rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 2
2022

In this issue:

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 2
2022

In this issue:

Istat

rivista di statistica ufficiale
n. 2/2022

Four-monthly Journal: registered at the Court of Rome, Italy (N. 339/2007 of 19th July 2007).

e-ISSN 1972-4829
p-ISSN 1828-1982

The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-anonymised approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale.

Editorial Preface

This issue N. 2/2022 of the *Rivista di statistica ufficiale/Review of official statistics* consists of four articles, the first of which is by Luigi Biggeri and Monica Pratesi, based on a scientific report prepared for the *MAKSWELL* Project (MAKing Sustainable development and WELL-being frameworks work for policy analysis). This Project is funded by the European Union's Horizon 2020 Programme: it proposes to extend and harmonise the indicators able to capture the main characteristics of the beyond-GDP approach suggesting a new framework that includes them in the evaluation of the public policies.

This paper has two main objectives: (*i*) verifying the methodological approaches for estimating Sub-National Spatial Price Indices, (*ii*) and how to assess poor-specific Sub-National Spatial Price Indices, without knowing what prices the poor pay for the different goods. The organisation and the results of some experiments conducted for the Italian provinces are illustrated using the Italian National Institute of Statistics – Istat scanner database on retail prices.

The authors' proposal to abandon the approach of close like-to-like comparability of products obtained good results as well as that of using the data from the first quintile of price distributions of each product to estimate poor-specific Sub-National Spatial Price Indices. From this work it clearly emerges that it is appropriate to continue to deepen these themes through further research and experimentation, also at an international level, *e.g.* bringing them to the attention of the National Statistical Institutes of other European countries.

In the following article, Ciro Baldi, Sara Gigante, Silvia Pacini, and Roberta Rizzi show an experimental integration at the micro level of each job position between an employer and an employee in the private non-agricultural sector, enabling the extension of the Labour Register stocks and flows for including those of the Compulsory Communications.

This was possible thanks to the relevant advances in labour statistics due on the one hand to the ongoing Istat's Labour Register Project, consisting of a micro-level employer/employee longitudinal database full of information about employment, hours, wages, social contributions, *etc*. On the other hand, due to the five-party agreement, a synergy aimed at producing integrated statistics, which involves the following Italian bodies: Istat, Ministry of Labour and Social Policies, National Social Security Institute - INPS, National Agency for Active Employment Policies - ANPAL, and National Institute for Insurance against Accidents at Work - INAIL.

This allowed correcting the Labour Register for the absence of very short-duration job positions, and projecting job stocks and flows up to the last available date of the Compulsory Communications, thus enhancing the results on the labour dynamics. The integration methodological details were also analysed in-depth to highlight strengths and weaknesses.

In the third article, Solange Leproux, Adriano Pareto, and Claudia Rinaldelli describe the Istat Economic Sentiment Indicator – IESI, suggesting a new method for calculating both the sectoral confidence climates and the IESI, thus ensuring consistency in the evolution of the indicators and guaranteeing an easy interpretation of the results.

The IESI is a measure of the confidence climate in the Italian production sector and represents the result of the aggregation of the variables used in the calculation of the confidence climate indices of manufacturing, construction, service, and retail trade sectors.

The current procedure adopted for the calculation of the IESI can determine discrepancies between the evolution of the composite index and the dynamics of the sector-level confidence climates. Although these differences are explicable from a methodological point of view, they can create substantial problems in terms of the interpretation and assessment of the results.

For this reason, the authors processed a new IESI, while preserving the essence of the existing methodologies and giving an easy explanation and communication of the confidence state both at a sectoral and total level. They also presented the new IESI results in terms of the ability to capture fluctuations in aggregate economic activity and carried out the performance tests in tracking the reference series, using the current version of the indicator.

Finally, this issue closes with an article by Diego Zardetto, Marco Di Zio, and Marco Fortini, who propose to apply methods usually adopted inside National Statistical Institutes for balancing large systems of national accounts to address the population’s stock and flow consistency problem in the integrated estimation system formed by the Italian Permanent Census and the Base Register of Individuals. This paper faces the aspects of Istat's modernisation programme related to the new way of producing data using statistical registries. By integrating survey and administrative information, these registries are certainly suited to meet the need of releasing timely, reliable, and coherent estimates of demographic stocks and flows. However, these data are currently obtained independently: the integrated estimation system provides the population size, and the municipal registries instead deal with births, deaths, and migration flows.

The authors designed and implemented a system that can handle this situation by obtaining good simulation results, which show that their balancing approach determines improved estimates of population counts. This can allow Istat to achieve stock and flow consistency, also for all the subnational estimates that need to be officially disseminated in compliance with Italian and European Regulations.

Patrizia Cacioli
Editor

Nadia Mignolli
Coordinator of the Editorial Board

Estimation of local Spatial Price Indices using scanner data: methods and experiments applied also for assessing poor-specific indices

Luigi Biggeri[1], *Monica Pratesi*[2]

Abstract

This paper presents the organisation and the results of some experiments conducted to estimate Sub-National Spatial Price Indices (SN-SPIs) for the Italian provinces, by using scanner data on retail prices. The objective is twofold: to verify which are the appropriate methodological approaches to estimate SN-SPIs and whether it is possible to estimate poor-specific SN-SPIs, without knowing what prices the poor pay for the various products. Our proposal to abandon the approach of close like-to-like comparability of products results valid and the use of the data from the first quintile of price distributions of each product to estimate poor-specific SN-SPIs is surely interesting. Further research and experiments are needed, which should be conducted also by the National Statistical Institutes of other European countries.

Keywords: Scanner data on retail purchases, sub-national Spatial Consumer Price Indices, poor-specific Spatial Consumer Price Indices, country product dummy model, GEKS method.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_2.2022.01

1 University of Firenze and Camilo Dagum Centre on Advanced Statistics for the Equitable and Sustainable Development – ASESD, luigi.biggeri@unifi.it.

2 University of Pisa and Camilo Dagum Centre on Advanced Statistics for the Equitable and Sustainable Development – ASESD, monica.pratesi@unipi.it.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

This paper builds based on a scientific report prepared for the MAKSWELL Project[3]. It refers to the first part of the report moving forward from the results obtained for the estimation of Sub-National Spatial Consumer Price Indices (SN-SCPIs) for each Italian province, by using scanner data on retail prices[4].

The objective is twofold: first, to verify which are the appropriate methodological approaches to estimate SN-SCPIs and, second, to verify whether it is possible to estimate poor-specific SN-SCPIs, without knowing what prices the poor pay for the various products, and if it is possible to track poor-specific consumption behaviour.

The availability of high-frequency scanner data is surely useful, in addition to traditional sources of price data, to compute the SN-SCPIs.

According to the experiments conducted by various National Statistical Institutes, particularly in Italy, the use of such data has certain advantages but also some limitations which will be summarised in Section 2.

In any case, for the computation and use of the SN-SCPIs, there are two main problems, which arise also by using scanner data, that need to be addressed and possibly solved (ICP-World Bank, 2021).

The first problem refers to the level of comparability of products among the different subnational areas to be achieved. The more tightly the products are defined, the more difficult it becomes to find products meeting the specifications in all the subnational areas. Therefore, to increase the number of comparable products that have to be also representative of each area, it can be appropriate to use loosening specifications of products (Biggeri and Rao, 2021, pag.19).

The second problem depends on the main use of the SN-SCPIs. To provide adequate information regarding the real incidence, nature, and extent of

3 Deliverable 3.2 of MAKSWELL project. MAKSWELL project (MAKing Sustainable development and WELL-being frameworks work for policy analysis), funded by the European Union’s Horizon 2020 Programme, proposes to extend and harmonise the indicators able to capture the main characteristics of the beyond-GDP approach suggesting a new framework that includes them in the evaluation of the public policies.
The authors of this deliverbale are: Pratesi, M., C. Giusti, S. Marchetti, L. Biggeri, G. Bertarelli, F. Schirripa Spagnolo, T. Laureti, I. Benedetti, F. Polidoro, F. Di Leo, and M. Fedeli.

4 We have adequately modified the content of various sections especially for the comment of the results obtained and written a new section for the estimation of poor-specific spatial price indices.

economic poverty, it is necessary to use and therefore compute poor-specific SN-SCPIs (Biggeri and Rao, 2021, pp.13-14).

This paper presents the organisation and the results of some experiments conducted to address the issues and achieve the two objectives mentioned above.

The first experiment refers to the general computation of the SN-SCPIs conducted by using the same scanner database and elementary data but applying a different principle of product comparability. One is computed considering the comparison of like-to-like products for the different subnational areas (World Bank approach). The other considers the principle of comparability at the level of very detailed groups of products, by loosening the “tight” specification of the elementary products given that the products of each group satisfy the same consumer needs (that we called the ASESD approach, considering that the idea was born by ASESD Centre’s researchers[5]).

For the computation of poor-specific SN-SCPIs - that is to calculate SPIs closer to the prices paid by the poor - we conducted a second experiment by using price values of the first quintile of the price distribution of each product, assuming for the time being that the poor purchase the less expensive items of a product and that these have prices whose values are below the first quintile of the price distribution.

However, to corroborate this last assumption, it would be important to have some information on the consumption behaviour of poor households. To this aim, we report here some results of a specific tentative survey and analysis on where people in a condition of absolute poverty purchase some large consumption products. This information is collected by the Italian Households Expenditure Survey and the analyses have been conducted by researchers of the Unit of Price Statistics for the MAKSWELL project.

The paper is organised as follows. In Section 2, the database is presented,

5 The ASEDS is a Tuscan Universities Research Centre (also called Camilo Dagum Centre) on Advanced Statistics for the Equitable and Sustainable Development-ASESD, which has 61 members.

highlighting the advantages and the limitations of the scanner data on prices for the computation of the SN-SCPIs. In Section 3, the procedure, the methodology, and the main results of the estimation of the indexes obtained by using the World Bank and the ASESD approaches are presented. As the results obtained by the two approaches are quite different, we propose, in Section 4, some explanation of the differences and an analysis of their internal consistency. In Section 5, the methodology and the results of the poor-specific SN-SCPIs computation are presented. Section 6 is dedicated to the main results of the analysis of the places where people in a condition of absolute poverty purchase some large consumption products. Final remarks and recommendations in Section 7 conclude the paper.

2. The available scanner database: advantages and limitations for the computation of Sub-National Spatial Consumer Price Indices

The Italian National Institute of Statistics - Istat has over ten years of experience in the usage of the scanner data on retail prices for constructing both CPIs (Consumer Price Indices) and SN-SCPIs (Sub-National Spatial Consumer Prices Indices). Scanner data is the scanning of bar codes at checkout lines of retail stores.

Since 2014, Istat has been receiving the scanner data on the retail prices from the market research company ACNielsen which is authorised to do it by the chains of modern distribution in the framework of an agreement with the Association of Modern Distribution. ACNielsen provides Istat with scanner data every week by uploading the data files on a dedicated Istat web portal.

Recently, Istat has introduced the scanner data in the computation of the official CPIs, while until now the data have been used in the construction of the SN-SCPIs only in an experimental way (Laureti and Polidoro, 2021).

These data benefit from an impressive coverage of transactions along with information on sales, expenditure, quantities, and quality with very detailed information on characteristics of products sold (brand, size, and type of outlet) provided at barcode level or, more precisely, at the GTIN (Global Trade Item Number) code. The scanner data of the modern distribution provide millions of prices for thousands of products identified by the GTIN code. They predominantly refer to supermarkets and hypermarkets, especially for food, beverages, and personal and home care products. After a process of data cleaning and trimming outliers, the unit value price per item code can be computed by dividing the total turnover for that item by the total quantity sold.

In October 2018, an agreement between Istat and ASEDS-Dagum Centre was signed to implement the tasks of the MAKSWELL project, and, for this purpose, Istat provided the Centre with the scanner database referred to the years 2017 and 2018. The database is a random sample of approximately 1,800 outlets, hypermarkets (more than 500), and supermarkets (almost 1,300), and contains data concerning food and grocery products sold in the most important retail chains (95% of modern retail chain distribution that covers 55.4% of total retail trade distribution for this category of products). More specifically, scanner data are obtained from 1,781 outlets of the main 16

Retail Trade Chains (RTCs), thus covering the process of the entire national territory. Outlets have been stratified according to provinces (107), chain distribution (16), and outlet types (hypermarket, supermarket), for a total of more than 800 strata. Probabilities of selection were assigned to each outlet based on the corresponding turnover value. For each GTIN, prices were calculated taking into account turnover and quantities: the weekly unit value price is equal to the weekly turnover divided by weekly quantities. Monthly and annual unit value prices are calculated by the arithmetic mean of weekly prices weighted with quantities.

Many authors have contributed with papers and reports on how scanner data can improve price measurements, highlighting the advantages and some difficulties in using data that allow for the simultaneous collection of price and quantity information (for a review of the literature see Feenstra and Shapiro, 2003); Biggeri and Rao, 2021; Laureti and Polidoro, 2022). Taking also into account the results of the experiments and the many discussions among the members of ASESD-Dagum Centre and the researchers of Istat's Price Statistics Unit, we can summarise here the various advantages and some limitations in using the scanner data for the computation of the SN-SCPIs.

The main advantage is that scanner data may help to overcome the issue of price data availability in the various areas involved in the comparisons by fulfilling the requirements of representativeness and comparability that emerge when compiling SCPIs. Due to the high territorial coverage which characterises scanner data, we can compare price levels among the various geographical areas within a country at a very detailed territorial level (provinces). In addition, it is worth noting that GTIN codes describe the products in detail and they are generally the same for each item at the national level. In this way, we can solve the issue of comparability. Since detailed information on turnover and quantities for each item code in every area is available, it is possible to account for the economic importance of each item in its market, thus fulfilling the representativeness requirement. Moreover, as different modern RTCs can sell products of different quality and offer additional services, information on the type of outlet and retail chain can be included in order to account for these quality characteristics that may influence the price of a product. Moreover, the availability of turnover weights (defined considering also sampling weights, when necessary) allows to correctly include the corresponding representativeness of the products in terms of the total turnover of the group to which the products in question belong (Basic Headings, in the World Bank approach).

Other advantages of the use of scanner data are: (*i*) the reduction of measurement errors. By using the unit value for each GTIN as a price concept we can refer to a more accurate measure of an average transaction price than an isolated price quotation as in the case of traditional price data collection (see Diewert, 1995). (*ii*) The reduction of conceptual uncertainty. The GTIN unit value is a more representative price over the reference period than the usual price collected using traditional on-field surveys. These prices include temporary price promotions and reflect the actual price paid by consumers. Moreover, by aggregating over a year it is possible to smooth out the effect of price and quantity bouncing behaviour. Using scanner data, (*iii*) we add a time dimension to multilateral spatial price comparisons since detailed data are usually available at the point of sale and at the time of transaction. Another advantage (*iv*) is the use of itemised information contained in scanner data. When using the unit value approach, indeed, items must be tightly defined at a fine level of aggregation to maximise homogeneity and prevent quality differences from affecting the unit values. Finally, (*v*) it is obvious that using scanner data to carry out spatial comparisons will increase cost efficiency, since price data collection may be limited to traditional stores and shops, thus lowering data collection costs for the National Statistical Institutes.

However, some limitations should be taken into account in the context of this study and the computation of complete SN-SCPIs. The available scanner data: (*i*) do not cover all the retail chains of modern distribution (95%); (*ii*) cover all the 107 Italian Provinces in 2018, but the small size of some newly established provinces allows some estimations only for 103 out of 107 provinces and the rural areas are not completely covered; (*iii*) cannot be used for perishables and seasonal products such as vegetables, fruit and meat, and fresh fish, since these products are sold at price per quantity and generally are not pre-packaged with GTIN codes.

Moreover, we have to consider that, as already said, all the scanner data available cover about 10.5% in terms of the total expenditures of families for consumption (Istat, 2020). In addition, this share is not uniform across the Italian territory.

Therefore, it is evident that to estimate a complete system of SN-SCPIs it is necessary to build up a database that could allow the estimation of these indices related to the entire universe of household consumption.

Istat indeed collects consumer prices by using different data sources: territorial surveys at the outlets by non-probability samples, administrative data, and scanner data (selected by probability sampling). However, a strategy to use and integrate all the data sources is still missing, considering also the fact that the data come both from probability and non-probability samples, as already discussed in a paper presented at a Workshop on non-probability samples held at the University of Trier (Biggeri *et al.*, 2020). Istat is working on this line.

3. The estimation of Spatial Consumer Price Indices by using two different approaches

The available and usable scanner data provided to the Dagum Centre refer mainly to food and grocery products concerning the categories and sub-classes of the ECOICOP - European Classification of Individual Consumption: they are 63 and the total number of products is 87,545. The annual price quotes are 2,032,574.

The sub-classes included in Food and non-alcoholic beverages are 46 (from sub-class 01.1.1.0 to sub-class 01.2.2.3.0). The sub-classes included in Alcoholic beverages are 7 and the sub-classes included in the other non-food categories are very few (only 10) and not enough to be representative of all the categories belonging to non-Food classes. Therefore, we decided to present our experiments only on the Food subclasses (which include also non-alcoholic beverages)

The analysis was conducted on the Italian provinces using the 2018 scanner database.

The two approaches apply different methods to satisfy the principle of product comparability. Methodologies, procedures, and results are presented in the next sub-sections.

3.1 World Bank approach

In this experiment, the principles and the construction procedure are quite similar to the one used in the ICP (International Comparison Programme) of the World Bank to compute the international PPPs (Purchasing Power Parities). In particular, the principle of comparability is applied in a very tight way by considering the comparisons of the like-to-like items (products) for the different sub-national areas (provinces, in our experiments). Under this approach, the lowest level of aggregation of the products is the so-called Basic Heading (BH) level, as defined by the World Bank (World Bank, 2013). Here we consider the 46 sub-classes belonging to the food category as BHs.

Within this approach, there is the risk that not all the products are available in all areas.

However, because the 2018 scanner dataset includes all the products identified by the corresponding GTIN, we can include in the comparisons also those products acquired by consumers with reduced quantities. The availability of turnover weights (defined considering also sampling weights) allows including the corresponding representativeness of these products correctly. For this reason, we provisionally call the computed SPIs as PPPs.

To improve the quality of price comparisons, defined by the strength of interconnections and overlaps in the priced items across different provinces, the following group of products was excluded since in these cases price data does not exhibit a spatial chain: whole milk, low-fat milk, olive oil, aged cheese, other cheese, lager beer, and frozen seafood.

As far as the methods for the PPPs computation are concerned, a two-step procedure is adopted (World Bank Group, 2015). In the first step, provincial PPPs are computed at the BH level using the Country Product Dummy (CPD) model by comparing price and quantity data referring to products sold in the various Italian provinces, while in the second step, we aggregate the results from the BH level comparisons to the higher level aggregate (Food category), using the GEKS procedure based on Fisher indices.

3.1.1 Aggregation method at BH level: BH PPPs

As underlined above, scanner data bring detailed information about the characteristics of the elementary product and information about the turnover of that specific product, allowing the comparison of like-to-like products. Weights for each specific product are based on turnover for that product.

For each GTIN, weight is obtained by dividing the total weighted turnover by the weighted turnover for that product for each province.

Since product overlaps exhibit a chain structure, the weighted CPD method exhibits some aspects of spatial chaining and therefore we selected this method for computed provincial PPPs for product aggregates. Aiming at taking into account the economic importance (representativeness) of each product expressed by expenditure weights w_{ijr} based on turnover, we used a weighted CPD model. In this way, the representativeness requirement can be achieved by computed weighted spatial index numbers.

Let's assume that we are attempting to make a spatial comparison of prices between *Mr* provinces, with $r = 1, 2, \ldots, R$ Regions. In the first stage of aggregation of price data at the item level, which leads to price comparisons at the BH level, p_{ij} and q_{ij} represent the price and quantity of $i-th$ item in $j-th$ province $i = 1, 2, \ldots, N; j = 1, 2, \ldots, Mr$. To compute provincial PPPs, we used the CPD model according to the approach followed by the World Bank.

Besides accounting for quality variations in the cross-area price data, CPD is a regression-based econometric methodology that can be extended and generalised to provide a comprehensive framework for carrying out both international and intra-national analyses. The literature is still expanding and a recent paper by Rao and Hajargasht (2016) further developed the CPD-based stochastic approach through the use of modern econometric tools. This method suggests that price levels are estimated by regressing logarithms of prices on provinces for each province and product dummy variables; the model is given for each BH by:

$$\begin{aligned} lnp_{ij} &= lnPPP_j + lnPPP_i + ln\mu_{ijr} \\ &= \pi_j + \eta_i + v_{ijr} \\ lnp_{ij} &= \sum_{j=1}^{M_r} \pi_j D^j + \sum_{i=1}^{n} \eta_i D^i + v_{ijr} \end{aligned} \tag{1}$$

where D^j is a provincial-dummy variable that takes value equal to 1 if the price observation is from the $j-th$ province and 0 otherwise, and D^i is a commodity dummy variable that takes value equal to 1 if the price observation is for the $i-th$ commodity and 0 otherwise and v_{ijr} are normally distributed errors with zero mean and constant variance σ^2. Parameters of this kind of model can be estimated once one of the parameters of the model is set at a specific value (Suits, 1984; Laureti and Rao, 2018). For example, if province 1 is taken as the reference or numerator province, then π_1 is set at zero and the remaining parameters are estimated. To estimate the parameters of this model, we impose normalisation $\sum_i^{Mr} \pi_j = 0$, thus symmetrically treating all provinces. If $\hat{\pi}_j = (1, 2, \ldots, Mr)$ are estimated parameters, PPP for the province j in region r is given by $PPP_j = e^{\hat{\pi}_j}$. The CPD method-based price comparisons are transitive and base-invariant.

Aiming at taking into account the economic importance (representativeness) of each product, expressed by expenditure weights w_{ij} based on turnover, we used a weighted CPD model, by running weighted least squares on the model (1). The weighted CPD is equivalent to applying ordinary least squares to the following model:

$$\sqrt{w_{ij}}lnp_{ij} = \sum_{j=1}^{M_r} \pi_j \sqrt{w_{ij}} D^j + \sum_{i=1}^{n} \eta_i \sqrt{w_{ij}} D^i + \sqrt{w_{ijr}} v_{ijr} \qquad (2)$$

The assumptions and procedures to obtain the weighted PPP_j are those above explained.

3.1.2 Aggregation above BHs: Provincial PPPs for Food category

The next and final step for compiling provincial price comparisons is to aggregate the results from BH level comparisons to higher level aggregates. Let's assume that there are L basic headings ($l = 1, \ldots, L$) and e^r expenditure for $i - th$ BH in province r. We decided to use the Fisher price index since it has a range of axiomatic and economic theoretic properties. The Fisher index is given by:

$$P_{rk}^{Fisher} = \sqrt{P_{rk}^{Laspeyres} \cdot P_{rk}^{Paasche}} \qquad (3)$$

Where:

$$P_{rk}^{Laspeyres} = \frac{\sum_{l=1}^{L} p_l^k q_l^r}{\sum_{l=1}^{L} p_l^r q_l^r} = \sum s_i^r \left(\frac{p_l^k}{p_l^r}\right) \qquad (4)$$

$$P_{rk}^{Paasche} = \frac{\sum_{l=1}^{N} p_l^k q_l^k}{\sum_{l=1}^{N} p_l^r q_l^k} = \left[\sum_l s_l^k \left(\frac{p_l^k}{p_l^r}\right)^{-1}\right]^{-1} \qquad (5)$$

with:

$$s_i^r = \frac{e_i^r}{\sum_{l=1}^{L} e_l^r} = \frac{p_i^r q_l^r}{\sum_{l=1}^{L} p_i^r q_l^r} \qquad (6)$$

As the Fisher binary index in eq. 3 is not transitive, it is possible to use the GEKS method (World Bank, 2013) to generate transitive multilateral price comparisons across different regions. The resulting index is given by:

$$P_{rk}^{GEKS-FISHER} = \prod_{r=1}^{R} \left[P_{rs}^{Fisher} \cdot P_{sk}^{Fisher}\right]^{1/R} \qquad (7)$$

The GEKS-Fisher-based formula is used in cross-country comparisons made within the ICP at the World Bank Group (2015) and the OECD-Eurostat comparisons. In order to obtain a set of $R\ PPP_s$ that refer to the group of regions (Italy), we standardised the GEKS-Fisher-based PPPs (S-GEKS).

As these PPPs are now transitive, the ratios between the PPPs for each base are the same. To achieve a set of PPPs that has the group of countries as a base – thereby ensuring a neutral presentation - it is necessary to standardise the PPPs in the matrix. This is done by dividing each PPP by the geometric mean of the PPPs in its column.

3.1.3 Results

Following the methodology illustrated in sub-section 3.1.1, we first run a CPD model for each available BH and each province by using weighted turnover.

As expected – having considered identical products in the various provinces - the results obtained show that for many food products, frequently purchased (for example dry pasta and non-alcoholic beverages) there are very small differences in the level of prices among the various provinces. However, the variability of provincial PPPs is different for the different BHs, and in some cases, a certain or high variability emerges. See for example the descriptive statistics reported in Table 3.1 respectively for Coffee, Fresh Pasta, and Eggs BHs, and in Table 3.2 for Food category (Italy=100). The PPPs are reported in Figure 3.1 (a, b, c, d).

Table 3.1 - Descriptive statistics based on provincial PPPs for Coffee, Fresh Pasta, and Eggs (Mean =100) - Year 2018

	Coffee	Fresh Pasta	Eggs
Min	92.69	83.03	75.90
Max	107.24	112.14	114.55
CV	3.32	6.45	7.38

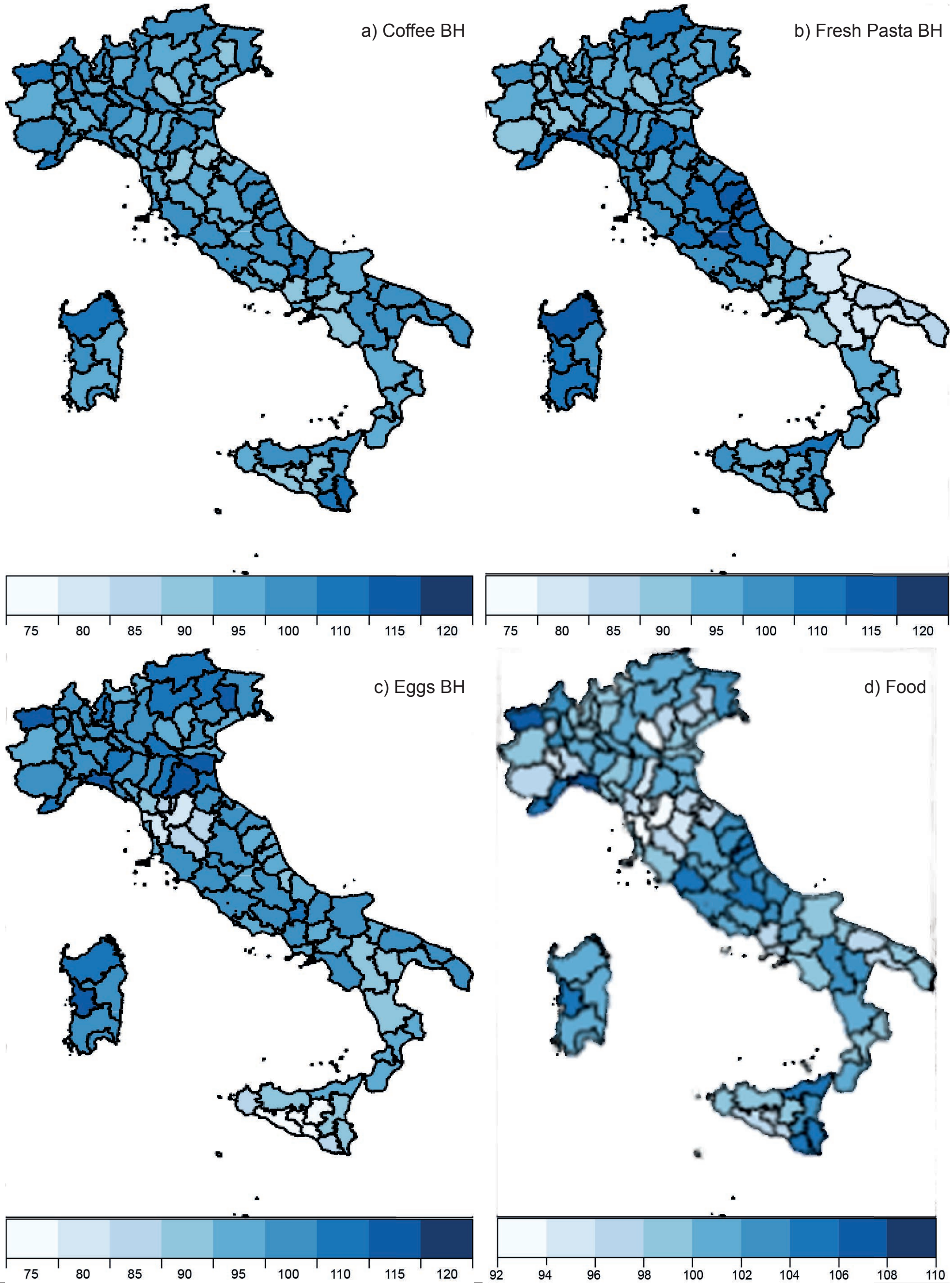
Source: Authors' processing on Istat scanner database

Table 3.2 - Descriptive statistics based on provincial PPPs for Food (Mean =100) - Year 2018

	Food	
Min	92.47	Province: Firenze
Max	107.83	Province: Aosta
CV	3.00	

Source: Authors' processing on Istat scanner database

Figure 3.1 - PPPs at provincial level for certain products and for Food (Italy=100) - Year 2018



Source: Source: Authors' processing on Istat scanner database

The North-South dualism is confirmed only for some BHs. As shown in Figure 3.1, a modest variability or near homogeneity is observed in PPPs for the Coffee BH; in this case, the coefficient of variation across Italian provinces is equal to 3.32%. In the case of Fresh Pasta, for which the coefficient of variation is equal to 6.45%, PPPs in the Northern provinces are generally higher than those in Southern Italy. The less expensive provinces are Matera (83.03), Potenza (83.31) located in the Basilicata region, and Foggia (83.67) located in Puglia, while the most expensive provinces are Genova (112.14) located in the Liguria region, Ascoli Piceno (111.06) and Fermo (111.02) located in Marche. Moreover, interesting results are provided for the Eggs BH, for which a high level of heterogeneity across Italian provinces is observed.

