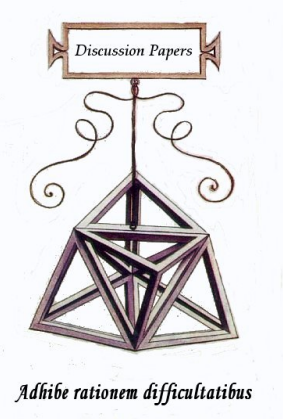
Discussion Papers
Collana di
E-papers del Dipartimento di Economia e Management – Università di Pisa



Monica Pratesi, Stefano Marchetti, Caterina Giusti, Gaia Bertarelli, Francesco Schirripa Spagnolo, Luigi Biggeri

Estimations of local spatial price indices using scanner data, to be used for the comparisons of economic poverty measures

Discussion Paper n. 281
2021

Discussion Paper n. 281, presentato: **Dicembre 2021**

Indirizzo dell'Autore:

Monica Pratesi, Dipartimento di Economia e Management, via Ridolfi 10, 56124 PISA – Italy. Email: monica.pratesi@unipi.it

Stefano Marchetti, Dipartimento di Economia e Management, via Ridolfi 10, 56124 PISA – Italy. Email: stefano.marchetti@unipi.it

Caterina Giusti, Dipartimento di Economia e Management, via Ridolfi 10, 56124 PISA – Italy. Email: caterina.giusti@unipi.it

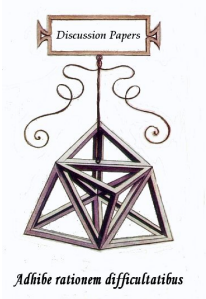
Gaia Bertarelli, Istituto di Management, EMbeDS, Scuola Universitaria Superiore Sant'Anna, via San Zeno, 2, 56124 PISA – Italy. Email: gaia.bertarelli@santannapisa.it

Francesco Schirripa Spagnolo, Dipartimento di Economia e Management, via Ridolfi 10, 56124 PISA – Italy. Email: francesco.schirripa@ec.unipi.it

Luigi Biggeri, Centro Interuniversitario ASESD Camilo Dagum, via Ridolfi 10, 56124 PISA – Italy. Email: luigi.biggeri@unifi.it

Si prega di citare così:
Pratesi M., Marchetti S., Giusti C., Bertarelli G., Schirripa Spagnolo F., Biggeri L. (2021), "Estimations of local spatial price indices using scanner data, to be used for the comparisons of economic poverty measure", Discussion Papers del Dipartimento di Economia e Management – Università di Pisa, n. 281 (http://www.ec.unipi.it/ricerca/discussion-papers).

Discussion Paper
n. 281



Monica Pratesi, Stefano Marchetti, Caterina Giusti, Gaia Bertarelli, Francesco Schirripa Spagnolo, Luigi Biggeri

Estimations of local spatial price indices using scanner data, to be used for the comparisons of economic poverty measures

Abstract

In the last decades, the fight against poverty is assuming a more and more central role in Europe. Indeed, among the goals of Agenda 2030, one of the priorities is to end poverty, in all its forms and dimensions. Moreover, the demand for poverty and living conditions data referred to local areas and/or sub-populations, has become urgent because policymakers need to know the spatial distribution for implementing policies and distributing resources. At the same time is essential to measure within-country differences in the cost of living, assessing and comparing poverty levels in real terms. Scanner data on retail price represent an interesting data source to renew the computation of sub national Spatial Consumer Price Indexes. Therefore, in this paper, we propose a new approach to construct sub national Spatial Consumer Price Indexes for Italian provinces by using scanner data with a particular focus on the poorest part of the Italian population.

Keywords: Local poverty indicators, Sustainable Development Goals (SDGs), Inequality, Official Statistics

