XXIV Convegno Nazionale di Economia del Lavoro - AIEL
Sassari 24-25 settembre 2oo9

Producing monthly estimates of labour market indicators exploiting the longitudinal dimension of the LFS microdata

By Riccardo Gatto[1], Silvia Loriga[2] and Andrea Spizzichino[3]

ISTAT, Italian National Institute of Statistics, Labour Force Survey Unit

Abstract

Availability of monthly estimates for employment, unemployment and other labour market indicators is obviously useful in Business Cycle Analysis as well as timeliness and robustness are appreciate qualities in the these estimates.

Eurostat currently releases monthly unemployment estimates based on the Labour Force Survey (LFS) data at both the national and the European Community level; Italy is still not producing monthly estimates.

In this paper results from a study project for the production of monthly estimates are presented. The main proposal is the adoption of a regression composite estimator that takes advantage of the overlapping structure of the sample over the time. Final figures show to be much more robust and compatible with the quality requested to an official survey.

JEL classification: J01, C42.

Keywords: Monthly estimates, Regression composite estimator, Rotation sample scheme, Robust estimates

[1]Riccardo Gatto, e-mail riccardo.gatto@istat.it

[2]Silvia Loriga e-mail silvia.loriga@istat.it

[3]Andrea Spizzichino e-mail andrea.spizzichino@istat.it

This research was supported, in part, by grants from the EUROSTAT. Views expressed in this paper are those of the author and do not necessarily represent those of the Italian Institute of Statistics.

1 Introduction

Since the beginning of 2007 a study project focused on the production of monthly estimates is being conducted in Istat, the Italian statistical institute; the main purpose is to study a proper methodology to produce robust monthly estimates based exclusively on Labour Force Survey (LFS from now on) data. This is a need since no reliable administrative sources on unemployment are available in Italy. The expected result of the project is the development of the chosen methodology and the production of monthly unemployment estimates and their regular release to Eurostat from the beginning of 2009. Several estimators have been taken into account to be tested on Italian LFS data: for more reasons the most interesting estimator is a particular kind of regression composite estimator (Singh, Kennedy, and Wu (2001)).

In the next paragraph the regression composite estimator is described in all its features. In the third paragraph results of the application to the Italian LFS are presented. Conclusions contain some final consideration.

2 The regression composite estimator

Few options are available to produce monthly estimates from a quarterly survey, and some of them are already adopted by some European national institutes. The first option is to produce direct LFS estimates with the same methodology used for the quarterly release. In this case some problems can arise since the sample design is not studied for monthly calculation and this can generate high sample errors and, in general, low robustness. Then this approach leaves space for improvements. The second road is the production of a three months moving average estimator using the last three available data, but this is not a proper monthly estimator and cannot be used for short term analysis. The third possibility is to use additional external information, such as monthly registered unemployment, in a direct estimator (using the LFS data as a benchmark) or in a model based estimator like in the Chow-Lin approach. Unfortunately in Italy reliable administrative data regarding the labour market are not available and in any case this approach cannot deals in the same way with employment and unemployment, intention of the Istat is to produce data both on employment and unemployment. The approach here proposed is the regression composite estimator which is more robust with respect the standard LFS one, it does not use external information so is

coherent enough with the quarterly release and can be used for any branch of the labour market, finally is a proper monthly estimator so produces useful data for short run analysis.

The generic regression composite estimator is a model based estimator that belongs to the class of the composite estimators; it may be used for repeated surveys with partially overlapping samples. It is based on the regression of the usual cross-sectional estimator on a set of predictors computed on the correlated observations on the overlapping sub-sample from other time points. Through this estimator both the estimates of level at a point in time and changes between two time points are improved. According to the Italian LFS sampling design, households participate to the survey four times during a period of 15 months; the households rotation scheme, 2-(2)-2, produces a 50% overlap of the sample between a quarter and the one preceding it and a 50% overlap between a quarter and the same quarter of the previous year. The regression composite estimator could then be applied to the Italian LFS trying to improve the quality of monthly estimates.