Using the results obtained in the first step, we computed the PPPs for the "Food" aggregate using as a weight the weighted turnover for each BH.

Descriptive statistics are reported in Table 3.2 and the PPPs for "Food" are reported in Figure 3.1d.

From Table 3.2 and Figure 3.4 we cannot observe a high level of price heterogeneity across Italians Provinces for food products. The most expensive provinces are Aosta and Genova (with PPPs equal to 107.83 and 107.5 respectively) located in Northern Italy, while the less expensive provinces are Firenze and Pisa (with PPPs equal to 92.47 and 93.57 respectively) located in Central Italy. This latter result is certainly not expected and maybe it could be partially explained by the prevalent presence of a specific retail chain, in those areas.

As already pointed out, there are various advantages to using scanner data to compute SPIs and PPPs. On the other hand, the like-to-like approach may have some limitations. To use strictly comparable products, indeed, some products may be excluded since they are produced and consumed at the local level. In our application, we had to exclude some BHs due to the insufficient overlap across provinces (*i.e.* whole milk, low-fat milk, olive oil, aged cheese, other cheese, lager beer, and frozen seafood). It is also worth noting that PPP results may be influenced by the characteristics of the modern retail trade which is not uniformly distributed across Italian territory in terms of types of retail chains and market share. In Table 3.3, although data are referred to the year 2016, it is clear that in some

Southern regions the share covered by the retail chains is lower than that observed in the North of Italy. In addition, consumer choice among the different distributional channels may be considered. In Southern regions, consumers tend to buy food products in open markets and traditional shops more frequently than consumers in Northern regions.

Moreover, we have to take into account the influence of the pricing policy adopted by the different commercial chains in the different areas of the country.

Table 3.3 - Scanner data: market shares (hypermarket and supermarket) (% values) – Year 2016

Retail Chains		Coop Italia	Conad	Esse-lunga	Selex commer-ciale	Gruppo Auchan	Gruppo Carrefour Italia S.P.A.	Finiper	Gruppo VéGé	Gruppo SUN	Agorà Network S.C.A.R.L.	Gruppo Pam	Aspiag	Bennet S.P.A.	Sigma	Crai	Despar servizi	**Total**
North-West	Piemonte	18.2	4.3	12.4	17.9	7.0	16.4	1.5		1.4	2.5	3.7		8.7	0.1	1.6		**95.9**
	Valle d'Aosta/ Vallée d'Aoste	-	22.3		8.6	-	45.1	-	-	-	-	-	-	-	-	-	-	**76.0**
	Liguria	42.2	17.0	3.9	4.8	0.7	8.8		1.5	3.2 2.6	13.5	2.7		1.3		0.3		**99.8**
	Lombardia	7.9	3.3	31.3	9.9	8.2	9.9	6.4	1.1		6.1	0.9		5.2	1.1	0.2	0.6	**94.8**
North-East	Trentino-Alto Adige/Südtirol	18.0	13.8	-	-	-	-	-	-	-	34.4	0.6	32.4	-	-	-	-	**99.1**
	Veneto	9.1	3.6	1.2	32.3	6.3	2.1	1.6	6.2	2.0	0.4	3.1	12.7	1.2	2.8	2.6		**87.0**
	Friuli-Venezia Giulia	21.3	7.7	-	9.4	1.1	4.2	2.9		1.2	-	8.0	29.9	4.0	2.6	2.1		**94.3**
	Emilia-Romagna	41.2	26.5	9.9	6.6	1.5	1.8	1.4	0.2	0.3	0.2	1.8	1.8	1.9	3.0	0.5		**98.5**
Centre	Toscana	51.2	14.8	22.1	1.1	1.9	2.8	-	0.1	-	0.2	5.4	-	-	0.3	0.0		**99.9**
	Umbria	30.8	29.9	0.9	22.1	2.7	0.7	-	0.2	2.4	-	3.1	-	-	0.3	-	-	**92.2**
	Marche	18.5	12.6	-	18.2	25.8	0.9	4.1		9.8	-	-	-	-	7.0	0.4		**97.4**
	Lazio	14.3	24.5	0.9	3.4	10.7	13.3	-	0.7	14.4	-	8.5	-	-	0.8	1.7	0.0	**93.2**
South and Islands	Abbruzzo	10.0	29.8	-	2.7	11.1	5.7	8.3	2.6	18.2	-	0.7	-	-	3.2	0.7		**92.9**
	Molise	-	30.9	-	23.4	-	1.6	-	5.7	27.6	-	-	-	-	6.4	0.9		**96.6**
	Campania	4.4	20.5	-	7.6	8.1	9.2	-	20.7	-	-	0.2	-	-	2.8	2.3	1.8	**77.5**
	Puglia	18.6	9.6	-	29.1	17.2	-	-	1.2	-	-	1.4	-	-	6.9	0.2	7.1	**91.3**
	Basilicata	6.9	10.3	-	6.0	10.4	0.9	-	5.0	-	-	-	-	-	5.3	5.4	17.6	**67.9**
	Calabria	-	30.2	-	3.3	17.3	8.9	-	4.0	-	-	-	-	-	1.6	3.5	18.4	**87.2**
	Sicilia	6.3	19.5	-	4.4	20.1	1.5	-	19.8	-	-	-	-	-	1.1	7.5	6.2	**86.4**
	Sardegna	-	30.6	-	12.8	12.6	5.6	-	13.8	-	-	3.8	-	-	5.0	9.7	4.3	**98.0**
Italia		**18.5**	**13.3**	**12.1**	**11.11**	**7.8**	**7.1**	**2.3**	**3.2**	**3.1**	**2.8**	**2.7**	**2.7**	**2.5**	**1.8**	**1.4**	**1.2**	**93.7**

Source: Authors' re-aggregation on the computation made by Istat's Price Unit on the scanner dataset

3.2 ASESD Approach

3.2.1 Estimation of Spatial Consumer Price Indices by loosening the tight comparability of the products

This second experiment has been conducted by using a different innovative approach regarding the definition and satisfaction of the principle of comparability decided by the ASESD Centre. The principle of comparability is applied at the level of each "Group of Products" of the ECOICOP classification, by loosening the "tight" specifications of the elementary products. The approach considers the unit value prices from the consumer side (or point of view). The hypothesis is that the elementary products (items) belonging to each group satisfy in any case the same consumer needs (and may give him the same utility), also if the brands, quality, *etc.* are different.

The comparison is therefore done by considering the average level of prices of the group of products purchased in the different areas (provinces), considering the basket of elementary products that the consumers of each area have really purchased[6]. Then the average level of prices of the different groups of products is aggregated to obtain the SPIs for each sub-national area. Therefore, these groups, and not the BHs, are the building blocks of the comparison, defined using 102 groups of the ECOICOP-8-digit classification.

In the first step, we computed $\overline{p}_{kj}$ which is the weighted mean price for each group of products in ECOICOP-8-digit k and province j. Let r_{ikj} and q_{ikj} be respectively the annual turnover and the total quantity sold[7] of item i belonging to ECOICOP-8-digit k in province j. The unit price p_{ik}j is equal to[8]:

$$p_{ikj} = r_{ikj} / q_{ikj}$$

and its relative weights in terms of turnover are equal to:

$$w_{ikj} = r_{ikj} / \sum_{i=1}^{n_{kj}} r_{ikj}$$

6 The value of the average level of prices of the different provinces could be affected by the different typologies of families (number of components, age, *etc*.) in the provinces (Istat, 2009; Biggeri and Laureti, 2018). To obtain a more precise comparison among the different averages, it could be necessary to make some standardisation of the provincial averages. This is an issue that the unit of research will deepen in a near future.

7 Which are the expenditure and the quantity purchased by consumers.

8 Obviously, taking account of the different size of the items' package.

where n_{kj} is the number of items in the *k-th* group of products in the *j-th* province. Finally, the weighted mean price is:

$$\bar{p}_{kj} = \frac{1}{n_{kj}} \sum_{i=1}^{n_{kj}} p_{ikj}\ w_{ikj}$$

The second step is devoted to the aggregation of the average level of prices of the group of products (102) to estimate the provincial SPI. It should be noted that not all 102 product groups are present in all provinces. Therefore, to compute the SPI at the provincial level we use a Country Product Dummy model.

Under the CPD model approach the logarithm of p_{kj} is considered as a function SPI_{jj}, the spatial price index of the *j-th* province relative to the other provinces, of P_k, the "provincial" average price of the *k-th* group of commodities, and of a random error term:

$$\begin{aligned} ln\bar{p}_{kj} &= lnSPIj + lnP_k + ln\ \mu_{kj} = \\ &= \pi_j + \eta_k + v_{kj} \\ ln\bar{p}_{kj} &= \sum_{j=1}^{M_r} \pi_j D^j + \sum_{i=1}^{n} \eta_k D^k + v_{kj} \end{aligned} \tag{8}$$

where D^j is a provincial-dummy variable that takes value equal to 1 if the price observation is from $j-th$ province and 0 otherwise; and D^k is a k – group of products dummy variable that takes value equal to 1 if the price observation is for $k-th$ group and 0 otherwise, and $\upsilon_{k,j}$. are normally distributed with zero mean and constant variance σ^2.

Following the estimation procedures already illustrated in sub-section 3.1.1 and using a weighted CPD model, where the weight is the ratio between the total turnover of one group of products in one province and the total turnover of the province, we can estimate $\hat{\pi}_j$ and then obtain the provincial Spatial Price Indices by SPIj = $e^{\hat{\pi}_j}$.

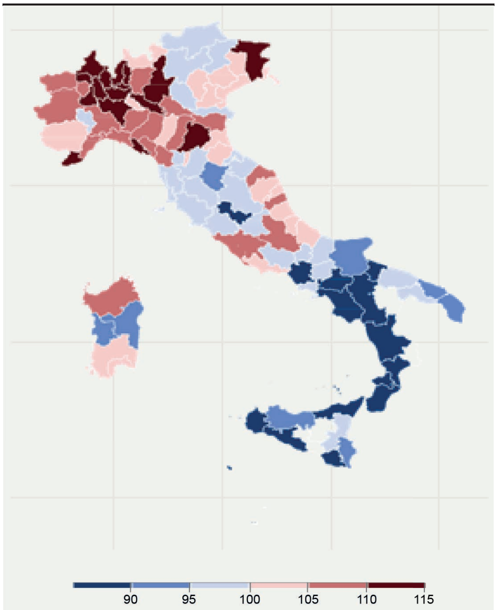
3.2.2 Results: the ASESD SPIs for food

The estimates of the SPIs for Food obtained on the basis of model equation (8) present high variability or heterogeneity as shown by the descriptive statistics reported in Table 3.4 and by the SPIs reported in Figure 3.5.

Table 3.4 - Descriptive statistics based on provincial SPIs for Food: ASESD Approach (Italy=100) - Year 2018

	Food	
Min	70.73	Province: Agrigento
Max	113.38	Province: Como
CV	7.51	

Figure 3.5 - SPIs for Food at Provincial level: ASESD Approach (Italy=100) - Year 2018



Source: Authors' processing on Istat scanner database

The results obtained are somehow expected. Indeed, provinces in the north of Italy show SPI values greater than 100, while provinces in the south show values smaller than 100. However, there are exceptions, *i.e.* provinces in the north-east Alps Mountains that show SPI below 100, even if they are close. Provinces in the centre of Italy have SPIs close to 100, with some evidence of SPI lower than 100 for provinces located in the Appennino mountains (middle of central Italy), and SPI greater than 100 for the provinces located at the seaside, both Adriatico (east), Ligure and Tirreno (west).

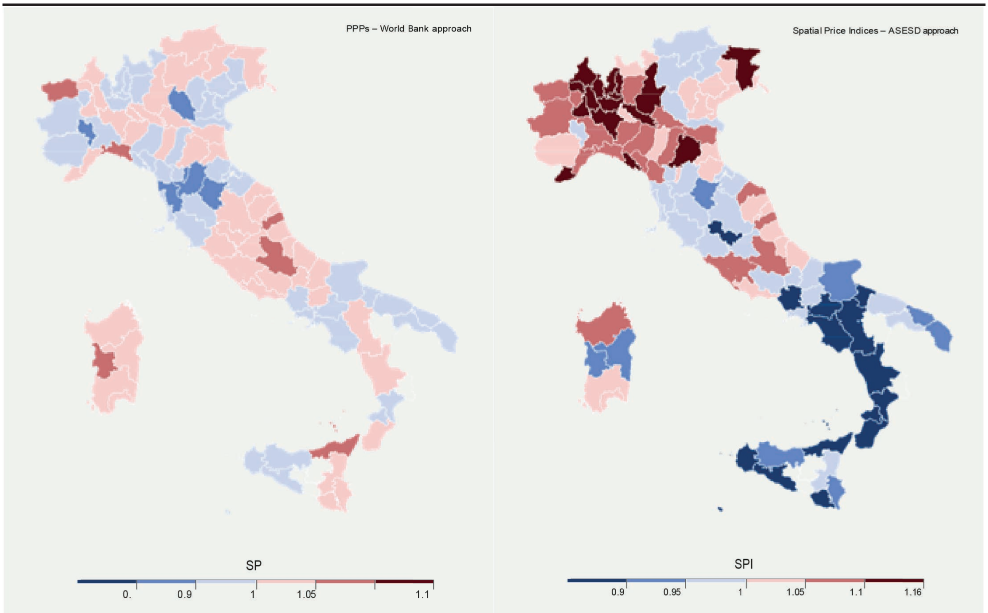
The lowest SPI is estimated for the province of Agrigento, in Sicily (south of Italy), while the highest is in the province of Como, in Lombardia region (north of Italy). The provinces with the highest SPI are all located in the north-west, but these are also over 100 for the city of Aquila (AQ) located in Abruzzo, a region in the south of Italy.

In conclusion, it must be observed that in many provinces the SPIs is around 100. In these cases the comparisons between the values of the different provinces are quite uncertain. In fact, following Suits (1984) we derived the p-values for the rescaled α_j, which are not reported here. However, setting an I type error equal to 0.1 we observed that for 43 provinces we can't reject the hypothesis $\alpha_j = 0$, which corresponds to SPI equal to 1 (in our case = 100).

4. A general analysis of the results of the experiments to compute the PPPs and SPIs: some concluding remarks

The results obtained from the two experiments are undoubtedly interesting. Actually, the estimations of the PPPs (World Bank approach) and SPIs (according to the ASEDS approach) at the provincial level are quite different with different variability, as we can see from the different CV reported in Table 3.2 and 3.4 (respectively 3.00 and 7.51). The differences are easily appreciated by examining Figure 4.1 which reports the indices computed with the two approaches using the same scale.

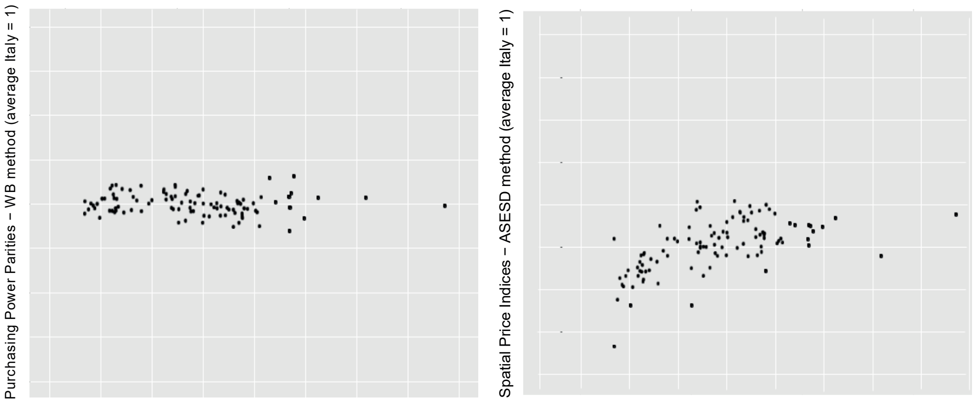
Figure 4.1 - PPPs and SPIs computed according to the World Bank and ASESD approaches - Year 2018



Source: Authors' processing on Istat scanner database

The results of the two experiments are therefore different, but we have to take into account that the followed procedures are different too. As suggested in the World Bank book (2013), when we compute the within-country PPPs one would expect some internal consistency. Price levels in poor areas should be generally lower than those in richer areas; they also show a similar pattern across the basic headings. To check this hypothesis, we have done a comparison between the two computed indices and the value-added per capita by Italian province as reported in Figure 4.2. It is evident that the Indices computed with the ASESD method satisfy in some way the previously mentioned consistency, while the PPPs do not satisfy it (in fact the correlation coefficients are about +0.60 for the first index and -0.10 for the second).

Figure 4.2 - Scatterplot of the PPPs (World Bank approach) and SPIs (ASESD Method) vs. the value-added per capita in the Italian provinces - Year 2018



Source: Authors' processing on Istat scanner database

As explained in the previous sub-sections, the two sets of SPIs values have been computed with different methodologies and with slightly different sets of data. The difference in the results most likely depends on the different methodologies used: the Word Bank approach is based on like-to-like product comparisons, and we expect that the range of the price for the same identical product cannot vary so highly in hypermarkets and supermarkets, although located in different areas of the Italian territory. The influence of the political prices of the different commercial chains should be analysed; to have a clearer picture of the reasons for the differences the analyses should be done at a disaggregated level. Unfortunately, a finer comparison at the Basic heading level is not possible at this moment, but we plan to better investigate it in the future. We will conduct further experiments to compute SPIs using other methods, for example, that used in the US: to make the products as comparable as possible, a hedonic regression was estimated, below the basic heading with class variables that include the product's characteristics, such as type, size, and brand (Aten, 2021).

In any case, the results already obtained are interesting and useful from a scientific and official statistics point of view. We recommend that the unit of research continues the experiments and that the same experiments are conducted also by the units of research in other European countries.

5. A tentative measure of poor-specific SPIs: SPIs for the first quintile of price distributions

To provide adequate information regarding the real incidence, nature, and extent of economic poverty, it became necessary to use and therefore compute poverty-related SN-SCPIs.

To compute poor-specific PPPs at the international level, Deaton (2006) and Deaton and Dupriez (2011), proposed to combine the average price levels with different specific baskets of goods and services for the poor, obtained by the BHS from 62 developing countries. This is an important step to calculate the global poverty lines. However, the method is not easily applicable to obtain adequate poor-specific spatial indices, especially at the sub-national level within a country. In fact, the consumer behaviour of poor households varies for quality of the commodities, channels of distribution, location of the markets, and above all the prices paid. So to compute the poor-specific SPIs, we have to use also the prices paid by the poor (Biggeri and Leoni, 2004; Giusti *et al.*, 2017)

Therefore, because we do not know the prices paid by the poor, for the computation of poor-specific SN-SCPIs following the ASESD approach, we conducted a preliminary experiment by using data of the first quintile of the price distributions of each product, assuming for the time being that the poor purchase the cheapest items of each specific product. After all, the lowest prices are often also used for the measurement of Absolute Poverty (ABSPO) as it is done by Istat (2009)[9] and it is written in a report on ABSPO published by the Joint Research Centre of the European Commission (Menyhért *et al.*, 2021).

To obtain such SPIs by using the lowest prices, the model equation (8) is modified as follows:

$$ln\bar{p}_{kj}^{\tau} = lnSPI_j^{\tau} + lnP_k^{\tau} + ln\mu_{kj}^{\tau} \tag{9}$$

where $\bar{p}_{kj}^{\tau}$ is the quantile of order τ of the unit prices (p_{ijk}) of items belonging to group k (ECOICOP-8-digit) and province j. The parameters of the model are estimated by regressing the logarithm of the quantile on province dummy variables and on a group of product dummy variables as done for model (8).

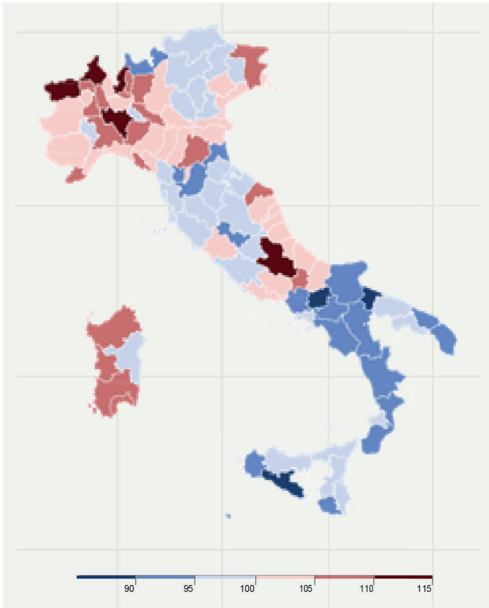
9 Actually, for Italy Spatial Price Indices for poor families can be obtained by using the data on absolute poverty (Biggeri and Pratesi, 2018; Biggeri and Laureti, 2018).

For example, setting τ = 0.2 we have obtained the estimates of spatial price indices related to the lowest prices for each Italian province, which we denote as SPIs(Q0.2), for which we report the descriptive statistics in Table 5.1, showing the results in Figure 5.1.

Table 5.1 - Descriptive statistics based on provincial SPIs (Q0.2) for Food: ASESD method (Italy=100) - Year 2018

	Food	
Min	82.00	Province: Agrigento
Max	114.30	Province: Aosta
CV	6.10	

Figure 5.1 - Spatial Price Indices SPIs Q0.2 at provincial level for Food: ASESD method (Italy=100) - Year 2018



Source: Authors' processing on Istat scanner database

Looking at the results of the estimation of the SPIs(Q0.2), many provinces in the North-west and on the Adriatico seaside, excluding Puglia provinc$_{es}$, show a SPIs($Q_{0.2}$) greater than 100, while many provinces in the North-east, Centre and South of Italy show SPIs($Q_{0.2}$) near or smaller than 100. Sardegna provinces show SPIs($Q_{0.2}$) greater than 100, with the exception of Nuoro.

If we compare this tentative measure of poor-specific SPIs with the general SPIs computed (as mean) considering all the prices, we observe that the former indices present, as expected, smaller values and a little bit less variability than the latter (CV respectively 6.10 and 7.51), but the differences are not uniform at the territorial level within the country.

This tentative measure of poor-specific SPIs has provided some interesting information. However, we must keep in mind that the comparisons between the SPIs($Q_{0.2}$) of different provinces and between SPIs($Q_{0.2}$) and the general SPIs are really uncertain in various cases, as the computations of p-values show.

Therefore, to improve the validity of these price indices some information on the consumption behaviour of poor households should be collected. A tentative survey to give some insight on that has been carried out by Istat.

6 Where people in a condition of absolute poverty purchase some frequent consumption products

As anticipated in Section 1, to obtain some information on the consumption behaviour of households, a first tentative survey and analysis were conducted by researchers of Istat. We present here a summary of the results. Since 2015, Istat has added a special section (a two-week diary) in the Household Expenditures Survey (HES) aimed at investigating the type of outlets where the households purchased a list of the 25 most frequently purchased products. The types of outlets listed are seven: traditional shops, open markets and street vendors, hard discounts, hypermarkets and supermarkets, department stores and outlet chains, farm or direct producers, and the Internet. After an experimental period, Istat researchers carried out an analysis of 2019 HES data, which produced the following results.

The results obtained from these preliminary analyses of 2019 HES data show some interesting differences between non-poor and absolutely poor households (identified by means of the poverty lines calculated and updated each year by Istat) in terms of choice of the type of outlet where purchasing a list of 25 large consumption products.

In particular, the purchase of some essential products such as bread, milk, and eggs is (relatively) more important for poor families. In general, the use of the traditional shop and open market is very similar for the poor and non-poor families; while the poor families made more frequent purchases in hard discounts and less in hypermarkets/supermarkets than the non-poor families.

In Table 6.1, we report in detail the types of outlets chosen by Italian families in absolute poverty who have made at least one purchase in the two weeks of observation. It is clear that the choice of the types of outlet is different for each kind of product.

This evidence is worth to be deepened also by breaking down the analysis at territorial level, overcoming the problem of a too small sample if we take into consideration only poor households. This line of research is aimed at improving the estimation of the actual prices paid by poor families in different Italian geographical areas by taking into account their different behaviour in the choice of the outlet where purchasing in particular large consumption products. The possible results obtained could enhance the spatial comparison of consumer prices by making reference to the poor part of the population.

Table 6.1 - Types of products by types of outlets where Italian absolute poor households make purchases (% distribution) - Year 2019

Products	Traditional shops	Open market and Street vendors	Hard discounts	Hypermarkets and supermarkets	Departments stores and outlet chains	Farms or direct producers	Internet	Total
Bread	41.5	1.7	19.1	37.6	0.1	0.0	0.0	**100.0**
Pasta	13.9	1.0	31.3	53.3	0.5	0.0	0.0	**100.0**
Biscuits, rusks, snacks	10.9	1.6	32.3	55.0	0.2	0.0	0.0	**100.0**
Fresh meat	30.2	1.3	23.9	43.9	0.3	0.3	0.0	**100.0**
Frozen meat	10.9	1.6	16.4	69.3	0.8	0.8	0.1	**100.0**
Cured meat	15.7	1.1	29.6	53.1	0.2	0.3	0.0	**100.0**
Fresh fish	41.8	13.6	14.0	30.3	0.0	0.3	0.0	**100.0**
Frozen fish	10.3	1.7	32.6	54.4	1.0	0.0	0.0	**100.0**
Milk	13.7	1.1	32.5	52.5	0.2	0.0	0.0	**100.0**
Cheeses	14.7	0.8	30.4	53.7	0.2	0.2	0.0	**100.0**
Yogurt	10.7	0.9	32.5	55.6	0.3	0.0	0.0	**100.0**
Eggs	14.3	2.4	32.6	49.7	0.2	0.7	0.0	**100.0**
Fresh fruit	25.1	10.8	24.9	38.8	0.1	0.2	0.0	**100.0**
Fresh vegetables, potatoes, and legumes	25.1	9.6	24.1	40.7	0.2	0.3	0.0	**100.0**
Dried or frozen vegetables, potatoes and legumes	10.3	5.0	30.6	53.4	0.4	0.4	0.0	**100.0**
Olive oil	8.5	1.3	33.2	55.4	0.3	1.4	0.0	**100.0**
Mineral water	11.5	0.7	29.1	58.1	0.4	0.1	0.0	**100.0**
Soft drinks	8.1	2.0	31.2	58.2	0.5	0.0	0.0	**100.0**
Wine	13.4	0.0	32.3	53.5	0.0	0.9	0.0	**100.0**
Coffee	10.1	1.1	31.4	56.5	0.1	0.3	0.5	**100.0**
Medicines	94.0	0.0	0.7	5.4	0.0	0.0	0.0	**100.0**
Personal hygiene products (soap, deodorant, baby diapers, etc.)	10.3	0.9	31.7	55.1	1.7	0.1	0.2	**100.0**
Cleaning products	12.8	2.3	31.1	51.7	2.1	0.0	0.0	**100.0**
Disposable items for the kitchen (napkins, dishes, etc.)	13.5	1.5	34.1	49.6	1.3	0.0	0.0	**100.0**
Toys and video games	27.2	9.1	17.3	34.4	11.9	0.0	0.0	**100.0**

Source: Authors' re-aggregation on the computation made by Istat's Price Unit on 2019 Istat HBS data

7. Concluding remarks

The goal of our work is twofold: first, to verify which are the appropriate methodological approaches to estimate the Sub-National Spatial consumer price Indices (SN-SCPIs) and second to verify whether it is possible to estimate poor-specific SN-SCPIs, without knowing what prices the poor pay for the various products and, consequently, if it is possible to track poor specific consumption behaviour.

