

THE NEW METHODOLOGY FOR THE STS INDICATOR ON THE “NUMBER OF PERSONS EMPLOYED”

Marco Lattanzio

1. Introduction

Starting from May 2018 Istat produces the quarterly indicator on the Number of Persons Employed required by the Eurostat-STS regulation (Reg. EU n.1165/98) according to a new compilation methodology. The regulation concerns the production of harmonized short-term statistics on businesses and this is the unique indicator on employment on the business side. The first innovation of this paper consists in the new exploitation of a short-term quarterly variable, the number of self-employees calculated in the context of the Labour Force Survey (LFS), used as indicator in the temporal disaggregation of the annual benchmark, represented by the number of self-employees from National Accounts (NA). This latter source of information better matches the definition requirements. Furthermore, the second aspect of the innovation of this paper is the new method for quarterly temporal and sectorial disaggregation of the NA source used. The LFS source is mainly used to give information about the short-term dynamic that with this new method is considerably improved, considering also that, as it will be described, the previous temporal disaggregation method used as indicator the time series of the employees, that is a source very poor in order to give information about the self-employee's short-term dynamic. Nevertheless, improvements are balanced with the unavoidable increasing of the variability in the series, consequence of the use of sample estimates for the self-employee component.

The document is organized as follows: paragraph 2 recaps some information about the objective of the indicator, the regulation framework and the STS requirements. Paragraph 3 describes briefly the source of information and the methodology used for estimating of the employee component. Paragraph 4 highlights on the methodological differences between the new and the old sources and new and old methods. Paragraph 5 shows some graphical comparisons between old and new time series, highlighting the improvements reached by the new method and reflecting what expected from the methodological point of view. Finally, paragraph 6 gives conclusions, describes critical aspects and suggests enhancements.

2. Requirements and definitions for STS variable 210 - Number of persons employed.

As written on the Eurostat web site, “...*STS indicators are tools for formulating and monitoring the economic and monetary policy of the European Union and the euro area... STS data are in great demand for economic analysis in the European Commission and European Central Bank (ECB), national governments and central banks, companies and financial market*”¹. The importance of these indicators at European level has encouraged Istat to improve the methodology and remove the confidentiality status, still characterizing the delivery of the Number of persons employed indicator. Until February 2019, even if the Italian indicator was used by Eurostat to calculate the European countries’ aggregated indicators for all the levels of detail of the economic activity classification required, it was not published. The indicator is now available in the EUROIND database². The STS regulation defines the statistical requirements of the variable 210: Number of persons employed, not only in terms of coverage of the population observed but also in terms of levels of detail by economic activity classification, timeliness and time interval of the series. The classifications used are the Italian version of the European Nace rev.2 nomenclature, that is the Ateco 2007 classification³, and the MIGs (Main Industrial Groupings) classification. In particular indicators are organized in four annexes (Industry without construction, Construction, Retail Trade, Services) and for each one different level of details are required, broken down until 2-digit Nace level. The target population are the persons employed, made by employees and self-employees, for which a regular contract exists between worker and employer. Fixed base indices are produced, on a quarterly basis, starting from the 1st quarter of 2000, and released 60 days after the end of the reference period. For each release, also revised data on the previous quarters are given following a defined revision policy.

3. Employees estimates: source and methodology

The employee component of the indicator is approximated by the Number of jobs produced in the context of the Oros⁴ survey, a process where administrative data

¹ <https://ec.europa.eu/eurostat/web/short-term-business-statistics/overview/sts-in-brief>

² at the link <https://ec.europa.eu/eurostat/web/euro-indicators/> in the *industry, trade and services/labour input* section

³ https://www.istat.it/it/files//2011/03/metenorme09_40classificazione_attivita_economiche_2007.pdf

⁴ Oros is the acronym for Occupazione, Retribuzioni e Oneri Sociali. For more information about the survey see Istat 2019, and for what concerns the number of jobs’ estimation see the work LATTANZIO et al 2019

related to the population of Italian firms with employees, acquired by the Italian National Social Security Institute (Inps), are integrated with survey data related to the Large Enterprises subpopulation (500+ employees) that come from an ad-hoc survey. The Oros process has also in charge the compilation of the STS indicator on the total wages and salaries.