Which is the underlying rationale? It is well known that a generic regression estimator is more efficient than the Horvitz-Thompson estimator whenever the auxiliary variables that are used in the regression are correlated with the variable we want to estimate; usual regressors are demographic variables (gender and age classes) for which population data are known from external sources. In longitudinal surveys the employment status observed in a previous point in time is correlated with the current employment status (usually correlation is higher for employment but it is relevant also for unemployment). The idea is then to use the past information on the employment status (observed on the overlapping sub-sample) in the current estimation, through regression. More precisely Singh et al. developed for the Canadian LFS a kind of regression composite estimator in which micro-level past information are used as predictors, through the usual methodology of calibration, which will be shortly described in the following (Deville and Särndal (1992)).

The calibration estimator is the estimator currently used to produce quarterly estimates for the Italian LFS; the estimator of the variable Y total may be expressed as:

$$\widetilde{Y} = \sum_{k \in s} y_k w_k \tag{1}$$

where, referring to the sample s, y_k is the observation of the Y variable on

the unit k, and w_k is its grossing weight.

Grossing weights are obtained solving a minimization problem under constraints; the constraints regard the estimates of some auxiliary variables that have to be equal to the totals in the reference population derived by external sources. Through the calibration estimator, applying grossing weights, the sample reproduces the same distribution of the population according to the chosen auxiliary variables. Grossing weights for each sample unit are computed in two steps: 1) initial weights d_k are obtained for all the selected units as the inverse of the inclusion probability; 2) final weights w_k are obtained solving the following minimization problem under constraints:

$$\begin{cases} \min\{\sum_{k\in s} dist(d_k, w_k)\} \\ \\ s.t. : \sum_{k\in s} \mathbf{x}_k w_k = \mathbf{t} \end{cases} \tag{2}$$

where *dist* is a distance function between d_k and w_k, $\boldsymbol{t}$ is the k-vector of the totals of the auxiliary variables (known from external sources) e $\boldsymbol{x}_k$ is the k-vector of the auxiliary variables observed on the k^{th} unit.

In the regression composite version of the calibration estimator developed to produce monthly estimates for the Italian LFS (following a similar approach as for the Canadian estimator) some additional auxiliary variables have been introduced, regarding the employment status observed at the individual level in a previous point in time (3 months ago and 12 months ago), for individuals in the overlapping sub-sample; constraints are derived from the previous estimates referred to 3 months ago and 12 months ago obtained with this same estimator. The proposed Italian regression composite estimator is different from the Canadian one also due to a different chosen treatment of no overlapping people and to the different rotation structure of the sample.

From a theoretical point of view the choice of this kind of regression composite estimator seems to be preferable to other estimators (such as other model based estimators which use a macro-level regression) to assure a better comparability among monthly and quarterly estimates, both based on a calibration procedure. Moreover the households rotation scheme adopted by the Italian LFS, which assures a high overlap (50%) both at 3 months and at 12 months, is particularly adapt for this kind of estimator. In some sense taking 3 months ago information we add some information involving the trend, while 12 months ago information are useful to take into account the seasonality.

3 Results

The chosen version of the regression composite estimator has been applied to the data of the level of Italian employment and unemployment from the ISTAT LFS, over the period from 2004-I to 2007-VI. As a comparison other two monthly estimators have been tested on the same data set and results are shown in the following.

The three tested estimators are:

QCE, The Quarterly Calibration Estimator per Month: it is the standard estimator used for the quarterly release, weights are computed on the standard quarterly base while the total population is computed on a monthly base.

MCE, The Monthly Calibration Estimator: it is the standard estimator but weights and population are computed on on a monthly base. This estimator is allowed since the sampling design has already a monthly stratification.

MRCE, The Monthly Regression Composite Estimator: it is the calibration estimator with weights and population computed on on a monthly base adding constraints on the condition at time t-3 and t-12. Actually this is the estimator proposed in this work.

Figure (1) shows the evolution of the total employment and total unemployment over the period 2004-I - 2007-VI according to the three estimators, but a comparison among the series is difficult at this level. In order to magnify the differences a decomposition of the series has been run separating the three temporal component: Cycle-Trend, Seasonality and Erraticity. To do this the algorithm TRAMO-SEATS has been used like implemented in the Demetra software (version 2 SP1)[4].

A reduction in the variability of the erratic component for an indicator, changing the estimator, depends presumably on a reduction in the sampling error and results in a more stable and robust series. In the figure (2) the seasonal and the erratic components of the QCE series of the employment are shown, the magnitude of the two series are similar and this is not a desirable feature. In the figure (3) the same series, but estimated by means of the MCE, are shown. Now the variability of the erratic component has been sensibly reduced but still some serious outliers affect the series. In figure(4), finally, results from the MRCE are shown. A further reduction in

[4]Demetra is a software with the copyright of European Communities, developed for Eurostat mainly by J. Dossè and S. Hoffmann.

