent of the survey (in terms of time and finance) and demografical changes of the enterprises.

5. Preparations made before backcasting

Some special conditions had to be fulfilled in order to perform backcasting successfully. At first the enterprises in the statistical business register on 31 December 2007 had to have both activities, the old one and the new one, likewise applied also for all the yearly states of the SBS back to 2000. For 2008 the survey sample was chosen according to both activity classifications: according to the old one for the current data publication and according to the new NACE in order to create a data base for the future (2009) indices calculations. Then the translation matrices had to be prepared for 2000-2007 and the new weights for 2004, 2005, 2006 and 2007 on the basis of the SBS data were calculated. Using the administrative sources (VAT data base) and some series from other surveys (turnover indices in retail trade and services) we also calculated the series for the activities that were previously out of scope.

Creation of the translation matrices
On SORS the decision to prepare the translation matrices (for 2000-2007) was made for the variables:
- according to the enterprise's *turnover,*
- according to the *no. of persons working in the enterprises* (also available on the level of LKAU),
- according to the *no. of enterprises* within the activity, (also available on the level of LKAU).

For the industrial activity the translation matrices according to the no. of employees were most suitable. Namely, in Slovenia, due to its small economy, besides the enterprises also the KAU's are observed in short-term industrial statistics, and for them we do not have the information on turnover available from final accounts. Thus, the translation tables (matrices) were made for all the years (starting from 2007 until 2000). In these tables the shares of old activities that occurred also in some new activities are stated.

Table 2: Part of the translation matrices

Year	NACE Rev. 2	NACE Rev. 1	Share	Source of indices
2008	05	10	1	1
2008	08	14	1	1
2008	09	11	0,945205	1
2008	09	14	0,054795	1
2008	10	15	0,999935	1
2008	10	51.310	0,000065	2
2008	11	15.9	0,961229	1
2008	11	01.131	0,017494	2
2008	11	51.340	0,021277	2
2008	13	17	0,994721	1
2008	13	36.6	0,001262	1
2008	13	52.740	0,004017	2
2008	14	17	0,121619	1
2008	14	18	0,878381	1
2008	15	18	0,004538	1
2008	15	19	0,995246	1
2008	15	36	0,000216	1

Sources of indices:
1 – monthly industrial statistical survey
2 – other statistical surveys or VAT data