4. Self-Employees estimates: old and new sources and methodology

4.1. The self-employee annual source from the National Accounts

The main source used to estimate the self-employee component is the annual number of self-employees calculated in the context of the National Accounts (NA). An ad-hoc elaboration prepared by NA experts is used in order to extract only the regular component of the self-employment, as required. These data, calculated only for the STS purposes, are used as a benchmark, as in the past, since they represent the best match with the statistical and timeliness STS requirements⁵. Actually, the main source used in NA for the calculation of the number of self-employees is the Business Register (BR), from which the Nace classification is also derived. In particular, say a the year of the release, the BR is used to estimate year $a-2$, while data for $a-1$ are estimated using other sources, as LFS data and administrative data from Inps, used also in order to integrate little typologies not covered by the BR. Data are available for the 2-digit level of detail of the Nace rev.2 classification, for the B to N sectors. At this level of detail, however, only annual data are available, so a procedure of quarterly disaggregation has to be performed.

4.2. The old temporal and sectorial disaggregation method

In the previous method used to estimate the number of self-employees, the sectorial disaggregation was not necessary and only a quarterly temporal disaggregation of the annual time series was performed, since data on self-employees were calculated on a 2-digit Nace level base. Until the delivery of February 2018, the procedure for the temporal disaggregation applied was very simple, aimed at preserving the quarterly dynamic of the indicator series chosen, that was the number of jobs. Let NPE denote the total number of employees, E denote the number of employees and SE denote the number of self-employees. Say $SE_{j,a}^{NA}$ the NA annual estimates of self-employees related to year a and sector j , say E_{j,t_a} the Oros quarterly estimates of Jobs related to the quarter t_a of year a and sector j . The quarterly estimates of the self-employees, \hat{SE}_{j,t_a} , was obtained in the following way:

⁵ Data on regular self-employment are available also from the Business Register but with a higher time delay.

$$\hat{S}E_{j,t_a} = SE_{j,a}^{NA} \frac{E_{j,t_a}}{\frac{\sum_{t_a=1}^4 E_{j,t_a}}{4}} = \alpha_{j,a} \cdot E_{j,t_a} \cdot SE_{j,a}^{NA}$$

where $\alpha_{j,a} = \frac{4}{\sum_{t_a=1}^4 E_{j,t_a}}$. Finally, estimates on the number of persons employed

$N\hat{P}E_{j,t}$ were obtained simply adding the number of jobs to the estimate of the number of self-employees. The most critical aspect of this method is that the quarterly dynamic of self-employees, and also that of the total number of persons employed, were strongly affected by that of the employees. This fact can be easily proved calculating analytically the dependence between the number of persons employed year-on-year growth rate ($V(NPE)_{t,t-4}$) and quarter-on-quarter growth rate ($V(NPE)_{t,t-1}$) and the self-employees and employees growth rate ($V(E)$). This can be done using the same analytical tools used in LATTANZIO 2019⁶:

$$V(NPE)_{t,t-1} = V(E)_{t,t-1} \quad (\text{for } t = Q2, Q3, Q4) \quad (1)$$

$$V(NPE)_{t,t-4} \approx V(E)_{t,t-4} + V(1 + \alpha_a \cdot SE_a^{NA})_{a,a-1} \quad (2)$$

Considering the results of the decomposition formula, the dynamic of the employee part on the total number of persons employed is of a great impact. This was the first reason that induced Istat to study a new method and exploit a new source in order to acquire information about the dynamic of the self-employee component.

4.3. The new additional source for the temporal disaggregation

Data related to the quarterly number of self-employees are calculated in the context of the Labour Force Survey (LFS), according to a two-stage sampling (first municipalities, then households) survey conducted on households. The sample is rotated every week and the calibration estimator is used to calculate quarterly estimates, available at 2-digit Nace level⁷. The main difference with the NA self-employees is that LFS register *heads* and not *jobs*. Furthermore, it is important to say that the Nace code is communicated by the person interviewed and translated by the interviewer, so an acceptable quality of this variable is achieved only at a high

⁶For the sake of simplicity the proof of these results is not presented but can be found in the reference paper.