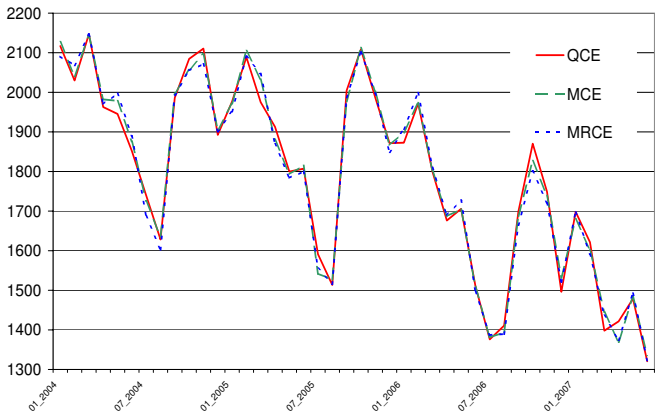

Figure 1: Employment in Italy: results from the application of different estimators. *(thousands of persons).*

the variability of the erratic component is evident and the outliers seem to be automatically correct by the MRCE. In the figures (5), (6) and (7) the same comparison can be run over the series of Italian Unemployment data.

The opinion of the authors it is possible to state that the MRCE is the most stable and robust estimators among the three studied.

4 Conclusive remarks

This work deals with the possibility to produce monthly figures about the labour market using data from the Labour Force Survey (LFS), produced by the ISTAT, the Italian National Institute of Statistics.

Even if the sampling design has already a monthly stratification the overall project of the survey is intended to produce quarterly estimates. Results from a straight use of the usual estimators show that results, even if consistent, can be improved in robustness and reliability.

In particular a revised version of the regression composite estimators, initially developed for the Canadian LFS, is here proposed. Applying the proposed

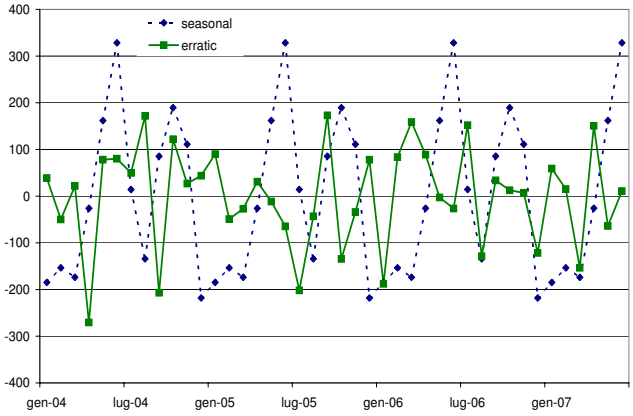

Figure 2: Employment in Italy estimated by means of Quarterly Calibration Estimator: seasonal and erratic component. *(thousands of persons)*.

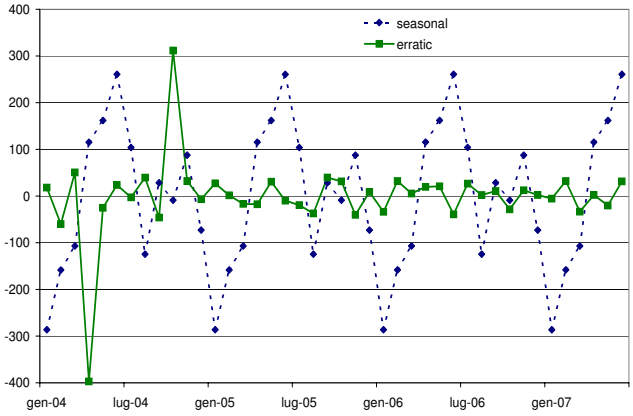

Figure 3: Employment in Italy estimated by means of Monthly Calibration Estimator: seasonal and erratic component. *(thousands of persons)*.