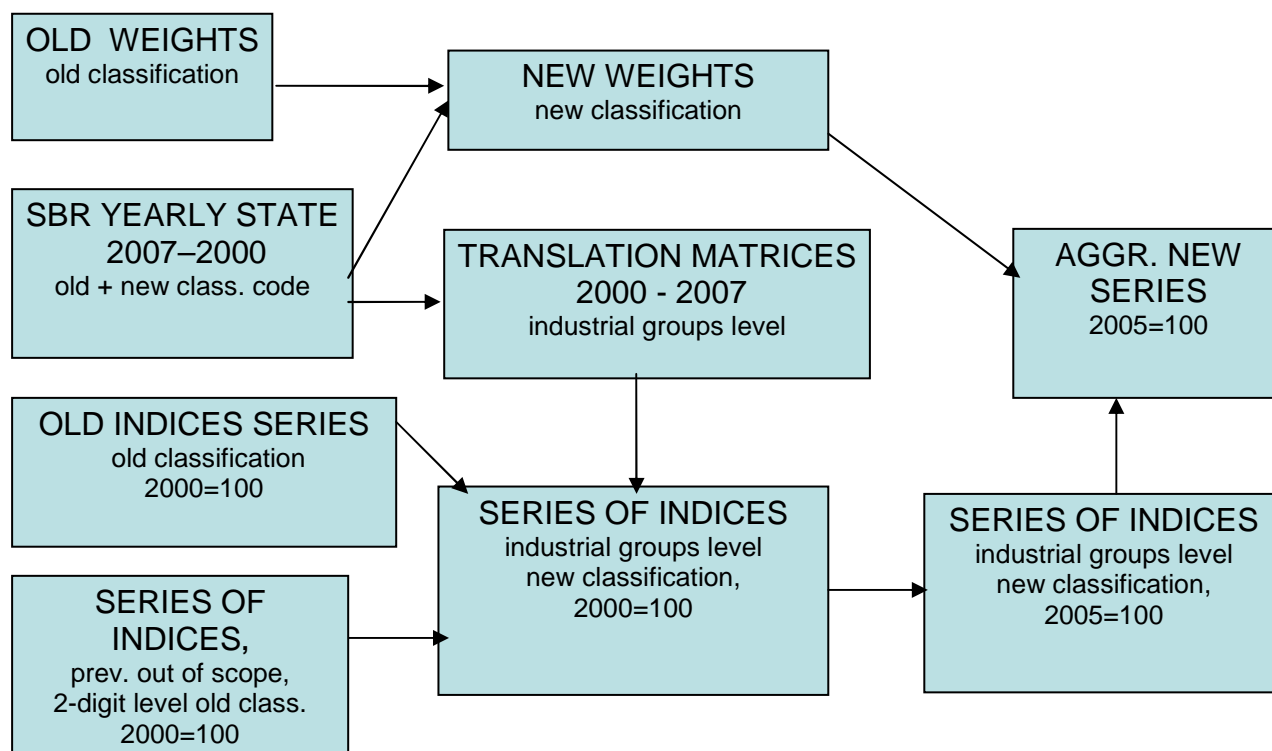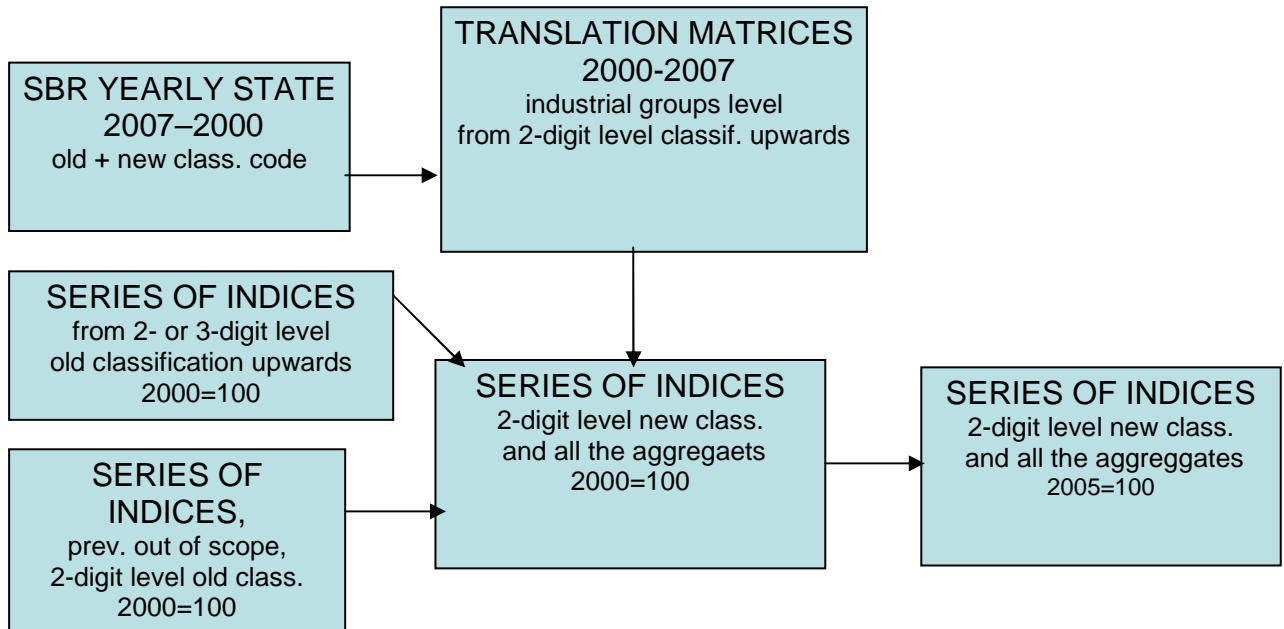## Figure 1: Flowchart of backcasting for the industrial production index