⁷ For more information about the survey and this source see Istat 2006.

⁸ Note that, although LFS data are available at 2-digit NACE level, they were not used to estimate directly the quarterly number of self-employees due to the issues described in paragraph 4.3 and the difference with respect to the NA source.

aggregated level. Last, but not for importance, LFS includes also the non-regular employment.

4.4. The new temporal and sectorial disaggregation method

The exploitation of the time series of the Number of self-employees calculated in the context of the LFS survey, allows to consider information about the quarterly dynamic of this population. The new methodology used to estimate the self-employee part of the indicator can be summarized in two steps:

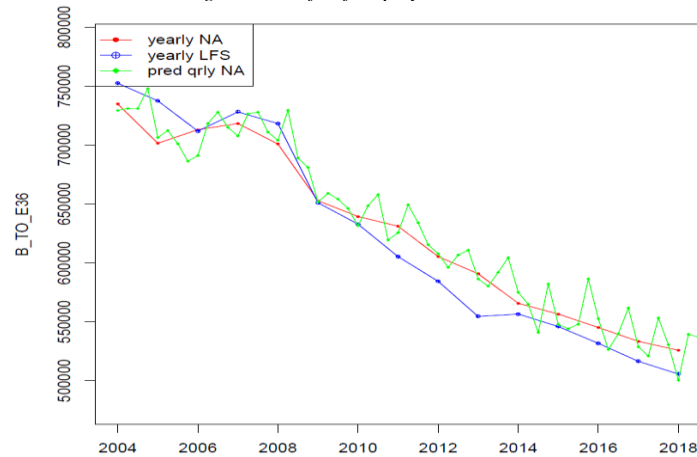
- 1st step - Temporal disaggregation

The first step is the temporal disaggregation of the NA annual time series of the Number of self-employees. For this purpose the Chow-Lin regression method (see the Appendix) is applied using the generalized R-package for temporal disaggregation “tempdisagg”⁹. The methodology, ordinarily used also at Istat in the context of the Quarterly National Accounts, is performed on data aggregated at a macro Nace level of detail, in order to ensure good estimates of the parameters and for which a linear relation can be resumed between the annual time series compared (NA and annualized LFS). More methodological details on the method are presented in the appendix of this work. The method is applied on data aggregated in 5 macro sectors. For the sake of simplicity, results regarding only the macro sector Industry without construction (indicated with the label B_TO_E36 in the graphics) are presented, showing a graphical comparison between the annual series and the predicted one (Figure 1), and output statistics obtained with the R package (Figure 2).

⁹ <https://cran.r-project.org/web/packages/tempdisagg/tempdisagg.pdf>.

Figure 1 – Comparison between NA annual number of self-employees (red line), LFS annualized number of self-employees (blue line) and the results of the quarterly disaggregation (green line) for the macro sector Industry without construction (B_TO_E36).

Period 2004-2018, average number of self-employees-



Source: Istat – Elaboration of LFS survey data and NA data.

Figure 2 – Output statistics of the R “tempdisagg” package for the application of the Chow-Lin regression method - sector B_TO_E36

```
fit statistics ~ se.lfs aggregate B_TO_E36
Call:
td(formula = NA[, k] ~ 1 + LFS[, k], conversion = average)

Residuals:
    Min       1Q   Median       3Q      Max
-13306.7  -7216.8  -654.9   8653.1  15632.0

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.230e+05  2.700e+04  4.558 0.000658 ***
LFS[, k]    8.173e-01  4.272e-02  19.133 2.34e-10 ***

'chow-lin-maxlog' disaggregation with 'average' conversion
14 low-freq. obs. converted to 57 high-freq. obs.
Adjusted R-squared: 0.9656  AR1-Parameter: 0.7081
```

Source: Istat – Elaboration of LFS survey data and NA data.