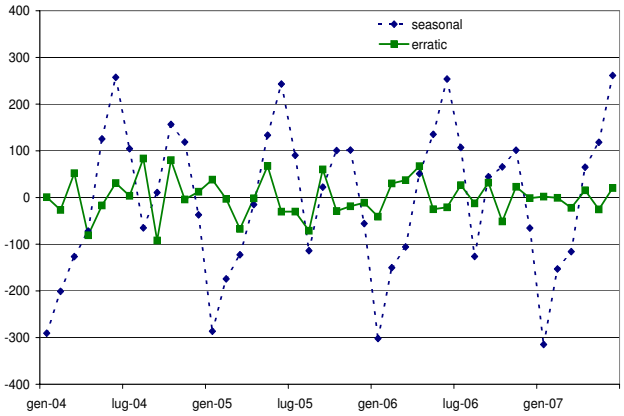

Figure 4: Employment in Italy estimated by means of Monthly Regression Composite Estimator: seasonal and erratic component. *(thousands of persons)*.

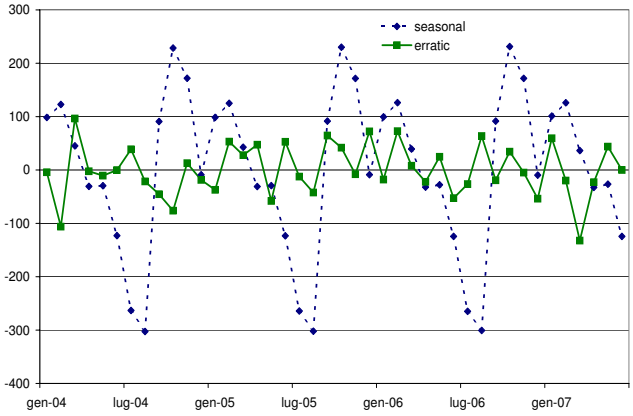

Figure 5: Unemployment in Italy estimated by means of Quarterly Calibration Estimator: seasonal and erratic component. *(thousands of persons)*.

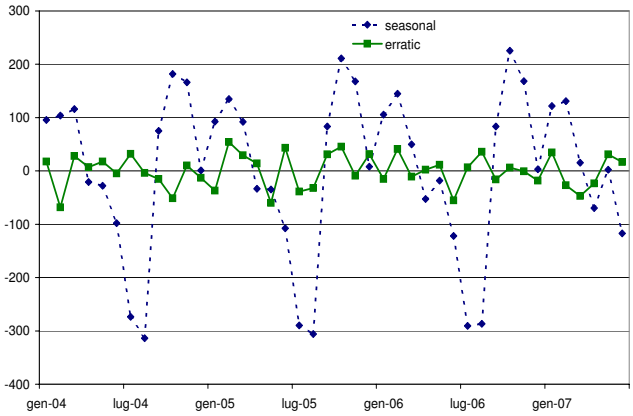

Figure 6: Unemployment in Italy estimated by means of Monthly Calibration Estimator: seasonal and erratic component. *(thousands of persons).*

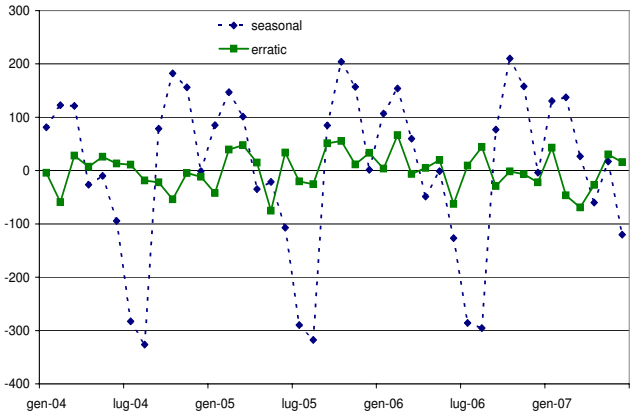

Figure 7: Unemployment in Italy estimated by means of Monthly Regression Composite Estimator: seasonal and erratic component. *(thousands of persons).*

methodology Istat will be able to produce good monthly estimates of the main labour market indicators as they appears in the PEEs. These estimates will be characterized by a good timeliness, a good coherency with the quarterly estimates, a good temporal coherency that allow for a robust pattern reducing the noise.

References

Deville, and Särndal (1992): "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87.

Singh, Kennedy, and Wu (2001): "Regression composite estimation for the Canadian labour force survey with a rotating panel design," *Survey Methodology*, 27.