Figure 2: Backcasting for turnover, new orders, stocks (non-weighted series)

```
┌──────────────────────────┐     ┌─────────────────────────────────┐
│   SBR YEARLY STATE       │     │   TRANSLATION MATRICES          │
│      2007–2000           │     │         2000-2007               │
│  old + new class. code   │───▶ │    industrial groups level      │
│                          │     │ from 2-digit level classif.     │
│                          │     │          upwards                │
└──────────────────────────┘     └─────────────────────────────────┘
                                               │
┌──────────────────────────┐                  │
│   SERIES OF INDICES      │                  ▼
│   from 2- or 3-digit     │     ┌─────────────────────┐   ┌─────────────────────┐
│   level old classifi-    │     │  SERIES OF INDICES  │   │  SERIES OF INDICES  │
│   cation upwards         │───▶ │ 2-digit level new   │──▶│ 2-digit level new   │
│       2000=100           │     │ class. and all the  │   │ class. and all the  │
└──────────────────────────┘     │     aggregaets      │   │     aggreggates     │
                                 │      2000=100       │   │      2005=100       │
┌──────────────────────────┐     └─────────────────────┘   └─────────────────────┘
│   SERIES OF              │        ▲
│   INDICES,               │        │
│  prev. out of scope,     │────────┘
│ 2-digit level old class. │
│       2000=100           │
└──────────────────────────┘
```

The main differences between the series broken down by the old classification and that broken down by the new classification:

- *Mining and quarrying* – change of designation from C to B (there are also some differences on the lower level of classification, of minor importance)
- *Manufacturing:*
    - change of designation from D to C,
    - because of the split of publishing activities and recycling of materials out of scope this area of observation is smaller by about 2.6% (in terms of no. of employees),
    - activities that are newly included into observation from the area of agriculture, wholesale trade and services comprise a small share and are scattered over the different groups of activities,
    - a lot of changes are to be found, also rearranged among the variety of groups within the manufacturing activities.
- *Electricity, gas and water supply*:
    - instead of the previous designation E, there is now **D**
    - water collection, treatment and supply is now within the division E, which is not relevant for us since it has always been out of scope,
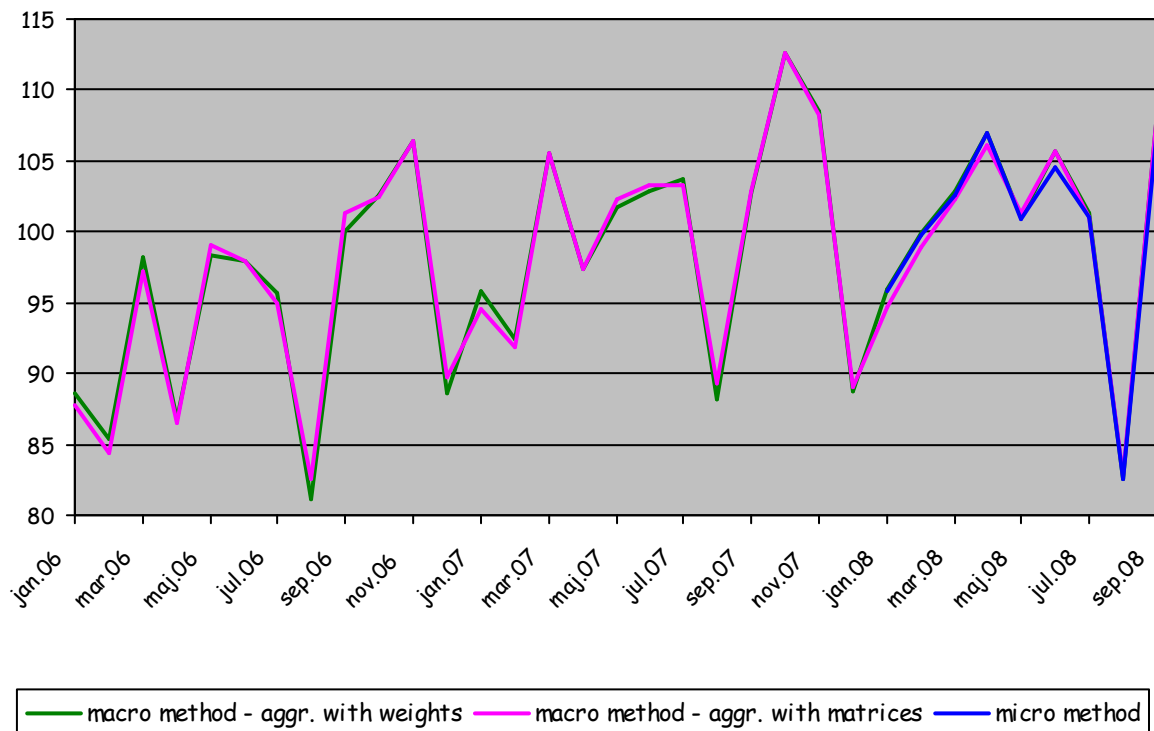   - some additional differences are also on the lower classification levels .

