The Italian Labour Force Survey consistency framework (draft version)

Antonio Discenza, Silvia Loriga, Alessandro Martini[1]
ISTAT – Italian National Statistical Institute

9th Workshop on Labour Force Survey Methodology
Rome, 15 – 16 May 2014

Abstract:

The Italian LFS is one of the most important statistical sources about the Italian labour market. It provides monthly, quarterly and yearly figures of the main aggregates and, in a longitudinal perspective, flows estimates as well. Due to the needs of short term analysis monthly and quarterly figures are produced also as seasonally adjusted time series.

Methodological approach and estimation procedures in IT-LFS take into account the need of providing consistent estimates for the different indicators.

The adopted approach to achieve consistency on both micro data and final figures is based on calibration estimators while benchmarking on macro data is used for seasonal adjusted time series.

Since IT-LFS is designed as a quarterly survey, quarterly figures are the "main benchmark" to achieve a consistent framework.

In the calibration process specific constraints are defined to guarantee coherence between different type of estimates (e.g. monthly vs quarterly or longitudinal vs quarterly).

For seasonal adjusted data a benchmarking procedure has been set up in order to take into account restrictions that can be of two types: contemporaneous constraints, assuming the form of linear combinations of the variables which should be fulfilled in every observed period, and temporal aggregation constraints, which require that the high frequency adjusted series (e.g. monthly) be in line with low frequency aggregates (e.g. quarterly).

Even if we do not apply wave approach yet, it could be necessary to introduce it considering the new modularization approach. In this context the IT-LFS consistency framework should be adapted in order to face this issue. A brief exercises has been conducted for evaluating the impact of wave approach in terms of biasness and accuracy of estimates on our rotational scheme.

1. Introduction:

LFS in Italy (IT-LFS) is designed as a quarterly survey, with rotational pattern 2-2-2, and the sample uniformly spread across all the weeks such that all territorial domains are represented in each month and in each of the 4 waves (rotational groups).

In the cross-sectional perspective it provides monthly and quarterly figures -both not-seasonally adjusted (NSA) and seasonally adjusted (SA) for the main aggregates - and yearly figures at NUT2 and NUTS3 level (as average of the quarterly figures). Moreover it is used as input for the small-area-estimator to obtain yearly estimates of employment and unemployment, levels and rates, for the 686 Local Labour Market Areas.

In a longitudinal perspective it provides 3 and 12 months flows estimates, as well as the yearly averages of the latter. Several collections of micro-data are traditionally disseminated to the users, both anonymized (for researchers, university, etc.), not-anonymized (for members of the National statistical system) , for both cross-sectional and longitudinal databases.

Methodological approach and estimation procedures in IT-LFS take into account the need of providing full consistency of the disseminated figures between the different sets of indicators and with their micro-data. This current approach is based on the use of calibration estimators for NSA figures, and benchmarking on macro-data for SA time series.

[1] Antonio R. Discenza: discenza@istat.it, Silvia Loriga: siloriga@istat.it, Alessandro Martini: alemartini@istat.it

Since IT-LFS is designed as a quarterly survey, quarterly figures are the "main benchmark" to achieve a consistent framework. In the calibration processes specific constraints are defined to guarantee coherence between different type of estimates (e.g. monthly vs. quarterly, longitudinal vs. cross-sectional).

For seasonal adjusted data a benchmarking procedure has been set up in order to take into account restrictions that can be of two types: contemporaneous constraints, assuming the form of linear combinations of the variables which should be fulfilled in every observed period, and temporal aggregation constraints, which require that the high frequency adjusted series (e.g. monthly) be in line with low frequency aggregates (e.g. quarterly).

For Local Labour Market Areas estimates, the implemented small area estimators[2] include a benchmark to the reference population by sex, age-groups, and municipality, and to the IT-LFS yearly estimates of employment and unemployment at NUTS3 level.

In Italy, the wave approach is not used so all the variables are collected in the quarterly questionnaire on the whole sample. However, taking into account the new modularization approach that is going to be used in the EU-LFS we start considering how the IT-LFS consistency framework would be affected and how it should be adapted. A brief exercises has been conducted for evaluating the impact of wave approach on our estimation procedures and in terms of coherence between yearly estimates and annual averages.

2. Demographic data on resident population

For the consistency framework of IT-LFS, and the timely dissemination of its figures, a fundamental role is played by the known totals (auxiliary information) available for the reference population. Demographic data on resident population in the Italian municipalities are collected on monthly bases from the Population Register Offices of each municipality, thus are updated by the Demographic Sector in Istat and are made available for weighting purposes of LFS. These data consists of the resident population in each municipality by sex, age and citizenship (Nationals/Non-Nationals).This population is used as known total in the weighting procedures for the monthly and for the quarterly estimates.

Istat started to produce Demographic monthly data on population when it started to produce LFS monthly estimates. From that time more effort has been put to enhance its quality and timeliness to suite this purpose. In fact, it was soon recognized that the month-to-month variation of labour market indicator was highly influenced by demographic totals when these were not continuously updated (especially in the past when, from one month to another, demographic figures showed very high variation of non-nationals in the populations).