JEL: C18, C80, C83,

1 Introduction

The fight against poverty is still nowadays one of the major concern over the world. However, although the 2030 Agenda for Sustainable Development have set a list of targets to eradicate poverty, the debate on which data and indicators are more useful to effectively measure poverty and living conditions is still very actively. Moreover, mapping the spatial distribution of poverty at a finer subregional and local level it is necessary for the policy makers in order to formulate and implement policies, distribute resources and evaluate the effect of the local policy actions (Pratesi, 2016). However, when the aim is to study the spatial distribution of poverty making comparisons among areas, it could be interesting taking into account the potential differences of prices in the local areas to be compared. Accordingly, in the last years, a new dimension has been more and more considered: the cost of living over the areas. Indeed, considering income in real terms allows to take into account for inflation and for differences in the price levels. Furthermore, the spatial price adjustment could be important for policy makers who need to know where poverty is concentrated. Therefore, it is essential to define spatial measures of price levels.
The construction of sub national Spatial Consumer Price Indexes (SN-SCPIs) is the basis to measure the changes in the prices of goods and services paid by consumers across different areas within a country. At international level, the International Comparison Programme (ICP) of the World Bank provides a statistical initiative for measuring the relative cost-of-living both over time and areas. Generally, to construct the SN-SCPIs national statistical offices (NSOs) collect data on prices of a large range of products representative of the consumption baskets of the households and the information on weights reflecting the importance of the different products. However, in the last years more and more NSOs are using scanner data for the production of the CPI for constructing both temporal consumer price indexes and sub national Spatial Consumer Price Indexes.

The availability of high-frequency "scanner data" in addition to other sources of data enables price statisticians to deal with the SN-SCPIs issue from a renewed approach. 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, the GTIN (Global Trade Item Number) code. The scanner data of the modern distribution can provide millions of prices for thousands of products (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, unit value price per item code can be computed by dividing total turnover for that item by the total quantity sold.

Regardless of which provider scanner data come from, NSIs must reclassify them in order to make them suitable for constructing the mentioned indexes. It should be noted that the SN-SCPIs are in essence direct spatial price level comparisons because within a country there is a common currency.

Therefore, the first aim of this paper is to propose a new approach to construct SN-SCPIs in Italy by using scanner data.

Moreover, in order to calculate SPIs closer to the prices paid by the poor, preliminary experiments have been conducted by using the data of the first quintile of the price distributions, assuming that the poor purchase the cheaper items of a product. To have more information on this field, the researchers of Istat (the Italian National Statistical Institute) have done a specific analysis of data collected with the Italian Households Expenditure Survey (HES) to know where people in condition of absolute poverty purchase some large consumption products.

The paper is organized as follows: in the Section 1, information on the characteristics of the available scanner data base is presented; in the Section 2 the procedure, methodology and the main results of the estimation of the indexes, by using the above mentioned approach, and also with reference to prices of the cheaper items, are presented. Section 4 a brief description of the differences in the purchasing attitudes between poor and non-poor household is presented. Final remarks and recommendations in section 5 conclude the paper.

2 The scanner data base

Scanner data on retail price used in this paper refers to 2018 and they have been provided to researchers by Istat (the Italian National Statistical Institute). These data contains the following information at the GTIN (Global Trade Item Number) code: sales, expenditure, quantities and detailed information on characteristics of products sold. After a process of data cleaning, for each GTIN the weekly unit value price is calculated as the ratio of the weekly turnover and the weekly quantities. The monthly and annual unit value prices are then obtained by averaging the weekly prices weighted with quantities.

Scanner data has been recently introduced by Istat in the official Consumer Price Indexes computation, moreover a series of experiments on the estimations of SN-SCPIs at a regional level on an annual basis have been implemented(Laureti and Polidoro, 2017, Laureti et al., 2017).

However, the scanner data have also some limitations. They cover 103 over 107 provinces, but the rural areas are not covered; they cannot be used for perishables and seasonal products such as vegetables, fruit and meat, and fresh fish. Accordingly to (Istat, 2020) the scanner data available cover about the 10.5% of the total expenditures for the consumption of the families and this is not uniform across the Italian territory.

3 Spatial Price Indexes: ASESD approach

3.1 Methodological approach

A standard approach, based on the one used in the ICP of the World Bank to compute the international Purchasing Power Parities (PPPs), should be based

on a very tight principle of comparability by considering the comparisons of the "like to like" items (products) for the different sub-national areas. However, there is the risk that not all the products are available in all the areas and/or in results are influenced by the characteristics of the modern retail trade which is not uniformly distributed across Italian territory.

We decide to embrace a different approach. The principle of comparability is applied at the level of group of products, by loosing the 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 be gives 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, considering the basket of elementary products that the consumers of each area have really purchased. Then the average level of prices of the group of products is aggregated to obtain the SB-SPIs for each sub national area. Therefore, these groups, and not the Basic Headings (BHs), are the building blocks of the comparison, defined using the ECOICOP-8-digit classes of products.

Following this approach we estimate the SPIs at provincial level (NUTS 3 level in EU classification) in Italy.