In order to check the correctness of the backast data we carried out a comparison of the data that were derived from the previously described macro method with the data derived from a micro method recalculation. Namely, we also applied that method to gain some aggregates for 2008 when we observed the population according to both classifications. We introduced the so called 'test data base' with the micro data for parallel calculation. Then we calculated the correlation coefficients which were

all between 0,96 and 0,99, and made some visual control of the diagrams, which all verified our hypothesis that the backcast series were reliable.

Graph 1: IIP - Industry total: comparison between the NACE Rev. 1 data and macro calculations – aggregation using the weights, indices, Slovenia, Jan. 2006 – Sept. 2008
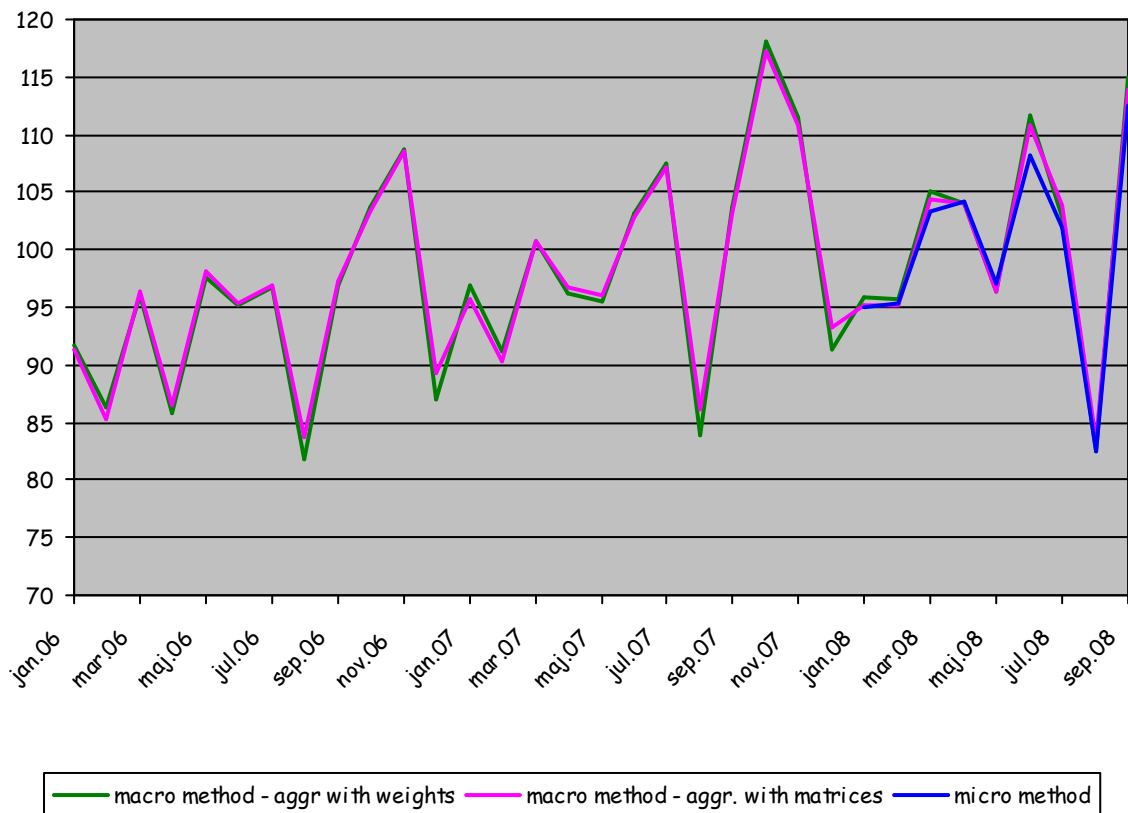**Quality check**
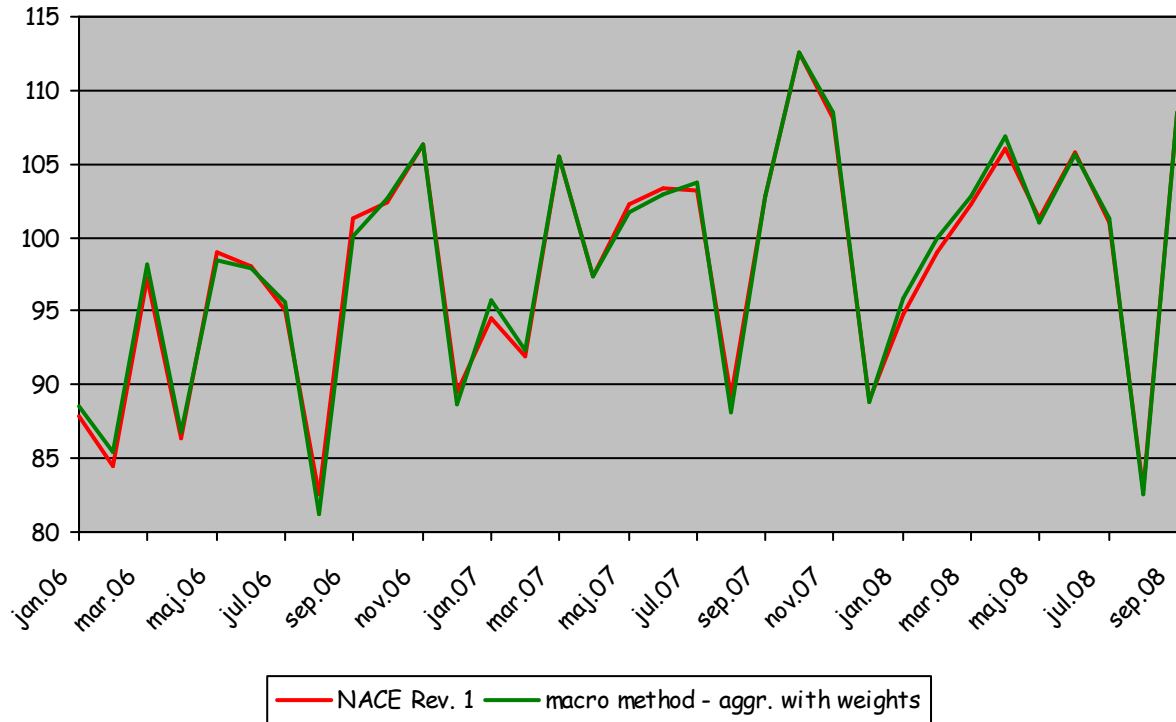
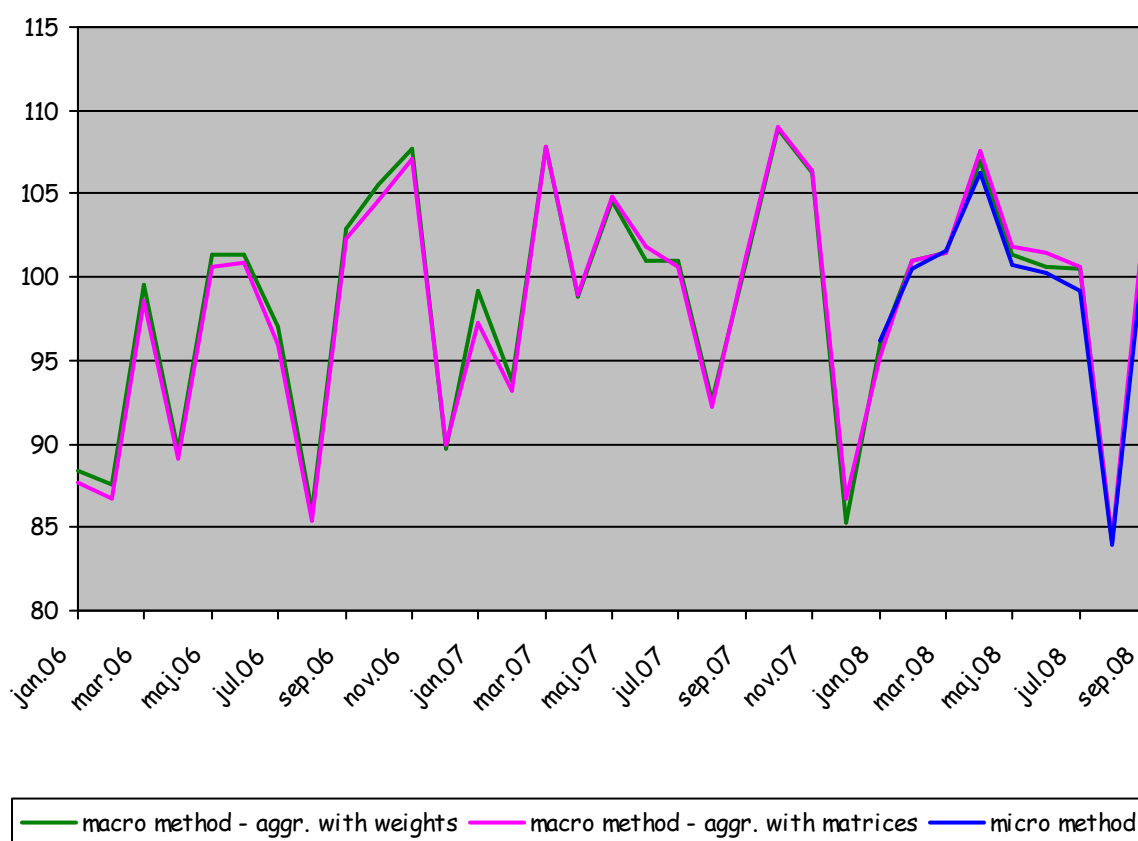**Graph 2: IIP - Industry total: various calculations, indices, Slovenia, Jan. 2006 – Sept. 2008**

