WORKING GROUP
LABOUR MARKET STATISTICS

Document for item 2.2 of the agenda

Correcting breaks in time series in Eurostat’s main indicator series

Practical aspects

13–14 June 2018
BECH Building
Room Quetelet
Luxembourg
1. PURPOSE

Due to the changes in the IESS Framework Regulation, most countries expect breaks for a number of data series derived from the LFS. So far, countries and Eurostat have agreed that a strategy for the production and publication of selected break-corrected series is necessary in order to provide stakeholders with the most relevant policy indicators. The minimal input to be delivered by countries to Eurostat at the end of the year in which the IESS enters into force has been defined as at least one correction factor for 14 series (see table 1, line 1 and 2).\(^1\) Alternatively, countries can send more correction factors, or correction factors covering more series, or of course complete data series. *Eurostat encourages all countries which are able to do so to send complete series for all indicators shown in table 1.*

Table 1: Data series to be corrected after IESS break

<table>
<thead>
<tr>
<th>Correction factor/series</th>
<th>Indicator (male, female)</th>
<th>Age breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Required</td>
<td>Employment</td>
<td>15-24,25-64,65+; 20-64</td>
</tr>
<tr>
<td>2 Required</td>
<td>Unemployment</td>
<td>15-24,25-64,65+</td>
</tr>
<tr>
<td>3 Derived</td>
<td>Employment</td>
<td>25-54,55-64</td>
</tr>
<tr>
<td>4 Derived</td>
<td>Unemployment</td>
<td>25-54,55-64,20-64</td>
</tr>
<tr>
<td>5 Derived</td>
<td>Long-term unemployment Unemployment broken down by education</td>
<td>15-24,25-54,55-74,20-64</td>
</tr>
<tr>
<td>6 Derived</td>
<td>Part-time employment Temporary contracts Employment broken down by education</td>
<td>15-24,25-54,55-64,20-64</td>
</tr>
<tr>
<td>7 Derived for use in denominator only</td>
<td>Inactivity (also by education), Employees</td>
<td>15-24,25-54,55-64,65-74,20-64</td>
</tr>
<tr>
<td>8 Derivation not anticipated, only done on request by policy DG</td>
<td>NEET (15-24 only) Supplementary indicators</td>
<td>15-24,25-54,55-74,20-64</td>
</tr>
</tbody>
</table>

\(^1\) In fact, until now Eurostat has referred to a minimum of one, and a maximum of five correction factors per series; however, as this restriction does not cover the possibility of delivering combinations of multiplicative and additive factors, Eurostat has decided to drop the restriction on the upper bound. At the same time, sending more than one factor per series might lead to one of them being dropped if they cannot be applied together in a meaningful manner.
Some concise proposals for the methodologies which are available to countries to produce these corrected series or correction factors have been presented to the LAMAS working group in document Eurostat/F3/LAMAS/58/16 in December 2016. They encompass parallel and pilot surveys, as well as modelling using external information. A combination of both is equally possible and explained in the LAMAS paper. Eurostat naturally has a preference for a large scale independent parallel survey to derive break corrected series; however, financial and logistical constraints will prevent this ideal situation in most, if not all, cases and independence of samples might have to be relaxed. Next to ideal conditions, minimum requirements for both approaches are detailed in table 2. Eurostat refrains from giving more detailed recommendations at this point, as countries are in the best possible position to judge the methodology that can be implemented in their national context. Of course, Eurostat offers countries the possibility to send their proposals to Eurostat for detailed feedback.

Table 2: Summary of possible approaches for derivation of corrected series/correction factors

<table>
<thead>
<tr>
<th>Eurostat preference</th>
<th>Minimum requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full parallel survey for a whole year</td>
<td>Small scale pilot survey for one quarter</td>
</tr>
<tr>
<td>• Independent sample, ideally same size as current and future LFS</td>
<td>• Run in quarter closest to annual average in the past</td>
</tr>
<tr>
<td>• Follow future rotation pattern in introduction of parallel sample (avoid rotation group bias)</td>
<td>• Explain and document in detail underlying assumptions made in the trade-off decision when running the pilot on part of the existing sample, or when deviating in any other aspect from the future LFS</td>
</tr>
<tr>
<td>• Sample selection based on future methodology</td>
<td></td>
</tr>
<tr>
<td>• Same collection modes as future LFS</td>
<td></td>
</tr>
<tr>
<td>• Same proxy rules, imputation standards, weighting procedure</td>
<td></td>
</tr>
<tr>
<td>Modelling in addition to parallel survey</td>
<td>Modelling without parallel survey</td>
</tr>
<tr>
<td>Modelling should preferably be used in addition to a parallel survey (see document Eurostat/F3/LAMAS/58/16 for details), in order to make up for any shortcomings in the parallel survey (i.e. parallel survey covers only one or two quarters, rotation groups have not the same size, etc.)</td>
<td>Method used should make use of all relevant available data</td>
</tr>
<tr>
<td></td>
<td>Exogenous variables should not be dependent on LFS (e.g. do not use national accounts if they use LFS data)</td>
</tr>
<tr>
<td></td>
<td>Ensure that breakdowns by age and sex add up to totals when modelling each series separately</td>
</tr>
</tbody>
</table>
The remainder of this short paper explains what Eurostat will do with the received correction factors and series, which additional series will be derived, and how this data will be published.