3.2 Estimation of Spatial Consumer Price Indexes for the Italian Provinces

We have computed the spatial consumer price indices (SN-SCPI) for Italian provinces for Italian provinces, by using the scanner data of the products sold in modern distribution chains referring to the year 2018 and only to the products (barcodes or GTINs) in food and beverages categories, excluding fresh food. Usually the information on products' quantities is reported in terms of grams and millilitre, but sometimes in units; given that we needed to use comparable prices, we discarded about 17,000 quotations expressed in units.

To estimates the SN-SCP for each of the 103 provinces two-step procedure has been followed.

In the first step, we computed the average unit price at level of province, by considering the unit value prices from the consumer side (or points of view). In applying the principle of comparability, we did not follow a very tight way by considering the comparisons of the 'like to like' items (products). Instead, we applied the principle at a different level, the level products' groups, and exactly at the level of the 102 groups of the ECOICOP-8-digit classification. The hypothesis is that the elementary products (items) within a group are sufficiently similar, so that consumers are generally indifferent about the choice of the product that in any case satisfy the same consumer needs (may be giving him/her the same utility), even if the brand, quality, etc. is different. The comparison is therefore done by considering the average level of prices of the group of products purchased in the different provinces, considering the basket of

elementary products that the consumers of each province have really purchased[1].

In what follow we define the weighted mean price $\bar{p}_{ij}$ for ECOICOP-8-digit j and province i. Let r_{ijk} and q_{ijk} be the annual turnover and the total quantity sold[2] respectively of item k belonging to ECOICOP-8-digit j in province i. These quantities are estimated by Istat using the scanner data and the sampling weights computed according to the survey design summarised in section 2. Let u_{ijk} be the quantity of the item ijk in terms of gr. or ml. For each item we define its annual price per gr. or ml. as

$$p_{ijk} = \frac{\frac{r_{ijk}}{q_{ijk}}}{u_{ijk}}.$$

Then, for each item we define its relative weights in term of turnover as

$$w_{ijk} = \frac{r_{ijk}}{\sum_{k=1}^{n_{ij}} r_{ijk}},$$

where n_j is the number of items in the jth ECOICOP-8-digit aggregation and the ith province. Finally, the weighted mean price is:

$$\bar{p}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} p_{ijk} w_{ijk}.$$

Therefore, $\bar{p}_{ij}$ is the weighted mean price per gr. or ml. for products in ECOICOP-8-digit j and province i.

The second step is devoted to the aggregation of 102 average level of prices to estimate the provincial Sn-SCPI. Note that not all the ECOICOP-8-digit aggregates are present in all the provinces.

To compute the SPIs at provincial level we adapt a Country Product Dummy model (Laureti and Rao, 2018). The products are aggregated by province and ECOICOP-8-digit classification, for a total of 103 provinces and 102 ECOICOP-8-digit. Note that not all the ECOICOP-8-digit aggregates are present in all the provinces. The CPD model we propose is as follows:

$$\log \bar{p}_{ij} = \alpha_0 + \alpha_i D_i + \beta_j I_j + \varepsilon_{ij}, \quad i = 1, \ldots, 103 \quad j = 1, \ldots, 102, \qquad (1)$$

where D_i is a vector equal 1 if the mean price is in province i and 0 otherwise, I_j is equal 1 if the mean price belongs to jth ECOICOP-8-digits and 0 otherwise. The index i is for the provinces and the index j is for the ECOICOP-8-digit. The error $\varepsilon_{ij} \sim N(0, \sigma^2)$.

[1] 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 more precise comparison among the different averages, it could be necessary to make some standardization of the provincial averages. This is an issue that the unit of research will deepen in a near future

[2] Which are the expenditure and the quantity purchased by consumers.

To take into account the different level of the turnover between the ECOICOP-8-digit aggregates we estimate the model (1) using weighted least squares, where the weights are computed as

$$wls_{ij} = \frac{\sum_{k=1}^{n_{ij}} r_{ijk}}{\sum_{k=1}^{n_i} r_{ijk}},$$

that is the ratio between the total turnover of one aggregate in one province and the total turnover in the province (n_i is the number of items in the ith province).