To this end, two experiments have been conducted. The first experiment refers to the general computation of the SN-SCPIs by using the same scanner database and elementary data but applying a different principle of product comparability. One is computed considering the comparison of like-to-like products for the different subnational areas (World Bank approach). The other considers the principle of comparability at the level of very detailed groups of products, by loosening the “tight” specification of the elementary products given that the items products of each group satisfy the same consumer needs (that was called ASESD approach). For the computation of poor-specific SN-SCPIs - that is to calculate SPIs closer to the prices paid by the poor, we conducted a preliminary experiment by using data of the first quintile of the price distributions of each product, assuming, for the time being, that the poor purchase the cheapest items of a product.

The results of the first experiment show that the provincial PPPs computed by using the World Bank approach present little variability among the different territorial areas; while the SPIs computed following the ASESD approach show a much more pronounced variability. In any case, the proposal to loosen the “tight” comparability of the elementary products provides spatial price indices which look better than those obtained by applying the World Bank approach, because they are more consistent with the fact that price levels in poorer areas are generally lower than those in richer areas. The results will be further verified and analysed at a disaggregated level to have a clearer picture of the reasons for the differences. Moreover, further experiments will be conducted to compute SPIs using other methods, for example making the products as comparable as possible by estimating a hedonic regression, below the basic headings.

In addition, the tentative experiment to measure poor-specific SPIs provides some interesting information. If we compare the poor-specific SPIs with the general SPIs considering all the prices, as expected, the former indices are a little smaller and present a little bit less variability than the latter, but the differences are not uniform at the territorial level within the country. Moreover, we have to keep in mind that their estimation is really uncertain in various cases, as the computations of p-values show. To improve the adequacy of the poor-specific SPIs it is necessary to know the consumption behaviour of poor households at the sub-national level (possibly at the provincial level) and their behaviour in order to choose cheaper products. The results of the survey and analysis conducted by Istat researchers provided important insights, but they are only at the national level. Therefore, it is necessary to break down the analysis at the territorial level, overcoming the problem of a too-small sample if we take into consideration only poor households. In this way, we could take into account their different behaviour in the choice of the outlet where they purchase products and the prices they pay.

Finally, we have to remember that by using only scanner data on retail prices of the outlet of modern distribution it is impossible to estimate a complete system of sub-national Spatial Consumer Price Indices. In fact, the purchases of products and services by consumers take place in various forms and in different types of outlets, as shown in Section 6.

Therefore, it is evident that to estimate a complete system of general SN-SCPIs it is necessary to build up a database that could allow the estimation of these indices related to the entire universe of household consumption.

Istat collects indeed consumer prices by using different sources: territorial surveys at the outlets by non-probability samples, use of administrative data, and use of scanner data (Big Data). Therefore, it is necessary to follow a strategy to use and integrate all the consumer price sources of data, considering also the fact that the issue has to be faced that the data come both from probability and non-probability samples. It is important to stress the need for further research and experiments in this field, also taking account of the recent studies and development presented at many Workshops and Webinars organised during the last months of the year 2021 and the first months of 2022 by the International Association of Survey Statisticians (IASS, 2022) and by the Survey Research Methods Section of the American Statistical Association.

References

Aten, B. 2021. "Country Case Studies: United States". In Biggeri, L., and D.S. Prasada Rao (Eds.). *A Guide to the Compilation of Subnational Purchasing Power Parities (PPPs)*, Appendix A1: 48-59. Washington, DC, U.S.: World Bank, International Comparison Program, Inter-Agency Coordination Group.

Biggeri, L., and T. Laureti. 2018. "Sub-national PPPs: Country case studies. Publications, experiments and projects on the computation of spatial price level differences in Italy". Paper presented at the *3rd Meeting of the Country Operational Guidelines and Procedures Task Force*, 27th September 2018. Washington, DC, U.S.: World Bank.

Biggeri, L., and L. Leoni. 2004. "Family of Consumer Price Indices for different purposes. CPIs for sub-groups of population". *Statistical Journal of the United Nations Economic Commission for Europe - UNECE*, Volume 21, N. 2: 157-165.

Biggeri, L., and M. Pratesi. 2018. "Misure di povertà economica a livello locale e loro aggiustamento per il differente costo della vita". *Neodemos*, 6 luglio 2018.

Biggeri, L., M. Pratesi, C. Giusti, and S. Marchetti. 2020. "Experiments and projects on the computation of spatial consumer price level differences in Italy by using Big Data: issues related to the use of non-probability samples". Presentation at the SIG *workshop on non-probability samples*, University of Trier, Germany, 27th-28th February 2020.

Biggeri, L., and D.S. Prasada Rao (Eds.). 2021. *A Guide to the Compilation of Subnational Purchasing Power Parities (PPPs)*. Washington, DC, U.S.: World Bank, International Comparison Program, Inter-Agency Coordination Group.

Deaton, A., and O. Dupriez. 2011. "Purchasing Power Parity Exchange Rates for the Global Poor". *American Economic Journal: Applied Economics*, Volume 3, N. 2: 137-166.

Diewert, W.E. 1995. "Axiomatic and Economic Approaches to Elementary Price Indexes". NBER *Working Paper*, N. 5104. Cambridge, MA, U.S.: National Bureau of Economic Research - NBER.

Feenstra, R.C., and M.D. Shapiro (Eds). 2003. *Scanner Data and Price Indexes*. Chicago, IL, U.S.: The University of Chicago Press, *National Bureau of Economic Research – NBER, Studies in Income and Wealth*.

Giusti, C., L. Masserini, and M. Pratesi. 2017. "Local Comparisons of Small Area Estimates of Poverty: An Application Within the Tuscany Region in Italy". *Social Indicators Research*, Volume 131, N.1: 235-254.

International Association of Survey Statisticians – IASS. 2022. *Calendar of events*. http://isi-iass.org/home/events/.

Istituto Nazionale di Statistica - Istat. 2020. "Prezzi al Consumo. Agosto 2020. Dati provvisori". Statistiche Flash. Roma, Italy: Istat. https://www.istat.it/it/archivio/246776.

Istituto Nazionale di Statistica - Istat. 2010. "La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane". *Comunicato Stampa*. Roma, Italy: Istat. https://www.istat.it/it/archivio/6279.

Istituto Nazionale di Statistica - Istat. 2009. "La misura della povertà assoluta". *Metodi e Norme,* N. 39. Roma, Italy: Istat.

Laureti, T., C. Ferrante, and B. Dramis. 2017. "Using scanner and CPI data to estimate Italian sub-national PPPs". In *Proceedings of 49th Scientific Meeting of the Italian Statistical Society* – SIS 2018: 581–588. Palermo, Italy, 20th – 22nd June 2018.

Laureti, T., and F. Polidoro. 2022. "Using Scanner Data for Computing Consumer Spatial Price Indexes at Regional Level: An Empirical Application for Grocery Products in Italy". *Journal of Official Statistics - JOS*, Volume 38, N. 1: 23-56.

Laureti, T., and F. Polidoro. 2021. "Country Case Studies: Italy". In Biggeri, L., and D.S. Prasada Rao (*Eds.*). *A Guide to the Compilation of Subnational Purchasing Power Parities (PPPs)*, Appendix A3: 94-114. Washington, DC, U.S.: World Bank, International Comparison Program, Inter-Agency Coordination Group.

Laureti, T., and F. Polidoro. 2017. "Testing the use of scanner data for computing sub-national Purchasing Power Parities in Italy". In *Proceedings of 61st World Statistics Congress – ISI 2017*. Marrakech, Morocco, 17th – 21st July 2017.

Laureti, T., and D.S. Prasada Rao. 2018. “Measuring spatial price level differences within a country: Current status and future developments”. *Estudios de Economía Aplicada*, Volume 36, N. 1: 119–148.

Menyhért, B., Z. Cseres-Gergely, V. Kvedaras, B. Mina, F. Pericoli, and S. Zec. 2021. “Measuring and monitoring absolute poverty (ABSPO). Final Report”. *JRC Technical Report* (Joint Research Centre – JRC; the European Commission’s Science and Knowledge Service). Luxembourg: Publications Office of the European Union.

Pratesi, M., C. Giusti, S. Marchetti, L. Biggeri, G.Bertarelli, F. Schirripa Spagnolo, T. Laureti, I. Benedetti, F. Polidoro, F. Di Leo, and M. Fedeli. 2021. “Estimations of local spatial price indices using scanner data and their impact on the measure of poverty incidence”. *MAKSWELL Project*, Deliverable 3.2.

Rao, D.S. Prasada, and G. Hajargasht. 2016. “Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP)”. *Journal of Econometrics*, Volume 191, Issue 2: 414-425.

Suits, D.B. 1984. “Dummy variables: Mechanics v. Interpretation”. *The Review of Economics and Statistics*, Volume 66, Issue 1: 177-180.

World Bank. 2015. *Operational Guidelines and Procedures for Measuring the Real Size of the World Economy: 2011 International Comparison Program-ICP*. Washington, DC, U.S.: World Bank.

World Bank. 2013. *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program - ICP*. Washington, DC, U.S.: World Bank.

The micro-level integration of Labour Register and Compulsory Communications to provide timely and consistent stock-flows accounts of the labour market

Ciro Baldi[1], Sara Gigante[1], Silvia Pacini[1], Roberta Rizzi[1]

Abstract

An experimental integration at the micro level of each job position between an employer and an employee in the private non-agricultural sectors enables the extension of the stocks and flows of the Labour Register (LR) to include the much-updated flows of the Compulsory Communications (CC). This analysis makes it possible to correct the LR for the absence of very short-duration job positions and project job stocks and flows up to the last available date of the CC. Daily stocks and the relative activation and cessation flows allow detailed studies of the labour dynamic. The integration methodological details are examined in depth to highlight the main evidence and the points to enhance.

Keywords: Labour Register, Compulsory Communication, Integration, Projection, Employment, Jobs stock-flow measurement.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_2.2022.02

1 Ciro Baldi (baldi@istat.it); Sara Gigante (gigante@istat.it); Silvia Pacini (pacini@istat.it); Roberta Rizzi (rizzi@istat.it), Italian National Institute of Statistics – Istat.

The authors would like to thank the members of "*Accordo a 5*" for sharing the outputs of this work and providing the Compulsory Communication data. Although this article is the result of all the authors' commitment, the paragraphs are attributed as following: 1, 6 and 7 to Ciro Baldi; 2 to Silvia Pacini; 3 to Roberta Rizzi; 4 and 5 to Sara Gigante.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

The current situation of administrative data and statistical registers in Italy allows the calculation of two sets of employment measures referred to jobs.

On one side there are stock-type statistics such as the number of active jobs on a specific day of the year and various kinds of averages over such figures over a given period (*i.e.* a month, a quarter, a year, *etc.*). These measures are by now produced by the Italian National Institute of Statistics - Istat trough the Oros and the Labour Register (LR)[2] processes based primarily on the Italian National Social Security Institute - INPS data. LR data are disseminated through the Statistical Business Register (BR) and its employment detailed version[3], the Labour Cost Survey, the National Accounts, *etc*.

On the other side, the Compulsory Communications (CC) administrative system allows the production of flow-type statistics such as the number of activated and ceased work relationships in any given period (a day, a month, a year, *etc*.). For instance, this type of statistics is what is actually disseminated by the Ministry of Labour in its quarterly publication or, after a certain data transformation[4], in the Quarterly Note and Annual Report conjointly produced by Istat, Ministry of Labour, INPS, *Agenzia Nazionale per le Politiche Attive del Lavoro* - ANPAL, Italian National Institute for Insurance against Accidents at Work – INAIL (aka "*Accordo a 5*"), and what has been monthly released, for almost 20 years now, by the *Veneto Lavoro* Research Centre.

The two sets of measures are, in theory, related to each other by the dynamic equation of job stocks and flows (see Section 6). This circumstance makes it possible to derive from each of them its own measures of gross changes, such as the difference in the job positions between two moments in time. However, since the actual statistics are based on independent sources, the estimated figures of change, even after having limited the comparison to a common target population, are often inconsistent. An obvious solution to this problem is to build a data system in which the different sources are integrated at the micro level and are thus able to provide at once consistent measures of job stocks and flows. But, up to now, there have been no attempts

2 The Labour Register is called (in Italian) *Registro Tematico del Lavoro* and its acronym is RTL.

3 The Business Register is called in Italy *Archivio Statistico Imprese Attive* - *ASIA* and its employment detailed version *ASIA Occupazione*.

4 See Methodological Note in the Quarterly Note of "*Accordo a 5*".

to build such a micro-based system. Two main obstacles have prevented this development. The first is the fact that the relevant administrative sources are held by different institutions. The second is that a number of methodological and data cleaning issues have to be faced and resolved to connect the two administrative sources.

However, in recent years, new conditions have arisen that are finally allowing this evolution of the labour statistics. The first is the ongoing Istat project of the Labour Register, a micro-level Employer/Employee Longitudinal Database with information on employment, hours, wages, social contributions, *etc*. The project of the LR entails that the main statistical unit of the Register is the job position, which is the association between an employer and an employee defined by a certain activation date. Among the various reasons to choose this statistical unit the possibility of linking the CC source was crucial since its data are implicitly based on the same unit.

The second circumstance was the birth of the "*Accordo a 5*", the coalition of the 5 above- mentioned institutions to produce statistics. This cooperation enables the transfer of macro and, more recently, microdata between the Ministry of Labour and Istat thus laying the foundations to extend the LR with the inclusion of CC information.

It is important to notice that it is already possible to produce consistent measures of job stocks and flows from the data of the Labour Register itself, without the integration of the CC source, for the very reason that the statistical unit of the LR is the job position and the activation and cessation dates are recorded in the Register. However, this operation has two main limitations. First, due to the nature of the input data source, the LR by itself is not able to measure the frequent intra-month job activations and cessations between a given employer and a given employee that are quite widespread in certain economic activities with the implication of overestimating the duration of some job positions. Second, and more important, the LR data is issued yearly with a time delay of about 8-10 months from the reference year with provisional data. The integration of the CC source should, in principle, provide the possibility of overcoming the two limitations thus allowing to produce the stocks and flows integrated statistics on a short-term basis.

2. The Labour Register's aims and principles

The Italian Labour Register is one of the components of the System of Integrated Registers that Istat has been implementing to face the challenges of the European Statistical System strategy for the current decade (*ESS Vision 2020*. ESSC, 2014).

The aims of improving the efficiency of the production processes and releasing more integrated and coherent indicators play a crucial role in the new system of registers. This is a system of micro-level statistical databases mainly derived from administrative sources, at the most detailed level of the information available, covering the whole population of statistical units (UNECE 2017).

In the context of labour statistics, the Labour Register aims to be the basis for a number of labour market indicators and analyses: from the construction of standard macro statistical indicators to the evaluation of labour policies and the release of longitudinal microdata standard files for the researchers (Baldi *et al.*, 2018).

With a complex and rich structure of information, the purpose of the LR is to cover all regular paid jobs active in the national territory in all sectors of economic activity, either private or public, dependent or independent.

The realisation of the LR implies that a number of different administrative sources have to be integrated. As for the population of employees, the primary sources on which it is based are social security data and, in a second order, tax data. Additional relevant sources, for the public sector, are payroll data and, for the population of self-employed workers, Chambers of commerce data and other sources. The employer-employee structure of the LR and its information at the maximum level of detail available in the input sources in terms of units, variables and time references, allow a modular use of the information. Depending on the target variable, it is possible to derive information referred to each day, week, month or year on three main statistical units: the individual, the economic unit, and the job position which is the specific unit of the LR.

In this way, the Register is linkable on the one hand to the Population Register, through the person ID, and on the other to the Economic Units Registers through the unit ID, with the obvious advantage of possibly matching a lot more information.

Moreover, the very specific unit of the LR is the job position, defined as the working relationship established between an economic unit and an individual with a starting date. It nables tracking the relationship between an employer and an employee since its inception and, over time, to construct the working career of each worker within and across employers and the status of employment. The LR, due to its longitudinal nature and statistical unit, allows measuring a very rich stock-flow accounting with figures on levels, gross and net changes of employment both in terms of workers and jobs. Moreover, each job position is characterised by a wealth of other attributes such as the type of contract, the working time, the occupational qualification, the workplace, *etc.* Other information available in the LR are gross wages, other labour costs, total labour cost, gross earnings, hours paid/worked, mainly with a monthly reference period.

In this analysis, we focus on the jobs stock-flows measurement for the subpopulation of employees in the non-agricultural sectors employed in private enterprises according to the Italian Business Register. This selection is mainly due to the advanced stage of development of the LR for this subpopulation[5]. In the context of this subpopulation, referred to dependent employment, the job position corresponds to the employment contract between an employer and an employee where the starting date is the date of activation of the contract and the ending date its cessation.

For the subpopulation we focus on, the main source of the LR is the UniEmens declaration of INPS which is integrated with the INPS-Dmag declaration for the agricultural workers employed in enterprises of the private non-agricultural sector. These two administrative declarations, made by employers for their employees with information on the job position and employer level, are used both for the social security payments and the individual pension schemes. Both these administrative sources have information to estimate either stocks or flows. To be more precise, the UniEmens dates of activation and cessation have to be integrated with the activation dates of tax

5 For the sake of precision the data for this subpopulation of the LR are derived from the RACLI Register. The RACLI Register is basically the frontline version of the module of the LR concerned with the population of employees with a job in enterprises of the private sectors covered by the BR. The standardised integration process of the sources and other production aspects like main metadata flowed into the LR process which is now an input sources of the RACLI process for the production of specific variables (worked and paid hours, labour costs, *etc.*) and their official release (see I.Stat database http://dati.istat.it/). Moreover the RACLI register is also the experimental area (for new variables, integration of new sources as CC, and so on) before introducing innovations into the LR.

sources for the job position activated before the first year of availability of the LR (2016). Moreover, in this social security source, when there are several job positions between the same employer and worker in the same month, only the earliest activation date and the latest closing date must be indicated in the declaration. This implies an undercounting of very short-term job positions within a month.

Regarding timeliness, both UniEmens and Dmag sources are available to Istat twice a year: one with provisional data and one with final data, both having a delay of 4 and 10 months from the end of the reference year respectively. Due to the timing of the production process and the availability of other sources, the first, provisional release of the LR data is available at about 10 months from the reference year and the final release at about 16 months.

In this work, the experimental analyses are carried out using the longitudinal LR information on job position stocks and flows related to the years from 2016 to 2019, with provisional data for this last year. As above mentioned from this data source alone it is possible to produce stocks and flows accounting of job positions. The analysis can be broken down across different characteristics of the job position, the employer and the employee. However, the analysis allowed by the LR alone has two main limitations. The first is the reference period: the timeliness of the administrative sources used at the moment in the LR process allows to produce preliminary estimation with a delay of about 10 months after the end of the reference year. This implies that estimates referred to January are released with a delay of 21 months. The second is related to the accuracy of the measurement of the flows of very short-term (intra-month) job positions for the reasons cited above. These limitations can be overcome by integrating into the LR the data of Compulsory Communication declarations.

3. The Compulsory Communications source: main features

The Compulsory Communications system, managed by the Ministry of labour and social policy, is a stream of declarations due by employers to notify relevant events of each employment relationship: the activation of a new employment relationship (contract) between an employer and a worker; the extension of a previous temporary contract; the transformation of the contract from one type to another (temporary to permanent, apprenticeship to permanent, part-time to full-time); and the termination of a contract before the natural deadline (not requested for contracts ending at the natural deadline) (Baldi *et al*., 2014). The events contained in the declarations, per se, allow measuring only employment flows and not stocks of jobs or persons. However, the events pertaining to the same contract between an employer and an employee can be linked together to construct the job position. Unfortunately, since the CC system includes only the set of jobs which have been interested by an event since the start of the system (March 2008), all the permanent jobs created before 1 March 2008 and not subject, after that date, to termination, an extension or a transformation are not observed. Thus, the number of job positions active at any given date measured by the system is severely underestimated. While the system is unable to measure the level of the stocks it should in principle provide quite accurate estimates of their change. In fact, the net change in the number of jobs between two dates is simply equal to the difference between the number of activations and the number of cessations.

The CC declarations have information on the employer - identified through the tax code and characterised by economic activity sector, registered office, place of work, the worker - identified through the tax code and characterised by biographical variables such as gender, age, educational level, nationality, and the job (connecting employer and worker) - characterised by the date of activation, and eventually termination, established duration, type of contract, working time. As above mentioned, the job positions are built by the Ministry of Labour by sequentially linking all the declarations referred to the same relationship: one activation, (possibly) one transformation of fixed-term employment to permanent, (possibly) one or more extensions, and (possibly) one termination. The key, which permits the reconstruction of the job position and identifies it, is threefold: composed of the ID of the worker, the ID of the employer and the starting date of the relationship. If one of the three key variables is affected by errors, a declaration is not matched with the correct job and the chain of related events is affected by errors.

The collection of this kind of data was initiated to provide an information system supporting actions to contrast irregular work but also to implement a database for monitoring and evaluating labour policies. Some features of this source make it valuable for carrying out a very innovative analysis of the labour market, especially when integrated with stock data offered for example by the Labour Register, as it is in the current work (see Section 5). The integration of the CC with the LR can contribute to improve the Register in a number of ways: full integration of LR with CC may allow to assess the quality of job definition, the most relevant statistical units of the Register, that have been defined through the threefold key (employer id, worker id, starting date) consistently with the CC source; checking and editing some variables already present in the Register with the corresponding variables available in CC; integrating variables absent in the Register (like professional qualification) or present with a lower level of detail (like the type of contract or place of work); improving metadata information thanks to further standard dictionaries and classifications. However for the current work the integration was limited to three main purposes (see Section 4): checking possible under coverage of the Register thanks to the very wide coverage in terms of employers and workers of the CC; identifying and measureing the labour input of the jobs with intramonth duration (those between the same employer and worker activated and terminated more than once in the same month) which cannot be identified by the LR *per se*; improving the timeliness of the information obtainable from the Register. This last point is due to the fact that since the CC declaration flows in the system daily, in principle they can be used to project the data of the LR, updated to the end of the previous year and provide preliminary, but very timely, estimation of the employment stocks with a very detailed breakdown.

4. Methodology

In order to improve the accuracy in estimating employment dynamics on LR basis, information declared in the administrative source of CC was integrated at the micro level of job position with the information estimated in the LR. Micro-integration is defined as "*the method that aims at improving the data quality in combined sources by searching and correcting for the errors on unit level*" (Bakker, 2011).

Integration at the level of individual employment (job position level) is also meant to obtain consistent and high-frequency stock and flow measures (number of active job positions every single day, for example), that enable to build indicators on the employment and job turnover as a synthesis of activation and cessation rates.

Another advantage in integrating the CC data with the Labour Register, in addition to the improvement in the estimation of the number of job positions and their duration, is producing very timely preliminary estimates of employment stocks and flows.

In terms of data quality, this exercise aims at improving *accuracy and timeliness* dimensions in LR (Eurostat, 2003; Statistics Canada, 2002; Rosén and Elvers, 1999), constrained to the maintenance of prior level quality in *consistency* (Wang and Strong, 1996; Batini and Scannapieco, 2006), *privacy and security and unique keys* (Daas et al., 2009) dimensions. According to the major experts in register-based statistics, quality represents indeed one of the main methodological issues to be tackled (Bakker and Daas, 2012).

The preliminary step to the exercise of micro-integration consisted in the harmonisation of reference periods, the completion of populations and the harmonisation of units between the two sources (Van der Laan, 2000).

In particular, the perimeter of the exercise is defined by all the job positions with an employee contract between a worker and a private enterprise whose economic activity belongs to NACE rev 2 sections B to S. In the exercise the temporary agencies and the job-on- call positions are excluded[6].

6 The exclusions of the temporary agencies depends on the unavailability, at the moment, of the CC information on temporary workers (that are collected via a dedicated form within the CC system), while the exclusions of the jobs on call depends on the fact that both in the LR and in CC these jobs positions have a large number of errors on the dates of cessations and on the fact that the duration of the job position for them is a poor measure of the labour input. However since these two segments of the labour force account for a relatively low share of total employment (respectively 1.2% and 2.7%) their exclusion does not affect the overall relevance of the work.

The population of enterprises is further refined by including only those units listed in the Statistical Business Register for at least one year from 2016 to 2019.

The operationalisation of the above perimeter has required to restrict both the LR and the CC source (completion of populations) on the set of enterprises listed in the BR (with the exception of temporary agencies) and on the employee contracts (with the exception of job-on-call) to be included. This last restriction has been based on the source-specific variables that characterise the employment contracts[7]. In LR the harmonisation of the unit has demanded, substantially, a simple operation of aggregation on the job position, the statistical unit derived from administrative data. In particular, the essential structure of both CC and longitudinal LR in a specific period can be described as a list of job positions between an employer and an employee active for at least one day in the period.

Table 4.1 - LR and CC basic structure - Theoretical scheme (a)

Employer	Employee	Job position	Activation date (AD)	Cessation date (CD)
A	John	A-John-1	A-John-1(AD)	A-John-1(CD)
…	…	…	…	…
A	John	A-John-N	A-John-N(AD)	
A	Sally	A-Sally-1	A- Sally-1(AD)	A- Sally-1 (CD)
…	…	…	…	…
H	John	H-John-1	H-John-1(AD)	H-John-1(CD)
…	…	…	…	…
Z	Jack	Z-Jack-1	Z-Jack-1(AD)	

(a) The data used are anonymised so they have only an ID number for employers and employees. Here proper nouns are used only for immediate comprehension.

The harmonisation of the reference period in the two sources has entailed the building of an *ad hoc* LR process for provisional data on 2019. The estimate of quantitative and qualitative variables to be attributed to each job and the harmonisation of other variables and classifications is postponed to a future extension of this work.

7 The selection in CC of positions with an employee type contract (excluding job-on-call workers) was carried out on the basis of variables that may differ from those present in the social security sources underlying the LR.

4.1 Longitudinal editing and imputation process and integration methodology

Since the administrative sources at the basis of the LR are not intended to record the job positions as above defined, the quality of the Activation (AD) and Cessation (CD) dates is not always optimal. Thus, the statistical unit needs to be derived (Wallgren and Wallgren, 2014; Daas and Ossen, 2011) with a process aimed at:

1. estimating the activation and cessation dates selecting only events compliant with the statistical definition (in line with what is stated in CC);
2. making dates consistent in the year and among years, in order to ensure a longitudinal coherence.

An iterative E&I process has been set to improve the quality of the LR unit estimation (about 2.5% of the positions are subject to correction and the incidence of jobs with no longitudinal inconsistencies post-E&I is 99.99%): in particular, for each employer-employee pair the process aims at making the dates consistent in the year and between years on the basis of the edit rules, defined by domain experts and summarised in Table 4.2.

Table 4.2 - LR rules for E&I on longitudinal coherence - Theoretical scheme

Edit rule number	Type of ER	Description
1	consistency	AD not null if CD previous not null
2		CD not null if AD next not null
3		AD <=CD
4	non-overlapping	AD >= CD_previous or CD<= AD_ next
5	non-duplication	AD = CD if AD_ next = AD
General		CD_previous <= AD <= CD <= AD_next

The editing and imputation process, while guaranteeing excellent results in terms of reducing logical violations in the longitudinal reading of the positions (Table 4.4), does not deal with the poor accuracy in measurement and detection for very short-term relationships due to:

1. Quality of UniEmens flow: in the presence of several job positions between the same employer and employee in a month, the earliest activation date and the most recent cessation date of the month must be indicated.

2. In LR's estimation of job positions the temporary interruption in sending administrative declarations does not automatically close the position. This implies that the register might record as one long-lasting job position, which in reality is a series of job positions for which no cessation date and subsequent activation dates were communicated.

These cases where the LR pair has a single long-lasting position instead of multiple positions, possibly interspersed with interruptions, represent the lack of accuracy in Register we are aiming to correct by integrating CC source. The desired result of integration is to identify all the active positions in the domain of interest in the period and attribute the most accurate duration to them. A particularly problematic aspect of this exercise is that the variables in the link key and the variables that are intended to be corrected through micro-integration are partially overlapped.

The criteria underlying micro-integration are based on simplicity (it is assumed that the sources are essentially not subject to errors, with the sole exception of the well-known LR underestimation of short-duration jobs) and standardisation (positions relating to the population were treated according to the same integration rules).

In particular, it is assumed that:

1. A job position between a given employer and a given employee is considered to be the same in the two sources if it presents in both the same activation date or the same cessation date (definition of link key).
2. In the event of a conflict of either cessation or activation dates for job positions linked in the two sources (on AD or CD) the one closest to the linked date is chosen (in this way it is intended to correct the lack of information regarding very short jobs in LR). Job positions linked on CD with different ADs are treated as one job position in the integrated register with the common CD and as AD the one which is closest to the CD. Job positions linked on AD with different CDs are treated as one job position in the integrated register with the common AD and as CD the one which is closest to the AD[8].

8 An additional validation of the duration of the jobs could also take into account the consistency with other register variables on labour input. This might furtherly reduce the risk of errors in the estimation of job duration, and particularly of underestimating the length of employment.