Table 3 gives a summary overview of the potential inputs received by Eurostat and how this input is going to be used to correct the 14 series agreed upon. Based on these series, Eurostat has developed two very simple approaches to correct further series needed for different policy purposes (Table 4). A practical example is developed in the excel sheet in the Annex.

2. **EUROSTAT APPROACH TO BACK-ESTIMATION EXPLAINED**

The first, and preferred option is for countries to send to Eurostat complete back-calculated series for the 14 agreed series, additional breakdowns of these series, as well as for all other indicators that Eurostat publishes under the "LFS main indicators" heading (and described in table 1) in the format provided by Eurostat. This way, Eurostat and countries are certain to publish the same information, and countries can use more advanced methodologies in the back-calculation as the simple "one size fits all" approach Eurostat is forced to adopt due to the large number of series that need to be treated. However, please note that also in this case, Eurostat will derive totals via aggregation of the underlying series, annual series from the quarterly series, and seasonally adjust and aggregate the series as this is the general approach for the LFS Main Indicator production. **If countries would like to send seasonally adjusted series, it is possible under the conditions that the indirect approach used by Eurostat is respected, series are sent for all indicators (also those not contained in the break correction exercise) in the format specified by Eurostat, and also sent in all future periods.** Please note that it is perfectly acceptable for countries to initially send correction factors based on first results to avoid publishing data with breaks, and then correct the factors or send complete series later as soon as they become available.

**Table 3: Eurostat approach to calculation of back-series based on delivered inputs**

<table>
<thead>
<tr>
<th>Country sends:</th>
<th>Eurostat derived estimated data $\hat{Y}_n$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full series back to 2010Q1</td>
<td>Eurostat publishes those series as received, performs seasonal adjustment and derives annual series. If only the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.</td>
</tr>
<tr>
<td>One factor</td>
<td>Eurostat derives quarterly series according to type of factor sent: For multiplicative factors, $c = 0$; for additive factors, $\beta = 0$ (subscript &quot;o&quot; refers to &quot;old&quot;, or current, LFS data, &quot;n&quot; to the new data based on the new IESS Framework Regulation). $\hat{Y}_n = \beta \ast Y_o + c$. If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.</td>
</tr>
</tbody>
</table>
Two factors per quarter: one additive, one multiplicative

Eurostat derives series accordingly:

\[ \hat{Y}_n = \beta \ast Y_o + c \]

If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.

Four factors, one per quarter; eight factors, two per quarter

Eurostat derives series for each quarter separately, using the available correction factors:

\[ \hat{Y}_{nq} = \beta_q \ast Y_{oq} + c_q \]

If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.

Two/three quarterly correction factors

Eurostat derives two/one missing quarterly factors as average of delivered factors. If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.

Two/three factors, one/two quarterly, one annual

Eurostat ignores annual factor, derives series according to quarterly factors only. If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below.

| Two factors per quarter: one additive, one multiplicative | Eurostat derives series accordingly: \( \hat{Y}_n = \beta \ast Y_o + c \) If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below. |
| Four factors, one per quarter; eight factors, two per quarter | Eurostat derives series for each quarter separately, using the available correction factors: \( \hat{Y}_{nq} = \beta_q \ast Y_{oq} + c_q \) If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below. |
| Two/three quarterly correction factors | Eurostat derives two/one missing quarterly factors as average of delivered factors. If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below. |
| Two/three factors, one/two quarterly, one annual | Eurostat ignores annual factor, derives series according to quarterly factors only. If only factors for the 14 required series are sent, Eurostat will derive the missing series in table 1, rows 3-7, as described below. |

2.1 Derivation of series for which countries have not sent inputs

In order to derive the missing indicators in the LFS Main Indicators database, several simple approaches will be used. Indicators which can directly be derived from the required inputs delivered by countries are grouped in list 1. Indicators which are derived by comparing simple forecasts of old LFS data to the delivered new LFS data are grouped in list 2. Breakdowns of series that have to fulfil the constraint that all breakdowns have to add up to a known total are those listed in list 3. They are derived by comparing old LFS shares to new LFS shares, correcting for seasonality. Methodologies for list 2 and 3 are explained below.

List 1: inactivity totals and age breakdowns (15-24, 25-64, 65+)

List 2: long-term unemployment totals (15-74) by sex, part-time employment totals (15-64) by sex, temporary contracts totals (15-64) by sex, employees (15-64)

List 3: missing age breakdowns for all series, employment broken down by education, unemployment broken down by education, inactivity broken down by education

2.1.1 Methodology applied to indicators in list 2

Step 1: using data up to the break ("old" data), Eurostat runs a simple OLS regression on the series to be estimated (Y), using the growth rate of either unemployment or employment (X) as regressor. Eurostat saves the coefficient \( \beta \) and the intercept c (for a list of notation, see the annex).
\[ Y_o = \beta \frac{\Delta X_o}{\Delta t} + c \]