3. Quarterly estimates and annual averages

The LFS sample used from 2004 onwards, was re-designed in order to satisfy the requested precision for the estimates and the other Eurostat requirements contained in Council Regulation 577/98. It takes into account, the organisational practicability, operational and budgetary restrictions, as well as other important national dissemination needs such as:

monthly representativity of the sample in order to produce monthly estimates of the main figures at national level

- quarterly estimates about the non-national population
- annual estimates of main indicators of labour market at NUTS 3 level;
- annual estimates of household indicators and household characteristics
- annual estimates for the 13 large municipalities (with more than 250 thousands residents)
- annual estimates of employment and unemployment at Local Labour Market Areas (using small area estimators)

The quarterly sample, from 2014, is composed by about 71,000 households and 150,000 persons, with an average sampling rate around 0.4%.The reference population is all persons residing in private households. The survey technique is a mixed CAPI /CATI.

[2] "Occupati residenti e persone in cerca di occupazione per SLL 2001. Medie 2004 e 2005." Relazione Metodologica. Istat, february 2006 - M. D'Alo', S. Falorsi, F. Solari

The sampling design is a two-stages, with stratification of primary units in the first stage (these are the Municipalities stratified by demographic size, at NUTS 3 level) and rotation of final units (private household) in the last stage. The sample of Municipality and households are randomly selected from population registers, and all the de-facto members of the selected households are interviewed.

Households follow a 2-2-2 rotation scheme; they participate to the survey for 2 consecutive quarters, then they temporally exit from the sample for the following 2 quarters, and then come back in the sample for 2 quarters. This rotation scheme produces a 50% overlap of the sample between a quarter and the previous one and a 50% overlap of the sample between a quarter and the same quarter of the previous year.

Each quarter the sample is composed by 4 groups of households (rotation groups) of equal size, being respectively at first, second, third and fourth wave. For each of the four rotation groups, the sample of Municipalities and households are uniformly distributed across the weeks of a quarter and of the whole year (a space-time allocation of the sample).

Moreover, all territorial strata are equally represented in each month of the quarter (at least in one week) and all the selected municipalities are involved in the survey at least three weeks during the quarter (one for each month).

Grossing weights for the IT-LFS are computed in three steps: In the first step, the initial weights are calculated as the inverse of probability of selection; in the second step, non-response adjustment factors are calculated by household characteristic; in the last step, final weights are calculated using a calibration estimator using the auxiliary demographic information regarding the reference population by sex, five-years age groups, nationality and region (NUTS 2 and NUTS 3 level).

Final weights are computed at household level, which means that each component of the same household has exactly the same final weights of all the others (household weight). This method allow us to produce coherent estimates at both individual level and household level (e.g. the number of full-time employed "mothers" living in an household "parents with children" is equal to the number of households "parents with children" in which the "mother" is full-time employed). In Italy, household data are traditionally and diffusely used for socio-economic analysis as matter of fact that individual attachment, behavior, and success in the labour market is also determined by his household characteristics, and on the contrary, new household creation and changes are determined by individual social-economic and labour status.

For the purpose of weighting the whole quarterly sample, a quarterly populations computed as weighted average of the three monthly populations, with weights given by the number of weeks in the month.

Thus

$$P_{psen}(q) = \frac{P_{psen}(m_1) \cdot z_{m_1} + P_{psen}(m_2) \cdot z_{m_2} + P_{psen}(m_3) \cdot z_{m_3}}{13}$$

where

- $P_{psen}(q)$ is the population referring to the a generic geographical domain p, gender s, 5-years age group$e$ and nationality n, for the generic quarter q;
- $P_{psen}(m)$ is the population referring to the a generic geographical domain p, gender s, 5-years age group$e$ and nationality n, for the generic month m;
- z_m is the number of weeks in the month (4 or 5);
- 13 is the number of weeks in the quarter.

As for the sample design, also the weighting procedure is more complicated than required by EU regulations. In fact, it includes many more constraints, from external sources, in order to satisfy the other national needs listed at the beginning of this paragraph, as listed below:

- distribution of population by sex and fourteen 5-year age groups (0-14, 15-19, …, 70-74, 75 and more years), at NUTS II level;
- distribution of non-national population divided as follow: male, females, other EU citizens, Non-EU citizens, at NUTS II level;
- number of households for each rotation group (1/4 of the total), at NUTS II level;
- distribution of population by sex and five age groups (0-14, 15-29, 30-49, 50-64, 65 and more years), at NUTS III level;

- distribution of population by sex and five age groups (0-14, 15-29, 30-49, 50-64, 65 and more years) for the thirteen large municipalities with more than 250.000 inhabitants;
- distribution of population by sex, for each of the three months of the quarter, at NUTS II level.