3. Job positions may exist and may not be present in LR: a job position between a given employer and a given employee found in the CC source for which no job position between the same employer and employee with the same AD and/or CD is found in the LR is treated as a job position to be added to the integrated register (this rule is consistent to employment projection on CC bases only as described in Section 5).
4. Job positions may exist and may not be declared in source CC: a job position between a given employer and a given employee found in the LR source for which no job position between the same employer and employee with the same AD and/or CD is found in the CC is treated as a job position to be added to the integrated register (this rule intend to correct the under coverage in the CC source that lacks all the job positions activated before the birth of the system - March 2008 - that had no change afterwards).

Table 4.3 - Theoretical scheme of LR and CC integration methodology to develop the Extended LR (E-LR)

	A: Job positions with the same activation and cessation date	
Advantage	None	
Risks	None	
LR jobs	**A- Sally – 1 $_{LR}$**	
CC jobs	**A- Sally – 1 $_{CC}$**	
Linkage	A – Sally – 1 $_{LR}$ Link (on AD/CD) A- Sally – 1 $_{CC}$	
AD estimate	A-Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$ = A- Sally -1 (AD) $_{CC}$	
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$ = A- Sally -1 (CD) $_{CC}$	
E-LR jobs	**A- Sally – 1 $_{ELR}$**	
	B: Multiple short-term job positions not correctly identified in LR	
Advantage	Better accuracy in estimation	
Risks	None	
LR jobs	**A- Sally – 1 $_{LR}$**	
CC jobs	**A- Sally – 1 $_{CC}$**	**A- Sally – 2 $_{CC}$**
Linkage	A – Sally – 1 $_{LR}$ Link (on AD) A- Sally – 1 $_{CC}$	A – Sally – 1 $_{LR}$ Link (on CD) A- Sally – 2 $_{CC}$
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$ = A- Sally -1 (AD) $_{LR}$	A- Sally -2 (AD) $_{ELR}$ = Max (A- Sally -1 (AD) $_{LR}$, A- Sally -2 (AD) $_{CC}$) = A- Sally -2 (AD) $_{CC}$
CD estimate	A- Sally -1 (CD) $_{ELR}$ = Min (A- Sally -1 (CD) $_{LR}$, A- Sally -1 (CD) $_{CC}$) = A- Sally -1 (CD) $_{CC}$	A- Sally -2 (CD) $_{ELR}$ = A- Sally -2 (CD) $_{CC}$ = A- Sally -1 (CD) $_{LR}$
E-LR jobs	A- Sally – 1 $_{ELR}$	A- Sally – 2 $_{ELR}$

Table 4.3 cont. - Theoretical scheme of LR and CC integration methodology to develop the Extended LR (E-LR)

	C: A long lasting job position not declared in CC		
Advantage	Accuracy in estimation of occupational stocks		
Risks	None		
LR jobs	A- Sally – 1 $_{LR}$		
CC jobs			
Linkage	No link		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$		
E-LR jobs	**A- Sally – 1 $_{ELR}$**		
	D: A short-term job position not present in LR		
Advantage	Accuracy in estimation of occupational stocks		
Risks	Incorrect classification of an out-of-domain position in CC (over coverage()		
LR jobs			
CC jobs	A- Sally – 1 $_{CC}$		
Linkage	No link		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{CC}$		
E-LR jobs	A- Sally – 1 $_{ELR}$		
	E: Job positions of different duration with the same activation date or cessation date		
Advantage	Correction of approximation in the estimation of a date after editing and imputation process in LR		
Risks	Underestimation of the length of employment		
LR jobs	A- Sally – 1 $_{LR}$		
CC jobs	A- Sally – 1 $_{CC}$		
Linkage	A – Sally – 1 $_{RLR}$ Link (on CD) A- Sally – 1 $_{CC}$		
AD estimate	A- Sally -1 (AD) $_{ELR}$ = Max (A- Sally -1 (AD) $_{LR}$, A- Sally -1 (AD) $_{Cc}$) = A- Sally -1 (AD) $_{CC}$		
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$ = A- Sally -1 (CD) $_{CC}$		
E-LR jobs	A- Sally – 1 $_{ELR}$		
	F: Job positions with different activation and cessation date (maybe partially overlapping)		
Advantage	Accuracy in estimation of occupational stocks and flows (when jobs are not partially overlapping)		
Risks	Duplication		
LR jobs	A- Sally – 1 $_{LR}$		
CC jobs		A- Sally – 1 $_{CC}$	A- Sally – 2 $_{CC}$
Linkage	No link	No link	No link
AD estimate	A- Sally -1 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{LR}$	A- Sally -2 (AD) $_{ELR}$ = A- Sally -1 (AD) $_{CC}$	A- Sally -3 (AD) $_{ELR}$ = A- Sally -2 (AD) $_{CC}$
CD estimate	A- Sally -1 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{LR}$	A- Sally -2 (CD) $_{ELR}$ = A- Sally -1 (CD) $_{CC}$	A- Sally -3 (CD) $_{ELR}$ = A- Sally -2 (CD) $_{CC}$
E-LR jobs	A- Sally – 1 $_{ELR}$	A- Sally – 2 $_{ELR}$	A- Sally – 3 $_{ELR}$

Table 4.3 presents a diagram of the possible cases that may arise when integrating LR and CC sources to derive an Extended LR (E-LR). Using the notation of Table 4.1 (employer-employee-job source), the possible cases in the relationship between employer A and employee Sally (see note (a) of Table 4.1) as declared in the sources are examined and the applied integration rule is listed highlighting its advantages and risks. In paragraph 4.2, all risks due to linkage errors (Fellegi and Sunter, 1969; Arts, Bakker and van Lith, 2000) or other micro-integration errors, and their impact on the integrated basis will be examined.

Another critical issue regards companies involved in legal changes and their employees. In order to simplify the methodology treatment, it was decided not to go into the matter in depth in this context. In this exercise, an attempt was made to minimise the impact of the differences in the enterprise's declarations in administrative sources by aligning the dates with the events reported by BR for transferred workers: in particular, for each event, it was necessary to align in both sources the cessation date within the transferring company and the activation date within the acquiring company[9].

4.2 Evaluation of methodological choices

With regard to the quality of the estimates in LR for the 2016-2019 ad hoc longitudinal data and the E-LR longitudinal data, some statistics on longitudinal consistency of positions are given below (Table 4.4).

The E&I process described above and applied on the LR should be performed also on the E-LR, which is after the micro-integration of the two sources. However, this was not done in this experimental work. This implies that the number of violations of the edit rules on the dates is higher in the E-LR than in the LR. Nonetheless, the incidence of violations remains very low: even summed up across rules this incidence amounts to 0.8% of the total number of job positions (Table 4.4), one-third of the incidence of positions with violations found in LR pre-E&I. In a similar way, the number of partial overlaps that are introduced by the integration (case F in Table 4.3) is very

9 In E-LR this justifies the possible change of dates for about 550 thousands of pairs (1.63% of total pairs) in case there is no consistency with the figure estimated by BR. In essence, the comparison leads to change in about 150 thousands of pairs (0.49% of the total) and finds alignment for the remaining 445 thousands (1.14%).

limited as well. This confirms the substantial consistency between the sources of social security, underlying the LR, and the CC source.

Table 4.4 - Number of jobs with E&I rules' violations in LR and in E - LR data - Years 2016-2019

Violated rule	LR		E-LR	
	N	%	N	%
Rule 1 (consistency)	267	0.001	118,841	0.349
Rule 2 (consistency)	1,756	0.005	0	0
Rule 3 (consistency)	0	0	14,490	0.043
Rule 4 (non-overlapping)	143	0	177,639	0.521
Rule 5 (non-duplication)	0	0	0	0
None	34,067,197	99.99	38,816,768	99.21
Total	34,069,363	100	39,127,738	100

Source: LR and extended LR to CC

Regarding the employer-employee pairs, in the considered period, the number of pairs found only in the CC source amount to 3% of the total number of pairs in the E-LR database (case D in Table 4.3). By construction of the database, these are workers present in the CC source as employees of co-present companies only. The lack of employment signals in the LR could be explained by the absence of contributory coverage for them in the period or by a classification of the type of contract not aligned between the one declared in CC with that derived in the LR on the basis of social security sources. Further study on this issue, especially through comparison with the tax data, may indicate the strategy to follow for them. On the opposite side, 22.53% of total pairs are not found in CC (case C in Table 4.3). These are likely the job positions activated before the birth of the CC system, which have not undergone changes afterwards, and thus are unobserved by the CC source. The evidence that the average ratio between the number of job positions and the number of employer/employee pairs is very close to 1 (1.06) might be an indication that supports the hypothesis.

Table 4.5 - Number of employer-employee pairs and number of job positions classified by the presence of employer-employee pairs in the sources - Years 2016-2019

	N° pairs	Incidence % pairs	N° job positions
Pairs with at least one job in both sources	21,464,468	74.29	30,792,010
Pairs with LR only jobs	6,510,693	22.53	6,894,091
Pairs with CC only jobs	917,459	3.18	1,441,637
Total	28,892,620	100.00	39,127,738

Source: Extended LR to CC

Looking at the outcome of integration among the 74.3% of couples present in both sources (Table 4.6), they developed about 31 million job positions: 72% of them have both cessation and activation dates coinciding in the two sources (case A in Table 4.3), 11.0% of the positions have the same activation date, but different cessation date, while 3.3% have the same cessation date, but different activation date. The total number of positions linked on activation or cessation date between the two sources is 86.3%. By their nature and construction, greater coverage and accuracy in the measurement of stocks in the LR and better quality in the measurement of flows in the CC source are expected. Positions activated/closed in the period (eventful positions), which are not in CC, represent 1.21% of positions for co-present pairs, while stable positions located only in source CC are 0.42%.

Table 4.6 - Outcome of job position's integration for employer–employee pairs present in both sources. Years 2016-2019

	No		**% job positions**	
	Which of the tota	Total	Which of the total	Total
Same activation and cessation date in LR and CC				
Total		22,162,397		71.97
Same activation and different cessation date in LR and CC				
Different cessation date: CC choice	2,841,828		9.23	
Different cessation date: LR choice	555,851		1.81	
Total		3,397,679		11.04
Same cessation and different activation date in LR and CC				
Different activation date: CC choice	830,980		2.70	
Different activation date: LR choice	188,381		0.61	
Total		1,019,361		3.31

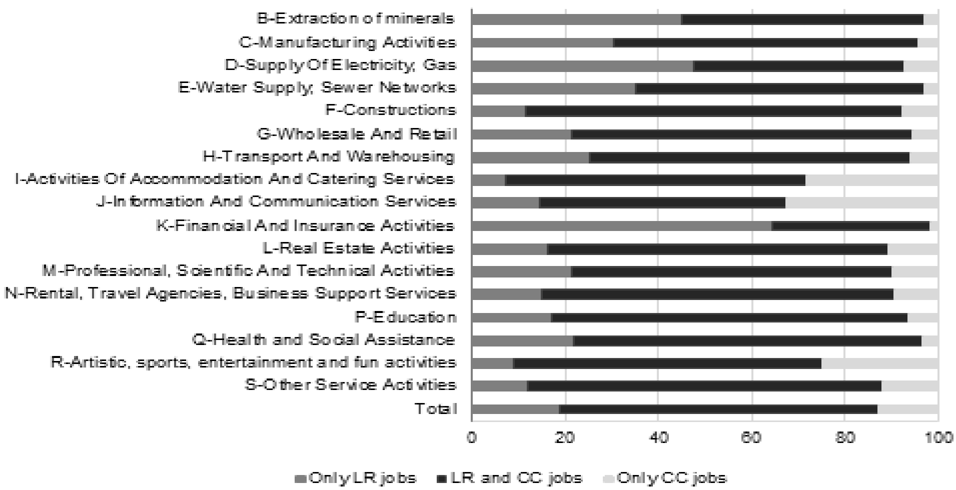
Table 4.6 cont. - Outcome of job position's integration for employer–employee pairs present in both sources. Years 2016-2019

	No		% job positions	
	Which of the tota	Total	Which of the total	Total
Different activation and cessation date in LR and CC				
LR only				
Stable positions in the period			0.62	
Eventful positions in the period			1.21	
Total		563,698		1.83
CC only				
Stable positions in the period			0.42	
Eventful positions in the period			11.43	
Total		3,648,875		11.85
Total		30,792,010		100.00

Source: Extended LR to CC

Looking now at the outcome of integration for all pairs[10] (Figure 4.1), the total number of positions linked between the two sources is the 67.93% on a total of 39,127,738 job position.

Figure 4.1 - Percentage composition of job positions in E-LR by source and economic activity of the employer. Years 2016-2019 (a)



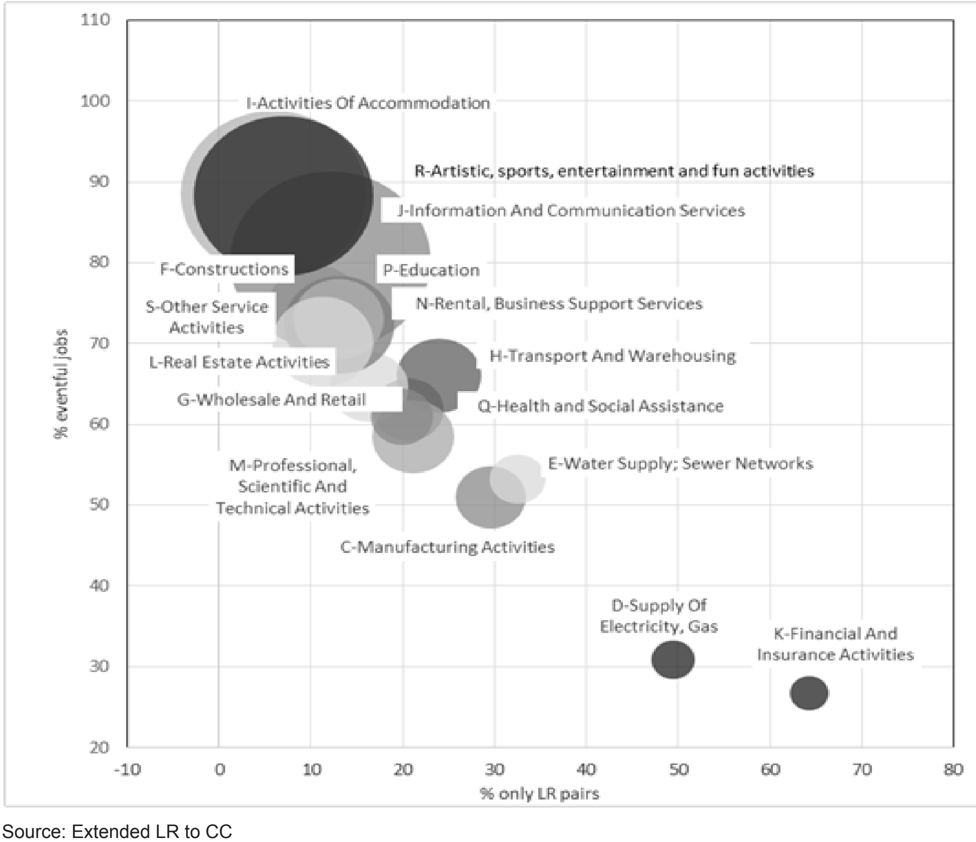
Source: Extended LR to CC

(a) The economic activity of the company is obtained using that attributed by BR for the most recent year available.

10 Correction made by merging positions that close on the same date even with different activation date is meant to correct also some overlapping position in CC source (0.3%)

The following figures show the correlation between the incidence of eventful positions, the incidence of only CC positions and the incidence of only LR pairs (Figure 4.2) and the correlation between the average duration of positions per pair and the incidence of only CC jobs. Figure 4.2 shows that as the incidence of eventful jobs increases the presence of only LR employer-employee pairs diminishes and the incidence of CC jobs only (size of bubbles) increases[11].

Figure 4.2 - Percentage incidence of eventful job position (Y value), only LR pairs (X value) and only CC job positions (circle area) by economic activity of employer. Years 2016-2019 (a)



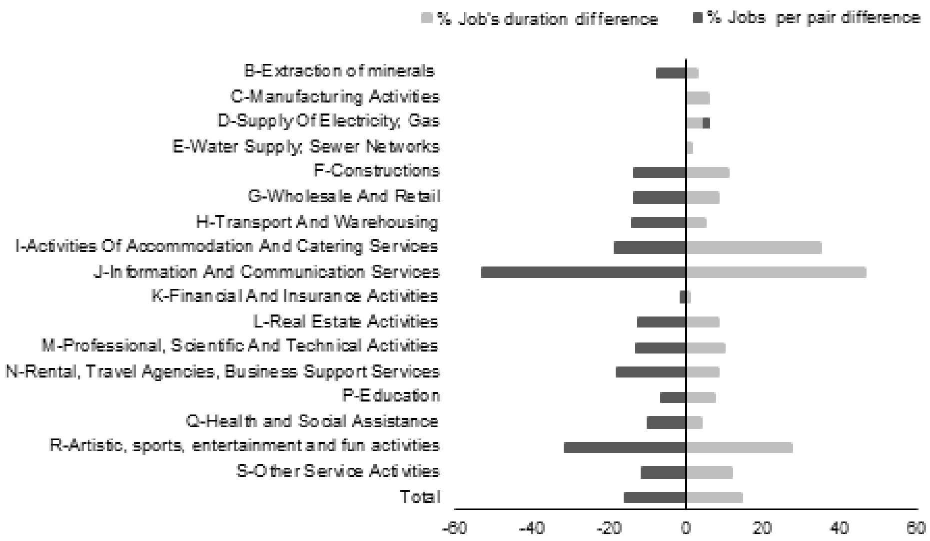
Source: Extended LR to CC

(a) The economic activity of the company is obtained using that attributed by BR for the most recent year available for observations in the LR and that declared in the CC source for the others.

11 From the figure are excluded some employer-employee pairs not present in the LR that declare in CC employer's activities not in the LR target population (Agriculture, Services to the person, Public Sector), underlining the problems of different classification by source due to misalignments in administrative declarations.

Figure 4.3 shows that the differences between LR and E-LR in the number of average positions per employer-employee pair and in the median duration of employment relationships are symmetrical. This evidence shows how the reduction in the duration of positions after integration with the CC source (16% in median) is due to a parallel increase in the number of relationships per pair (15%). Hence, the integration's aim to improve the accuracy in measuring short-term position seems to be accomplished.

Figure 4.3 - Percentage differences between LR and extended LR-CC in the number of average positions per employer-employee pair and in the median duration of job relationships. Years 2016-2019 (a)



Source: Extended LR to CC
(a) The economic activity of the company is obtained using that attributed by BR..

A joint reading of the two graphs highlights that, compared to LR, the percentage increase in the number of job positions per pair due to the integration is equal on average to 16% with the sectors with particularly high job turnover that record the largest increases. Sector R, for example, has an incidence of eventful jobs equal to 88% (Figure 4.2) and an increase in the number of job positions per pair of about 30% (Figure 4.3) aside from an incidence of only CC jobs of about 30%. The CC source, on the other hand, underreports the active positions in the low turnover sectors in the period (Supply of Electricity, Financial and Insurance Activities).

Below some evidences concerning the impact on integrated and corrected job positions (11.68% of total positions) of the methodological choices underlying integration are described. In particular, it is aimed at evaluating the possibility that the choice, in job positions characterised by the same pair and activation date, of the cessation date closest to the date of activation (*point a*)[12] or the choice, in job positions characterised by the same pair and cessation date, of the activation date closest to the cessation date (*point b*)[13] may underestimate the duration of the relationships (case E in Table 4.3). In order to assess the possible underestimate, the cessation date (activation date) in the source not chosen was compared with the maximum cessation date (minimum activation date) in the integrated base for the employer-employee pair in order to check whether several short positions in the integrated basis corresponded to a single position in the source.

Basically, the methodological choices seem to correct errors present in one of the two sources in most cases, although, of course, possible errors remain for a limited number of positions (about 183,612 positions corresponding to 0.47% of total positions and 4.02% of corrected positions). In general, the choice in LR to impute the cessation date (or activation date) only in the presence of a strong administrative signal (declaration of an activation date for the pair in the following months to impute the cessation date and the presence of a closing declaration to impute an activation date in the following months) has simplified and improved the integration of information from the most suitable source for making this attribution (CC source). Net of the few unresolved cases in the presence of inconsistencies between the sources the integration indeed ensured:

12 Considering *point a*, first we observe the positions for which the cessation date declared in CC prevailed (7.26%). Deepening each case, integration choices seem to be improving: remains only 1.7% (49,164 positions) for which the chosen method could underestimate the duration of the position of 45 days in median. Of them, 70% are fixed-term and, looking at some specific cases, there is a seemingly unjustified conflict between what is stated in UniEmens and what is present in CC. On the other hand, looking at the job positions for which the LR cessation date is chosen (1.42%), the possible approximation within the month for the dates imputed in LR could lead to a small error in the estimate (on average of 1 day, since however the approximation in LR concerns the days of the first week with contribution coverage in UniEmens) for 19% of the positions.

13 In deepening *point b* (the job positions combined on the cessation date presenting a conflict on the date of activation in the sources), it is studied below how this choice may have led to a potential underestimate of the duration of the positions in the integrated base. Looking initially at the positions gathered on the closing date for which the date of CC activation was chosen (2.12%), the possible underestimate in the duration of the position compared to that indicated in LR concerns 43.3% of the positions subject to correction, but for most of them (32.7%) the date of activation present in LR is before 2016 so imputed thought tax source and it has probably been corrected to update it with the most recent CC date (any active position previously is excluded from the database). Looking now at the merged job positions on the closing date for which the date of intake of LR was chosen (0.48%), the possible underestimate in the duration of the position compared to that indicated in CC concerns 54.2% of the positions submitted for correction, and for most of them (40%) as previously seen, the date present in CC is after 2016.

1. an improvement in the estimate of the activation date for all cases of positions activated before 2014 (for which the LR uses the date present in the tax source, which is arguably less precise than the UniEmens);
2. an improvement in the identification of short-term relationships that were collapsed into a single position in the LR;
3. an improvement in the estimation of cessation dates, sometimes not declared in the sources used in the LR and not corrected in the editing and imputation process.

The analysis of the effect of the integration of the CC source on job flows shows, as expected, a large increase in the estimated number of activations and cessation. On average on the four years considered the first increases of 18.5% and the second of 20.5% (Table 4.7). The effect on the stocks of employment, is more limited, since the main issue of the LR is that two or shorter job positions between an employer and an employee are misrepresented as one longer job position. In substituting the long job position of the LR with two or shorter job positions, accounted for in the CC, the measure of the employment stock, that is the overall labour input, does not change much. It is also interesting to notice that while there is an increment on the level of the employment stock, the difference in the change of it over two consecutive years is negligible, especially when we exclude the change due to the business demographics.

Table 4.7 - Differences on employment stocks and flows between LR and extended LR - Years 2016-2019 (Percentage points) (a)

	Years				
	2016	2017	2018	2019	2016-2019
	Extended LR *vs.* LR				
Job positions ceased	20.2	20.0	21.2	22.6	20.5
Job positions activated	18.0	18.1	19.2	20.7	18.5
Stock at 1/01 year t (a)	1.2	1.6	2.3	3.0	2.3
Stock at 1/01 year t+1 (b)	0.7	1.2	1.6	2.3	0.7
Changes (a-b)	-0.5	-0.5	-0.6	-0.7	-1.7
	Extended LR *vs.* LR - only for employers always active in the period				
Job positions ceased	19.9	19.4	19.6	21.1	20.0
Job positions activated	17.2	17.9	18.0	20.0	18.7
Stock at 1/01 year t (a)	0.5	0.5	0.3	0.5	0.3
Stock at 1/01 year t+1 (b)	0.4	0.8	0.5	0.3	0.7
Changes (a-b)	-0.2	0.4	0.3	-0.2	0.5

Source: LR and Extended LR to CC
(a) Positions active on a given date include all positions activated or closed on that day.

5. The projection of Labour Register microdata based on the integration of the Compulsory Communications

As reported in Paragraph 1, since the information related to all the months of year *t*-1 of UniEmens and the other social contributions sources are available at Istat in March of year *t*, what can be done with the Labour Register, due to the timing of the production process, per se is to produce preliminary estimates of year *t*-1 with a delay of about 8-10 months. Up to August of year *t*+1 hence those are the most up-to-date estimates derived by the LR. The use of CC, whose information flows day by day into the system, may greatly improve this timeliness by projecting the set of job positions in the register referred to the year *t*-1 potentially up to a very recent time.

The methodology underlying the projection is the same described in the previous paragraphs in relation to the integration process. The projection on a CC basis can be seen in fact as the integration between the positions in the register updated to year *t*-1 and those present in the CC updated to year *t*.

The aim of this paragraph is to describe the effects of the integrated methodology on projecting the set of job positions with the use of CC up to the very last period covered by the CC source. As above mentioned, all the analyses are based on LR data 2016-2019 and CC data that cover from 2012 up to July 2020. The final scope is to project the LR referred to 2019 up to July 2020 in order to measure the most recent evolutions of employment in the first COVID-19 year. This estimation can be compared and evaluated on the basis of external data sources such as the employment estimates of the Labour Force Survey (LFS) and those of the Oros process (see also Anastasia, 2016*a*). In addition, the projections have been simulated also for the year 2019 on the basis of the LR data up to the year 2018 and CC data up to 2019. In this case, the projection can also be evaluated using all the data of the E-LR for 2019 in order to control any source of distortion due to the use of CC information not integrated with LR in year *t*. The target population of enterprises and jobs is the same one described for the integration: the set of firms belonging the sections B to S of classification NACE Rev 2, with the exceptions of temporary agencies, and all employee jobs in the target population of enterprises, with the exceptions of jobs-on-call. Operationally, the basis of the projection is the set of job positions in the E-LR belonging to the enterprises *active with employees* at year *t*-1, according to the BR referred to that year (set A). Moreover, to provide a full account of the employment changes, the projections include the set of job positions belonging to the enterprises included in the target population that have become *active with employees* in year *t* (that is those actually born in year *t* with at least one employee or those that have passed from not having employees in year *t*-1 to having employees in year *t*) (set B). In the absence of the BR for year *t*, the *newborn* enterprises are identified thanks to information from the Oros process[14].

The application of the integration methodology in projecting employment up to year *t* (each job position is updated to the last available state in year *t* using the CC information only) implies that:

a. all job positions still active at the end of time *t*-1 in the E-LR that have no further information for *t* in the CC source are still considered active at the end of time *t* (since there has been no communication of cessation in CC);
b. all job positions still active at the end of time *t*-1 in the E-LR for which the CC signals a cessation at year *t* are ceased on that day;
c. all job positions activated in year *t* are accounted active up to their cessation date in year *t* or up to the end of the observation period if no cessation date has been communicated.

The methodology implies assuming that for all job positions belonging to enterprises ceased in year *t*-1 or year *t* the CC system has received communication of cessation. This assumption does not prove true as the database seems to lack the cessations of job positions for a certain number of enterprises that are ceased according to the information available in the external administrative source used by the Oros survey (in 2020 about 28%

14 The simulation exercise on 2019 and 2018 has, however, shown that to choose the set of companies to be involved in the projection cannot be considered only the presence of the company in BR in year *t*. In fact, the estimates obtained considering all the companies present in the BR target population in the reference year show an underestimated structural in the CC projection for both 2018 and 2019. The difference in the number of individuals declared in CC compared to those declared in LR in some companies in 2018 is due to the entry of non-new companies in the BR. When the company enters the field of observation the cut on BR in the two years generates new virtual relationships (all the company's relationships although active for years). If these relationships are cut off by BR from LR in year *t*-1 and have been active and stable for many years, they are not found in the CC data in year *t*, leading to an underestimate of the activated positions and an underestimate of variations in projection. To avoid this problem, the projection must include only the companies present in year *t*-1 that have not ceased and the *newborn* ones, without considering the statistical variations in economic activity.

of companies ceased according to Oros administrative information have still active job positions in CC). In addition, a distortion in the estimates can be seen if compared with external sources. This distortion is compatible with an under-declaration of the termination of relationships in companies no longer active with employees in 2020, an underestimation probably more problematic in 2020 and not evident in the comparisons carried out so far. In order to overcome this issue, the cessation dates have been corrected for all the job positions in companies that are no longer active (according to Oros process) and that seems to be still active in CC. The cessation date was set equal to the closing date of the company.

After this correction, the changes in 2020 approaches those estimated in LFS at least at the level of the total economy (while the differences at the level of economic activity might be due to classifications errors) and in Oros survey[15].