Graph 3: IIP - Capital goods industries (B), indices, Slovenia, Jan. 2006 – Sept. 2008

Legend: macro method - aggr. with weights | macro method - aggr. with matrices | micro method

Graph 4: IIP - Consumer goods industries (C), indices, Slovenia, Jan. 2006 – Sept. 2008

Legend: macro method - aggr with weights | macro method - aggr. with matrices | micro method

Graph 5: IIP – Intermediate goods industries (A), indices, Slovenia, Jan. 2006 – Sept. 2008

Legend: macro method - aggr. with weights | macro method - aggr. with matrices | micro method

6. Conversion of the output price indices for the domestic and the non-domestic markets and import price indices

In 2005 we decided to do some revisions on price indices. At that time in Slovenia we only published data about Output price indices for the domestic market. The first step of the revision was to prepare a new computer program for calculating data. We had two choices. One was to prepare a program to calculate the indices in two stages, firstly to calculate the basic indices (current month compared to December of previous year) on the level of class and then all the indices that are published. Eurostat announced that a new classification would be introduced. The best choice was to fill the data base with all basic data. Our data base was filled with all data needed for calculating the indices. We prepared single files, for every month of every year from January 1998 forward.

The single file included data for the selected month about:
- reporting unit (serial number, register number, name),
- product (NIP[1] code, price code[2], specification of product, unit of measure),
- prices (for December of previous year, past month, current month, code for currency and code for country[3], weight).

The preparation of several single files was comprehensive work that took a great deal of time and demanded a lot of precision. Parallel with preparation of single files about price data the expansion of the sample of reporting units was in progress. The sample of units that were reporting prices for the domestic market was enlarged from 309 reporting units up to 345. In 2007 Slovenia has introduced

---

[1] NIP – National Nomenclature of Industrial Products
[2] Price code – internal code, that split diferent products that belong under the same NIP code
[3] Only for Output prices for non-domestic market and for Import prices

the Euro. That consequently meant the obligation to distinct the Output price indices for the domestic market to Euro and nonEuro. The need to enlarge the sample was on the spot. So the sample of reporting units was enlarged from 74 units in 2001 when SORS started to collect export producer prices, up to 220 units. The purpose of all that work was to insure better quality of the calculated data and maximum availability of calculated data. The sample of reporting units for Output prices was enlarged. Within Output price indices the coverage of NACE divisions was up to 76%. Then we decided to change the base year from 2000 to 2005 in January 2006.