Step 2: using a (S)ARIMA model, Eurostat forecasts the quarter b (quarter of the break of series) for Y based on the old data:

\[ \hat{Y}_{ob} = f(Y_o) \]

Step 3: Eurostat derives the difference between the actually collected data for quarter b using the new methodology and the forecast based on the old data for quarter b:

\[ \delta_b = Y_{nb} - \hat{Y}_{ob} \]

To capture the trend of the series, in addition the difference \( \delta_t \) is calculated for all t up to b-1:

\[ \delta_t = Y_{ot} - \hat{Y}_{ob} \]

Step 4: in order to derive the new back series from 2010Q1 to quarter b-1, apply the estimated coefficient \( \beta_o \) and estimated constant \( c_o \) for all t from 2010Q1 up to b-1. To capture the level shift, add \( \delta_b \), and to capture the trend, add \( \delta_t \):

\[ \hat{Y}_n = \hat{\beta}_o \frac{\Delta X_n}{\Delta t} + \hat{c}_o + \delta_b + \delta_t \]

### 2.1.2 Methodology applied to indicators in list 3

The series estimated using methodology 3 are those which have an additional constraint, e.g. age breakdowns which have to add up to the total. Therefore, the logic is slightly different to the methodology for list 2. First, Eurostat estimates the share \( s \) of each series from the total for the old data for quarter b-1, and then the same statistic from the new data for quarter b. Eurostat then calculates the difference \( \delta \) of these two shares, and corrects them for seasonality by subtracting the difference between the two shares of quarters b-5 and b-4.

Following this logic, the newly derived shares will still add up to one, and can now be applied to the total which is already corrected for the break. Step by step, Eurostat does the following:

Step 1: Eurostat calculates the shares \( s_t \) up to \( s_m \) (where m is the number of breakdowns, and \( \sum s_m = 1 \)) for all breakdowns and relevant time periods t (b, b-1, b-4, b-5) and data types (o,n). Specifically, we assume to have information on the breakdowns for \( Y_n \) only for t=b. Thus,

\[ s_m = \frac{Y_m}{Y} \]

\(^2\) Of course, these two steps can be combined into one.

\(^3\) Depending on developments in the economies and labour markets of the countries concerned, averages of the seasonal effect of several years might be used. This can only be decided once the data for 2018, 2019 and 2020 has been collected and processed. Please also note that the validity of all proposed methodologies relies on the assumption that no major event with a large impact on labour market data as e.g. the financial crisis of 2008 occurs between now and the implementation of the IESS FR.
Eurostat then determines the differences $\delta_{mb}$ and $\delta_{mb-4}$ for all $m$ such that:

$$\delta_{mb} = s_{mb} - s_{mob-1} \text{ and }$$
$$\delta_{mb-4} = s_{mob-4} - s_{mob-5}$$

Step 2: Eurostat now calculates the differences of shares, and adds them to the shares known from the current LFS, thus correcting for seasonality; levels can then be derived directly.

$$\Delta \delta_{mb-t} = \delta_{mb} - \delta_{mb-4}$$

$$\hat{Y}_{mnb-t} = (\Delta \delta_{mb-t} + s_{mob-t})\hat{Y}_{nb-t}$$

3. **Publication of Break-Corrected Series**

For the dissemination on the Eurostat website, detailed LFS results will always be disseminated as collected (and flagged the first year of implementation of the IESS Framework Regulation).

For the LFS Main Indicators tables of high policy relevance, there will be a duplication of tables:

- Country level data based on the current LFS methodology will be continued to be published and revised in the current LFS Main Indicators tables. Once IESS entered into force, current tables will only be corrected if revised current LFS data becomes available.
- Copies of these tables will be created and filled with data using the current LFS from 2010 on. With entry into force of the IESS Framework Regulation, and delivery of the first micro-data based on the new regulation, the copied tables will be extended to cover also data collected under IESS. Initially, all copied series will be published with a break flag; as soon as correction factors or back-casted series are available, the back data will be overwritten and the break flag removed. EU and EA aggregates will always be published based on the available data. The break flag in the aggregates will only be removed once all back series are corrected, i.e. at the latest in May of the year following the introduction of the IESS FR.

Please note that implementation of the break-correction for the monthly unemployment series as well as publication of these series is discussed separately, and not covered by this note.

*The LAMAS delegates are invited to:*

- comment on this strategy;
- send their plans for the break correction to Eurostat.
ANNEX:

Y – series to be estimated

X – explanatory series backcasted by country, i.e. unemployment or employment

o – subscript for "old" series, i.e. series before break-correction

b - quarter of break, i.e. first quarter of data delivered under IESS

b+1 – one quarter after break

b-1 – one quarter before break

n – subscript for "new" series, i.e. series after break-correction

β – coefficient estimated using OLS

\( \hat{\beta}_o \) – coefficient estimated on "old" series,

\( \hat{c}_o \) – constant estimated on "old" series

\( \delta_b \) – difference between data based on the new methodology, and forecast based on old methodology for quarter b

\( \delta_t \) – difference between forecast based on old methodology for quarter b and "old" series quarters t

s – share of breakdown in total

1,…,m - total number of breakdowns