Table 5.1- Employment level at 1st semester of 2020 and changes of the first semester 2020 on the first semester of 2019 in the extended LR (a), LFS and Oros (b)

NACE	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)
	Oros survey		LFS survey		E-LR	
Total (B-S excluding O)	12,602	-1.6	12,212	-1.3	12,564	-1.5
Industry and market services (B-N)	11,484	-1.7	11,572	-1.5	11,451	-1.6
Total (B-F)	4,472	-0.3	5,075	1.0	6,842	-0.6
Industry in the strict sense (B-E)	3,621	-0.5	4,235	0.2	3,736	-0.7
Manufacturing activity (C)	3,325	-0.6	-	-	3,408	-0.9
Supply of electricity, gas, steam and air conditioning (D)	83	-0.5	-	-	87	-0.6
Water supply, sewerage, waste management and rehabilitation activities (E)	195	1.2	-	-	210	1.7
Construction (F)	851	0.6	841	5.1	861	0.8
Total (G-S excluding O)	8,130	-2.3	7,136	-2.9	7,966	-2.1
Market services (G-N)	7,013	-2.5	-	-	6,853	-2.4
Retail sale repair of motor vehicles and motorcycles (G)	2,194	-0.1	2,061	-1.4	2,244	-0.8
Transport and storage (H)	1,033	-1.5	975	-3	1,029	-1.5
Accommodation and catering activities (I)	989	-10.5	918	-12.7	1,057	-9.6
Information and communication services (J)	506	-0.6	501	4.3	522	0.6

15 Furthermore, it can be observed that the variations calculated on the same domain for unweighted levels (active positions at 31 July, for example) are not significantly different from those for average levels.

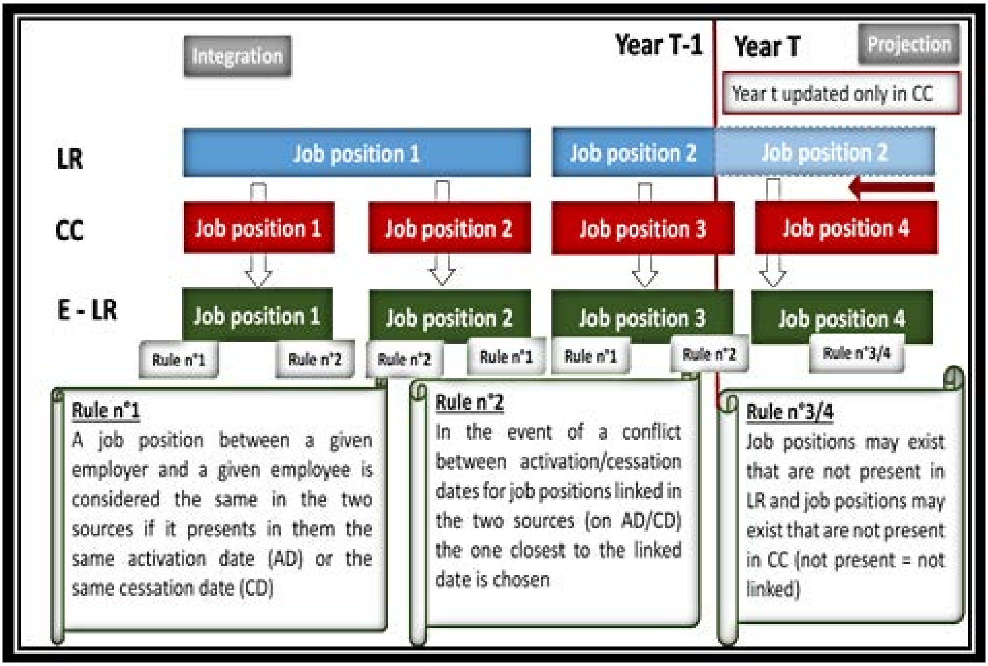
Table 5.1 cont. - Employment level at 1st semester of 2020 and changes of the first semester 2020 on the first semester of 2019 in the extended LR (a), LFS and Oros (b)

NACE	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)	Stock at I semester of 2020 (thousands)	Change (%)
	Oros survey		LFS survey		E-LR	
Financial and insurance activities (K)	455	-0.3	531	2.1	459	-1.4
Real estate, professional and rental activities (L-N) (c)	1,836	-2.3	-	-	1,540	-1.2
Of which: job-on-call	296	-7.8	-	-	-	-
Education, health and social work, arts and other service activities (P-S)	1,118	0	-	-	1,113	-0.5

Source: Extended LR to CC, Labour force survey (LFS), Oros process
(a) Changes calculated on the basis of E-LR refer to positions weighted by their duration in the period to make them comparable with those on the basis of Oros survey.
(b) Oros and LFS survey data are ad hoc elaborations carried out in order to make the different sources comparable.
(c) Job-on-call and job positions in temporary agencies are not included.

In order to summarise the methodology adopted, the main steps and choices are synthetically listed in the following table/prospect.

Figure 5.1- Main methodological rules of LR and CC integration to develop the Extended LR (E-LR) and its projection



6. Measuring stocks, flows and changes in employment: the evolution of the labour market in the first part of the pandemic crisis

The procedures of integration of CC data into the Labour Register and the projection of job positions, described in the above paragraphs, enable to analyse the stocks and flows dynamics up to a very recent period measured with the precision of daily intervals.

Notice that this full stock-flows accounting of job positions is not possible with any of the existing data sources. The Labour Force survey, in fact, can only represent the worker dynamics and not the job position dynamics and not with the precision of daily intervals. Moreover, being a sample survey, these estimates would be accurate only for quite large aggregates. As for the CC data, for instance in the form of the CICO dataset, they can provide a daily account of the job flows and of the change of the job stocks, but not of the level of the job stocks.

In this paragraph, in order to provide an example of the analysis possibilities, the database built is used to study the evolution of the labour market up to July 2020, using the LR data up to December 2019 and the CC data up to July 2020.

But before describing the results it is useful to recollect the basic relationships in stocks and flows accounting (Baldi *et al.*, 2018; Anastasia, 2016*b*). In any day, *t*+1, the number of job positions (at the end of the day), is equal to the number of job positions (at the end of the day) of day *t* plus the activations of jobs recorded in day *t+1* minus the cessations of jobs recorded in day *t*[16]. In formula:

$$J_{t+1} = J_t + A_{t+1} - C_t \tag{1}$$

The change in the number of jobs between two consecutive days is thus equal to the number of activations of day *t*+1, minus the number of cessations of day *t*.

$$\Delta J_{t+1} = A_{t+1} - C_t \tag{2}$$

16 The difference in timing between Activations and Cessations depends on the fact that while the activations are counted as active job positions from the very day in which they are recorded (that is these jobs start to be active from the activation day), the cessations are still in the count of jobs in the day they are recorded and are no longer in the set of active jobs from the day after.

By recursively substituting in [1] it is easy to obtain the relationship over a given period of time. For instance over a year:

$$\Delta J_{(t,t+365)} = J_{t+365} - J_t = A_{(t+1,t+365)} - C_{(t,t+364)} = \sum_{i=1}^{365} A_{t+i} - \sum_{i=1}^{365} C_{t+i-1} \tag{3}$$

And the relative change over a year of any given day t can be written as:

$$\frac{J_{t+365} - J_t}{J_t} = \frac{A_{(t+1,t+365)}}{J_t} - \frac{C_{(t,t+364)}}{J_t} \tag{4}$$

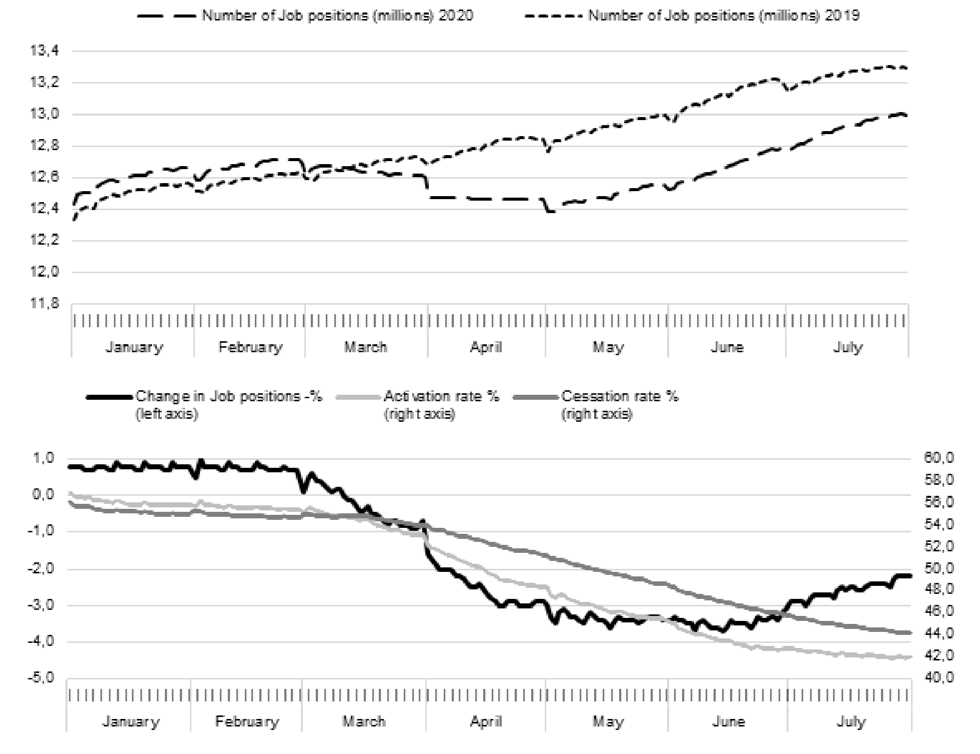
That is the relative (net) change of job positions is equal to the (gross) activation rate minus the (gross) cessation rate.

In Figure 6.1, where the daily dynamics of the first 7 months of 2020 is shown compared to the same period of 2019, the number of job positions (the stock measure) shows an upward trend during 2019 (also in the months not shown in the graph) and the beginning of 2020. The average stock for the first two months of 2020 is indeed higher than the equivalent figure for 2019 of about 95 thousand job positions. Starting from March 2020, when due to the policies of lock-down the economic effect of the pandemic crises has begun, there has been a sudden trend reversal that rapidly brought the daily number of job positions below the respective figure of 2019. The maximum distance is recorded for the 14th of June when the figure for 2020 is 484 thousand job positions lower than that of 2019, with a value of -3.7% in terms of per cent change. In the subsequent period there has been a slow but progressive recovery of the trend up to the end of the period, the 31st of July when the level of job positions reaches 13 million with a difference of -294 thousand (-2.2%) with respect to the same day of the year before (Baldi *et al.*, 2020).

The study of the job flows in the period January – July, which lie behind the evolution of the stocks shows that both activations and cessations started decreasing in the last days of February 2020. In particular, the drop in the levels of job positions is driven by an accelerated decrease of the job activations whose speed diminished only by mid-June. This dynamic of activation of new jobs is likely due the immediate worsening of the economic expectations of the enterprises at the outset of the crisis. The cessations also dropped but at a much lower rate due, on one side, to the policy of prohibition of dismissals that halted all termination of permanent jobs not due to the will of the workers and, on the other side, to the reduced inflows of short-term jobs explained above.

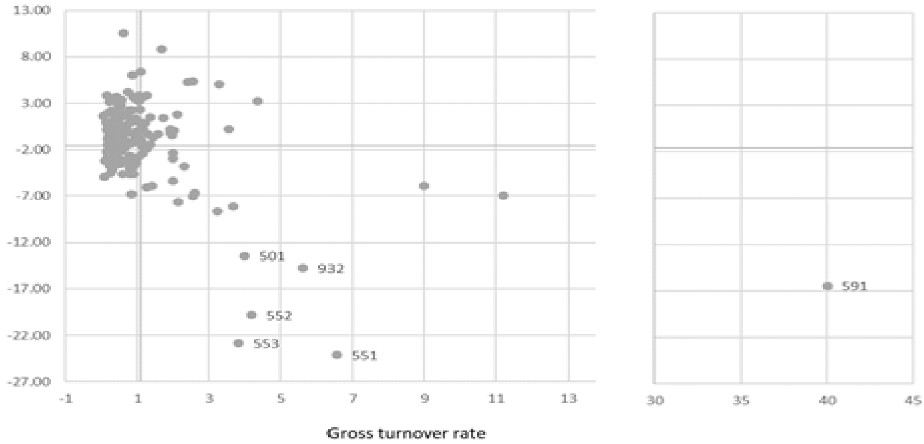
The availability of a Register that exhaustively covers the whole population of jobs in the private sector allows to break down the analysis, on various dimension, at a very detailed level. Figure 6.2 plots, at the level of 3-digit NACE sectors, the per cent change in the number of job positions against the 2019 job turnover rate (equal to the sum of the job activation rate and the job cessation rate). The figure very clearly shows that the sectors most heavily hit by the crises are those with a very high job turnover such as some activities of *Accommodation and food services*, some sectors of the *Information and communication services and Arts entertainment and recreation*. Besides having been particularly targeted by the lockdown policies, these sectors show the most dramatic decrease in the number of job positions due to their very large share of fixed-term contracts for which providing an on-the-job protection policy was not possible.

Figure 6.1 - Daily stocks of job positions and year-to-year daily changes – total, activations and cessation rate. January-July 2019 *vs.* January-July 2020



Source: Extended LR to CC

Figure 6.2 - Job positions net change rate (January-July 2020 vs. January-July 2019) vs. Gross turnover rate of 2019 (a)



Source: Extended LR to CC
(a) The X-axis is reported with a break in it. This gap is necessary to show the exceptional outlier value of the NACE group 591 without compressing the scale of the axis.

7. Concluding remarks

The experimentation performed in this work has allowed the integration of two of the main sources for the analysis of the labour market in Italy, extending the Labour register to include Compulsory Communications.

For the very first time in Italy and, to our knowledge, in the whole world, it was built a very timely employer-employee register from which it is possible to derive a full stocks-flows accounting of employment in terms of jobs. The database has been used to shed a light on the evolution of the labour market in the first part of the COVID-19 crisis in 2020 allowing us to display the exact daily timing of the crisis not only in terms of absolute changes in the number of jobs but also in terms of relative numbers and highlighting the contribution of gross flows on net flows. Moreover, it has uncovered, at a very detailed breakdown in terms of NACE sectors, the relationship and the probable causes between the job turnover and the reduction in employment. These analyses are only examples of the potential of such a database. A number of more refined studies are possible by using the characteristics of the employer, the employee and the job and breaking them down into finer classes.

In brief, the integration has allowed reaching two main purposes: the correction of the LR for the very short duration job positions and the projection of the job stocks and flows up to the last date available in the CC source. Regarding the first purpose, in the considered period 2016-2019, a total of about 3,648,875 job positions (11.85% out of the total) were added to the LR.

As explained in the text, since the bulk of this addition is likely to consist in fragmenting in job positions of short duration what has beforehand accounted as long job positions, it does not affect much the level of average labour input measured in a year by the LR (in the period the average increase is of 1.7%). Regarding the second purpose, considering the very low number of records for which at least one edit rule failed, the projection is expected to be quite accurate. Moreover, the comparison with the dynamic shown by other stock indicators is very encouraging.

The methodology can be improved in order to try to lessen even further the micro-integration errors also with the application of the longitudinal E&I corrections of the dates of activations and cessations to the Extended LR.

Moreover, the difficult challenge constituted by the corporate transformations may be enhanc ed in order to align the two input sources before the linkage procedure.

A number of new analytical challenges can be taken up.

Firstly, it can be studied whether it is possible to project other measures of labour input, such as those traditionally calculated in the Labour Register.

A second and very interesting area of the study refers to the possibility to pass from stocks and flows accounting in terms of jobs to one in terms of workers in a unified framework.

Thirdly, with the enlargement of the LR, extending the exercise and the analysis to other sectors starting from the public one will be possible.

The fourth domain for future work is the integration of CC variables in the LR and the use of the first source to correct the information of the statistical register.

References

Anastasia, B., L. Bertazzon, M. Gambuzza, e M. Rasera. 2016*a*. "Grammatica delle Comunicazioni Obbligatorie /4. Guida ai confronti con le altre fonti statistiche sul mercato del lavoro". *Working Paper*, N. 4. Venezia, Italy: Veneto Lavoro - Osservatorio & Ricerca.

Anastasia, B., G. Emireni, M. Gambuzza, S. Maschio, e M. Rasera. 2016*b*. "Grammatica delle Comunicazioni Obbligatorie /3. Guida alle elaborazioni a partire dai dati di flusso". *Working Paper*, N. 3. Venezia, Italy: Veneto Lavoro - Osservatorio & Ricerca.

Arts, K., B.F.M. Bakker, and E. van Lith. 2000. "Matching administrative registers and household surveys". In Boon, M., P.G. Al, B.F.M. Bakker, and L. Hoeksma (*Eds*.). "Integrating administrative registers and household surveys". *Netherlands Official Statistics,* Volume 15, Summer 2000, *Special Issue*: 16–22.

Bakker, B.F.M. 2011. "Micro-integration. State of the art". In *ESSnet on Data Integration. WP1 - State of the Art on Statistical Methodologies for Data Integration*: 77-107. Luxembourg: Eurostat.

Bakker, B.F.M., and P.J.H. Daas. 2012. "Methodological Challenges of Register-based Research". *Statistica Neerlandica*, Volume 66, Issue 1: 2-7.

Baldi, C., C. Ceccarelli, S. Gigante, and S. Pacini. 2018. "The development of the Italian Labour register: principles, issues and perspectives". In Abbruzzo, A., E. Brentari, M. Chiodi, and D. Piacentino (Eds.). *Book of Short Papers SIS 2018* (49th Scientific Meeting of the Italian Statistical Society, Palermo, Italy, 20th – 22nd June 2018): 514-522. London, UK: Pearson.

Baldi, C., G. De Blasio, G. Di Bella, A. Lucarelli, and R. Rizzi. 2014. "Turning the Compulsory Communication Data into a Statistical System". In Crescenzi, F., and S. Mignani (*Eds*.). *Statistical Methods and Applications from a Historical Perspective. Selected Issues*: 217-226. Cham, Switzerland: Springer International Publishing, *Studies in Theoretical and Applied Statistics, Selected Papers of Statistical Societies*.

Baldi, C., S. Gigante, S. Pacini, e R. Rizzi. 2020. "Approfondimento: La dinamica integrata di stock e flussi occupazionali nelle imprese dei settori dell'industria e dei servizi". In Ministero del Lavoro e delle Politiche Sociali –

ML, Istituto Nazionale di Statistica – Istat, Istituto Nazionale della Previdenza Sociale – INPS, Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro – INAIL, e Agenzia Nazionale per le Politiche Attive del Lavoro – ANPAL (*a cura di*). *Il mercato del lavoro 2020. Una lettura integrata*: 57-62. Roma, Italy: Istat. https://www.istat.it/it/archivio/253812.

Baldi, C., D. Ichim, F. Pintaldi, M.E. Pontecorvo, F. Rapiti, and R. Rizzi. 2018. "Digging into labour market dynamics: toward a reconciliation of stock and flows short-term indicators". In Abbruzzo, A., E. Brentari, M. Chiodi, and D. Piacentino (Eds.). *Book of Short Papers SIS 2018* (49th Scientific Meeting of the Italian Statistical Society, Palermo, Italy, 20th – 22nd June 2018): 523-530. London, UK: Pearson.

Batini, C., and M. Scannapieco. 2006. *Data Quality: Concepts, Methodologies and Techniques*. Berlin, Germany: Springer, *Data-Centric Systems and Applications*.

Daas P.J.H., and S. Ossen. 2011. "Report on methods preferred for the quality indicators of administrative data sources". *Deliverable 4.2*, Dissemination level: PU. BLUE-Enterprise and Trade Statistics - BLUE-ETS. Brussels, Belgium: European Commission, *European Research Area*.

Daas, P.J.H., S. Ossen, R. Vis-Visschers, and J. Arends-Tóth. 2009. "Checklist for the Quality evaluation of Administrative Data Sources". *Discussion paper* (09042). The Hague, The Netherlands: Statistics Netherlands – CBS.

European Statistical System Committee – ESSC. 2014. "ESS Vision 2020. Building the future of European statistics". *21st Meeting of the European Statistical System Committee*, Luxembourg, 14th and 15th May 2014.

Eurostat. 2003. "Item 4.2: Methodological Documents. Definition of quality in statistics". *Working Group* "Assessment of quality in statistics". Sixth meeting, Luxembourg, 2-3 October 2003.

Fellegi, I., and A. Sunter. 1969. "A theory for record linkage". J*ournal of the American Statistical Association*, Volume 64, N. 328: 1183–1210.

Rosén, B., and E. Elvers. 1999. "Quality Concept for Official Statistics". In Kotz, S., C.B. Read, and D.L. Banks (Eds.). *Encyclopedia of Statistical Science. Update Volume 3*: 621-629. Hoboken, NJ, U.S.: John Wiley and Sons.

Statistics Canada. 2002. *Statistics Canada's Quality Assurance Framework*. Ottawa, Ontario, Canada: Statistics Canada.

United Nations Economic Commission for Europe - UNECE. 2007. *Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics*. Geneva, Switzerland: United Nations.

Van der Laan, P. 2000. "Integrating administrative registers and household surveys". In Boon, M., P.G. Al, B.F.M. Bakker, and L. Hoeksma (*Eds*.). "Integrating administrative registers and household surveys". *Netherlands Official Statistics*, Volume 15, Summer 2000, *Special Issue*: 7–15.

Wallgren, A., and B. Wallgren. 2014. *Register-based Statistics: Statistical Methods for Administrative Data, 2nd Edition*. Hoboken, NJ, U.S.: John Wiley and Sons, *Wiley Series in Survey Methodology*.

Wang, R.Y., and D.M. Strong. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers". *Journal of Management Information Systems*, Volume 12, N. 4: 5-33.

Zhang, L.-C. 2012. "Topics of statistical theory for register-based statistics and data integration". *Statistica Neerlandica*, Volume 66, Issue 1: 41-63.

The Istat Economic Sentiment Indicator: a new proposal

Solange Leproux[1], Adriano Pareto[1], Claudia Rinaldelli[1]

Abstract

The Istat Economic Sentiment Indicator (IESI) is a measure of the confidence climate in the Italian production sector. It is the result of the aggregation of the variables used in the calculation of the confidence climate indices of manufacturing, construction, service, and retail trade sectors. The current procedure adopted for the calculation of the IESI can determine discrepancies between the evolution of the composite index and the dynamics of the sector-level confidence climates. Although these discrepancies are explainable from a methodological point of view, they can create considerable problems in terms of interpretation and communication of the results. This work proposes a new method for calculating both the sectoral confidence climates and the IESI that ensures consistency in the evolution of the indicators and guarantees an effortless interpretation of the results

Keywords: Confidence Indicators, Composite Indicators, Business Cycle.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_2.2022.03

1 Solange Leproux (sleproux@istat.it); Adriano Pareto (pareto@istat.it); Claudia Rinaldelli (rinaldel@istat.it), Italian National Institute of Statistics – Istat.

Although this article is the result of all the authors' commitment, the paragraphs are attributed as following: 1, 6, 7 and 8 to Solange Leproux; 3, 4 and 5 to Adriano Pareto; 2 to Claudia Rinaldelli.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

The Istat Economic Sentiment Indicator (IESI) is disseminated on a monthly basis with the Istat press release “Consumer and Business Confidence” starting from June 2012. It was conceived in 2011 when the direction of the business and consumer surveys, which are all part of a harmonised European programme coordinated by the European Commission[2], was passed from the Italian Institute of Economic Studies and Analyses (Isae) to the Italian National Institute of Statistics (Istat)[3].

This indicator was created with the aim of providing a composite measure of the state of the entire Italian production sector confidence, but over time it has also proved capable of providing coincident or even leading signals of our national economic cycle movements (Leproux and Matera, 2015).

The Istat produces a range of indices on sentiment in the Italian production sector. 11 series are available. These are currently aggregated into the IESI. As well there are four sectoral indices: the manufacturing confidence climate, the construction confidence climate, the market services confidence climate and, finally, the retail trade confidence climate. These use groups from the same 11 series. There are 3 each for manufacturing, market services, retail trade, and 2 in construction.

A difficulty that has been found is that movements in the aggregate index may not agree with the sectoral indices. For example, in May 2016, the IESI showed a slight increase (+0.7 percentage points) against a slight decrease in all the sector-based confidence climate indices. Later, in November 2019, the IESI again showed a slight increase (+0.2 percentage points) against a substantial stability in the retail trade and market service confidence indices and a slight reduction in the manufacturing and construction sector indices.

This does not seem satisfactory. As will be shown it occurs because the 11 series are modified in different ways when forming the sectoral indices to that when forming the aggregate index.

---

2 The programme is governed by the Decision of the European Commission C (97) 2241 of 15 July 1997 and by the Communication of the Commission COM (2006) of 12 July 2006.

3 The Isae suppression and the merger of the Institute with Istat were provided for by art. 7, paragraph 18, of the decree-law n. 78 of 2010, converted with amendments by law n. 122 of 2010.

In this paper a new IESI is computed that ensures that such a discrepancy does not happen. The construction of the new index preserves the essence of the existing methodologies. So it is a parsimonious[4] solution that gives an easy interpretation and communication of the state of confidence both at a sectoral and aggregate level. The results of the new IESI in terms of ability to capture fluctuations in the aggregate economic activity, are also presented. Finally, to complete the analysis, the performance tests in tracking the reference series are carried out using the current version of the indicator (current IESI).

The organisation of the work is as follows: Section 2 points at the criteria and the elaboration phases to be followed for the construction of a composite index in the existing methodological framework of the IESI; Section 3 shows the current calculation scheme. It describes in detail the present methodologies for the elaboration of both the sector-based confidence climates and the IESI index; Section 4 provides both the description and an example of the new methodologies proposed to replace the current ones; Section 5 presents a comparison between the new IESI and the current one in order to confirm the agreement between the results; Section 6 shows the results of the performance analysis carried out on the new versions of the indicator. For the sake of completeness, the same Section also reports the results obtained by subjecting the current version of the IESI to the same performance tests. Finally, Section 7 illustrates the authors' opinions on how much the effects of the COVID-19 pandemic play on the performance test results. Some conclusions are presented in Section 8.

4 The principle of parsimony states that the composite index must be as simple as possible, to allow easy interpretation of the results. See Mazziotta and Pareto (2020).

2. The revision of the Istat Economic Sentiment Indicator (IESI)

The IESI is part of a complex framework, because its methodology has always been discussed and partially guided by the Joint Harmonised EU BCS programme, and it has been disseminated for some years now.

This work on the IESI could not be configured as the ordinary construction of a composite indicator (as for other composite indicators already built in Istat), but rather as the corrective intervention of some phases of the current procedure, to avoid new inconsistencies, without however drastically intervening in a pre-existing methodological framework discussed at European level.

Therefore we tried to insert the criteria and the construction phases of a composite index in the existing methodological framework of the IESI. In particular, in order for the composite index to be as simple as possible and provide results consistent with the performance of the individual components, the processing to be carried out on the data must be reduced to the bare minimum[5].

Here these phases are summarised briefly (Mazziotta and Pareto, 2017; 2020):

1. *Defining the phenomenon to be measured*. The definition of the concept should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators. In this case, we aim to measure the state of confidence of the entire Italian productive sector, based on 4 sub-groups of confidence climate indicators of manufacturing, construction, services and retail trade sectors.
2. *Selecting a group of individual indicators*. The selection is generally based on theory, empirical analysis, pragmatism or intuitive appeal. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility and so on. The selection phase is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. A group of 11 individual indicators of confidence climate were considered to calculate the IESI.

5 Only one normalisation method must be applied to the data matrix and no further transformation of the obtained scores should be carried out, as they are already normalised (Terzi *et al.*, 2021).

3. *Normalising the individual indicators*. This phase aims to make the indicators comparable. Normalisation is required before any data aggregation as the indicators in a data set often have different measurement units. Therefore, it is necessary to bring the indicators to the same standard, by transforming them into pure, dimensionless, numbers. Besides, since some indicators may be positively correlated with the phenomenon to be measured (positive polarity), whereas others may be negatively correlated with it (negative polarity), we have to transform the indicators so that an increase in the normalised indicators corresponds to increase in the composite index. The main normalisation methods are: standardisation (or z-scores), re-scaling (or Min-Max) and distance to a reference (or index numbers) (OECD and JRC, 2008). Standardisation and re-scaling are more commonly used when indicators have different measurement units and/or magnitude (*e.g.* GDP per capita and Life Expectancy); whereas index numbers are commonly used when indicators are of the same nature (*e.g.* prices or quantities). The new IESI is based on a normalisation of individual indicators by index numbers.
4. *Aggregating the normalised indicators*. It is the combination of all the components to form one or more composite indices. This phase requires the definition of the importance of each individual indicator (weighting system) and the identification of the technique (compensatory or non-compensatory)[6] for summarising the individual indicator values into a single number. Different aggregation methods can be used, such as additive methods (compensatory approach) or multiplicative methods and unbalance-adjusted functions (non-compensatory or partially compensatory approach). The methodological framework of the IESI is based on a compensatory approach and individual indicators are aggregated by a weighted arithmetic mean. The weighting system uses as weights the corresponding of Value Added as defined by National Accounts of each sector. Since March 2015, 2012 Value Added data are used.

6 Compensability among individual indicators is defined as the possibility of compensating any deficit in one dimension with a suitable surplus in another (OECD and JRC, 2008). Thus we can define an aggregation approach as ‘compensatory’ or ‘non-compensatory’ depending on whether it permits compensability or not (Casadio Tarabusi and Guarini, 2013).