As matter of fact that quarterly weights already incorporate the benchmark to the NUTS 3 population, the annual datasets are obtained using all the quarterly interviews and annual weights are computed simply dividing the quarterly weights by four. Thus, the full coherence between annual and quarterly estimates is automatically ensured.

With reference to the monthly constraint applied in the quarterly weights, this was introduced at the beginning of 2004 for two reasons.

Firstly, the weekly and monthly allocation of the quarterly sample implies that the weighting system (which includes the total-non-response adjustment) must contains specific constraints to be sure that the actual sample (which could be affected by differential weekly-related-non-response[3]) reflects the theoretical allocation. Of course it was not possible to put constraints on the single weeks but it was easily done at monthly level.

Secondly, the possibility to produce direct estimates of the monthly figure at national level. The current weighting procedure provides fully consistent monthly and quarterly weights, although the former are available only when all the interviews of the quarter have been completed and thus, the quarterly weights computed. Monthly estimates can be directly obtained using the monthly sample and its monthly weights. Monthly weights can be easily obtained from the final quarterly weighs using an adjustment factor equal to 13/4 or 13/5 depending on the number of weeks in the month.

One of the problem with this approach is the inability to produce timely results as required by Eurostat. Moreover, the monthly time series obtained with direct estimates showed too much fluctuations, thus these direct estimates were never published. For few years Istat studied the possibility to improve timeliness and quality of the monthly estimates and it found that a regression composite estimator would have suite the purpose[4]. After evaluating the results and tuning the model for a long period, monthly estimates where finally disseminated. The method used is explained in the next paragraph

4. Quarterly and monthly estimates

In In December 2009 Istat started the production of official monthly data of the main labour market indicators. Data are firstly disseminated on a provisional basis, being computed over a partial sample, about 30 days after each reference month. A release calendar is arranged (and published well in advance on the website) in order to be concurrent with the Eurostat release on monthly unemployment.

Both not seasonally adjusted (NSA) and seasonally adjusted (SA) data are disseminated, even if the analysis on the press release is focused on SA. In particular, the release concerns the level of the main aggregates (employed, unemployed and inactive) by gender and their corresponding rates. A specific focus is devoted to the youth unemployment.

Monthly estimates are purely based on LFS data. They are computed using only the sample units allocated and interviewed in the month and are benchmarked to the monthly population $P_{psen}(m)$. The production process for the provisional estimates starts about 22 days after the end of the reference month. At that moment the fieldwork for that month is not completed yet, thus these estimates are based on a partial sample (over 80% of response rate). However, the fieldwork continues (up to 5 weeks) and thus the final estimates are computed at a later stage, based on the full monthly sample (net of total non-response).

[3] In the IT-LFS happens that, in some periods of the year (summer or Christmas holidays), response rates are much lower than in the rest of the reference quarters. If the labour market dynamics vary a lot from month to month (and this is the case in Italy) then the bias in the temporal distribution of the actual sample may produce bias in the quarterly estimates.

[4] Producing monthly estimates of labour market indicators exploiting the longitudinal dimension of the LFS microdata. Paper presented at the NTTS conference 2009. R. Gatto, S. Loriga, A. Spizzichino

Final monthly data are then produced when the corresponding quarterly data are available, that is about 60 days after the reference quarter. Together with quarterly data, final monthly figures (of the three months of the quarter) and the provisional estimates for the first month of the following quarter are released. Final monthly data are based on the whole actual sample and they are consistent with quarterly data.

Estimation procedure of IT monthly estimates is based on a regression composite estimator (see for instance Singh Kennedy and Wu, for a detailed description of the estimator and an application to the Canadian LFS data). The 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 improving the estimates of levels at a point in time and changes between two time points. This kind of estimator is based on the regression of the usual cross-sectional estimator (the calibration estimator computed on monthly data in our case) on the new predictors provided on the correlated observations on the overlapping sub-sample from other time points.

Given the longitudinal structure resulting from the 2-2-2 rotation scheme, the regression composite estimator can then be applied to the IT-LFS improving the quality of monthly estimates. In particular, we adopt a version of the regression composite estimator based on the technique of calibration: the main feature of this estimator is that the information available on individuals already interviewed in the past are used as auxiliary information in the calibration.

In essence, besides the constraints usually used in calibration (distribution of population by age classes and gender at different levels of geographical detail, for which the data are known from external sources) we have additional constraints based on the labour status of individuals 3 months before and/or 12 months before (respectively known for about 50% of the sample, unless the non-response).

Once every three months, when quarterly figures are made available, we produce and disseminate the final monthly estimates for each month of the quarter.

The production of final monthly data is slight different from that of provisional data. The sample used to compute the estimates is not partial but complete, the sample size is therefore higher compared with provisional data. The check and imputation phase is skipped because the data used are already imputed by the procedure applied to quarterly data. The estimator is the same as the provisional one, the regression composite estimator.

The core of the transition from provisional to final monthly data is represented by a specific step in the estimation procedure that assures that monthly data are consistent with the quarterly ones: for the main aggregates, the weighted average of the three monthly figures (with weights proportional to the number of weeks in each month: 4/13 or 5/13) is equal to the corresponding quarterly figures.

