7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

G. Data quality: Quality issues on monthly figures, timeliness and consistency between quarterly and monthly data

G3 – Providing monthly estimates based on quarterly LFS surveys

Miguel Angel García, Manuel Mateo – Spain

Providing monthly estimates based on LFS
A description of the procedures of consistency between
monthly, quarterly and annual figures
Workshop on LFS Methodology. Madrid, May 2012.
Miguel Ángel García Martínez (Instituto Nacional de Estadística.)
Manuel Mateo (Instituto Galego de Estatistica)
SPAIN

LFS is a continuous survey

According to the current requirements (Council Reg. 577/1998), the LFS is a 'continuous survey' that will provide quarterly and annual results. The quarters (years) are defined as 13 (52) consecutive weeks that contain the correspondent calendar period.

Under this definition (52x7 = 364 days a year) after several years, it is needed to introduce an additional week to match the LFS quarters to the calendar quarters.

The concrete weeks for each quarter is also determined by regulation (See Annex II of the Commission Reg.377/2008)

Typically we have sample for every week of the year. It seems natural to see if it is possible to produce 'monthly estimates' based in the LFS taking advantage of the availability of sample for every week.

A first obstacle comes up since the beginning. The link between the main reference period in the LFS (the week) and the month is less steady comparing to the quarters and years. The definition of 'months' using the homogeneous criteria as quarters/years is from weekly units means that we have always months of 4 and 5 weeks every quarter. Concerning the sample size of the months this means a potential variation of 25% of the sample, which is not negligible at all.

What do we call 'monthly estimate'?

The LFS surveys in the EU are designed (generally speaking although there are exceptions) to produce quarterly estimates and monitoring the 'change' between consecutive quarters (see article 3 of the Reg.577/1998 about precision requirements).

A first characteristic of a monthly estimate from a quarterly LFS survey should be the **coherence** with the quarterly (and annual) estimates based in the LFS survey. In other words, how the monthly estimates are related to the quarterly (and annual) estimates of the survey.

Nevertheless, to 'zoom' the 'quarterly approach' to a monthly basis is not straightforward. In the case of Spain calculating directly monthly estimates from the LFS sample produce erratic results due to the reduction of the sample and the inexistence of common sample between consecutive months. I suppose this situation is rather usual among the EU LFS surveys (unless the original design of

the survey focused from the beginning in producing monthly estimates to monitor the monthly variations).

Nowadays in Spain the distribution of the LFS sample attends to the optimization of the fieldwork organization by province (NUTS3 unit) in order to balance the workload according to the available resources, and to the objective of producing a full quarterly micro data dataset. In particular, the overlap between consecutive months never happens. In fact, there are provinces where some sub samples are missing some months.

Faced to this problem we can assume that we had a kind of 'small domains estimation' problem in the temporal dimension; i.e. we have to produce estimates not included as objective in the design of the survey. In this case, the 'small domain' would be the shorter time period (month) than the one for the survey is designed to provide reliable estimates (quarter).

Other key aspect of the 'monthly estimate' is the **timeliness**. It is obvious that the utility if the monthly estimate is related to the timely availability of the data.

In the case of Spain, the quarterly results are available before the end of the following month after the end of the reference period (i.e. April for the 1st quarter, July for the 2nd, October for the 3rd and January of the following year for the 4th). The quarterly estimates are weighted to the population now cast calculated for the central month of the quarter (the first quarter is weighted to mid February population now casts and so on). We produce also 'moving averages of three months' similarly to quarters. The problem is that, assigning as reference period the one of the now cast populations, the estimate would be available two months after the reference (which is too late).

For monthly estimates it is rather 'natural' to state the timeliness deadline within the following month to the reference of the data. In our case, the tests we have carried out have been based in the 'monthly' processing of the sample obtained weekly and providing the monthly data the third week of the following month after the reference period.

And finally, the **definition of the output** for the monthly figures is fundamental. The output could be the reproduction of the complete set of data and indicators available for the quarterly dataset (which would mean a monthly LFS). But the usual situation is to limit to a number of indicators that would be produced in a monthly basis.

Then, the outlining of the output becomes a key issue. Depending on the stability of the output defined, some procedures of estimation can be judged as more suitable. For those estimations based in models, provided that the coherence among the set of indicators must be ensured, the flexibility to modify the output is much more limited.

In the Spanish case, the tests were based in a limited but rather complete set of indicators (see Annex 1) but the selected procedure permit to increase the set of monthly information in a very flexible way. This was considered a crucial advantage. In fact, if the 'product' is successful, further desegregations of the 'monthly indicators set' are predictable (regions, main economic sectors). A EU

harmonised monthly output could help to determined a fixed set of indicators. A starting point to define the list of indicators is the current harmonised unemployment data produced by Eurostat (using a variety of methodologies depending on the data available in each country)

Another issue for discussion is if the indicators will be provided as such or they would be seasonally adjusted. Including seasonal adjustments introduce ad additional element of intervention that made the whole process more complex.