5. *Validating the composite index*. Validation phase aims to assess the robustness of the composite index, in terms of capacity to produce a correct and stable measure, and its discriminant capacity (Influence Analysis and Robustness Analysis). A comparison of the new IESI with the current IESI and the Italian GDP was performed.

In the next Section, we show how the current IESI is calculated, whereas in Section 4 the new calculation scheme is described.

3. The current calculation scheme

The steps for calculating the current IESI and the confidence climate indicators of manufacturing, construction, services and retail trade sectors are the following. They were developed at different times by different teams.

3.1 Computing the confidence climate indicators

Let $\mathbf{X}^k_{n,m_k} = \{x^k_j\}$ be the matrix of seasonally adjusted balances of the climate k, where:

$$-100 \leq x^k_{ij} \leq 100$$

and x^k_{ij} is the balance value for month i (i=1, …, n) and variable j (j=1, …, m_k) of climate k (k=1, 2, 3, 4).

To have positive values, we move to the transformed matrix $\tilde{\mathbf{X}}^k_{n,m_k} = \{\tilde{x}^k_{ij}\}$, with: $\tilde{x}^k_{ij} = x^k_{ij} + 100$, and then construct the matrix of the means $\overline{\mathbf{X}}_{n,4} = \{\overline{x}_{ik}\}$, where:

$$\overline{x}_{ik} = \frac{1}{m_k} \sum_{j=1}^{m_k} \tilde{x}^k_{ij}$$

The 4 confidence climate indicators are given from the normalised matrix of index numbers with base 2010[7] $\mathbf{C}_{n,4} = \{c_{ik}\}$ where:

$$c_{ik} = \frac{\overline{x}_{ik}}{\frac{1}{12} \sum_{l \in 2010} \overline{x}_{lk}} 100$$

Note that in this scheme the order of the phases 3 (normalisation) and 4 (aggregation) of Section 2 is not respected, as individual indicators are first aggregated into means, and then the means are transformed into index numbers.

7 The base update has been planned for 2024, subject to available data.

3.2 Computing the current IESI

Given the original matrix $\mathbf{X}^k_{n,m_k} = \{x^k_{ij}\}$, the normalised matrix of z-scores is constructed $\mathbf{Z}^k_{n,m_k} = \{z^k_j\}$, with:

$$z^k_{ij} = \frac{x^k_j - \mathrm{M}^k_j}{\mathrm{S}^k_j} \tag{3.1}$$

where M^k_j and S^k_j are respectively the mean and standard deviation of variable j of climate k.

Let $\mathbf{W}_4 = (w_1, w_2, w_3, w_4)$ be the array of the weights of the 4 confidence climates, with:

$$\sum_{k=1}^{4} w_k = 1 \qquad \text{and} \qquad 0 < w_k < 1\,.$$

The weighted mean of the 11 normalised indicators is calculated as follows:

$$\overline{\mathrm{Z}}_i = \sum_{k=1}^{4} \frac{w_k}{m_k} \sum_{j=1}^{m_k} z^k_{ij}$$

and then it is normalised again by the formula:

$$\mathrm{Z}_i = \frac{\overline{\mathrm{Z}}_i - \mathrm{M}}{\mathrm{S}} 10 + 100$$

where M and S are respectively the mean and standard deviation of $\overline{\mathrm{Z}}$.

Finally, the IESI with base 2010, for month i, is given by:

$$\mathrm{IESI}_i = \frac{\mathrm{Z}_i}{\frac{1}{12}\sum_{l \in 2010} \mathrm{Z}_l} 100$$

There are a number of points of interest in this scheme. First, two different normalisation methods are used: standardisation (*i.e.* z-scores) and distance to a reference (*i.e.* index numbers). Standardisation is used twice, first for individual indicators and then for the weighted mean of standardised indicators; whereas distance to a reference is used for the standardised weighted mean. However, index numbers describe percentage distances and computing percentage distances on z-scores does not make sense (Mazziotta and Pareto, 2021). Second, when a new month of data becomes available, the mean and standard deviation in formula (3.1) change, therefore standardised indicators must be recalculated for all series. Third, the current IESI cannot be computed as a weighted mean of the 4 confidence climate indicators, and this can cause inconsistent results. Last but not least, the calculation procedure is not based on the principle of parsimony.

4. The new calculation scheme

This *Section* describes, in detail, the new method for calculating the IESI and the confidence climate indicators, which allows the elimination of the current inconsistencies. The new IESI can be computed both as a weighted mean of the 4 confidence climate indicators and as a weighted mean of the index numbers of the 11 balances of the original variables, thereby obtaining a consistent result.

The calculation scheme is based on the guidelines of the literature which envisage a first step of normalisation of the individual indicators - the balances – by the distance to a reference method[8] and subsequent aggregations for the construction of partial indices or pillars - the 4 confidence climate indicators - and of the global index - the IESI (Aureli Cutillo, 1996; Salzman, 2003; OECD, 2008).

4.1 Computing the confidence climate indicators

Given the matrix of seasonally adjusted balances of the climate k $\mathbf{X}^k_{n,m_k} = \{x^k_{ij}\}$, we move to the transformed matrix $\widetilde{\mathbf{X}}^k_{n,m_k} = \{\widetilde{x}^k_{ij}\}$, with: $\widetilde{x}^k_{ij} = x^k_{ij} + 100$. Then, the normalised matrix of index numbers with base 2010 $\mathbf{Y}^k_{n,m_k} = \{y^k_{ij}\}$ is constructed, where:

$$y^k_{ij} = \frac{\widetilde{x}^k_{ij}}{\frac{1}{12}\sum_{l \in 2010} \widetilde{x}^k_{lj}} 100$$

and y^k_{ij} is the index number with base 2010 for month i and variable j of climate k.

Finally, the 4 confidence climate indicators are given from the matrix $\mathbf{C}_{n,4} = \{c_{ik}\}$, where:

$$c_{ik} = \frac{1}{m_k} \sum_{j=1}^{m_k} y^k_{ij}$$

4.2 Computing the new IESI

Let us consider the array of the weights of the 4 confidence climates $\mathbf{W}_4 = (w_1, w_2, w_3, w_4)$.

8 Note that the balances were not normalised by standardisation because they have the same nature and the same range.

The IESI with base 2010, for month *i*, can be obtained – as a function of the confidence climate – by the formula:

$$\mathrm{IESI}_i = \sum_{k=1}^{4} c_{ik} \cdot w_k \tag{4.1}$$

or, alternatively – as a function of the index numbers of the 11 transformed balances of the original variables – by applying the formula:

$$\mathrm{IESI}_i = \sum_{k=1}^{4} \frac{w_k}{m_k} \sum_{j=1}^{m_k} y_{ij}^k \tag{4.2}$$

4.3 An example of computation

This paragraph shows an example of computation of the 4 climates and the new IESI for the year 2010.

Table 4.1 illustrates the matrices $\mathbf{X}_{n,m_k}^k$ of seasonally adjusted balances of the 4 climates and Table 4.2 the matrices $\tilde{\mathbf{X}}_{n,m_k}^k$ of the transformed (positive) values[9].

In Table 4.3 the matrices $\mathbf{Y}_{n,m_k}^k$ of index numbers with base 100 = 2010 are reported. In fact, the mean value of the index numbers, for the year 2010, is equal to 100. Lastly, Table 4.4 shows the matrix $\mathbf{C}_{n,4}$ of the 4 climates - with their weights based on 2012 Value Added data - and the IESI. As can be seen, the composite indices (partials and global) also have a base of 100 = 2010.

9 The variable labels are as follows. Retail trade: R_1=assessments on sales, R_2= assessments on stocks (negative polarity, the balance is taken with the sign reversed), R_3=expectations on sales. Services: S_1= assessments on orders and on demand in general, S_2=expectations on orders and on demand in general, S_3=assess$_m$ents on business trend. Construction: C_1=assessments on orders and/or construction plans, C_2=employment expectations. Manufacturing: M_1=assessment on stocks of finished products (negative polarity, the balance is taken with the sign reversed), M_2=assessment on the overall order books, M_3=expectations on production level.

Table 4.1 - Matrices of seasonally adjusted balances

Month	Retail trade			Services			Construction		Manufacturing		
	R_1	R_2	R_3	S_1	S_2	S_3	C_1	C_2	M_1	M_2	M_3
Jan-2010	8.4	2.5	14.2	1.2	7.1	0.7	-53.7	-19.4	3.5	-40.2	5.6
Feb-2010	-2.5	-3.1	8.8	1.8	15.2	4.7	-57.4	-28.7	3.0	-38.2	8.6
Mar-2010	-0.7	-0.2	19.5	-1.7	9.2	-2.2	-50.2	-21.8	3.1	-38.0	8.1
Apr-2010	0.6	-3.5	15.7	6.0	13.8	8.0	-49.1	-16.9	0.2	-30.9	10.3
May-2010	0.6	-7.2	15.5	0.9	10.5	3.7	-55.2	-23.2	3.1	-26.7	10.8
Jun-2010	-6.6	-7.0	13.9	-2.9	6.7	-0.3	-58.7	-12.0	4.5	-31.3	11.1
Jul-2010	-10.7	-11.3	12.9	-3.3	11.8	0.4	-56.8	-14.5	3.1	-23.1	10.9
Aug-2010	-3.6	-16.8	15.4	-0.5	8.6	0.2	-46.0	-9.1	0.5	-21.9	10.7
Sep-2010	-14.2	-8.4	10.8	-0.4	6.2	1.2	-45.6	-19.0	-0.4	-22.4	12.7
Oct-2010	-0.9	-8.0	16.2	-1.3	7.7	-0.6	-41.9	-19.3	-1.3	-18.4	13.7
Nov-2010	-5.9	-10.5	18.5	1.7	7.1	2.3	-40.6	-14.5	-1.4	-19.4	12.7
Dec-2010	12.6	-8.2	27.8	0.0	7.2	2.1	-46.3	-14.8	0.4	-14.3	14.1

Source: Authors' own processing

Table 4.2 - Matrices of transformed balances

Month	Retail trade			Services			Construction		Manufacturing		
	R_1	R_2	R_3	S_1	S_2	S_3	C_1	C_2	M_1	M_2	M_3
Jan-2010	108.4	102.5	114.2	101.2	107.1	100.7	46.3	80.6	103.5	59.8	105.6
Feb-2010	97.5	96.9	108.8	101.8	115.2	104.7	42.6	71.3	103.0	61.8	108.6
Mar-2010	99.3	99.8	119.5	98.3	109.2	97.8	49.8	78.2	103.1	62.0	108.1
Apr-2010	100.6	96.5	115.7	106.0	113.8	108.0	50.9	83.1	100.2	69.1	110.3
May-2010	100.6	92.8	115.5	100.9	110.5	103.7	44.8	76.8	103.1	73.3	110.8
Jun-2010	93.4	93.0	113.9	97.1	106.7	99.7	41.3	88.0	104.5	68.7	111.1
Jul-2010	89.3	88.7	112.9	96.7	111.8	100.4	43.2	85.5	103.1	76.9	110.9
Aug-2010	96.4	83.2	115.4	99.5	108.6	100.2	54.0	90.9	100.5	78.1	110.7
Sep-2010	85.8	91.6	110.8	99.6	106.2	101.2	54.4	81.0	99.6	77.6	112.7
Oct-2010	99.1	92.0	116.2	98.7	107.7	99.4	58.1	80.7	98.7	81.6	113.7
Nov-2010	94.1	89.5	118.5	101.7	107.1	102.3	59.4	85.5	98.6	80.6	112.7
Dec-2010	112.6	91.8	127.8	100.0	107.2	102.1	53.7	85.2	100.4	85.7	114.1

Source: Authors' own processing

Table 4.3 - Matrices of index numbers (base 100=2010)

Month	Retail trade			Services			Construction		Manufacturing		
	R_1	R_2	R_3	S_1	S_2	S_3	C_1	C_2	M_1	M_2	M_3
Jan-2010	110.5	110.0	98.6	101.1	98.0	99.0	92.8	98.0	101.9	82.0	95.3
Feb-2010	99.4	104.0	94.0	101.7	105.4	103.0	85.4	86.7	101.5	84.7	98.0
Mar-2010	101.2	107.1	103.2	98.2	99.9	96.2	99.8	95.1	101.6	85.0	97.6
Apr-2010	102.6	103.6	99.9	105.9	104.2	106.2	102.1	101.1	98.7	94.7	99.6
May-2010	102.6	99.6	99.8	100.8	101.1	102.0	89.8	93.4	101.6	100.5	100.0
Jun-2010	95.2	99.8	98.4	97.0	97.7	98.0	82.8	107.0	102.9	94.2	100.3
Jul-2010	91.0	95.2	97.5	96.6	102.3	98.7	86.6	104.0	101.6	105.4	100.1
Aug-2010	98.3	89.3	99.7	99.4	99.4	98.5	108.3	110.5	99.0	107.1	99.9
Sep-2010	87.5	98.3	95.7	99.5	97.2	99.5	109.1	98.5	98.1	106.4	101.7
Oct-2010	101.0	98.7	100.4	98.6	98.6	97.8	116.5	98.1	97.2	111.9	102.6
Nov-2010	95.9	96.0	102.4	101.6	98.0	100.6	119.1	104.0	97.1	110.5	101.7
Dec-2010	114.8	98.5	110.4	99.9	98.1	100.4	107.7	103.6	98.9	117.5	103.0
Mean	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' own processing

Table 4.4 - Matrix of the 4 climates and new IESI (base 100=2010)

Month	Retail trade (w=0,12)	Services (w=0,39)	Construction (w=0,09)	Manufacturing (w=0,40)	New IESI
Jan-2010	106.4	99.4	95.4	93.1	97.3
Feb-2010	99.1	103.4	86.1	94.7	97.9
Mar-2010	103.8	98.1	97.5	94.7	97.4
Apr-2010	102.0	105.4	101.6	97.7	101.6
May-2010	100.6	101.3	91.6	100.7	100.1
Jun-2010	97.8	97.6	94.9	99.1	98.0
Jul-2010	94.6	99.2	95.3	102.4	99.6
Aug-2010	95.7	99.1	109.4	102.0	100.8
Sep-2010	93.8	98.7	103.8	102.1	100.0
Oct-2010	100.0	98.3	107.3	103.9	101.6
Nov-2010	98.1	100.1	111.5	103.1	102.1
Dec-2010	107.9	99.5	105.6	106.5	103.8
Mean	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Source: Authors' own processing

5. Comparing the new and the current IESI

Table 5.1 shows the correlation coefficients and the mean absolute differences between the time series obtained with the new method and the current series, for the period January 2010 - December 2020 (131 observations), in order to assess the agreement between the results of the proposed method and the current one[10].

As for the trend of the time series, the correlations for the 4 climate indicators are all above 0.99; whereas the correlation between the new and the current IESI is equal to 0.9763. This value is slightly lower due to the inconsistencies present in the current computation method (for example, see Table 5.2).

Instead, as regards the 'distances' between the time series, the series that differs most from the current one is the construction climate (2.8%); whereas the mean absolute difference between the series of the new IESI and that of the current IESI is 4.2%.

The impact of the new computation method appears to be limited, considering that the time series of the balances are recalculated every month, through seasonal adjustment.

Finally, the reconstruction of the time series was carried out for the month of May 2016, a month in which the current IESI highlighted a misalignment between its trend and that of the confidence climate indicators. On that occasion, in fact, an increase in the IESI was observed (from 102.7 to 103.4, with a variation of +0.7), against a reduction in all 4 confidence climate indicators. This is due to the fact that, as explained in *Section* 3, two independent procedures are used for constructing the current IESI and the 4 climates. The 4 climates are computed as index numbers of a simple mean of seasonally adjusted balances plus 100 (transformed balances); whereas the current IESI is computed as an index number of a weighted mean of the 11 balances trasformed into *z*-scores.

10 The weights used to compute the current and the new IESI are the same (*i.e.* 2012 Value Added data).

Table 5.1 - Comparing the methods. Jan-2010 - Dec-2020 (base 100=2010)

Time series	
Correlation coefficient	
IESI	0.9763
Retail trade	0.9995
Services	0.9999
Construction	0.9979
Manufacturing	0.9973
Mean absolute difference	
IESI	4.2
Retail trade	0.3
Services	0.1
Construction	2.8
Manufacturing	1.3

Sourse: Autors' own processing

The results obtained with the two methods are shown in Table 5.2. As can be seen, the new IESI - thanks to the associative property of the arithmetic mean that allows it to be expressed both as a function of the 4 climates (formula 4.1), and as a function of the 11 transformed balances of the original variables (formula 4.2) - is consistent with the trend of the 4 confidence climate indicators. In this case, in fact, the index shows a decrease (from 107.0 to 106.7, with a variation of -0.3%), against the reduction of all climates. This decrease is equal to the weighted mean of the changes of the 4 climates.

Table 5.2 - Comparing the methods. May-2016 (base 100=2010)

Composite indicator	Apr-2016	May-2016	Percentage change
Current method			
IESI	102.7	103.4	0.7
Retail trade	101.9	100.9	-1.0
Services	107.9	107.4	-0.5
Construction	121.2	120.4	-0.8
Manufacturing	102.7	102.1	-0.6
New method			
IESI	107.0	106.7	-0.3
Retail trade	101.5	101.1	-0.4
Services	108.3	107.7	-0.6
Construction	122.9	122.7	-0.2
Manufacturing	103.9	103.8	-0.1

Sourse: Autors' own processing

6. The performance of the new IESI compared with the Italian GDP

Aggregate indices expressed as synthesis of the information provided by the monthly surveys on consumer and business confidence, represent a useful tool for monitoring the evolution of the aggregate economic activity, through their ability to provide coincident or even leading signals of the economic cycle movements.

The aim of this Section is to illustrate the results of the tests to which the new IESI -obtained following the methodological proposal explained in Section 4 - has been subjected to verify its specific characteristics in terms of ability to capture cyclical fluctuations of the national economic activity.

Although the session is mainly focussed on the new IESI, all the performance tests have also been extended to the current IESI[11]. This has been done simply to make the analysis more complete and not to conclude which indicator between the two is the best in terms of the ability to capture fluctuations in aggregate activity. In fact, this is not the purpose of this study.

The analysis was carried out using the sample period 2005Q2-2021Q3.

In this regard, it is important to highlight that the quarterly series of both the new IESI and the current IESI[12] have missing data for 2020Q2, since the Coronavirus health emergency meant that the Istat surveys for that date were not carried out[13].

Finally, in order to evaluate the information capacity of the indicators with respect to the cyclical evolution of the Italian economy, the analysis was conducted having chosen the Italian GDP series as the benchmark series[14].

Following Moore and Shiskin (1967), the behaviour of the two indicators,

11 For this analysis, both the new IESI and the current IESI were calculated using the data (the seasonally adjusted balances of the variables included in the index definition) published by Istat in February 2022. All the results of the Istat surveys on the business confidence climate are available for consultation in the Institute data warehouse I.STAT (web page: https://www.istat.it/it/dati-analisi-e-prodotti/banche-dati).

12 To facilitate the comparison with the reference series, the original monthly series of both indicators have been transformed into quarterly series. In particular, the average of the three months has been used for the transformation.

13 More specifically, in the original monthly series the missing data is that of April 2020. In this work the statistic and econometric tests were carried out having replaced it with the whole number immediately below the lowest value reached by each monthly series. In particular, t was evaluated to be 60.0 for the new IESI and 54.0 for the current IESI. In fact, the minimum values that the series reached were 60.4 and 54.2, respectively (in May 2020, in both cases). As a consequence, also the quarterly series of the two indicators have presented the lowest value in correspondence to the 2020Q2.

14 The seasonally adjusted series. Chain linked - reference year 2015. November 2021 edition.

with respect to the reference series, was evaluated by looking at the time consistency, the conformity and, finally, at the economic significance of the relationship existing between the series.

More specifically, the first analysis was carried out evaluating the time profile and the average lead/lag of the indicators for turning points in the reference series.

As for the identification of the turning points, both for the series of the indicators and for the one of the reference series, the Bry-Boschan procedure was used (Bry and Boschan, 1971)[15].

Following the growth cycle approach[16], the phases of expansion and recession of the reference series were identified after having removed the long-term trend from the Italian GDP series. This latter, in particular, was estimated using the Hodrick and Prescott (HP) filter in its standard version for quarterly series[17].

The classical NBER approach was adopted, instead, for the quarterly data of both the new IESI and the current IESI.

The second analysis was conducted calculating the directional coherence coefficients. They indicate the percentage of the times in which the new indicator and the current one move in the same direction as the chosen reference series[18] does.

Finally, the empirical relationship between the two indicators and the reference series was explored by resorting to both the cross-correlation test and the causality Granger test[19].

To start, Figure 6.1 shows the evolution of the new IESI indicator and of the

15 Specifically, the adaption of the Bry-Boschan procedure proposed by Harding and Pagan (2002).

16 Among the various methodologies for determining the turning points, the one based on the concept of growth cycle appeared to be the most appropriate given the stability characteristics of the economic growth path of our Country. Actually, following the approach of the cycle in growth rates, the analysis was conducted also using the Italian GDP transformed into the first differences of the logarithms. This allowed the identification of a higher number of turning points, but, as a whole, the results were considered less interesting for the purposes of the analysis.

17 Hodrick-Prescott (1997).

18 The Hodrick and Prescott detrended GDP series.

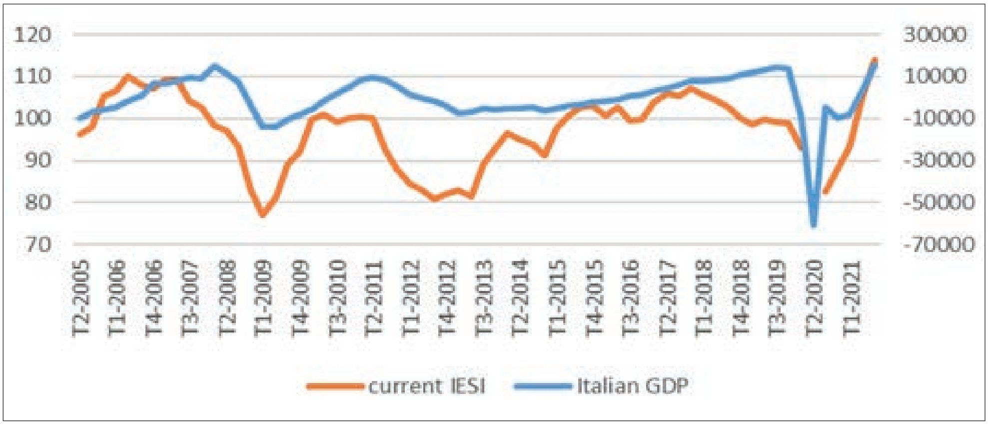
19 As for the reference series, these tests were performed having chosen the series of the Hodrick-Prescott detrended GDP, for the cross correlation analysis, and the one of the first differences of the GDP logarithm, for the Granger test, as will be set out below.

Italian GDP over the period 2005Q2-2021Q3. The similar pattern displayed by the two series leads one to think that the new indicator may have a quite good ability in tracking the Italian GDP. Besides, this relationship seems as close as the one that can be observed between the current IESI and the Italian GDP (Figure 6.2).

Figure 6.1 - New Istat Economic Sentiment Indicator (new IESI) and Italian GDP (a) - 2005Q2-2021Q3

Source: Authors' own processing
(a) GDP cyclical component obtained using the Hodrick and Prescott filter.

Figure 6.2 - Istat Economic Sentiment Indicator (current IESI) and Italian GDP (a) - 2005Q2-2021Q3



Source: Authors' own processing
(a) GDP cyclical component obtained using the Hodrick and Prescott filter.

As can be seen from a more careful examination of Figure 6.1, the new IESI series appears to anticipate the one of the HP filtered GDP in the first years of the sample period.

Furthermore, it seems to show higher volatility than the reference series used between 2013 and the end of 2017.

During the following two-year period (2018-2019), the indicators show opposite behaviours: the aggregate activity exhibits moderate growth, while the confidence indicator decreases.

Finally, in the last quarters of the period under observation, after the economic growth collapse caused by the COVID-19 pandemic in 2020Q2, both the variables rise again.

Looking now at the results reported in Table 6.1, in which the chronology identified by the Bry-Boschan procedure is reported, they display a good consistency with what emerged from the graphical inspection.

The lower dynamism of the Italian GDP series, particularly evident in the central years of the sample period, allows the procedure to identify for this series only two complete cycles from peak to peak. These latter appear characterised by a rather wide average length (23 quarters) because of the long central cycle 2011Q2-2019Q3.

Moreover, it is interesting to notice how the new IESI actually turns out to be able to track the reference series with a leading behaviour (-2.3 quarters, on average).

In particular, it appears coincident around the upturns, but decisively leading in correspondence to the downturn points (-4.7 quarters, on average).

As confirmation of what the graphical analysis highlighted, it anticipates the HP filtered GDP series in correspondence to both the peak present in the first part of the sample period (2008Q1) and the peak present at the end of the GDP moderate growth period (2019Q3).

As for the current IESI, it seems to track the reference series used even better than the indicator obtained by the new methodology. Initially, it shows some little variations which were not identified for the Italian GDP series and so the procedure located a higher number of complete cycles for this series (4).

Moreover, it anticipates the cyclical profile of the HP filtered GDP series of around four quarters on average (-3.7) with a lead in correspondence both of the upturn points (-1.3) and of the downturn points (-6.0).

Concerning the directional coherence analysis, a satisfactory result was obtained from the calculation of the related coefficient. In fact, as reported in Table 6.1, the new IESI appears to be able to correctly capture the sign of the reference variable in 68% of cases (the coefficient shows a value of 0.65 for the current IESI series).

At this point, as mentioned above, the empirical relationship between the indicators and the reference series was further verified resorting to the cross-correlation analysis and to the Granger causality test.

However, before proceeding with these tests, a preliminary study of the stochastic properties of the two indicators was carried out.

The results of this check showed that these series are stationary in the period 2005Q2-2021Q3. After all, these indicators are considered stationary by construction[20].

On the basis of this consideration, the time series of the two indicators were not subjected to any transformation.

Regarding the first test, the results of the cross correlation between the HP filtered GDP and the new IESI for the period 2005Q2-2021Q3 showed that the higher correlation coefficient between the two series was reached at time 0 (0.68)[21].

20 With reference to the presence of unit roots in series deemed to be stationary, see Brunello et al., 2000; Bruno, and Malgarini, 2002.

21 The 2-year moving correlation highlighted the empirical relationship between the HP filtered GDP and the indicator is particularly high in the two-year periods 2018Q2-2020Q2 (0.97) and 2009Q1-2011Q1 (0.93).

Table 6.1 - Turning point chronology and directional coherence (2005Q2-2021Q3)

Bry-Boschan Routine	Italian GDP	new IESI	current IESI
Number of cycles (from peak to peak)	2	3	4
Average duration (from peak to peak)	23.0	15.3	11.5
Average length of an expansion	17.0	7.7	5.0
Average length of a recession	5.0	8.3	7.2
Torning points			
Trough	/	/	/
Peak	2008Q1	2006Q2	2006Q2
Trough	2009Q2	2009Q1	2009Q1
Peak	2011Q2	2011Q2	2010Q2
Trough	2013Q1	2013Q2	2012Q3
Peak	/	2015Q3	2014Q1
Trough	/	/	2014Q4
Peak	/	/	2015Q4
Trough	/	2016Q3	2016Q3
Peak	2019Q3	2017Q4	2017Q4
Trough	2020Q2	2020Q2	2020Q2
Mean lead (-) /lag (+) at turning points (in quarters)			
Total	/	-2.3	-3.7
Upturns	/	0.0	-1.3
downturns	/	-4.7	-6.0
Directional coherence			
	/	0.68	0.65

Source: Authors' own processing

A similar result was obtained looking at the relationship between the reference series used and the current IESI. Also in this case, in fact, the maximum correlation coefficient (0.68) was reached at lag 0 (Table 6.2).

In the light of considerations pertaining to the adequacy of the regression model used, it was decided to make the next in sample forecasting exercise (the Granger causality test) using the first differences of the logarithms of the Italian GDP instead of the corresponding cyclical components[22].

22 Such a decision, in fact, allowed us to obtain the best results in terms of the regression model goodness. In the latter, in particular, having inserted up to 2 lags, all the parameters were statistically significant (0.000 the related probabilities), the R-squared was very high (0.83) and the F-statistic led to the clear rejection of the null hypothesis that the coefficients, except the intercept, are jointly equal to zero (0.000 the probability). Finally, using this model the residuals appeared homoscedastic and not serially correlated. On the contrary, both the Breusch-Pagan-Godfrey test and the White test led to the rejection of the null hypothesis of homoskedasticy of residuals when in the model the HP filtered GDP series, was introduced as the dependent variable. Moreover, on the basis of the LM test, the model was found to be characterised by serial correlated residuals. Lastly, the R-squared (0.51) and the adjusted R-squared (0.49) were lower than the ones of the model in which the first differences of logarithms of the Italian GDP was considered as the dependent variable (0.83, as reported above, and 0.82, respectively).