To ensure this consistency a further calibration step is applied, that is a correction factor is applied to the weights of the final monthly data obtained by the regression composite estimator. In this step, the data of the three months are jointly considered, the initial weight in the calibration procedure is the weight previously calculated, the constraints are related to both single months (total population by sex and age groups at different levels of geographical detail), and the quarterly estimates of the main aggregates: employed, unemployed and inactive by gender and three age groups (15-24, 25-64, 65+).

The procedure for seasonal adjustment and the reconciliation procedure are substantially the same as for provisional monthly data. First we apply a univariate seasonal adjustment procedure to each time series using the algorithm TRAMO-SEATS (Demetra). Then, in order to ensure consistency between different aggregates and the total population and between monthly and quarterly SA series, we apply a time series reconciliation procedure in two steps as proposed by Di Fonzo and Marini[5], based on a dual system of constraints: inter-temporal constraints and contemporary constraints.

This methodology of benchmarking is based on the "movement preservation principle" in order to maintain the temporal profile of the original series.

The inter-temporal constraints ensure consistency across monthly and quarterly SA estimates, in particular for each disseminated aggregate the average over the three months of a quarter (using the number

[5] "Simultaneous and two-step reconciliation of systems of time series: methodological and practical issues". Applied Statistics (2011) 60 Part 2. T. Di Fonzo, M. Marini.

of weeks as weights) is consistent with the quarterly figure. The SA quarterly time series (starting from the IV quarter 1992), are long and reliable series, that is in addition to the consistency they also carry robustness to the series.

The contemporary constraints ensure consistency of the several series with population data. In the first step a univariate benchmarking is performed (Denton modified, which preserves the dynamic of the original series) in order to meet intertemporal constraints. The second step performs a reconciliation quarter by quarter of monthly data with quarterly data, through redistribution of differences proportionally to the square of the level of the series, in order to meet contemporary constraints.

The result is a set of monthly SA time series all consistent with population data and with quarterly SA series.

Point of discussion about consistency between the SA monthly and quarterly time series.

A reconciliation is applied to the monthly Seasonal Adjusted series in order to ensure consistency between different aggregates (Employed + Unemployed + Inactive = Population) and with the quarterly Seasonal Adjusted series. Is this need for consistency just an IT-LFS priority or could it be a sort of guideline for European statistics?

5. Cross-sectional and Longitudinal estimates

Longitudinal data can be derived from IT-LFS, linking individual record from different quarters, obtaining quarter-on-quarter dataset at 3, 9, 12 and 15 months. They can be used to estimate the gross flows between two quarters, i.e. the movements between labour market statuses which determine the dynamic of the most important figures. It is important to bear in mind that yearly or quarterly net changes are the final result of a high number of gross flows of different nature and different magnitude (death, migration and labour status transitions).

The longitudinal micro-data files constitute a "by-product" of the survey itself; thus, we should point out that it is not a proper panel referred to the whole population. Transition matrices containing estimates of gross flows can be produced, by using these longitudinal datasets, taking into account the following aspects and limitations:

a) Labour market flows could in principle provide estimates of all flows regarding the population living in a country at the beginning of the period of observation. However, this initial population modifies during the period of observation due to demographic events such as deaths and internal/international migration. LFS is not a "real" panel survey thus persons which move out of the selected households, or household which move out of the municipality, are not re-interviewed, thus we cannot gather all the information that would be ideally desirable.
b) Longitudinal estimates can refer only to a specific longitudinal reference population which anyway is a substantial part of the initial population,
c) Panel attrition may occur at subsequent waves due to household non-contact, refusal, etc.;
d) Weights should reflect the longitudinal population, should account for the panel attrition (usually not at random) and have to ensure consistency with the other LFS quarterly estimates.

Given that the longitudinal data concerns only individual residing in the same municipality both at the beginning and at the end of the period, the longitudinal component (sub-sample) of the Italian LFS requires thus the specification of a suitable reference population[6] like follow: the population which is resident in the same municipality for the period of 12 months (or 3 months), thus net of deaths and internal or international migration. Considering the following quantities referring to the a generic geographical domain p, gender s, 5-years age group e, for a period going from a generic first quarter q_1 and a last quarter q_2: ${}^{q_1}P_{pse}$ and ${}^{q_2}P_{pse}$ are the quarterly cross-sectional population at the beginning and end of the period; ${}^{q_{1,2}}m_{pse}$ are the

[6] See "Longitudinal data for the analysis of Italian labour market flows" Statistica Applicata - Italian Journal of Applied Statistics Vol. 22 (2), B. Boschetto, A. R. Discenza, C. Lucarelli, S. Rosati, F.Fiori)

deaths occurring during the period, ${}_{q_{1,2}}n_{pse}$ are the births occurring during the period, ${}_{q_{1,2}}c_{pse}$ is the emigration from province p during the period, ${}_{q_{1,2}}i_{pse}$ is the immigration into province p during the period; we have the following population equation

$$ {}_{q_1}P_{pse} - {}_{q_{1,2}}m_{pse} - {}_{q_{1,2}}c_{pse} + {}_{q_{1,2}}n_{pse} + {}_{q_{1,2}}i_{pse} = {}_{q_2}P_{pse} $$

and thus the longitudinal population defined as

$$ {}_{l_{1,2}}P_{pse} = {}_{q_1}P_{pse} - {}_{q_{1,2}}m_{pse} - {}_{q_{1,2}}c_{pse} $$

The longitudinal population, by sex and age, at NUTS3 level, is updated on quarterly bases by the Demographic Sector in Istat and made timely provided for weighting purpose of LFS longitudinal.