In the first figure a high correlation between the annual series of the LFS and NA number of self-employees can be seen, same results are obtained from the output statistics looking at the goodness of fit for the parameters of the regression, the high adjusted R-squared and the estimate of the autoregressive parameter.

- 2nd step - Sectorial disaggregation

The second step consists in splitting the estimates of the number of self-employees from the macro aggregation to the 2-digit Nace level detail, as required by the regulation. A procedure of sectorial disaggregation is then performed, that can be formulated as follows. Say $SE_{j,t}^{LFS}$ the LFS number of self-employees observed at quarter t in the Nace division j , $SE_{S,t}^{LFS}$ the same number for the macro sector S , then the final quarterly estimate of the number of self-employees, $\hat{SE}_{j,t}$, is obtained as:

$$\hat{SE}_{j,t} = \hat{SE}_{S,t} \frac{SE_{j,t}^{LFS}}{SE_{S,t}^{LFS}} \quad (3)$$

This method is aimed to preserve the LFS quarterly sectorial structure, represented by the ratio estimate $\frac{SE_{j,t}^{LFS}}{SE_{S,t}^{LFS}}$, that is applied to the quarterly estimate derived from the regression model, related to the macro sector S . It worth to note that the LFS time series on self-employees were available only starting from the 1st quarter of 2004. For the time interval ranging from the 1st quarter of 2000 to the 4th quarter of 2003 only data on the total number of persons employed and at a higher aggregated level with respect to the 2-digit Nace detail were available from LFS. For this reasons a different disaggregation level of detail was used for the temporal disaggregation and a process of reconciliation with the time series ranging from the 1st quarter of 2004 to the 4th quarter of 2017 applied. Apart from the reconciliation procedure, the temporal and sectorial disaggregation method used for the first time window reflect entirely that used for the other part of the time series and described here. In the following paragraph final results on the number of persons employed will be presented, in comparison with the results of the old method.

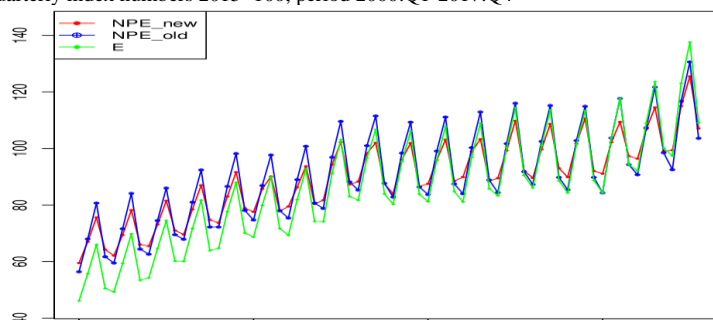
5. Results comparison

In order to show the improvement reached in the short-term dynamic of the indicator, indices and growth rates of the number of persons employed calculated in new (red line) and old (blue line) method were compared with that calculated on the number of employees (green line). In particular in figures 3 and 4 the series for the section I are compared, sector for which the self-employees represent a share of almost 38% on the total number of persons employed, in average on the entire time series, and for which can be observed a strong seasonal pattern for the number of employees. Looking at figure 4, apart from all the first quarters, the pattern is entirely reflected by the total number of persons employed calculated according to the old

method, consequence described in the quarter-on-quarter growth rate formulation in (1). Different is instead the dynamic for the series calculated with the new methodology.

Figure 3 – New (*NPE_new*) and old (*NPE_old*) number of persons employed and number of employees (*E*) in the Nace section I

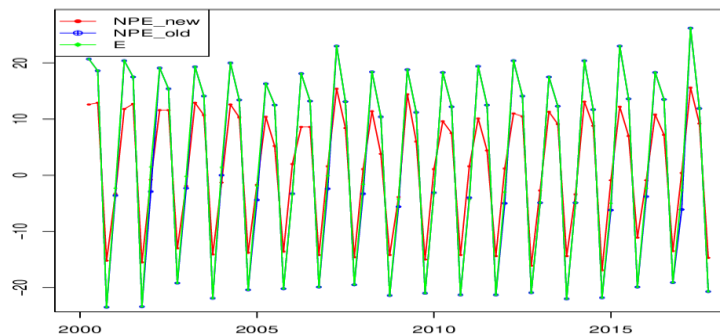
Quarterly index numbers 2015=100, period 2000:Q1-2017:Q4



Source: Istat – Elaboration of Oros and LFS survey data and NA data.

