



OECD Short-term Economic Statistics Working Party

23-24 June 2008

Synthetic outline paper on the development of flash estimates for certain PEEIs

In this short paper, we provide some information on the ongoing project on “Flash estimation for certain PEEIs” launched in 2007. No final result (e.g. no model and flash estimate) is included in this paper as the simulations and the model selection process are not completed.

Introduction

In order to meet the challenges of the need for improved timeliness of PEEIs, the 2005 Status report on Information Requirements in EMU by the EFC suggested that “more use of flash estimation techniques for European aggregates should be considered”.

In response, in 2006 Eurostat launched a call for proposals to develop methods and tools to produce flash estimates¹ of 3 short-term economic indicators: the monthly Industrial Production Index, the first GDP estimate of the Quarterly National Accounts and the Labour Cost Index, which should meet the following requirements:

- The estimation of the monthly industrial production index for the Euro-zone and the EU-25 within 30 days after the end of the reference month (actual delay, approx. 48 days). Monthly data is as far as possible broken down by main industrial groupings (MIGs);
- The estimation of the quarterly GDP at constant prices for the Euro-zone and the EU-25 within 30 days after the end of the reference quarter (actual delay, approx. 45 days). Whereas the main aim is total GDP (mainly growth rates are envisaged, level only if possible as by-product), quarterly data are as far as possible broken down by groups of industries according to A6 classifications;
- The estimation of the quarterly labour cost index for the Euro-zone and the EU-25 within 45 days after the end of the reference quarter (actual delay, approx. 70-80 days). Whereas the main aim is total Labour cost index, quarterly data is as far as possible broken down by NACE sections.

A proposal by four NSIs has been accepted by Eurostat. These NSIs work in close cooperation to meet the challenge: Institut National de la Statistique et des Études Économiques (INSEE, France, coordinator), Istituto Nazionale di Statistica (ISTAT, Italy), Office for National Statistics (ONS, United Kingdom) and Statistisches Bundesamt (DESTATIS, Germany).

Background

The Joint Report of the Council and the Commission on Eurozone statistics to the March 2003 European Council called for the development of infra-annual Principal European Economic Indicators (PEEIs) needed for economic analysis and monetary policy, to be made available with a timeliness and other important quality features that match the highest international standards. The Joint Report required the PEEIs to be fully implemented by 2005. In this context, the ECOFIN Council of 2 June 2004 invited the Economic and Financial Committee, with the assistance of Eurostat and the European Central Bank, to monitor closely progress on PEEIs including benchmarking against the statistical system of the United States. The 2005 Status Report on

¹ A flash estimate is an early estimate for an economic variable of interest over the most recent reference period and is normally calculated on the basis of a statistical or econometric model. The flash estimate should have a release date appreciably earlier than the first release date of the actual data for that variable.

Information Requirements in EMU by the EFC assesses the progress achieved on the availability and quality of PEEIs and formulates objectives for future work.

Due to the commitments of, and close-cooperation between, Eurostat and national statistical institutes (NSIs), major progress has been achieved in several fields over recent years. Notably, some new European indicators, such as GDP flash estimates, industrial new orders statistics and job vacancy rates, have been released. The timeliness of GDP and its breakdowns, of the turnover index for retail trade and repair and of the labour cost index has improved substantially. Moreover, the Eurozone HICP flash estimate is an effective reference indicator for monetary policy, although further increases in its country coverage are still desirable. As regards legislation, new Regulations were adopted in the areas of European quarterly non-financial sector accounts, quarterly public finance, short-term business statistics and external trade statistics, paving the way for further progress in those areas.

Overall, nine out of a total nineteen PEEIs currently fulfill or are close to the PEEI timeliness and coverage targets, thus improving considerably the information base for economic analysis and monetary policy. This progress notwithstanding, important challenges remain. As for short-term business statistics, the availability of timely and high-coverage indicators for services is currently inadequate and this represents a major weakness given the importance of services in the euro area and the EU. Finally, the timeliness of employment data should be improved as a matter of priority in order to enable an adequate assessment of the labour market situation in Europe. Generally, despite some important improvements in timeliness, most PEEIs still lag far behind the US indicators and major efforts are still required to catch up with US timeliness.

Consumer prices and GDP flash estimates are successful examples of improving timeliness of EU aggregates, combining the use of national flash estimates and other techniques. These are the outcomes of specific projects carried out in the last five years under the auspices and the co-ordination of Eurostat.

Outline Plan and Timetable

The work of the 4 NSIs is undertaken in two phases.

Phase 1

This phase started on January 2007 and ended on September 2007. The main achievements were:

- an inventory of available sources
- setting up a database of historically available indicators;
- Developing the methodology, including:
 - a review of existing methods used in the EU;
 - a variable selection strategy to establish a priori criteria for the suitability of models

In this table, the letters indicate the institutes involved in the sub-task: F (France), DE (Germany), I (Italy) and UK (United-Kingdom)

Task	T+3	T+6	T+9
Coordination	F	F	F
Detailed inventory of available sources of information	DE	DE	
Constitution of 3 (historical) databases (one for each selected variable)		F	F
A review of the literature and experiences on flash-estimation with a selected bibliography		I	I
A technical report presenting the adopted strategies to produce flash estimates		F	F
Intermediate technical reports presenting the methodologies		F,DE,I,UK	F,DE,I,UK
Intermediate technical reports presenting the variable selection methods		F,DE,I,UK	F,DE,I,UK
Intermediate technical reports presenting the quality measures		F,DE,I,UK	F,DE,I,UK

Phase 2

This phase began on completion of Phase 1 and will last 12 months. Therefore it should be completed by August 2008. It will include:

- Simulation studies of the possible models (evaluation of models and proposals for basis of flash estimates);
- Preparation of a toolbox for use by Eurostat
- Preparation of a final report and relevant bibliography.