To take into account the coherence between cross-sectional and longitudinal estimates, we have to consider that the longitudinal component produces both cross-sectional and longitudinal estimates referred to the longitudinal population. These cross-sectional estimates obtained by the longitudinal data have to be consistent with the "other" estimates provided by the cross-sectional samples (the full quarterly samples) at the beginning and at the end of the observed period.

The weighting procedure used for the production of the IT-LFS longitudinal data consists of three steps and uses two calibration[7] stages (see Boschetto et al. , Statistica Applicata, 2010):

- the first has been used to account for the differences in the rotation groups and to ensure consistency with a large number of quarterly estimates[8];
- the last has been used to adjust for the bias due to individual/household non-response and to make weighted longitudinal sample totals coherent to the longitudinal population totals.

This approach allow us to compute several kind of longitudinal micro-data files and transition matrices, producing estimates of gross flows and transition rates, assuring consistency of a large number of stock/flow results, by sex and age groups, and at NUTS2 and NUTS3 level.

Thus, it is straightforward to calculate quarter-on-quarter flow estimates such as:

- quarterly flows: from one quarter to the subsequent one (3 months);
- yearly flows: from one quarter to the same quarter of the subsequent year (12 months);

and year-on-year flow estimates such as:

- average yearly flows: as average of the 4 yearly flows[9], referring to the 4 quarters of the calendar year. In this case, the flow estimates are consistent with two yearly cross-sectional estimates.

Figure 1 shows the complete matrix in which the transition matrix containing stock and flow estimates for the longitudinal population (in yellow) is nested with the cross-sectional quarterly estimates (yellow vectors) and the other components of the system.

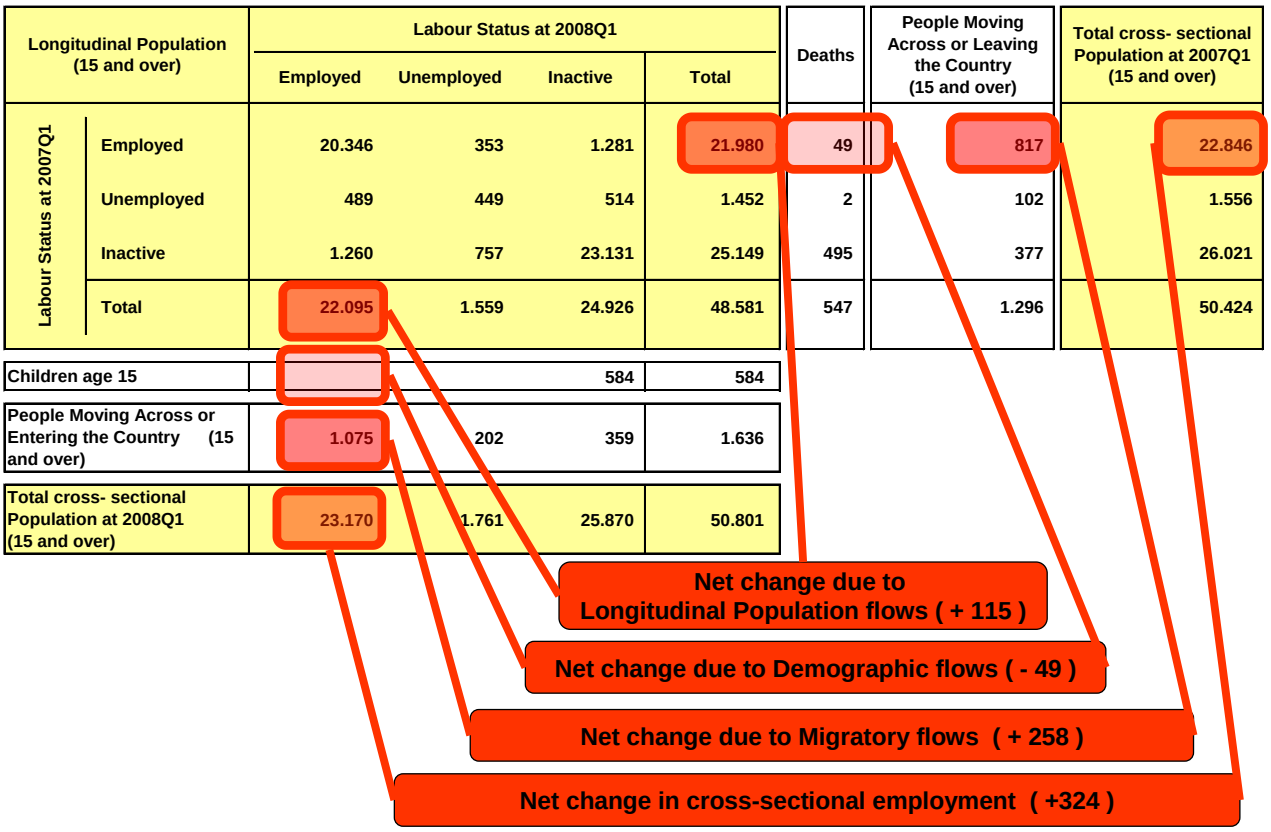
The complete matrix above shows that a net change of +324thousands employed has been observed between 2007Q1 and 2008Q1. This overall increase is the result of the net changes of several components (+258thousands due to migrants, -49thousands due to demographic changes, +115thousands due to the longitudinal component). Moreover, the transition matrix shows all the flow estimates related to the longitudinal population and reveals that the net increase of 115thousands employed is due to 1.634thousands of persons leaving the employment (353+1281) and 1.749 thousands entering in it (489+1260).