Figure 4 – New (*NPE_new*) and old (*NPE_old*) number of persons employed and number of employees (*E*) in the Nace section I.

q-o-q growth rates, period 2001:Q1-2017:Q4.



Source: Istat – Elaboration of Oros and LFS survey data and NA data.

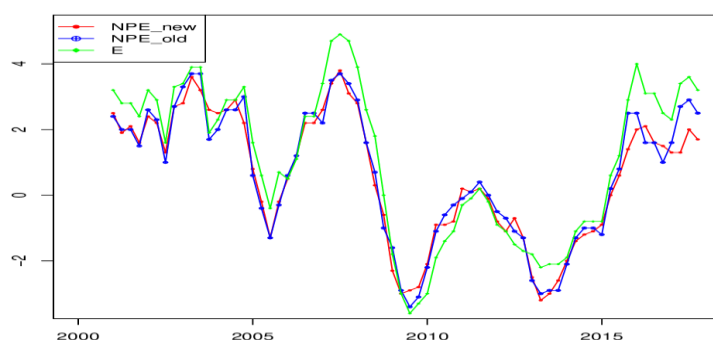
In figure 5 and 6 are shown series of the indices and the related year-on-year growth rates for the total industry and market services economy (indicated with the label B_TO_N in the graphics), for which the share of the self-employees is about 34% on the total number of persons employed, in order to show the improvement reached.

Figure 5 – New (NPE_new) and old (NPE_old) number of persons employed and number of employees (E) in the total economy B_TO_N for STS. Quarterly index numbers 2015=100, period 2000:Q1-2017:Q4.



Source: Istat – Elaboration of Oros and LFS survey data and NA data.

Figure 6 – New (NPE_new) and old (NPE_old) number of persons employed and number of employees (E) in the total economy B_TO_N for STS. y-o-y growth rates, period 2001:Q1-2017:Q4.



Source: Istat – Elaboration of Oros and LFS survey data and NA data.

6. Conclusions

The methodological improvement in the estimation of the number of persons employed has effects not only in terms of accuracy of the indicator produced but also in the use of a more statistically rigorous approach. Furthermore, it represents a case of exploitation of different administrative and statistical sources and an attempt of their integration in order to obtain new statistical information. An accurate analysis of the revisions implied by the new methodology and by the use of sources subjected

to revisions, such as NA data and employee Oros data, was subsequently performed and results compared with the same obtained analyzing Euro aggregate time series and those of the European countries. The results encouraged the release of the indicator on the number of persons employed as free instead of confidential, as flagged until the delivery of February 2019¹⁰. To conclude, some still critical aspects in the method remain and are object of study for possible future enhancements. In particular, the main aspects to take a closer look at are:

- NA data are subject to important revision, at Nace level, observed once a year and that affects totally the estimation process causing revision on the entire time series, through the modification of the temporal disaggregation's parameters. An attempt in order to have a lower impact could be to aggregate the data at a different level of Nace detail for which good estimate of the parameters in the regression model and lower revision of NA data are observed.
- An important aspect is the management of the accuracy, accounting also for the variability induced by the 2-digit Nace level LFS data in the sectorial disaggregation method described in (3). Another improvement could be the formulation and estimation of the dependence of the total variability with that of the LFS data, for which can be used the same approach described in LATTANZIO et al. 2017, including that induced by the temporal disaggregation.
- LFS data on self-employees contain also the non-regular component. An evaluation of the impact of this component should be done, in order to support an eventual procedure of decomposition.
- The NA annual number of self-employees covers also the outworker component of the self-employment, that, in order to satisfy the STS definition, should be excluded.

¹⁰ Results of this revisions analysis are not described in this work but a documentation is available at request.