Model (1) – as it is specified – is not identified, because the D_is vectors are a linear combination of the constant. Therefore, we impose the constraint $\alpha_1 = 0$ so that the model is identified. Once the model is estimated, from the data we obtain the estimates of the SPIs at provincial level by $\exp(\hat{\alpha}_i)$, where $\hat{\alpha}_i$ is the estimate of α_i. The coefficient α_i is the difference of fixed effects connected with the province i compared with the base province $i = 1$. To use as a reference Italy instead of area 1, the coefficient $\hat{\alpha}_i$ has been adjusted following Suits (1984). In this way, α_i represent the fixed effect of province i compared to Italy. Thus, the quantity $\exp(\hat{\alpha}_i)$ represents the spatial price index for food in province i with respect to Italy, and it is also called purchasing power parity of province i (PPP_i).

An advantage of the use of CPD models is that we can obtain p-values for the estimated coefficients. Following Suits (1984) we derive the p-values for the rescaled $\hat{\alpha}_i$s, which are not reported here. Setting a I type error equal to 0.1 we observed 43 provinces for which we don't reject the hypothesis $\alpha_i = 0$, which correspond to a SPI equal 1. Out of these 43 provinces 17 are located in the north, 18 in the center and 8 in the south of Italy.

The SPIs estimated at the province level can be used for many purposes. One of these purposes is to adjust the national poverty line at the province level, by this way relative poverty estimates take into account the different purchase power within the country. An application to Italian data is shown later in the Deliverable. The SPIs estimated according to model (1) are based on mean prices of specific headings (ECOICOP-8-digit), therefore the adjustment of the national poverty line is not poor specific.

As an alternative, our method can be easily extended to produce SPIs related to the first quintile of the distribution of the price of each specific product, assuming that poor purchase the cheaper items of the product. By this way we can adjust the national poverty line using SPIs based on lower prices instead of mean prices, which are reasonably related to poor households. To obtain such SPIs the model (1) is modified as follows:

$$\log Q(\tau, p)_{ij} = \gamma_0 + \gamma_i D_i + \beta_j I_j + \varepsilon_{ij}, \quad i = 1, \ldots, 103 \quad j = 1, \ldots, 102, \quad (2)$$

where $Q(\tau, p)_{ij}$ is the quantile of order τ of the unit prices (p_{ijk}) belonging to heading j (ECOICOP-8-digit) and province i. To estimate γ_i we use the same method used to estimate α_i in model (1).

For example, setting $\tau = 0.2$ we can obtain the estimates of spatial price indices related to the cheaper prices for each Italian provinces, which we denote as SPI($Q_{0.2}$)'s, as it is shown in the next sub-section where the results of the estimations are presented.

3.3 Results of the estimation of SPIs and SPI($Q_{0.2}$)'s

The estimates of SPI computed according to based on model (1) are reported in Figure 1: SPI obtained using mean unit prices on the left and quantile 0.2 of unit prices on the right.

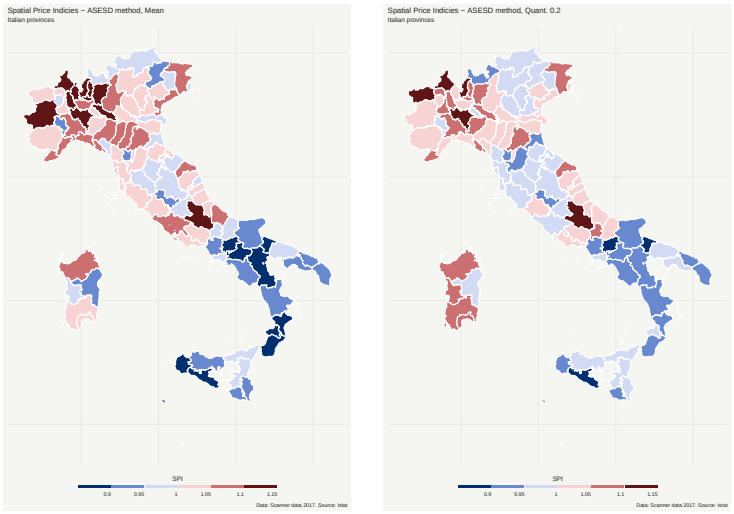


Figure 1: Choropleth map of SPI computed according to ASESD method. SPI obtained using mean unit prices (left) and quantile 0.2 of unit prices (right).