[7] Calibration involves modifying the original weights in order to simultaneously satisfy several marginal constraints (or to control totals) while minimizing the distance between original and adjusted weights.

[8] Longitudinal estimates have higher variability than quarterly estimates and it is not possible to control completely their consistency with quarterly estimates. However, the weighting strategy that we use for the longitudinal sample highly reduce the risk of obtaining inconsistent results.

[9] Append of the quarterly datasets and quarterly weights divided by four.

Figure 1. Complete Matrix with stocks and gross flows. Quarter 1 2007 – Quarter 1 2008. (Thousands)



Point of discussion about consistency between stock and flows

The use of this methodological approach requires the availability of data on longitudinal population of good quality and details, and this is the case for Istat. It would be interesting to study the possibility to use it in other countries, or at European level.

6. Modularisation of LFS and Wave Approach

Taking into account the new modularization approach proposed in the Eurostat modernisation programme we conducted a brief and not exhaustive exercise for evaluating the impact of wave approach in terms of additional complexity in estimation procedures (weighting system) and on the IT-LFS consistency framework.

First of all, being the LFS a quarterly survey, we would say that "the LFS use a standard questionnaire for quarterly variables and a longer one for a sub-sample from which it derives structural yearly estimates".

Wave approach has never been used in IT-LFS, but as all the variables are collected in the quarterly questionnaire on the whole sample, we have the possibility to simulate a wave approach on past data and compare results with the annual averages already disseminated.

Commission Regulation (EC) No 377/2008, sets "conditions for the use of a sub-sample for the collection of data on structural variables which optionally need only to be surveyed as annual averages." It states that "Consistency between annual sub-sample totals and full-sample annual averages shall be ensured for employment, unemployment and inactive population by sex and for the following age groups: 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 +" and moreover "The sample used to collect information on ad hoc modules shall also provide information on structural variables".

For this purpose we have selected the four 1st waves of the 4 quarters 2012 (yearly sub-sample). In this way, the sub-sample is made of independent observations and has the same theoretical sample size of a quarterly sample. Then we have reweighted the sub-sample benchmarking to the averages of the 4 quarters of the year.

The ideas behind this simulation exercise, are the following:

Every year the annual sub-sample has to be used also for the Supplementary Annual Modules (the actual ad-hoc modules) and the set of constraints should be similar to the one used for quarterly weights (and thus, for example, taking into account also the regional distribution of the population and its citizenship). In fact it is important to take into account also the differences between the theoretical and the actual sample in terms of distribution over time and space (to compensate for a possible different total-non-response in different quarters and different regions).

Some yearly variables in the sub-sample could be strictly correlated with those collected quarterly, not only with ILO status. Thus, if the sub-sample is biased with respect to the quarterly variable then the estimate of the yearly variable could be biased. For example, "income", "second job" and "looking for another job" are probably related to STAPRO, FTPT, TEMP, NACE, ISCO. For this reason, the minimum set of requirements in the regulation 377/2008 could be not sufficient to achieve coherent results, thus we introduced many more different constraints.

Moreover, it seems obvious that, the higher is the total non-response and the bias in the different waves or quarters, the higher is the risk of inconsistencies between the two kinds of annual averages (from the full and the sub-sample).