Appendix

The Chow-Lin regression approach used for temporal disaggregation

The method (see Chow and Lin - 1971) consists in fitting a linear regression model between the aggregated quarterly series y , to be estimated, and the indicator series x , available, supposing the residuals to follow an autoregressive process at lag 1¹¹

$$y_t = \alpha + \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

Since the quarterly values for the aggregated series are not available, the model is fit on the annualized data that are, instead, available both for the aggregated and the indicator series. This implies that the linear relation observed on the annualized data is reproduced on the quarterly disaggregated data. The regression is made using the GLS (Generalized Least Squared) regression approach. The parameters to be estimated are α, β and the auto-correlation parameter ρ . In fact the variance-covariance matrix has a known structure and depends only on this value, that is calculated using the maximum log-likelihood of the GLS regression method (see BOURNAY and LAROQUE 1979). This is the form of the variance-covariance matrix.

$$\Sigma_{CL}(\rho) = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \begin{bmatrix} 1 & \dots & \rho^{n-1} \\ \vdots & \ddots & \vdots \\ \rho^{n-1} & \dots & 1 \end{bmatrix}$$

Riferimenti bibliografici

CHOW G., LIN A. 1971. Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series, *The Review of Economics and Statistics*, No. 53, pp. 372-375.

BOURNAY J., LAROQUE G. 1979. Reflexions sur le methode d'elaboration des comptes trimestriels, *Annales de L'Insee*, Vol. 36, pp. 3-29.

ISTAT 2006. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Roma: Collana Metodi e Norme, Vol. 32.

SAX C., STEINER P. 2013. Temporal Disaggregation of Time Series, *The R Journal*, Vol. 5(2), pp. 80-87.

¹¹ Other versions of the Chow-Lin regression approach, such as the version by Litterman and that by Fernandez, propose different hypothesis on the distribution of the residuals, that is respectively a non-stationary process and a random walk.

ISTAT. AA.VV. 2015. Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione, *Istat Working Papers*, No. 19.

LATTANZIO M. et al. 2017. Quality improvements in variance estimation for the labour force survey, *Rivista SIEDS*, Vol. 71, No. 4.

ISTAT. 2019. *La rilevazione trimestrale Oros su occupazione e costo del lavoro: indicatori e metodologie*. Roma: Collana Metodi.

LATTANZIO M., POGELLI F.R., TUZI D. 2019. L'imputazione delle mancate risposte sulle posizioni lavorative dipendenti nei dati provvisori di fonte Inps, *Istat Working Papers*, No. 4.

LATTANZIO, M. 2019. Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella rilevazione Oros, *Istat Working Papers* (submitted to the review).

SUMMARY

The new methodology for the Italian STS indicator on the “Number of Persons Employed”

Since May 2018 Istat delivers to Eurostat quarterly business based indicators on the Number of Persons Employed, required by the EC-Regulation on Short-Term Statistics (STS), according to a new compilation methodology. The innovation concerns the calculation of the self-employed component. While for the employee part quarterly figures are calculated using data from the Istat Oros survey, for the self-employed part official statistics do not provide estimates at the frequency and business detail level required by the Regulation. Actually, figures on self-employed persons are available only at annual level from the National Accounts (NA) or from the Business Register (BR). For timeliness reasons, NA data were preferred. The approach at the basis consists in the quarterly disaggregation of the annual source. A simple disaggregation method was used also in the past methodology, as described in the paper, but the results were poorly meaningful, particularly in sectors with a significant self-employment influence. The new methodology includes two major improvements: 1) the exploitation of a new source for the estimation of the quarterly dynamic of self-employees, that is the Labour Force Survey (LFS) number of self-employed persons, used to temporally disaggregate the NA annual number of self-employees; 2) a new method for the temporal disaggregation, based on a mixed use of the Chow-Lin regression model, applied at an aggregated level, and a method aimed at preserving the LFS quarterly structure, for the Nace disaggregation (Eurostat requires data at 2-digit Nace level for industry and services). This work focuses on the innovations introduced, stressing on the differences with the old method showing the improvement in the short-term dynamic. Finally, some still critical aspects are described and possible enhancements suggested.

Marco LATTANZIO, Istat, lattanizio@istat.it