Also we need to bear in mind the availability of now cast estimates of population needed as population basis for weighting the sample. In our case, the calculation of population now casts is organised in a quarterly routine (precisely to feed the LFS needs). To move to calculate monthly now casts in not an exercise exempt of difficulties. The different 'maturity' of the now cast estimates under a monthly or quarterly routine has an impact on monthly estimates and their coherence to quarterly and annual estimates.

A short reference to the estimate procedure tested in Spain

The estimates we tested recently were based in a composite estimator (Empirical Best Linear Unbiased Predictors - EBLUP family) integrated by a 'direct' and a 'synthetic' components. The model assumption is that efficiency in the estimation for a sub period (month) can be substantially improved combining samples from present and former sub periods, as there is an strong structural component in the target variables within a domain in periods not too far from the reference period.

The expression for the composite estimator for a monthly reference period i (CME) in a domain   is of the type:

$$\theta_{CMEi} = \gamma_i \theta_{DMEi} + (1 - \gamma_i)(\theta_{MQEi}); \qquad 0 \le \gamma_i \le 1$$

θ_{DME} (Direct Monthly Estimator) is the preliminary estimator, playing here the role of 'direct' component', based on the available sub sample for the target month i.

θ_{MQE} (Moving Quarter Estimator) refers to the estimator based on the full available sample for the quarter that finalises in target month i ('synthetic' component).

The weight of each component is proportional to the sample size of the sample. In this way the construction of the estimate is determined completely by the sample available (without additional intervention). See references for further details.

The treatment of the monthly datasets are basically the same that the one of the quarterly datasets, except that the 'manual editing' is missed (the few problematic cases are deleted / ignored, as the editing affect to 'household' variables).

Some comments on the monthly data produced from quarterly LFS in Spain and Galicia.

This procedure fulfils many of the requirements we thought necessary for monthly estimates. The estimates are timely enough, the output is completely flexible (in fact could be the same that the quarterly output), the methodology is transparent (is would be known in advance and could be reproduced by a skilled enough user),

it is integrated in the usual processes of the quarterly survey (although some detail aspects need to be adjusted) and is completely based in LFS data.

Nevertheless, we have not been able to solve what we could call the 'communication to users problem'. Probably we overdo the users expected reaction, but the problem we found was to explain the relation between the monthly estimates of the three months of a quarter and the quarterly figures. Obviously, there is no way to calculate the quarterly estimates (the reference figures) from the monthly estimates. In fact, the 'synthetic' part of the composite estimator introduces an 'inertial behaviour' that seems to delay the detection of changes in the trend. (See charts in annex 2). The estimate, in fact, works like an advance of the quarterly figure that will be publish in the future.

Remarks on the production of monthly estimates based in the LFS

The timeliness of the monthly results is the key point. If we cannot produce the monthly estimates timely enough the exercise became almost completely useless.

The link between quarterly and monthly results is not straightforward and depends on the survey design. The communication to users on the relation between the monthly and quarterly (annual) estimates should be carefully assessed and developed.

The 'nature' of the monthly results requested must be clarified:
- Are they results that represent indicators of the reference month?
- Are they monthly advance estimates of the quarterly results?

The output must be clearly defined and be stable enough to fix suitable procedures and manage possible data revisions.

The determination of a harmonised set of EU monthly indicators of the labour market can play an important role to extend and develop the construction of monthly estimates based in LFS at national level.

Disclaimer clause:
This paper expresses the personal opinion of the authors on the topic. In particular, the content of the paper does not stated institutional positions from INE neither IGE about the monthly estimates based on the LFS.

References:

May 2009. *Monthly advanced estimates on ES-LFS quarterly data*. Workshop on LFS methodology Ljubljana. Honorio Bueno

2005. *Small area estimation in the Spanish Labour Force Survey*. INE & centro de investigación Operativa Universidad Miguel Hernández de Elche. Eurarea project papers. Jorge Saralegui, Montserrat Herrador, Domingo Morales, Agustín Pérez
2003. *Avance Mensual EPA: propuesta de estimación asistida por modelos*. Internal paper INE SG Metodología y Tecnología Estadística. Jorge Saralegui.

Annex 1: Preliminary set of indicators selected for testing

1. Total Employed
- BOTH SEXES
  - 16-24 year
  - 25 and over
- MALES
  - 16-24 year
  - 25 and over
- FEMALES
  - 16-24 year
  - 25 and over

2. Total Unemployed
- BOTH SEXES
  - 16-24 year
  - 25 and over
- MALES
  - 16-24 year
  - 25 and over
- FEMALES
  - 16-24 year
  - 25 and over

3. Employment rate
- BOTH SEXES
  - 16-24 year
  - 25 and over
- MALES
  - 16-24 year
  - 25 and over
- FEMALES
  - 16-24 year
  - 25 and over

4. Unemployment rate
- BOTH SEXES
  - 16-24 year
  - 25 and over
- MALES
  - 16-24 year
  - 25 and over
- FEMALES
  - 16-24 year
  - 25 and over

Annex 2. Charts for Spain 2006-2009