EUROPEAN COMMISSION
DIRECTORATE GENERAL
ECONOMIC AND FINANCIAL AFFAIRS
Policy strategy and co-ordination
Forecasts and economic situation

Brussels, 24 June 2011

Validation of NACE Rev.2 back-cast business survey series (EU Joint Harmonised Programme of Business and Consumer Surveys)

This note describes DG ECFIN's validation strategy for Business and Consumer Surveys (BCS) back-cast series at aggregate and Main Industrial Grouping (MIG) level. The aim of the validation strategy is to make available NACE2 series that are both: i) continuous (i.e. without spurious breaks due to nomenclatures' change), and ii) as long as possible.

1. BACKGROUND

Since May 2010, business survey data have been collected according to the NACE Rev. 2 (NACE2) classification and DG ECFIN has published its business survey results according to this new nomenclature. The move to NACE2 entails a change in the identification and grouping of some economic activities in the business surveys. As this change of classification constitutes *per se* a break in the survey time series, partner institutes participating in the Joint Harmonised Programme of Business and Consumer Surveys (BCS) have been invited to provide DG ECFIN with NACE2 back-cast series (back to 2000 for detailed 2-digit codes, and back to 1985 for totals and Main Industrial Groupings-MIGs).

DG ECFIN elaborated a detailed validation strategy related to the NACE2 changeover. This validation process had a twofold objective: to check the quality of the back-cast series, and to make available continuous and long series duly validated.

As a result of the validation process, DG ECFIN makes available a NACE2 validated dataset, which contains series both at aggregate and MIG level.

The validation method used for the aggregate back-cast series is different from that used for the MIG level. In the first case, a model-based methodology has been adopted together with statistical tests of structural break (see section 2.1).This methodology is reasonably robust for series which are not very volatile. Due to the higher volatility of MIG series compared with main aggregate series, its use for the MIG series would have entailed risks of detecting spurious breaks. Therefore, a lighter validation method has been put in place for the MIGs (see section 2.2).

2. THE VALIDATION STRATEGY

The overall purpose of the back-casting exercise is to re-constitute historical series according to the new (NACE2) classification, from the former existing series coded according to the old classification (NACE1.1).

It is worth noticing that, in most cases, the NACE1.1 time series are available with a longer time span than the NACE2 back-cast series received from partner institutes (hereafter called "original NACE2 back-cast series"). For the purpose of the validation, the NACE1.1 series are taken as the benchmark, against which the back-cast series were assessed, validated and then possibly extended further in the past.

In this respect, each series can be considered as consisting of two segments:

- a historical time segment where only NACE1.1 data are available,
- an overlapping time segment where both NACE1.1 and the original NACE2 back-cast data are available (in some cases the overlapping time segment coincides with the whole length of the historical NACE1.1 series).

2.1. Main aggregates (totals) – step by step

For the purpose of the validation, the series (both NACE2 back-cast and NACE1.1) are examined in backward order (e.g. starting from April 2010 and going back in the past).

Step 1

The difference d_t between the original NACE2 back-cast series and the old NACE1.1 series is modelled as a realisation of an autoregressive (AR) process on the overlapping period: $D_t \sim \mathrm{AR}(p)$. The inspection of the empirical autocorrelation functions for a large sample of series confirms that this is a reasonable assumption.

Within the class of AR models, the lag order p of the AR process is identified choosing the model for the series d_t which minimizes the Bayesian Information Criterion (BIC). The identification of p, specific to each series, is done through maximum likelihood estimation.

Step 2

The possible presence of breaks in the difference series d_t is tested through the Bai-Perron procedure[1], which allows identifying the dates on which the breaks occur, too. This tool is widely used in the econometric and financial literature, as it relies on general enough assumptions and yields robust results.

The Bai-Perron procedure makes use of a dynamic programming approach, by means of which structural changes in the mean (occurring at unknown dates) are detected. In a nutshell, the procedure aims at estimating the set of break-dates that split the series into homogeneous intervals, with different means. The estimation method is based on a least

[1] Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, 66, 47–78. Bai J. and Perron P. (2003). Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, 18, 1–22.

squares principle, so that the break-dates are those that minimize the residual sum of squares over all the possible partitions of the series.

One of the main features of the Bai-Perron procedure is the ability to deal with and detect multiple breaks simultaneously. For the purpose of validating the NACE2 back-cast series, the first break (going back to the past) is selected. However, any break occurring after January 2008 is discarded, in order to avoid identifying spurious breaks over the last 3 years, due to the higher volatility of the series during the crisis.

Step 3

The break date, found for each time series, is used to split the original NACE2 back-cast series in two sub-series, before and after the break. The back-cast data for the sub-period from the break date up to April 2010 are therefore considered as validated, whereas those for the sub-period from the beginning of the back-cast series up to the break date are not.

Step 4

The AR(p) model, which has been identified in Step 1, is re-estimated on the validated sub-period. The estimated AR coefficients are then used to convert the NACE1.1 series into a NACE2 series -<u>for the sub-period before the break date</u>.

As $d_t = b_0 + \Sigma_{j=1, \ldots, p} b_j(d_{t-j})$, this is done recursively through the following relationship:

$NACE2_t = NACE1.1_t + b_0 + \Sigma_{j=1, \ldots, p} b_j(d_{t-j})$,

where b_0 and bj (j = 1, …, p) are the estimated AR(p) coefficients and d_{t-j} is the difference between the NACE2 back-cast series and the NACE1.1 series at time (t-j).

The logic behind this step rests in replicating the same autoregressive structure, which has been estimated on the validated back-cast series, in the sub-period before the break date. This is achieved by applying the estimated autoregressive coefficients to the series NACE1.1 that is taken as benchmark through the whole validation process. This allows smoothing the transition between the two series. Furthermore, the adopted approach allows to have NACE2 series as long as the original NACE1.1 series, even when the available back-cast series is shorter, which is very often the case.

2.2. Main Industrial Groupings (MIGs) – step by step

Step 1

In order to assess co-movement between the original NACE2 back-cast series and the old NACE1.1 series, the following correlation coefficients are computed:

- coincident, lagged (t-1) and leading(t+1) correlation over the whole overlapping period,

- coincident correlation over the last 3 years, in order to capture any change in the co-movement between the series over the time.

Step 2

A threshold in the correlation coefficient of 0.4 is used to identify possible different dynamics in the original NACE2 back-cast series and the old NACE1.1 series. More

specifically, if 3 out of 4 correlation coefficients are higher than the threshold, the back-cast series is considered as sufficiently close to the old NACE1.1 series[2]. Then, the back-cast data are considered as validated and susceptible of being further extended in the past, to have the same time coverage provided by the former NACE1.1.

Step 3

The validated NACE2 back-cast series are extended back in the past by applying recursively the month-on-month changes observed on the NACE1.1 series, as:

$$NACE2_t = NACE2_{t+1} + NACE1.1_t - NACE1.1_{t+1}, \qquad \text{for } t \leq T$$

where T is the first month for which NACE2 series need to be extended (no back-cast data having been provided).

The logic behind this step rests in keeping the same month-on-month dynamics in both the NACE2 and the NACE1.1 series, without any level shift.

Finally, the adopted approach allows to have NACE2 series as long as the original NACE1.1 series, even when the available back-cast series is shorter, which is often the case.

[2] If not, the series undergo a deeper analysis and in presence of significant discrepancies (found in around 2% of the analysed series) the corresponding partner institute has been asked to check the data, and then either to correct them or to explain the source of the discrepancy.