The results we obtained are somehow expected. Indeed, provinces in the south of Italy show SPI smaller than 1, while provinces in the north show values greater than 1. However, there are exceptions, provinces in the north-east Alps mountains show SPI below 1, even if they are close. Provinces in the center of Italy have SPIs close to 1, with some evidence of SPI lower than 1 for provinces located in the Appennino mountains (middle of the central Italy), and SPI greater than 1 for the provinces located on the seaside, both Adriatic (east), Ligure and Tirreno (west). The lowest SPI is estimated for the province of Agrigento (AG), in Sicily (south of Italy), while the highest is in the province of Como (CO), in Lombardia region (north of Italy). The provinces with the highest SPI are all located in the north-west, but Aquila (AQ) located in Abruzzo, a region in the south[3] of Italy.

[3]Actually the Abruzzo region is in the central Italy territorial division, however, for historical reasons it is included among the southern regions

As it concerns the estimates of SPI($Q_{0.2}$) mapped in Figure 1 on the right, we recall that they are obtained modifying model (1). In the results we don't reject the null hypothesis that $\gamma_i = 0$, that is SPI($Q_{0.2}$)=1, for 13 provinces in the north, 15 in the center and 5 in the south of Italy. Looking at the results of the estimation of the SPI($Q_{0.2}$)'s we observe that many provinces in the north-west and on the Adriatic seaside, excluding Puglia provinces, show a SPI($Q_{0.2}$) greater than 1, while many provinces in the north-east, center and south of Italy show SPI($Q_{0.2}$) smaller than 1. Sardegna provinces show SPI($Q_{0.2}$) greater than 1, but Nuoro.

When we build spatial price indices using lowest prices we observe a similar behaviour of indices built with mean prices, however there are differences, for example the province of Rome has SPI($Q_{0.2}$) = 0.986 (pval = 0.4) and SPI = 1.06 (pval = 0.007), Isernia has SPI($Q_{0.2}$) = 1.07 (pval = <0.0001) and SPI = 0.957 (pval = 0.046). Other 19 provinces show discordance between point estimates of SPI($Q_{0.2}$) and SPI, 3 in the south and 16 in the north and central Italy.

In order to check the validity of our method, we have done a comparison between the computed index and the value added per capita by Italian province as reported in the figure 2. This analysis allows us to evaluate the internal consistency of the subnational indexes. Price level in poor areas should be generally lower than those in richer areas and showed a similar patter cross the basic headings. It is evident that the Indexes computed with ASESD method satisfy, in some way the previous mentioned consistency, in fact the correlation coefficients is about + 0,60.

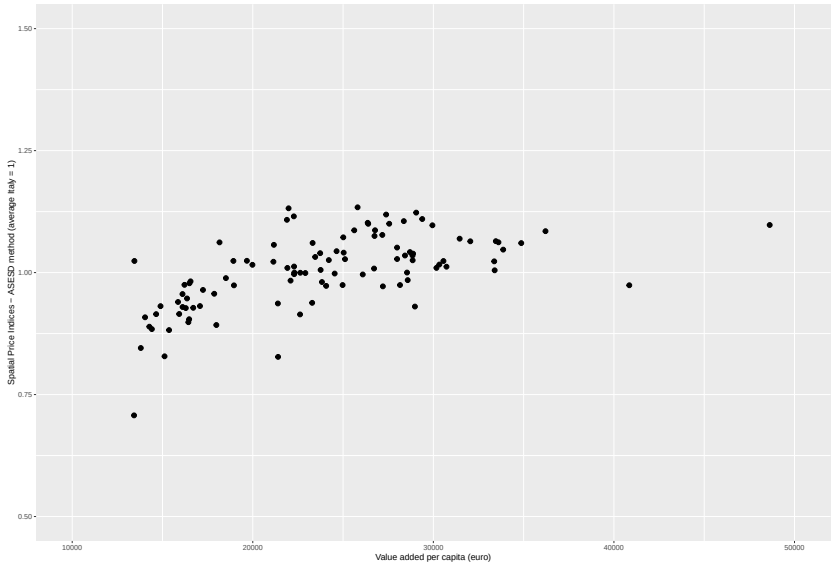


Figure 2: Scatterplot of the SPIs and PPPs versus the value added per capita, Italian provinces.