In 2006 we started to collect data for Import price indices. Similarly as for output prices for the domestic and the non-domestic markets, the data were loaded in the data base. The volume of the data was the same as it were for Output prices for the domestic and the non-domestic markets.
By the end of 2007 the preparations for conversion of the output price indices for the domestic and the non-domestic markets and import price indices began.

From the beginning we planned to recode all the data on the lowest level, that is micro level. Firstly we prepared instructions for the IT experts. The instructions contained all the notes on how the translation should be done. The first exercise did not give us the best results. The most effective way to insure good quality of data was to translate data on the micro level. Our plan was to recode all the data, on the micro level, with the help of a correspondence table (in the same way that all the data for STS variables were converted). But the results were not as good as we had expected. The next step was preparation of a special conversion table on the basis of the new Prodcom. Data for output and import prices are collected and stored in a data base under NIP codes. The NIP code is av combination of nine numbers, and the first four numbers are codes from the NACE classification. Prodcom codes have only eight numbers, so we prepared a translation table with artificial Prodcom codes. The first eight numbers were from Prodcom and at the end we added a zero, so the codes were the same as NIP codes (they had nine numbers). The instructions for IT experts were prepared one more time with new rules. The translation of the data for output prices and import prices was conducted by the following system: in the first round of translation, the program searched for a match between the old NIP code that was in the data base and the artificial old Prodcom code from the translation table. The matching had to be on the lowest level of all nine numbers. If the matching was not 100%, the data were translated in the next round. Then in every next round of translation the matching was performed on a higher level. Mainly the data were translated on the level of matching of the first six numbers. If matching was only on the first four numbers, the data were individually checked. For energy related products a special translation table was prepared. With a translation table for the translation of old NACE to new NACE a translation table for energy-related products was prepared.

During the translation process some rules had to be considered. The main rule was that during the translation process, only the NIP code could be changed from old to new. All other data had to remain the same. The combination of the most important data about the reporting unit and the information about the product (weight and prices) had to stay unchanged. The combination of important data about the reporting unit and the product was the key for calculating indices. Finally all data for output prices and import prices were translated according to new NACE Rev.2 classification. The next task was to prepare all the required instructions for calculating indices according to the new classification.

When the new program for calculating output price indices and import indices was prepared we took in account that the new classification will step into force. The program was prepared in such a way that it was possible to fill the program with all data translated to the new classification. There was a need to prepare new instructions how to calculate the indices according to the new classification. Some levels were excluded from the amount of data that were included in the calculation of output prices and import prices. New code tables for the calculation and aggregation of indices were prepared. In the second half of 2008 three calculations were in progress. One calculation was for calculating indices according to the old classification. Secondly we calculated translated data according to the old code tables. And finally the third calculation was for calculating the translated data according to the new code tables, which excluded some levels from the amount of data that were included in calculating the output price indices and import price indices. The first and second calculations were one way of checking if all the translated data were included in calculation. The third calculation was for checking if the new code tables were correctly prepared.

Our aim was to insure as much as possible data for our users. From the user side a long seria of data was needed so we insured a series as long as it was possible and in time when the First Release for Output put prices and Import prices was published.

7. Conlusion

With January 2009 we started to publish all STS variables according to the new classification. As Eurostat requested, the series of data was prepared and published in our data base named SI-STAT (http://www.stat.si/pxweb/Dialog/statfile1.asp).

Although the major part of the backcasting process was centrally planned by the experts in charge for the organization of execution, it was sometimes inevitable to improvise or substitute or estimate some data. The realization of such a huge task requires close cooperation of the entire group of co-workers, in fact of nearly the entire statistical office. Therefore we are thankful to all our co-workers, as without their contribution we could not have completed this process efficiently.