The results of the test showed that the lagged values of the new IESI (the independent variable of the regression model) did not improve the in-sample prediction of the Italian GDP values (Table 6.2). In fact, Granger's causality test led to the acceptance of the null hypothesis of no-Granger-causality between the new IESI and the reference series expressed in terms of the first differences of the logarithms (1.2, the value from the F statistic; 0.3 the value of the corresponding probability)[23].

To conclude, at the level of significance of 5% and of 10% also the current IESI seemed not to cause the Italian GDP (0.6 the related p-value of the F-statistic)[24].

Table 6.2 - Correlation function and Granger Causality test (2005Q2-2021Q3)

New IESI - Italian GDP (cyclical components)		New IESI - Italian GDP (first differences of logarithms)	
Correlation function		Granger Causality test (2 lags)	
ρ	0.68	F-Statistic	1.19
max	0	Probability	0.31
Current IESI - Italian GDP (cyclical components)		**Current IESI - Italian GDP (first differences of log)**	
Correlation function		Granger Causality test (2 lags)	
ρ	0.68	F-Statistic	0.45
max	0	Probability	0.64

Source: Authors' own processing

23 Having inserted up to 4 lags in the model, the F-statistic was 2.6 and the probability associated turned out to be just 0.05.
24 In this case, having inserted up to 4 lags in the model, the F-statistic was 1.6 and the probability associated 0.18.

7. A recent circumstance

According to the results, on the basis of which both indicators appear to not have any forecasting capabilities against the Italian GDP, some observations are really necessary regarding the latter two years due to the consequences of the COVID-19[25].

The values reported in Table 6.1 appear in fact so distant from those that would have been expected[26] to induce the authors to analyse the time series more closely considered in the regressive model.

In particular, it has been verified that the IESI series are stationary but a Granger causality test also requires that whatever GDP series is used as the dependent variable should be checked for stability of its moments across time.

Actually, the GDP series used[27] presents a serial correlation equal to 0.61 if the period 2005Q2-2019Q4 is considered, but equal to -0.24, if the entire sample period 2005Q2-2021Q3 is considered. This result (the negative correlation that the series presents in the entire period) shows how much these last two years have affected the series. The sudden changes in the state of the economy have introduced outliers in the Italian GDP series.

At this point, the data formation model of the series has changed and each of its values can no longer be partially explained by the data that precedes it, nor can it be indicative of the value that follows it.

This means that in the regression used for the investigation of causality in the Grangerian sense between the indicators and the Italian GDP, the results may be affected by the instability in the moments of the latter. Obviously, that makes the use of this model for the verification of the null hypothesis totally risky.

As proof of the weight that the anomalous data present in the last part of the Italian GDP time series had in determining the unexpected outcome of the tests, the cross-correlation and Granger's causality test were repeated taking into consideration the sub-period 2005Q2-2019Q4 (Table 7.1).

25 The authors thank an anonymous referee for having raised this point.

26 In fact, it is common opinion that the indicators of the business and consumer surveys, although subject to accidental cyclical fluctuations or other types of influences, are able to provide coincidental or even anticipatory signals of the cyclical fluctuations of the aggregate economy (Istat, 2022).

27 The Italian GDP series transformed into the first differences of the logarithms.

Table 7.1 - Correlation function and Granger Causality test (2005Q2-2019Q4)

New IESI - Italian GDP (cyclical components)		New IESI - Italian GDP (first differences of log)	
Correlation function		Granger Causality test (2 lags)	
ρ	0.64	F-Statistic	4.48
max	-2	Probability	0.02
Current IESI - Italian GDP (cyclical components)		**Current IESI - Italian GDP (first differences of log)**	
Correlation function		Granger Causality test (2 lags)	
ρ	0.68	F-Statistic	4.40
max	-2	Probability	0.02

Source: Authors’ own processing

8. Concluding remarks

This paper proposes an alternative methodology for calculating the confidence climate indices that are monthly disseminated by the Istat. In particular, a new methodology has been proposed for the calculation both for the sectoral confidence indices (in detail, the manufacturing, construction, service and retail trade confidence climate index) and for the Istat Economic Sentiment Indicator (IESI).

This methodology ensures the consistency between the evolution of the composite indicator (IESI) and the dynamics of the sector-based indices. Consequently, it allows the overcoming of the possible discrepancies such as those that occurred in the past (in May 2016 and in November 2019).

In the new procedure, in fact, the normalisation of the seasonally adjusted variables composing the sectorial indices (more specifically, their transformation into 2010 indices) is the first phase of the sector-based climate index processing.

The new IESI obtained on the basis of this methodology, being a weighted average of the seasonally adjusted and standardised series composing the sector-level confidence climates, is necessarily consistent with the evolution of these indices.

After having illustrated in detail the current methodology and the new one, also providing an application example of the calculation scheme proposed, the work presents some empirical assessments.

First, the concordance between the results obtainable using the new methodology and the ones obtained following the current methodology has been verified. Subsequently, the new indicator performance, with respect to the cyclical trend of the Italian GDP, has been assessed.

As for the first verification, the results indicate that the new calculation scheme would produce the new series of the sectoral indices very similar to the ones obtained with the current procedure. Furthermore, it would generate the series of the new IESI consistent with the dynamic of the sectoral indices and, precisely for this reason, slightly different from the one of the current IESI.

As for the second one, the composite indicator, developed according to the new methodology, shows quite a good cyclical profile with respect to

GDP and seems capable of providing leading signals of the movements in the national economic cycle. Overall, it seems to have characteristics rather similar to the ones of the current IESI.

Nevertheless, the study highlight how much the period taken under observation (2005Q2-2021Q2) is difficult to analyse because of the COVID-19 pandemic effects on the macroeconomic variable trends. In fact, the well-known relationship between the confidence climate indicators and the GDP (in particular, their ability to provide coincidental or even anticipatory signals of the business cycle movements) is actually confirmed only when the sub-period 2005Q2-2019Q4 is considered. That is, only when the outliers determined by COVID-19 are not considered.

In light of the results obtained, we recommend changing the current methodology for calculating the IESI and the four sectoral indices with the new proposal.

References

Aureli Cutillo, E. 1996. *Lezioni di Statistica Sociale. Parte seconda: Sintesi e graduatorie*. Roma, Italy: CISU – Centro Informazione Stampa Universitaria.

Brunello, G., C. Lupi, and P. Ordine. 2000. "Regional Disparities and the Italian NAIRU". *Oxford economic papers*, Volume 52, N. 1 (Special Issue on Unemployment Dynamics and the Natural Rate): 146-177.

Bruno, G., and M. Malgarini. 2002. "An Indicator of Economic Sentiment for the Italian Economy". *ISAE Working Papers*, N. 28. Roma, Italy: ISAE.

Bry, G., and C. Boschan. 1971. "Cyclical Analysis of the Time Series: Selected Procedures and Computer Programs". *NBER Technical Papers*, N. 20/1971. Cambridge, MA, U.S.: National Bureau of Economic Research – NBER.

Casadio Tarabusi, E., and G. Guarini. 2013. "An Unbalance Adjustment Method for Development Indicators". *Social Indicators Research*, Volume 112, Issue 1: 19-45.

Crosilla, L. (*a cura di*). 2022. "Le Indagini Qualitative sulla Fiducia delle Imprese e dei Consumatori. Caratteristiche e metodologia". *Letture statistiche – Metodi*. Roma, Italy: Istat. https://www.istat.it/it/archivio/266068.

European Commission, Directorate-General for Economic and Financial Affairs. 2021. *The Joint Harmonised EU Programme of Business and Consumer Surveys. User guide (Updated October 2021*). Brussels, Belgium: European Commission.

Harding, D., and A. Pagan. 2002. "A comparison of two business cycle dating methods". *Journal of Economic Dynamics and Control*, Volume 27, Issue 9: 1681-1690.

Hodrick, R.J., and E.C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation". *Journal of Money, Credit, and Banking*, Volume 29, N. 1: 1-16.

Istituto Nazionale di Statistica - Istat. 2022. "Consumer and Business Confidence. February 2022". *Statistics Flash*. Roma, Italy: Istat. https://www.istat.it/it/archivio/266742.

Kónya, L. 2004. “Unit Root, Cointegration and Granger Causality Test Results for Export and Growth in OECD Countries”. *International Journal of Applied Econometrics and Quantitative Studies*, Volume 1, Issue 2: 67-94.

Leproux, S., and E. Matera. 2015. “The Revised ISTAT Economic Sentiment Indicator: its Structure and Verification of Performance Against the Cyclical Movements of the Italian GDP”. *Istat working papers*, N. 23/2015. Roma, Italy: Istat. https://www.istat.it/it/archivio/180673.

Mazziotta, M., and A. Pareto. 2021. “Everything you always wanted to know about normalization (but were afraid to ask)”. *Italian Review of Economics, Demography and Statistics*, Volume LXXV, 1: 41–52.

Mazziotta, M., and A. Pareto (a cura di). 2020. *Gli indici sintetici*. Torino, Italy: G. Giappichelli Editore.

Mazziotta, M., and A. Pareto. 2017. “Synthesis of Indicators: The Composite Indicators Approach”. In Maggino, F. (*ed.*). *Complexity in Society: From Indicators Construction to their Synthesis*. Cham, Switzerland: Springer.

Moore, G.H., and J. Shiskin. 1967. “Indicators of Business Expansions and Contractions”. *NBER Occasional Paper*, N. 103. New York, NY, U.S.: National Bureau of Economic Research - NBER.

Organisation for Economic Co-Operation and Development – OECD, and the Joint Research Centre – JRC of the European Commission in Ispra, Italy. 2008. *Handbook on Constructing Composite Indicators. Methodology and user guide*. Paris, France: OECD Publications.

Salzman, J. 2003. *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*. Ottawa, Ontario, Canada: Centre for the Study of Living Standards - CSLS.

Terzi, S., A. Otoiu, E. Grimaccia, M. Mazziotta, and A. Pareto A. 2021. *Open issues in composite indicators. A starting point and a reference on some state-of-the-art issues*. Roma, Italy: Roma TrE-Press.

Balancing population stocks and flows in the Italian demographic estimation system: a proposal

Diego Zardetto[1], Marco Di Zio[1], Marco Fortini[1]

Abstract

Estimates of population counts ('stocks') should be consistent with counts of demographic events ('flows'). In particular, the Demographic Balancing Equation (DBE) should be satisfied. Istat's modernised production system, underpinned by statistical registers that integrate survey and administrative data, seems ideally positioned to overcome the challenge of producing timely, reliable, and coherent estimates of demographic stocks and flows. However, in Italy, estimates of stocks and flows entering the DBE are currently obtained independently. Population size estimates at subsequent reference times are provided by the integrated system formed by the Permanent Census and the Base Register of Individuals (BRI), whereas birth, death, and migration figures are derived from municipal civil registries. Therefore, owing to sampling and non-sampling errors affecting these "raw" estimates of stocks and flows, the DBE is not trivially fulfilled. Consistency of official estimates can, however, be attained through a suitable macro-integration process that optimally adjusts both stocks and flows to ensure compliance with the DBE. To this end, we propose to use 'balancing' methods that National Statistical Institutes routinely adopt to reconcile large systems of national accounts. We designed and implemented a system that can handle this task at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational (i.e. domain) estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Moreover, simulation results show that our balancing approach determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to "raw" ones. Finally, we provide considerations and guidance for using balanced population counts as control totals to adjust individual weights of the BRI.

Keywords: Data integration, macro-integration, demographic balancing equation, census estimates, demographic events, data quality, coherence and consistency, statistical registers.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_2.2022.04

1 Diego Zardetto (zardetto@istat.it); Marco Di Zio (dizio@istat.it); Marco Fortini (fortini@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Background information

The Italian National Institute of Statistics - Istat is currently engaged in completing the transition of its production processes to the model envisioned by the modernisation programme launched some years ago (Istat, 2016). The backbone of the new production system is the infrastructure formed by the ‘Integrated System of Statistical Registers’ (ISSR), namely a system of connected registers used as a reference for all the statistical activities carried out by Istat (Alleva *et al.*, 2019; Alleva *et al.*, 2021). A pivotal role within the ISSR is played by the ‘Base Register of Individuals’ (BRI), a comprehensive statistical register that integrates and stores data gathered from disparate sources about people usually or temporarily residing in Italy.

One of the most important achievements of the new statistical production system is the modernisation of the Italian population census. Until the year 2011, traditional population censuses were conducted in Italy every ten years, and their outcomes were employed to correct municipal civil registries once a decade. Starting from 2018, the Italian population census is no longer a complete enumeration survey but rather results from a *twofold* large-scale sample survey that is carried out each year, whose outcomes are integrated with the BRI. Istat has named this new census design ‘Permanent Census’. The Permanent Census involves two simultaneous sample surveys: the ‘A’ survey and the ‘L’survey. The L component relies on a list sample: its main objective is to observe variables that are either of insufficient quality or not available at all in the BRI. The A component is instead based on an area sample: it is designed to provide yearly estimates of the under-coverage and over-coverage rates of the BRI, evaluated at national and local levels for different sub-population profiles defined by variables like ‘sex’, ‘age class’, ‘nationality’.

The new production system enables Istat to deliver official population size estimates more frequently than happened before through traditional censuses. To this end, crude estimates of population counts derived from the BRI are corrected by using individual weights that are functions of the over- and under-coverage probabilities estimated through the A sample linked with the BRI (see Falorsi, 2017, and Righi *et al.*, 2021). This exercise counteracts coverage errors of the BRI, yielding estimates of population counts which, albeit improved, we still consider *“raw”* in this paper.

Official estimates of population counts (*stocks*) should be consistent with civil registry figures about demographic events (*flows*), in such a way that the resulting data system exactly fulfills the Demographic Balancing Equation (DBE).

The DBE states that the population counts at time $t + 1$ must be equal to the population counts at time t plus the sum of the natural increase and the net migration that occurred between t and $t + 1$:

$$P^{(t+1)} = P^{(t)} + N + M \tag{1}$$

where the natural increase, N, is the difference between births and deaths, and the net migration, M, is the difference between immigrants and emigrants:

$$\begin{cases} N = B - D \\ M = I - E \end{cases} \tag{2}$$

In Italy, *raw* estimates of stocks and flows entering the DBE are currently obtained *independently*. Birth, death, and migration figures are derived from municipal civil registries[2], whereas population size estimates at subsequent reference times are provided by the BRI and the Permanent Census, as already noted. Therefore, owing to sampling and non-sampling errors affecting raw estimates of stocks and flows, the DBE is *not* trivially satisfied. Consistency of *official* estimates can, however, be achieved through a suitable process that simultaneously adjusts *both* stocks and flows to ensure compliance with the DBE.

Note that the complexity of this task is inextricably linked to the sample-survey nature of the Permanent Census. Indeed, census estimates of population counts are now affected by sampling uncertainty, which prevents solving the stocks and flows consistency problem with the classical '*flows-first*' approach typically adopted in countries where traditional censuses are conducted. In the *flows-first* approach – called 'component method' (Eurostat, 2003) – the census provides "true" population counts at year t, to which estimates of population flows for the period $[t, t + 1]$ are added, thus obtaining estimates of population stocks at year $t + 1$ that satisfy the DBE by *construction*[3].

2 Indeed, the BRI is not able to correctly register demographic events (births, deaths, internal and cross-border migrations).

3 Note, however, that the flows-first approach has a serious weak point, as any errors or bias in estimating flows will be carried forward from timepoint to timepoint during the intercensal period.

To solve the stock and flow consistency problem in the integrated estimation system formed by the Permanent Census and the BRI, we propose to adjust raw estimates by using methods that are commonly adopted inside National Statistical Institutes (NSI) for *balancing* large systems of national accounts. Indeed, the National Accounts divisions of most NSIs routinely make use of independent initial estimates, which (*i*) are characterised by different degrees of reliability (as is also the case of demographic stocks and flows), and (*ii*) must be adjusted to satisfy a large set of accounting identities (as is the system of DBEs associated to any partition of the overall population into estimation domains). Relevant papers on modelling demographic figures as accounting matrices are Rees (1979), Stone and Corbit (1997), Bryant and Graham (2013), and Bryant and Graham (2015).

We designed and implemented a macro-integration procedure that ensures consistency between official (*i.e. adjusted*) estimates of demographic stocks and flows using balancing methods. The system we developed can handle this task at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational (*i.e.* domain) estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Moreover, simulation results show that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to raw ones.

While our proposal is directly targeted at improving the quality of official demographic statistics at *macro-level*, we see significant scope for leveraging its results at *micro-level* too. Once balanced population counts are obtained through our macro-integration procedure, one could think of using them as control totals to adjust the individual weights tied to records of the BRI. This would determine two beneficial effects. First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by our procedure. Second, BRI weights would receive a second layer of protection against bias, "borrowing strength" from both official population counts disseminated the year before and balanced demographic flows.

The rest of the paper is structured as follows. Section 2 provides the mathematical formulation of the problem, highlighting its computational complexity and the challenges of its extension to domain estimates, along with technical countermeasures put in place by our system. The same Section also introduces the main statistical properties of balanced estimates under ideal conditions. Section 3 illustrates the results of an experimental study based on simulations, where the statistical properties of balanced estimates are investigated under realistic (*i.e.* non-ideal) settings. Section 4 discusses details for the downstream application of our proposal to the BRI. Lastly, Section 5 offers some final remarks.

2. Problem formulation

We formalise the task of finding a system of consistent estimates of demographic stocks and flows as a constrained optimisation problem. This is accomplished along the lines of (Stone *et al.*, 1942) and (Byron, 1978), by suitably reformulating the models and algorithms introduced in those classical papers.

Given *initial* (=raw) estimates of all the aggregates entering the demographic balancing equations (1) defined for all the geographic areas of a given territorial level, we search for *final* estimates that are *balanced*, *i.e.* (*i*) satisfy all the DBEs, and (*ii*) are *as close as possible* to the initial estimates. Therefore, the objective function to be minimised is an appropriate distance metric between the final and initial estimates, while the constraints acting on the final estimates are the area-level DBEs. Moreover, we adopt a *weighted* distance metric such that aggregates whose initial estimates are more *reliable* will tend to be changed less.

Let us suppose we have initial estimates of the population size of k Italian regions[4] U_i at times t and $t + 1$, as well as initial estimates of births, deaths, and natural increase that occurred for each region between time t and $t + 1$[5]:

$$\begin{cases} P^{(t)} = \left(P_1^{(t)}, \dots, P_k^{(t)}\right)' \\ P^{(t+1)} = \left(P_1^{(t+1)}, \dots, P_k^{(t+1)}\right)' \\ B = (B_1, \dots, B_k)' \\ D = (D_1, \dots, D_k)' \\ N = (N_1, \dots, N_k)' \end{cases} \tag{3}$$

Moreover, let us suppose we have initial estimates of the *Migration Flows Matrix F*, whose generic element F_{ij} equals the number of people who *moved* from region i to region j between time t and $t + 1$:

4 The subpopulation notation U_i is adopted because, as we will explain later, “regions” can actually be any partition of the Italian population U, *e.g.* obtained by crossing variables ‘provinces’, ‘sex’, and ‘age classes’.

5 In equation (3), and throughout the whole paper, we denote matrix transposition with a single quote, $'$, to avoid confusion with time superscripts, ^{t}.

$$F = \begin{pmatrix} 0 & F_{1,2} & \cdots & F_{1,k} & F_{1,k+1} \\ F_{2,1} & 0 & \cdots & F_{2,k} & F_{2,k+1} \\ \cdots & \cdots & 0 & \cdots & \cdots \\ F_{k,1} & F_{k,2} & \cdots & 0 & F_{k,k+1} \\ F_{k+1,1} & F_{K+1,2} & \cdots & F_{k+1,k} & 0 \end{pmatrix} \tag{4}$$

Note that the $(k + 1)^{\text{th}}$ row and column of F represent migrations from and to any territory *outside* the nation, thus $k + 1$ means '*abroad*'. Note also that matrix F is not, in general, symmetric nor antisymmetric.

Let us indicate with M the *Net Migration Matrix*, whose generic element M_{ij} equals the count of people who *immigrated* in region i from region j *minus* the count of people who *emigrated* from region i to region j, $M_{ij} = F_{ji} - F_{ij}$:

$$M = \begin{pmatrix} 0 & M_{1,2} & \cdots & M_{1,k} & M_{1,k+1} \\ -M_{1,2} & 0 & \cdots & M_{2,k} & M_{2,k+1} \\ \cdots & \cdots & 0 & \cdots & \cdots \\ -M_{1,k} & -M_{2,k} & \cdots & 0 & M_{k,k+1} \\ -M_{1,k+1} & -M_{2,k+1} & \cdots & -M_{k,k+1} & 0 \end{pmatrix} \tag{5}$$

Note that matrix M is *antisymmetric* and actually equal to minus twice the antisymmetric part[6] of F:

$$\begin{cases} M = -M' \\ M = F' - F = -2F \end{cases} \tag{6}$$

Furthermore, let us assume we can attach to each *atomic* initial estimate involved in (3), (4), and (5) a measure of *reliability,* $R \in [0, \infty)$. These reliability measures could be either based on proper statistical measures (*e.g.* proportional to inverse estimated variances) or derived from an assessment made by subject matter experts. For instance, we will indicate the reliability measure of a generic element M_{ij} of the Net Migration Matrix M as $R[M_{ij}]$. Note that $R[\cdot] \to \infty$ will signal *absolute reliability*, and thus *prevent* the corresponding initial atomic estimates from being altered.

Lastly, let us denote *raw estimates* with a *tilde* (*e.g.* $\widetilde{M}_{ij}$) and *balanced estimates* with a *circumflex hat* (*e.g.* $\widehat{M}_{ij}$). Given (3), (4), and (5), we define the objective function, L, for the constrained optimisation problem as follows:

6 Any square matrix, F, can be decomposed into a symmetric part, F^S, and an antisymmetric part, F^A, such that: $F^S = (F^S)'$, $F^A = -(F^A)'$, and $F = F^S + F^A$. Of course: $F^S = (F + F')/2$, $F^A = (F - F')/2$.

$$
\begin{aligned}
L&\left(\hat{P}^{(t+1)}, \hat{P}^{(t)}, \hat{B}, \hat{D}, \hat{N}, \hat{F}, \hat{M}\right)\\
&= \sum_{i=1}^{k}\left(\hat{P}_i^{(t+1)} - \tilde{P}_i^{(t+1)}\right)^2 R\left[\tilde{P}_i^{(t+1)}\right] + \sum_{i=1}^{k}\left(\hat{P}_i^{(t)} - \tilde{P}_i^{(t)}\right)^2 R\left[\tilde{P}_i^{(t)}\right]\\
&+ \sum_{i=1}^{k}\left(\hat{B}_i - \tilde{B}_i\right)^2 R[\tilde{B}_i] + \sum_{i=1}^{k}\left(\hat{D}_i - \tilde{D}_i\right)^2 R[\tilde{D}_i]\\
&+ \sum_{i=1}^{k}\left(\hat{N}_i - \tilde{N}_i\right)^2 R[\tilde{N}_i] + \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\left(\hat{F}_{ij} - \tilde{F}_{ij}\right)^2 R[\tilde{F}_{ij}]\\
&+ \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}\left(\hat{M}_{ij} - \tilde{M}_{ij}\right)^2 R[\tilde{M}_{ij}]
\end{aligned}
\tag{7}
$$

where $\hat{P}^{(t+1)}, \hat{P}^{(t)}, \hat{B}, \hat{D}, \hat{F}$ and $\hat{M}$ are the final (*i.e.* adjusted and balanced) estimates we are looking for. Function L is simply the (squared) weighted Euclidean distance between the vectors of raw and balanced estimates of stocks and flows.

Note that the objective function (7) involves both *gross* and *net* migration flows, and both *gross* and *net* natural flows, as they are all very significant demographic statistics in their own right, which we would like to modify the least during the balancing procedure.

Therefore, the constrained optimisation problem we propose to solve is the following:

$$
\begin{cases}
\text{Argmin } L\left(\hat{P}^{(t+1)}, \hat{P}^{(t)}, \hat{B}, \hat{D}, \hat{N}, \hat{F}, \hat{M}\right) & \text{subject to:}\\
\hat{P}_i^{(t+1)} = \hat{P}_i^{(t)} + \hat{N}_i + \sum_{j=1}^{k+1} \hat{M}_{ij} & \text{for } i = 1, \dots, k\\
\hat{N}_i = \hat{B}_i - \hat{D}_i & \text{for } i = 1, \dots, k\\
\hat{M}_{ij} = \hat{F}_{ji} - \hat{F}_{ij} & \text{for } i, j = 1, \dots, k+1
\end{cases}
\tag{8}
$$

The constraints acting on the problem (8) are, of course, the area-level DBEs, plus structural constraints expressing the relation between births, deaths, and natural increase, and the antisymmetry of the Net Migration Matrix. The solution to problem (8) results in time and space-consistent estimates of population counts, natural flows, and migration flows for all the subpopulations U_i. In addition, owing to the linearity of the DBE, it is evident that the solution to problem (8) ensures consistency of stocks and flows for any other domain that can be obtained by aggregation of the balancing domains U_i, therefore, in particular, for the whole population U.

Problem (8) involves $2(k + 1)^2 + 5k$ unknowns and $(k + 1)^2 + 2k$ linear constraints. If we were to consider as "regions" the subpopulations U_i determined by cross-classifying 'NUTS 3'*'sex'*'5 years age classes', we would need to handle approximately 37,000,000 unknowns. For problems of this size the closed-form solution proposed by Stone *et al*., 1942, which is essentially derived from the generalised least squares method, is so computationally demanding that cannot be applied in practice. As a viable alternative, an iterative constrained optimisation approach was proposed by Byron, 1978, which exploits the Conjugate Gradient algorithm, see also (van der Ploeg, 1982) and (Nicolardi, 1998). The iterative Conjugate Gradient algorithm is indeed computationally very efficient, see (Greenbaum, 1997), and proved a perfect fit for the stocks and flows reconciliation task (8). To fully automate the solution of this task, we implemented a dedicated software system, based on R (R Core Team, 2022). This software system is publicly available on GitHub[7].

NSIs need to publish population counts by domains that cross territory with covariates like 'sex', 'age class', 'nationality', and so on. We remark that the macro-integration method described here can produce consistent estimates in this context as well. Indeed, when covariates like 'sex', 'age class', and 'nationality' are introduced, we can still write down *generalised DBEs* constraining cell counts of the corresponding N-way classification at subsequent times t and $t + 1$. However, these covariates bring into play:

i. A *more abstract notion of migration flows, e.g.* people can "migrate" from a given 'age class' to the subsequent one or from one 'nationality' to another.

ii. *New structural constraints* (*i.e.* "illicit migrations"), *e.g.* since people cannot get younger, they can only get stuck in their original 'age class', move to the next one, or die[8].

7 https://github.com/DiegoZardetto/Stocks-AND-Flows.

8 Note that structural constraints arising from variable 'age class' can actually be greatly simplified by trading variable 'age class' for variable 'class of cohort'. When using cohorts, the only residual constraint is that the modality of variable 'class of cohort' cannot change between t and $t + 1$.

Fortunately, we can leverage *reliability weights* in equation (7) to prevent "illicit migrations" from being generated within the balanced solution. Since illicit cells have 0 *raw counts*, all we have to do is to let $R[\cdot] \to \infty$ and the corresponding *balanced counts* will still be 0.

The software system we developed can indeed handle the additional complexity illustrated above at scale, thus allowing Istat to achieve stock and flow consistency for all the subnational estimates that must be officially disseminated in accordance with Italian and European statistical regulations. This has been extensively tested in practice using officially released estimates of population counts and demographic flows referred to the period [2018, 2019]. For instance, our system managed to obtain perfect consistency for all the DBEs related to the k = 8,640 subpopulations U_i determined by crossclassifying the following variables:

'NUTS 3' *	'citizenship (Italian/non-Italian)'	* 'sex'	* '5 years age classes'
[108 modalities]	[2 modalities]	[2 modalities]	[20 modalities]

Although in the setting above the optimisation problem (8) involves nearly 150,000,000 unknowns, completion of the balancing task only required about 40 minutes using an ordinary Windows server machine equipped with sufficient RAM (memory usage peaked at 32 GB).

Coming to the statistical properties of the balanced (*i.e.* final) estimates of population stocks and flows, (Theil, 1961) has shown that they are BLUE (best linear unbiased estimates) if:

1. *Errors* affecting raw (*i.e.* initial) estimates are *uncorrelated* and have *zero mean.*
2. *Reliability weights* are equal to *inverse variances* of raw estimates.

When the above assumptions do not hold, *e.g.* because raw estimates are *biased* or reliability weights are *misspecified*, the general properties of balanced estimates are no longer under theoretical control. Yet, of course, they can still be investigated through Monte Carlo simulations, as will be shown in Section 3 (see also Di Zio *et al.*, 2018). Experiments on real Italian data reported herein suggest that, under reasonable assumptions, the proposed approach determines improved estimates of population counts: bias and variance of demographic stocks and flows are both greatly reduced by the balancing process. As will be illustrated in Section 3, key assumptions underpinning this result are the following: (*i*) errors affecting civil registry figures of births and deaths are negligible; (*ii*) high-quality estimates of population counts are available at time *t*.