Thus, starting from the quarterly weights, we used a calibration approach, with several different sets of constraints (SoC) to get the new final weights. In this paper we present results for some of them only, pointing out that all the constraints have been defined at NUTS 2 level:

SoC_1:	Only the minimum set of constraints in the regulation 377/2008, but at NUTS2 level.
SoC_2:	The same constraints on the populations as in the regular quarterly weights; not those in the 377/2008 regarding labour status
SoC_3:	The same constraints on the populations as in the quarterly weights; plus labour status by sex and age groups (the same traditionally used at NUTS3 level[10])
SoC_6:	The same constraints on the populations as in the quarterly weights; plus labour status by sex and age groups (the same traditionally used at NUTS3 level); plus STAPRO (employee/self-employed), FTPT, TEMP, NACE (3 groups), ISCO (3 groups).
SoC_7:	The same constraints on the populations as in the quarterly weights; plus labour status by sex and age groups (the same traditionally used at NUTS3 level); plus STAPRO (employee/self-employed), FTPT, TEMP, NACE (3 groups), ISCO (3 groups); plus population 15 and over, by sex and labour status, by quarter.

Table 1 below shows the differences, for some of the most relevant variables, between the yearly estimates obtained from the full sample (as average of the quarterly estimates, our benchmark) and the yearly estimates obtained from the sub-sample using the different sets of constraints.

For INCDECIL (see Table 1), the sub-sample provides higher relative frequencies for lower monthly pay than the full-sample, especially for the first decile. The opposite happens for higher monthly pay. The differences became bigger in Soc_7 where constraints are put on the characteristics of the employment also.

For MAINSTAT (see Table 2), the sub-sample provides a lower number of employed (about 100 thousands) and a higher number of unemployed than the full-sample (100 thousands). The greater difference occur with Soc_2, where no constraints are put on labour statuses. No much difference between the other SoC's.

Table 3 shows the results for some of the variables related to the SECOND JOB. The sub-sample provides a much higher number of employed with a second job (+30%), and a much higher incidence (from 1.4% to 1.9%). As consequence, the number of total hours worked is higher (about 20%) providing a much smaller number of hours worked per employees (from 23.5 to 18.6). The estimates are higher for both employees and self-employed, and in all the main NACE sectors. However, the sub-sample tends to reduce

[10] Unfortunately, when we tried to put contemporary constraints on age, both at NUTS3 level (using the "traditional" groups 0-14, 15-29, 30-49, 50-64, 65+) and at NUTS2 level (using the groups required for the wave approach) the procedure did not converge. Thus, for this brief exercise, waiting for more deep analysis, the "traditional" age groups were used.

the incidence of employees and of the employed in the Service sector, and increase the incidence of self-employed and of the employed in Agriculture and industry.

Table 1 – INCDECIL: Annual averages obtained from the full sample and the sub-sample using different sets of constraints. Year 2012. (Percentages)

INCDECIL	Full-Sample	Sub-Sample				
		SoC_1:	SoC_2:	SoC_3:	SoC_6:	SoC_7:
01	*10.1*	*11.0*	*11.2*	*11.1*	*11.3*	*11.3*
02	*9.9*	*9.8*	*9.9*	*9.9*	*10.0*	*10.0*
03	*11.4*	*11.8*	*11.9*	*11.9*	*11.9*	*11.8*
04	*8.8*	*8.8*	*8.8*	*8.8*	*8.7*	*8.7*
05	*11.2*	*11.5*	*11.4*	*11.4*	*11.4*	*11.4*
06	*9.8*	*9.6*	*9.5*	*9.6*	*9.5*	*9.5*
07	*8.9*	*8.5*	*8.5*	*8.5*	*8.5*	*8.5*
08	*9.5*	*8.9*	*8.9*	*8.9*	*8.8*	*8.8*
09	*10.5*	*10.0*	*9.9*	*9.9*	*9.8*	*9.8*
10	*10.1*	*10.1*	*10.1*	*10.1*	*10.0*	*10.0*

Table 2 – MAINSTAT: Annual averages obtained from the full sample and the sub-sample using different sets of constraints. Year 2012. (Absolute values, Percentages)

MAINSTAT	Full-Sample	Sub-Sample				
		SoC_1:	SoC_2:	SoC_3:	SoC_6:	SoC_7:
Employed	22,455	22,345	22,273	22,343	22,331	22,329
Unemployed	5,194	5,320	5,566	5,314	5,321	5,319
Pupil, student	4,330	4,270	4,210	4,272	4,271	4,273
In retirement	10,624	10,542	10,478	10,517	10,516	10,516
Fulfilling domestic tasks	7,885	7,861	7,807	7,869	7,880	7,882
Others	1,508	1,658	1,662	1,680	1,676	1,676
% Employed	*43.2*	*43.0*	*42.8*	*43.0*	*42.9*	*42.9*
% Unemployed	*10.0*	*10.2*	*10.7*	*10.2*	*10.2*	*10.2*
% Pupil, student	*8.3*	*8.2*	*8.1*	*8.2*	*8.2*	*8.2*
% In retirement	*20.4*	*20.3*	*20.2*	*20.2*	*20.2*	*20.2*
% Fulfilling domestic tasks	*15.2*	*15.1*	*15.0*	*15.1*	*15.2*	*15.2*
% Others	*2.9*	*3.2*	*3.2*	*3.2*	*3.2*	*3.2*