4 Differences in purchasing attitudes of poor and non-poor households

A first experiment to study the differences between the poor and non-poor household is to evaluate the dissimilarities in terms of frequency of purchase and types of outlets. The Istat researchers carried out this analysis on 2019 HES data. The detailed results are available in the Deliverable 3.2 of the European Union's Horizon 2020 project MAKSWELL Zimmermann et al. (2020). Here we summarize the main results. Regarding the type of products, those that show a greater difference between poor and non-poor households in terms of frequency of purchase are cured meats, fresh fish, and medicines. These are most purchased by non-poor households. In contrast, the products with the smallest differences are bread, fresh meat and fresh vegetables, potatoes and legumes. Also, regarding the types of outlet chosen by families some clear differences emerges. Indeed, for the 25 products considered, only 10.6% of non-poor households made a hard discount purchase, while this share rises to 27.2% for households in absolute poverty (+16.6%). The difference in the case of hypermarkets/supermarkets is almost identical but, of course, of the opposite sign (from 65.5% to 48.8%; - 16.%).

5 Final remarks

Integration of traditional data with Big data sources is the line followed by many modern statistical agencies in the production of official data. Here we have described the assumptions and the steps necessary to innovate the production of poverty indicators, which are relevant SDGs indicators, using scanner data on prices of RTCs.

We are convinced that two major innovations are important when focusing on the possible estimates of people vulnerability: the first one concerns the inclusion of the measurement of cost of living or differences in the level of price in these, the second is extending their geographical notation to offer measures related to the places where people live. This means allowing for estimates which refer also at a subregional level individuated by NUTS3 level in European classifications.

The results of our study are useful to either looking at the measurement of poverty and inequality and/or to the measurement of differences in the level of prices.

The proposed methodology is applicable in European countries as it is based on current sample surveys as EU–SILC and HES and on scanner data on prices of RTCs, that are generally available for NSIs in western countries.

Acknowledgement
This work was supported by the Jean Monnet Chair "*Small Area methods for Multidimensional Poverty and living conditions Indicators in EU- SAMPIEU*"

funded by European Commission Erasmus+ Programme – Jean Monnet Action (Project reference: 600494-EPP-1-2018-1-IT-EPPJMO-CHAIR) and the EU H2020 project *"MAKing Sustainable development and WELL-being frameworks work for policy analysis - MAKSWELL"* (G.A. no. 770643).

References

Biggeri, L. and T. Laureti (2018). Publications, experiments and projects on the computation of spatial price level differences in italy. Technical report, Paper presented at the 3rd Task force meeting the ICP, World Bank held the 27th September 2018, Country case studies: Italy.

Istat (2009). *La misura della povertà assoluta*. Roma, Italy: Metodi e Norme.stat, Italian national statistical office.

Istat (2020). *Prezzi al Consumo, Agosto 2020, Statistiche Flash, August, 31st 2020*. Roma, Italy: Istat, Italian national statistical office.

Laureti, T., C. Ferrante, and B. Dramis (2017). Using scanner and cpi data to estimate italian sub-national ppps. In *Proceeding of 49th Scientific Meeting of the Italian Statistical Society*, pp. 581–588.

Laureti, T. and F. Polidoro (2017). Testing the use of scanner data for computing sub-national purchasing power parities in italy. In *Proceeding of 61st ISI World Statistics Congress, Marrakech*.

Laureti, T. and D. Rao (2018). Measuring spatial price level differences within a country: Current status and future developments. *Estudios de economia aplicada 36(1)*, 119–148.

Pratesi, M. (2016). *Analysis of poverty by small area estimation methods*, Chapter Introduction on measuring poverty at local level using SAE methods, pp. 2–7. New York: Wiley.

Suits, D. (1984). Dummy variables: Mechanics v. interpretation. *Review of Economics and Statistics 66*, 177–180.

Zimmermann, T., F. Polidoro, F. di Leo, M. Fedeli, J. Burger, J. van den Brakel, M. Pratesi, C. Giusti, S. Marchetti, L. Biggeri, G. Bertarelli, F. S. Spagnolo, T. Laureti, I. Benedetti, P. A. Smith, J. Dawber, N. Tzavidis, A. L. Hernandez, J. O'Donoghue, T. Flower, H. Thomas, N. Wurz, T. Schmid, C. Articus, J. P. Burgard, C. Caratiola, H. Dieckmann, F. Ertz, J. Krause, R. Munnich, and A.-L. Wolwer (2020, September). Deliverable 3.2 - guidelines for best practices implementation for transferring methodology. Project report, Makswell. Available at `https://www.makswell.eu/attached_documents/output_deliverables/deliverable_3.2.pdf`.

Discussion Papers
Collana del Dipartimento di Economia e Management, Università di Pisa

Email della redazione: lfanti@ec.unipi.it