Task	T+3	T+6	T+9	T+12
Coordination	F	F	F	F
Report on the results of the simulations Intermediate technical reports		F F,DE,I,UK	F F,DE,I,UK	F F,DE,I,UK
A set of documented routines Intermediate technical reports and provisional programs.		F F,DE,I,UK	F F,DE,I,UK	F F,DE,I,UK
Timetable for the production and dissemination of the flash estimates				F
Strategy for the transfer of knowledge				F
Update of the bibliography		I	I	I

Dissemination

One of the objectives of the action is to disseminate knowledge about flash estimation techniques in the European Statistical System. This will be done in 3 different and complementary ways:

- the dissemination of technical reports on the methodologies used in the project;
- the dissemination of generic programs, written in a standard statistical language, permitting the application of the proposed techniques
- the participation of the network members to conferences and working parties.

Some important remarks on the philosophy of the proposal

- Flash estimation is mainly based on statistical and econometric techniques. However, the use of these techniques cannot replace the ongoing works for structural improvements in the processes of data collection and indicator compilation at Member State level;
- The objective is to produce reliable flash estimations; the reports will clearly state on the possibility to do so.
- The proposed flash estimation models should be easy to use and to maintain. This implies that the explanatory variables would have to be easily available (and if possible already in Eurostat databases) and that the programs would have to be developed in standard statistical softwares;
- The proposed flash estimation techniques will be based on existing and observed related indicators. Pure econometric models will only be used as benchmarks in order to evaluate the quality of the proposed methods;
- Even if the project aims at providing flash estimates at the European level, the work will be reported in a way that Member States could easily make use of the proposed methodologies.

What has been achieved until now?

1. *A survey among the ESS to take stock of the various experiences on flash estimation of IPI, LCI and GDP.*

This large scale survey gave us information on the existing published or non-published flash estimates, on the strategies used to build the models and on the coincident or leading variables used. Lots of experiences have

been done on the flash estimation of the GDP, quite a few on IPI and none on the LCI. It has to be noted that Business tendency survey data (soft data) are often used in the various models.

2. *Constitution of a bibliography on flash estimation*

The main point has been the selection of a user-friendly environment permitting to edit, import and export entries. At this stage of the project we propose to use the JabRef program in conjunction with the bibtex standard together. JabRef is freely distributable under the terms of the GNU General Public License, version 2. Jabref has been used to build up a selected bibliography that contains 76 entries most of them (56) associated with the article in pdf file format. The update of this bibliography is still ongoing.

3. *Selection of a few variable selection strategies*

As there are a large number of possible indicators, when lags are also considered, the number of candidate models can be measured in terms of millions. Choosing between them is subject to high search costs, so the modeller cannot be confident that they are selecting the best representation of the true data generating process. The problem is further complicated by a lack of degrees of freedom as the number of indicators approaches or exceeds the number of time observations, and multicollinearity due to a high correlation between rival variables. The approaches considered in the project fall into two broad categories:

- **Data reduction:** This is based on the notion that a given pool of indicators can be adequately represented by a subset of the data. Cluster analysis suggests that indicators can be grouped together according to measures of similarity, and hence smaller numbers of candidate variables can be selected from each cluster. Factor analysis implies that many variables might be driven by a reduced number of common factors or shared trends. These can be typically extracted from the data set using principal components or dynamic factor models.
- **General-to-specific modelling (Gets):** Starts from a general and over-specified statistical model that captures the essential characteristics of the underlying data set, and then uses standard testing procedures to reduce its complexity by eliminating insignificant variables to form a more parsimonious model of the data.

4. *A real-time database*

The constitution of an historical database has been an important step of the project, both for the estimation and the stability analysis of the selected models. For the indicators available in the Euro-Indicators database, Eurostat took advantage of the daily snapshots of the database recorded by the Eurostat team in charge of NewCronos to get precise historical data including revisions of raw and seasonally adjusted data.

5. *First simulations and preliminary results*

The performances of the various strategies are illustrated through the first results obtained for Euro-area IPI. As it is known, comparing to GPD applications, not a lot of attention has been paid to IPI. Exceptions are Bacchini et al. (2005), Bruno and Lupi (2003), Ladiray and OBrien (2003) and Zizza (2002). Our applications extend the precedent contributions along three lines. Firstly, inside the bridge model approach we analyze the effects of a selection variables strategy based on cluster analysis and upon a dynamic factor model. Secondly, we evaluate how the countries preliminary estimation can improve the Flash estimation of the monthly Euro-area IPI. Finally, we evaluate the feasibility of approach consists in combining predictors from a small number of 'smooth factors' which are then used as regressor in a bridge equation (see Altissimo et al. 2007).

Business survey data as well as new orders appear to have a good “predictive power” of the IPI.

Other experiments have been done, using the Gets strategy, for the LCI. For these simulations, a large bunch of variables has been used:

- **Price data:** GDP deflator, Consumption deflator, Import price deflator, HICP
- **Wage data:** Compensation of employees (NA), Wages and salaries (NA), Earnings-manufacturing/industry (MEI database), Earnings- private sector (MEI), Earnings- all activities (MEI)
- **Business and Consumer Survey data:** Prices- last 12 months (Consumer), Prices- next 12 months (Consumer), Unemployment- next 12 months (Consumer), Employment- next 12 months (Services), Productive capacity (Industry), Capacity utilization (Industry), Labor shortages (Industry), Competition- Domestic, Competition- EU, Competition- Rest of the world
- **Other data:** GVA per head, Unemployment