Table 3 – EXIST2J-STAPRO2J-NACE2D2J-HWACTUA2: Annual averages obtained from the full sample and the sub-sample using different sets of constraints. Year 2012. (Absolute values, Percentages, averages)

EXIST2J - STAPRO2J - NACE2J2D - HWACTUA2	Full-Sample	Sub-Sample				
		SoC_1:	SoC_2:	SoC_3:	SoC_6:	SoC_7:
Employed with a second job	331	434	433	436	435	435
- Employees	142	172	177	177	177	177
- Self employed	189	262	257	259	258	258
-- Agricolture	16	24	23	23	24	23
-- Industry	21	30	30	30	30	29
-- Services	295	380	381	383	382	383
Number of hours worked	7,781	8,016	8,059	8,116	8,133	8,045
% of Employed with a second job	*1.4*	*1.9*	*1.9*	*1.9*	*1.9*	*1.9*
% Employees	*42.8*	*39.7*	*40.8*	*40.6*	*40.7*	*40.7*
% Self employed	*57.2*	*60.3*	*59.2*	*59.4*	*59.3*	*59.3*
% Agricolture	*4.7*	*5.5*	*5.3*	*5.3*	*5.4*	*5.4*
% Industry	*6.3*	*7.0*	*6.9*	*6.8*	*6.8*	*6.7*
% Services	*89.0*	*87.5*	*87.9*	*87.9*	*87.8*	*87.9*
Average Number of hours worked per employee	*23.5*	*18.5*	*18.6*	*18.6*	*18.7*	*18.5*

Points of discussion about the wave approach

When looking at these differences we don't feel we can use the term "bias" because; although the estimators should produce unbiased estimates, is indubitable that a panel attrition exist and the estimates from the full sample could be biased.

On the other hand, it seems also reasonable that estimates from the sub-sample should be "in principle" less biased than those from the full-sample, but with a lower precision.

However, having said that, some important questions arise:

a) Is it methodologically correct to benchmark the sub-sample estimates to the full sample ones if we suspect that the latter are more biased then the former ?
b) Are we sure that the benefits of a reduction in respondents burden are so high that they compensate, or exceed, the much bigger effort needed for the organization, for the continuous management of questionnaires and micro-data, and for the implementation of a more complex methodology?
c) Time series for the yearly variables could have breaks, given the differences in the annual averages obtained with the two systems. How to manage this issue?
d) What would be the dissemination strategy, given the new limitations in dissemination of results due to the consistency problem? What kind of yearly indicators can be produced: levels or percentage distributions?

For sure, the introduction of the wave approach will have some consequences on the dissemination of yearly results and their micro-data, for instance:

- a small reduction of the published quarterly and yearly indicators due to the impossibility to guarantee full consistency;
- a reduction in the contents of the sub-sample micro-data file, where some of the variables should probably be recoded with the same classification used in the weighting procedure (and for which consistency is guaranteed).

References

Albisinni, M., Discenza, A.R. (2002). La mobilità nel mercato del lavoro: principali risultati – aprile 1998-aprile 2002. Approfondimenti, ISTAT, Roma.

Boschetto, B., Discenza, A. R., Lucarelli, C., Rosati S. and Fiori F. (2009) "Longitudinal data for the analysis of Italian labour market flows", Statistica Applicata - Italian Journal of Applied Statistics Vol. 22

Ceccarelli C. and Rosati S. (2005) "Data editing for the Italian Labour Force Survey", Conference of European Statistician, Ottawa.

Ceccarelli, C., Discenza, A.R., Fiori, F. and Lucarelli, C. (2012). Weighting issues in LFS longitudinal data. Rivista Italiana di Economia Demografia e Statistica, LXVI (1): 77-84.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, (87): 376-382.

Deville, J.C., Särndal, C.E. and Sautory, O. (1993). Generalized raking procedures in survey sampling. Journal of the American Statistical Association, (88): 1013-1020.

Gatto, R., Loriga, S. and Spizzichino A. (2009) "Producing monthly estimates of labour market indicators exploiting the longitudinal dimension of the LFS microdata." Paper presented at the NTTS conference 2009.

Istat (2006). La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione. Metodi eNorme, ISTAT, Roma.

Rosati, S. (2004). Longitudinal imputation for the quarterly labour force survey. Atti della XLII Riunione Scientifica, Cleup, Bari: 335-338.

Rosati, S. and Boschetto, B. (2013). Longitudinal data editing for the Italian LFS. Proceedings of Statistics Canada International Symposium 2009. Gatineau, Quebec. In press.

Särndal, C.E. (2007). The calibration approach in survey theory and practice. Survey Methodology, (33): 99 -119.

Särndal, C.E. and Lundström, S. (2005). Estimation in Surveys with Non-response. Wiley Series in Survey Methodology, Chichester.

Tate, P.F. and Clarke, P.S. (1999). Methodological Issues in the Production and Analysis of Longitudinal Data from the Labour Force Survey, GSS Methodology Series, 17. Office for National Statistics, London