3. Simulation study

In this Section, we present a simulation study designed to investigate the behaviour of balanced estimates in a more general setting than the ideal one addressed in Theil (1961) and introduced in Section 2.

We start with official demographic figures (P^t, B, D, N, F, M) of administrative Italian regions (NUTS 2) in 2015 so that $k = 20$. From these data, we compute P^{t+1} using the DBE: this way the set $(P^{t+1}, P^t, B, D, N, F, M)$ *exactly* fulfills the DBE *by construction*. Then, we use such figures as *ground-truth* and perturb them to generate *raw estimates*. Note that in the following, as already stated, a *tilde hat* denotes *raw estimates* (*e.g.* $\tilde{P}_i^{t+1}$), a *circumflex hat* denotes *balanced estimates* (*e.g.* $\hat{P}_i^{t+1}$), whereas *no hat* denotes *ground-truth* values (*e.g.* P^{t+1}). The simulation goes as follows:

a. We assume that *births*, *deaths*, *natural increase,* and *population counts* at time *t* are *known without errors*, *i.e.* $\tilde{B} = B$, $\tilde{D} = D$, $\tilde{N} = N$, and $\tilde{P}^t = P^t$. Therefore, to prevent them from being changed by the balancing algorithm, we set their *reliability weights* to infinite:

$$R[\tilde{B}_i] = R\big[\tilde{D}_i\big] = R\big[\tilde{N}_i\big] = R[\tilde{P}_i^t] \to \infty \tag{9}$$

b. We generate the vector of raw estimates of population counts $\tilde{P}_i^{t+1}$ by adding to P^{t+1} a *Gaussian noise* with a *relative bias* β and a *coefficient of variation α*:

$$\tilde{P}_i^{t+1} = \mathcal{N}(\,\mu = (1+\beta)P_i^{t+1},\, \sigma^2 = (\alpha P_i^{t+1})^2\,) \tag{10}$$

c. We generate the perturbed Migration Flows Matrix $\tilde{F}$ from a *Negative Binomial* distribution centred around F with a *relative bias γ and dispersion parameter δ*:

$$\tilde{F}_{ij} = \mathcal{NB}\left(\,\mu = (1+\gamma)F_{ij}\,,\, v = \mu + \delta\mu^2\,\right) \tag{11}$$

d. We derive the perturbed Net Migration Matrix $\tilde{M}$ from $\tilde{F}$ as generated in (11) following identity (6):

$$\tilde{M} = \tilde{F}' - \tilde{F} \tag{12}$$

e. For the *reliability weights* of $\tilde{P}_i^{t+1}$, $\tilde{F}$ and $\tilde{M}$, we deliberately assume a *naïve and misspecified model*[9], setting their value to the reciprocal of the absolute value of the corresponding raw estimate:

$$R[\tilde{\cdot}] = 1/|\tilde{\cdot}| \tag{13}$$

f. Lastly, we compute *balanced estimates* $(\hat{P}^{t+1}, \hat{F}, \hat{M})$ by solving the constrained optimisation problem (8).

We repeated all the steps above 5,000 times ($s = 1, \ldots, S$, with S = 5,000) and compared the resulting Monte Carlo distributions of *raw estimates, balanced estimates, and ground-truth figures*. For evaluation, we used standard global accuracy measures:

- **MARB** (Mean Absolute Relative Bias): the average over regions of absolute values of Monte Carlo estimated relative biases (see equations (14)).
- **MRRMSE** (Mean Relative Root Mean Squared Error): the average over regions of absolute values of Monte Carlo estimated relative square roots of MSEs (see equations (15)).

For instance, setting for notational convenience $t \stackrel{\text{def}}{=} 0$ and $t+1 \stackrel{\text{def}}{=} 1$, the accuracy measures for the *balanced population counts* have been computed as follows:

$$RB_i = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{P}_i^{1(s)} - P_i^1}{P_i^1}\right) \qquad MARB = \frac{1}{k}\sum_{i=1}^{k}|RB_i| \tag{14}$$

$$RRMSE_i = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{P}_i^{1(s)} - P_i^1}{P_i^1}\right)^2} \qquad MRRMSE = \frac{1}{k}\sum_{i=1}^{k}RRMSE_i \tag{15}$$

We have studied 5 different simulation scenarios: S1, …, S5. The main features of these scenarios are summarised in Table 3.1. Note that different simulation scenarios have been highlighted in Table 3.1 using different colours so to make the presentation of results easier.

9 Model (13) is misspecified in the sense that it defines reliability weights that are clearly *different from the inverse of the variances* of the errors generated by equations (10), (11), and (12). Note that the *naïve* model (13) would instead be appropriate in case raw stocks and absolute values of raw flows follow a Poisson distribution.

Table 3.1 - The investigated simulation scenarios

Scenario	Main Features
S1	No Bias
S2	Only Migration Bias
S3	Both P1 and Migration Biases
S4	Overdispersed Migrations
S5	High Bias - High Variance

Source: Authors' construction

Note also that simulation scenarios S1, …, S5 have to be intended as a *hierarchy*, in the sense that each scenario actually *adds* its main features to those characterising the previous one (*e.g.* S4 switches on overdispersion in perturbed migration flows, but both migration flows and population counts $\hat{P}_i^1$ are already biased owing to S3).

Within each scenario, two combinations of the simulation parameters $(\beta,\ \alpha,\ \gamma,\ \delta)$ defined in (10) and (11) have been investigated. The simulation parameters and the corresponding simulation results expressed in terms of MARB(%) and RRMSE(%) for the *balanced population counts* $\hat{P}_i^1$ are reported in Table 3.2. The rows of Table 3.2 have been consistently highlighted with the same colours that have been used in Table 3.1 to differentiate the simulation scenarios. This makes it straightforward to visually link a given combination of simulation parameters of Table 3.2 to the simulation scenario it belongs to.

Table 3.2 - Main results of the Monte Carlo simulation (5 scenarios, 10 combinations of simulation parameters, 5,000 runs for each combination)

Simulation Parameters						Evaluation Criteria					
P1 Raw		Raw Migration Figures				P1 MARB (%)			P1 MRRMSE (%)		
RBias (%)	CV (%)	Matrix	RBias (%)	Disp (%)	Avg|CV (%)	Bal	Raw	Bal/Raw	Bal	Raw	Bal/Raw
0	10	F	0	0	8	0.0	0.1	-	0.0	10.0	0.2
0	10	M	0	0	15	0.0	0.1	-	0.0	10.0	0.2
0	10	F	-50	0	11	0.1	0.1	-	0.1	10.0	1.1
0	10	M	-50	0	21	0.1	0.1	-	0.1	10.0	1.1
-5	10	M	-50	0	21	0.1	5.0	2.1	0.1	11.2	1.0
5	10	M	-50	0	21	0.1	5.0	2.1	0.1	11.2	1.0
-5	10	F	-50	20	47	0.1	5.0	2.1	0.2	11.2	1.6
-5	10	M	-50	20	53	0.1	5.0	2.1	0.1	11.2	1.0
-10	20	F	-50	20	47	0.1	10.0	1.1	0.2	22.4	0.8
-10	20	M	-50	20	53	0.1	10.0	1.1	0.1	22.4	0.5

Source: Authors' construction

The left panel of Table 3.2 reports simulation parameters used to generate raw values of *population counts* ($\tilde{P}^1$) and *migration figures* (both $\tilde{F}$ and $\widetilde{M}$) for the k = 20 administrative Italian regions: 'RBias' columns indicate β and γ respectively, 'Disp' indicates δ, and 'Avg|CV|' indicates average CVs (in absolute value) of migration figures resulting from a given choice of (γ, δ) in the Monte Carlo. Note that, for migration matrices $\tilde{F}$ and $\widetilde{M}$, we computed 'Avg|CV|' figures from the outcomes of the Monte Carlo simulation and showed them in Table 3.2 because such CVs are not directly controlled by parameters of the simulation (see equations (11) and (12)), at odds with population counts (equation (10)).

The right panel of Table 3.2 reports the MARB(%) and RRMSE(%) for the *balanced* and the *raw estimates* of P^1 (columns 'Bal' and 'Raw' respectively), as well as the ratios of the corresponding accuracy measures (column 'Bal/Raw', given in percentages once again). The main results of Table 3.2 can be summarised as follows:

- Even though we injected *substantial bias* inside *raw estimates* of population counts ($\tilde{P}^1$) and migration figures ($\tilde{F}$ and $\widetilde{M}$), *balanced estimates* of population counts ($\hat{P}^1$) *are always nearly unbiased*: balancing removed *at least* 98% of the original bias.
- In all simulation scenarios, *balancing dramatically increased the efficiency of* P^1 *estimates*: the MSE of balanced estimates $\hat{P}^1$ is only ~1% of raw estimates' one.

Based on these findings, we are led to the conclusion that the benefits of balancing go beyond the coherence dimension of data quality and extend to the accuracy dimension. In fact, the simulation study showed that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to raw ones, and the accuracy gain seems robust against misspecification of reliability weights. Clearly, the results obtained here pertain to the adopted simulation settings and the investigated simulation scenarios. Real-world applications may exist such that the errors affecting raw estimates of population stocks and flows exhibit distributions that structurally differ from (10) and (11). Furthermore, the errors in the stocks might in principle be somewhat correlated with those in the flows. In the Italian demographic estimation system, for instance, municipal civil registries not only directly provide raw estimates of flows, but also feed the BRI before it is integrated with (and adjusted by) the Permanent Census, thus indirectly contributing to the estimates of stocks as well. In this regard, we plan to conduct further research and explore different simulation settings to improve the robustness of our findings.

The specification of reliability weights adopted in the simulation deserves a few dedicated remarks. Equation (13) implies that larger changes are expected in population counts than in flows, but – in a sense – this is a desirable side-effect, as it tends to more evenly distribute relative changes while restoring the DBEs. Moreover, whenever good information about the variances of error terms affecting raw estimates is unavailable, it is a common choice to set reliability weights to reciprocals of raw estimates (after taking absolute values, if needed). This choice dates back to Stone's seminal works and is often applied in the National Accounts directorates of NSIs for GDP estimation. It is also important to acknowledge that our approach, besides hopefully reducing biases and variances of stocks and flows in simulations, must lead to consistent (*i.e.* compliant with the DBEs) and well-behaved (*i.e.* non-pathological) estimates in concrete applications. The extensive large-scale tests we touched upon at the end of Section 2 (which we conducted using Italian officially released estimates of population counts and demographic flows referred to the period from 2018 to 2019) showed that the discussed specification of reliability weights (13) is indeed beneficial. For instance, balancing methods in themselves cannot protect against the occurrence of sign-changing adjustments. This could, in principle, lead to negative balanced estimates of positive-definite quantities. Arguably, the risk of such a pathological outcome is much higher for raw estimates that are smaller in size, *e.g.* the vast majority of the flows $\tilde{F}_{ij}$ in any highly disaggregated application. From our tests, we gathered compelling empirical evidence that the specification $R\left[\tilde{F}_{ij}\right] = 1/\left|\tilde{F}_{ij}\right|$ of equation (13) plays an important role in preserving the positivity of balanced flows' estimates. In future extensions of this work, we plan to more deeply involve demographers and subject matter experts who can propose different specifications of the reliability weights and test their impact under additional simulation settings.

4. Downstream effects on the Base Register of Individuals

Once balanced population counts are obtained through our macro-integration procedure (8), they are ready to be released as official estimates of population stocks. In addition, we propose to use them as control totals to adjust the individual weights tied to records of the BRI. This would determine two beneficial effects:

- First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by the balancing procedure.
- Second, since balancing has been shown to greatly reduce bias that could possibly affect raw estimates of population counts, the weights of the BRI would receive a second layer of protection against bias, "borrowing strength" from both official population counts disseminated the year before and balanced demographic flows.

In the following, we illustrate how this proposal can be implemented in practice. We also point out some technical details that might become relevant in case, in the foreseeable future, the process that currently generates BRI weights (see Falorsi, 2017 and Righi *et al.*, 2021 for details) is improved to accommodate integrated individual- and household-level weights.

In practice, *raw* estimates of population stocks in equations (7) and (8):

$$\begin{cases} \tilde{P}^{(t)} = \left(\tilde{P}_1^{(t)}, \dots, \tilde{P}_k^{(t)}\right)' \\ \tilde{P}^{(t+1)} = \left(\tilde{P}_1^{(t+1)}, \dots, \tilde{P}_k^{(t+1)}\right)' \end{cases} \tag{16}$$

to be fed as input to the balancing procedure, come from *internally released* BRI versions referred to times t and $t + 1$. As noted in Section 1, each released version of the BRI – referred to a generic time τ – will contain undercoverage and over-coverage corrected weights associated to *individual* records:

$$d_q^{(\tau)} \qquad \mathrm{q} = 1, \dots, \mathrm{N}_{\mathrm{BRI}}^{(\tau)} \tag{17}$$

and raw population counts for each balancing cell (= subpopulation) U_i will be obtained by simply adding the weights of BRI's individuals belonging to the cell:

$$\tilde{P}_i^{(\tau)} = \sum_{q \in U_i} d_q^{(\tau)} \qquad \text{i} = 1, \ldots, \text{k} \qquad (18)$$

Once the balancing procedure has been successfully executed, output *balanced* estimates of population stocks will be available. These balanced estimates can easily be exploited to *adjust* individual weights of the BRI in such a way that estimates of population counts derived from *adjusted BRI weights* fulfill all the DBEs in equation (8):

$$\begin{cases} d_q^{(\tau)} \xrightarrow{\text{BALANCING}} w_q^{(\tau)} & \text{q} = 1, \ldots, \text{N}_{\text{BRI}}^{(\tau)} \\ & \qquad\qquad \text{such that:} \\ \sum_{q \in U_i} w_q^{(\tau)} = \hat{P}_i^{(\tau)} & \text{i} = 1, \ldots, \text{k} \end{cases} \qquad (19)$$

To show how this appealing result can be achieved in practice, let us start with a very important distinction between population stocks referred to time t and those referred to time $t + 1$ (namely, the population counts reported in (16) and involved in equations (7) and (8)).

1. At the time the balancing procedure takes place $t^{BAL} > t + 1$, *Istat will have already disseminated to the external audience official population counts referred to time* t. Therefore:

   i. The balancing procedure taking place at time $t^{BAL} > t + 1$ will be performed in such a way that these official population estimates will *not* be altered:

   $$\hat{P}_i^{(t)} \equiv \tilde{P}_i^{(t)} \qquad \text{i} = 1, \ldots, \text{k} \qquad (20)$$

   which will simply be obtained by letting $R\left[\tilde{P}_i^{(t)}\right] \to \infty$.

   Note that this will allow Istat to produce balanced estimates on a yearly basis without the need to revise ever again any already disseminated official population counts.

   On the other hand:

   ii. We can assume BRI weights referred to time t to have *already been adjusted* by a balancing procedure run *the year before* and involving stocks and flows referred to times t – 1 and t.

2. At the time the balancing procedure takes place $t^{BAL} > t + 1$, Istat will have *not* disseminated official population counts referred to time $t + 1$ *yet*, however *a released version of the BRI referred to time t* + 1 *will be available, along with undercoverage and over-coverage corrected (but still raw) individual weights*. Therefore:
   iii. After successfully executing the balancing procedure, we will use its outputs to *adjust* individual weights of the BRI referred to time $t + 1$ in such a way that estimates of population counts derived from the BRI will henceforth be consistent and fulfill all the DBEs:

$$d_q^{(t+1)} \xrightarrow{\text{BALANCING}} w_q^{(t+1)} \qquad \mathrm{q} = 1, \ldots, \mathrm{N}_{\mathrm{BRI}}^{(t+1)} \tag{21}$$

   iv. Istat will use the *post-balancing adjusted individual weights* $w_q^{(t+1)}$ in (21) to compute *official estimates of population counts for arbitrary estimation domains* $\mathrm{D_e}$:

$$\hat{P}_e^{(t+1)} = \sum\nolimits_{q \in D_e} w_q^{(t+1)} \qquad \mathrm{e} = 1, \ldots, \mathrm{E} \tag{22}$$

The way to obtain *post-balancing adjusted* individual weights $w_q^{(t+1)}$ appearing in equation (21) is straightforward. It will only take to post-stratify raw individual weights $d_q^{(t+1)}$ using the balanced population estimates $\hat{P}_i^{(t+1)}$ as calibration benchmarks:

$$w_q^{(t+1)} = d_q^{(t+1)} \left[\frac{\hat{P}_i^{(t+1)}}{\tilde{P}_i^{(t+1)}}\right] = d_q^{(t+1)} \left[\frac{\hat{P}_i^{(t+1)}}{\sum_{q \in U_i} d_q^{(t+1)}}\right] \qquad \forall\, \mathrm{q} \in U_i \tag{23}$$

for $i = 1, \ldots, k$ and $q = 1, \ldots, N_{BRI}^{(t+1)}$

Although equation (23) is formally the solution of a calibration problem (which uses the unbounded Euclidean distance, one single auxiliary variable whose values are identically equal to 1, and calibration domains identified by the U_i cells of the balancing partition), its expression is so simple that a trivial PL/SQL Data Base procedure will be enough to accordingly update the weights of the BRI in a real production setting.

Now suppose that undercoverage and over-coverage corrected weights (17) associated to individual records of the BRI were constructed in such a way that *all the individuals belonging to the same household share the same weight*. This property would ensure consistent estimates of individual-level and household-level aggregates computed from the BRI[10]. Should this be the case, it would be desirable that *post-balancing adjusted* individual weights (21) retain that same property. In order to achieve such a goal, the simple formula (23) would *not* in general be suitable. More specifically, formula (23) would produce constant individual weights within households *only if* the balancing cells (= subpopulations) U_i *do not* cut-across households; of course, this would *not* be the case whenever individual-level variables like 'sex' or 'age class' are involved as classification variables in the balancing procedure.

For arbitrary settings – *i.e*. whatever might be the choice of balancing cells (=subpopulations) U_i – the goal of obtaining *post-balancing adjusted* individual weights that are *constant within households* can be easily attained[11] by means of standard calibration software available in Istat, *e.g. ReGenesees* (Zardetto, 2015). Even though the size of the BRI (~6·107 rows) may seem at first sight prohibitively large when compared to typical survey samples for which calibration procedures are routinely carried out in Istat, actually this will not pose any serious computational or technical issues. Indeed, suitable territorial domains can always be identified that would allow us to divide the overall calibration problem into smaller, computationally-affordable subproblems, which can be solved independently, one by one. In general, however, these subproblems will no longer entail a simple post-stratification, but rather require a more complex calibration of BRI weights. Consequently, the mathematical relation between weights $w_q^{(t+1)}$ and $d_q^{(t+1)}$ will be more complex than (23) and, in general, no longer expressible in analytic closed-form.

10 For instance, the estimated number of people living in households of 4 members would be equal to 4 times the estimated number of households of 4 members.

11 Methods to achieve integrated household-individual calibration weights are discussed in, *e.g.* Lemaître and Dufour (1987) and Heldal (1992). Istat's software ReGenesees adopts the approach of Heldal (1992). Put briefly, first the individual-level calibration model-matrix (which can encode simultaneously both household-level and individual-level auxiliary variables) is aggregated at houshold-level, then the calibration task is performed on this aggregated dataset, lastly the obtained household-level calibration weights are re-expanded to the individual-level, attaching to each individual the calibration weights of the household it belongs to.

5. Final remarks

Istat's production system, underpinned by statistical registers that integrate survey and administrative data, is ideally positioned to overcome the challenge of producing timely, reliable, and coherent estimates of demographic stocks and flows. To this end, official estimates of demographic stocks and flows should be consistent and satisfy the demographic balancing equation (DBE). Although this objective is inherently hard in a multi-source estimation system, we showed it can be attained using balancing methods, akin to the Stone-Byron approach often adopted to reconcile estimates in national accounting systems.

We designed and implemented a macro-integration procedure that can handle the demographic balancing task at scale, thus allowing Istat to achieve stock and flow consistency for all the domain estimates that must be officially disseminated in accordance with Italian and European statistical regulations. Consistency promotes credibility in published statistics, thereby enhancing the reputation of the National Statistical Institute. What is more, simulation results showed that our balancing approach also determines improved estimates of population counts: besides gaining consistency, they exhibit lower bias and variance as compared to "raw" estimates (*i.e.* those derived from the integrated system formed by the Permanent Census and the Base Register of Individuals (BRI)). Since the performance of our balancing approach was indeed very good in the illustrated simulation study, it is in our plans to investigate different simulation settings and scenarios in order to make our conclusions even more robust, *e.g*. by considering alternative error models and reliability weights.

In this work, simulation evidence was obtained under two fundamental assumptions (see Section 3, point A): (*i*) errors affecting civil registry figures of births and deaths are negligible; (*ii*) high-quality estimates of population counts at time t = 0 are available. Condition (*i*) is surely realistic in Italy. Condition (*ii*) has far-reaching practical consequences. Indeed, as illustrated in our simulation, unbiased estimates $\tilde{P}^t$ induce balanced estimates $\hat{P}^{t+1}$ which are still nearly unbiased. This would allow Istat to produce and publish balanced estimates on a yearly basis, without the need to revise ever again any already disseminated official estimates.

Beyond the simulation context, we did not discuss the uncertainty of balanced estimates. This is essentially for two reasons. First, as of today, Istat neither formally evaluates nor disseminates measures of the uncertainty affecting the inputs to our procedure, namely raw estimates of demographic stocks and flows. We note that this would be a challenging endeavor in itself, considering the variety of sampling and non-sampling errors affecting all the involved data sources (municipal civil registries, statistical registers, and probability sample surveys with complex sampling designs). Second, the macro-integration procedure we proposed does not rely on explicit model assumptions concerning the distributions of the errors affecting raw stocks and flows (although some implicit assumptions admittedly enter the picture via the choice of reliability weights). Some authors (Bryant and Graham, 2013, and Bryant and Graham 2015) adopted a purely model-based approach to demographic accounts (more specifically, a Bayesian approach) that naturally leads to uncertainty measures of output estimates. Further studies are needed to assess the pros and cons of model-based approaches to the Italian demographic estimation system.

Finally, we provided guidance on the scope of, and considerations for, using balanced population counts as control totals to adjust individual weights of the BRI. This downstream feedback would indeed determine two beneficial effects. First, estimates of population counts derived from adjusted BRI weights would exactly match officially disseminated estimates, and thereby satisfy the DBE for all the domains addressed by our procedure. Second, BRI weights would receive a second layer of protection against bias, “borrowing strength” from both official population counts disseminated the year before and balanced demographic flows.

References

Alleva, G., P.D. Falorsi, O. Luzi, and M. Scannapieco. 2019. “Building the Italian Integrated System of Statistical Registers: Methodological and Architectural Solutions”. *ESS Workshop on the use of administrative data and social statistics*, Valencia, Spain, 4th and 5th June 2019.

Alleva, G., P.D. Falorsi, F. Petrarca, and P. Righi. 2021. “Measuring the Accuracy of Aggregates Computed from a Statistical Register”. *Journal of Official Statistics - JOS*, Volume 37, Issue 2: 481-503.

Bryant, J.R., and P.J. Graham. 2015. “A Bayesian Approach to Population Estimation with Administrative Data”. *Journal of Official Statistics - JOS*, Volume 31, Issue 3: 475-487.

Bryant, J.R., and P.J. Graham. 2013. “Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources”. *Bayesian Analysis*, Volume 8, N. 3: 591-622.

Byron, R.P. 1978. “The Estimation of Large Social Account Matrices”. *Journal of the Royal Statistical Society*, Series A (General), Volume 141, N. 3: 359-367.

Di Zio, M., M. Fortini, and D. Zardetto. 2018. “Reconciling Estimates of Demographic Stocks and Flows through Balancing Methods”. Presented at the *9th European Conference on Quality in Official Statistics - Q2018,* Krakow, Poland, 26th – 29th June 2018.

Eurostat, and Statistics Netherlands - CBS. 2003. “Demographic statistics: Definitions and methods of collection in 31 European Countries”. *Working Papers and Studies - Population and social conditions*, 3/2003/E/n. 25. Luxembourg: Office for Official Publications of the European Communities.

Falorsi, S. 2017. “Census and Social Surveys Integrated System”. United Nations Economic Commission for Europe – UNECE, *Conference of European Statisticians - CES, Group of Experts on Population and Housing Censuses,* Nineteenth Meeting, Geneva, Switzerland, 4th – 6th October 2017.

Greenbaum, A. 1997. *Iterative methods for solving linear systems*. Philadelphia, PA, U.S.: Society for Industrial and Applied Mathematics, *Frontiers in Applied Mathematics*.

Heldal, J. 1992. “A Method for Calibration of Weights in Sample Surveys”.

Working Papers from Department for Statistics on Individuals and Households. Methods for Collections and Analysis, N. 3/1992. Oslo, Norway: Central Bureau of Statistics.

Istituto Nazionale di Statistica/Italian National Institute of Statistics – Istat. 2016. *Istat's Modernisation Programme*. Roma, Italy: Istat. https://www.istat.it/en/files/2011/04/IstatsModernistionProgramme_EN.pdf.

Lemaître, G., and J. Dufour. 1987. "An integrated method for weighting persons and families". *Survey Methodology*, Volume 13, N. 2: 199-207.

Nicolardi, V. 1998. "Un sistema di bilanciamento per matrici contabili di grandi dimensioni". *Quaderni di ricerca*, N. 4: 5-31. Roma, Italy: Istat.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rees, P. 1979. "Regional Population Projection Models and Accounting Methods". *Journal of the Royal Statistical Society*, Series A (General), Volume 142, N. 2: 223–255.

Righi, P., S. Daddi, E. Fiorello, P. Massoli, M.D. Terribili, and P.D. Falorsi. 2021. "Optimal Sampling for the Population Coverage Survey of the New Italian Register-Based Census". *Journal of Official Statistics - JOS*, Volume 37, Issue 3: 655-671.

Stone, R., D.G. Champernowne, and J.E. Meade. 1942. "The Precision of National Income Estimates". *Review of Economic Studies*, Volume 9, N. 2: 111-125.

Stone, R, and J.D. Corbit. 1997. "The Accounts of Society". *The American Economic Review*, Volume 87, N. 6: 17-29.

Theil, H. 1961. *Economic Forecasts and Policy*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Van der Ploeg, F. 1982. "Reliability and the adjustment of sequences of large economic accounting matrices". *Journal of the Royal Statistical Society*, Series A (General), Volume 145, N. 2: 169-194.

Zardetto, D. 2015. "ReGenesees: an advanced R system for calibration, estimation and sampling error assessment in complex sample surveys". *Journal of Official Statistics - JOS*, Volume 31, Issue 2: 177-203.

The *Rivista di statistica ufficiale* publishes peer-reviewed articles dealing with cross-cutting topics: the measurement and understanding of social, demographic, economic, territorial and environmental subjects; the development of information systems and indicators for decision support; the methodological, technological and institutional issues related to the production process of statistical information, relevant to achieve official statistics purposes.

The *Rivista di statistica ufficiale* aims at promoting synergies and exchanges between and among researchers, stakeholders, policy-makers and other users who refer to official and public statistics at different levels, in order to improve data quality and enhance trust.

The *Rivista di statistica ufficiale* was born in 1992 as a series of monographs titled "*Quaderni di Ricerca Istat*". In 1999 the series was entrusted to an external publisher, changed its name in "*Quaderni di Ricerca - Rivista di Statistica Ufficiale*" and started being published on a four-monthly basis. The current name was assumed from the Issue N. 1/2006, when the Italian National Institute of Statistics – Istat returned to be its publisher.

La Rivista di statistica ufficiale pubblica articoli, valutati da esperti, che trattano argomenti trasversali: la misurazione e la comprensione di temi sociali, demografici, economici, territoriali e ambientali; lo sviluppo di sistemi informativi e di indicatori per il supporto alle decisioni; le questioni metodologiche, tecnologiche e istituzionali relative al processo di produzione dell'informazione statistica, rilevanti per raggiungere gli obiettivi della statistica ufficiale.

La Rivista di statistica ufficiale promuove sinergie e scambi tra ricercatori, stakeholder, policy-maker e altri utenti che fanno riferimento alla statistica ufficiale e pubblica a diversi livelli, al fine di migliorare la qualità dei dati e aumentare la fiducia.

La Rivista di statistica ufficiale nasce nel 1992 come serie di monografie dal titolo "Quaderni di Ricerca Istat". Nel 1999 la collana viene affidata a un editore esterno, cambia nome in "Quaderni di Ricerca - Rivista di Statistica Ufficiale" e diventa quadrimestrale. Il nome attuale è stato scelto a partire dal numero 1/2006, quando l'Istituto Nazionale di Statistica - Istat è tornato a esserne l'editore.

ISSN 1828-1982 € 15,00