# Estimations of local spatial price indices using scanner data and their impact on the measure of the poverty incidence

M. Pratesi, C. Giusti, S. Marchetti, G. Bertarelli, F. Schirripa Spagnolo[1]
T. Laureti, I. Benedetti[2]
L. Biggeri[3]
F. Polidoro, F. Di Leo, M. Fedeli[4]

February 2021

Scientific Report prepared for the deliverable of the Project Makswell (Making Sustainable development and Well-being frameworks for Policy analysis) `www.makswell.eu`
Deliverable 3.2 - Guidelines for best practices implementation for transferring methodology:
`https://www.makswell.eu/attached_documents/output_deliverables/deliverable_3.2.pdf`

**Content**



[1]University of Pisa-Centro Camilo Dagum (ASEDS Centre)
[2]University of Tuscia-Centro Camilo Dagum (ASEDS Centre)
[3]University of Florence-Centro Camilo Dagum (ASEDS Centre)
[4]Istat-Unit of Price Statistics

# 1 Introduction

Over the last decade, there has been a growing interest in using scanner data on retail prices for constructing both temporal consumer price indexes and sub national Spatial Consumer Price Indexes (SN-SCPIs). The latter indexes are important above all to make adequate intra-country comparisons of economic poverty indicators. This report refers to the second field of research.

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 (National Statistical Institutes) 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.

The unit of research implemented two experiments by using a scanner data base provided by Istat (Italian National Statistical Institute), thanks to an agreement between Istat and Dagum Centre.

As it will be better explained in the following sections 3 and 4, the two implemented experiments use the same data base and elementary data, but differ in the application of the principles of comparability and representativity and in the methods of construction of the SN-SCPIs.

In the first experiment, the principles and the construction procedure is quite similar to the one used in the ICP (international Comparison Program) 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. In this case there is the risk that not all the products are available in all the areas. However, the information on the turnover of each product allows to give a weight to each product within every area. Under this approach the lowest level of aggregation of the products is the the Basic Heading (BH) level, as defined by the World Bank (World Bank, 2013).

The second experiment follows a complete 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 BHs, are the building blocks of the comparison, defined using the ECOICOP-8-digit classes of products.

Both experiments estimate the SPIs at provincial level (NUTS 3 level in EU classification).

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 specification of the scientific report's content follows. In the section 2, information on the characteristics of the available scanner data base is presented, highlighting the advantages and the disadvantages in using them for the computation of the SN-SCPIs. In the sections 3 and 4, the procedure, methodology and the main results of the estimation of the indexes, by using the above mentioned approaches, and also with reference to prices of the cheaper items, are presented. As the results obtained by the two approaches at aggregate level are quite different, in the section 5, we propose some explanation of the differences, also if the two approaches are based on different hypotheses, and difficult to compare at aggregate level. Section 6 is dedicated to the exploration of the places where people in condition of absolute poverty purchase some large consumption products. Section 7 describes how to measure the impact of cost of living of the poor on the estimation of local poverty rates. Final remarks and recommendations in section 8 conclude the report.

# 2 The scanner data base: advantages and limitations for the computation of sub-national spatial consumer price indexes

Since 2014, Istat (the Italian National Statistical Institute) received the scanner data on retail price by the market research company ACNielsen that 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 on a weekly basis by uploading the data files on a dedicated Istat web portal.

Istat was interested in using these scanner data on retail prices for constructing both temporal consumer price indexes and sub national Spatial Consumer Price Indexes (SN-SCPIs). Scanner data has been recently introduced by Istat in the official CPI computation, while until now they have been used in the construction of the SN-SCPI only in an experimental way.

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. However, as already underlined in the introduction, only after the 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.

More recently, Istat has initiated a series of experiments on the estimations of SN-SCPIs at a regional level on an annual basis[1]. In October 2018, an agreement between Istat and ASESD-Dagum Centre was signed to implement the tasks of the Makswell project.

[1] The experiments were conducted in subsequent improved versions of the scanner data base constructed for experimental CPI computation, analyzing the potential advantages and the empirical issues deriving from the use of the scanner data to estimate the SPIs (Laureti and Polidoro, 2017, Laureti et al., 2017).

To do these activities, Istat provided to the ASESD-Dagum Centre the scanner data base referred at the years 2017 and 2018. The data base refers to a random sample of approximately 1,800 outlets, hypermarkets (more than 500) and supermarkets (almost 1,300), and contains data concerning the 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 for 1,781 outlets of the main 16 Retail Trade Chains (RTCs) covering the process to entire national territory. Outlets have been stratified according to provinces (107), chains 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. The research group decided to use the 2018 data base for the analysis. For each GTIN, prices were calculated taking into account turnover and quantities: 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.

Taking account of the results of the experiments and of the many discussions among the members of ASESD-Dagum Centre and the researchers staff of the Istat' Price Statistics Unit, we can summarize 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 characterizes scanner data, we are able to compare price levels among the various geographical areas within a country. In addition, it is worth noting that GTIN codes describe the products in detail and they are generally the same for each item at 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 own 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, if some products acquired by consumers with reduced quantities, the availability of turnover weights (defined considering sampling weights) allows us to correctly include the corresponding representativeness of these products in terms of the total turnover of the BH or the group at which the products in question belong.

Other advantages of the use of scanner data are: (i) the reduction of measurement errors. By using as price concept the unit value for each GTIN we can refer to a more accurate measure of an average transaction price than an isolated price quotation Diewert (1995) as in the case of traditional price data collection; (ii) The reduction of conceptual uncertainty. The GTIN unit value is a more representative price paid by consumers over the reference period than the usual price collected using traditional on-field surveys. These prices include temporary price promotion 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 behavior; (iii) using scanner data we add 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 itemized information contained in scanner data. Indeed, when using the unit value approach, 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 NSI.

However, some limitations must be noted in the context of this study. The available scanner data: (i) do not cover all the retail chains of modern distribution; (ii) practically cover almost all the provinces (103 over 107), but the rural areas are not 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 are not pre-packaged with GTIN codes.

Moreover, we have to consider that, in any case, all the scanner data available cover about the 10,5% in terms of the total expenditures of families for the consumption (Istat, 2020). In addition, this share is not uniform across the Italian territory.

Therefore, it is evident that to estimate a complete set of SN-SCPIs it is necessary to build up a data base that could allow the estimation of these indexes related to the entire universe of household consumption. In fact, Istat collects consumer prices by using different sources: territorial surveys at the outlets by non-probability samples; use of administrative data, use of scanner data (Big data). Therefore, a strategy to use and integrate all the consumer price sources of data must be followed, considering also the fact that we need to face the issue due to the fact that the data come both from probability and non-probability sample. Istat is working on this line.

For the time being, this unit of research implemented two experiments to compute SN-SCPIs at provincial level in order to use them to adjust the local economic poverty indicators taking into account of the differences in the cost of living in the different areas by using a scanner data base. The first experiment is conducted following in tight way the principles and procedure followed by the ICP (International Comparison Program), coordinated by the World Bank, to compute the International PPPs (purchasing Power Parities). The second experiment has been conducted, by using a different innovative approach regarding the definition of the principle of comparability to verify its validity. Moreover, in order to calculate SPIs closer to the prices paid by the poor, preliminary experiments have been conducted by using the data for the first quintile of the price distribution, assuming that poor purchase the cheaper items of a product.

The procedures followed and the results obtained are presented in the following sections.

# 3 Spatial Price Indexes: World Bank approach

In this experiment we followed the principles and the construction procedure similar to that used in the ICP (international Comparison Program) of the World Bank to compute the international PPPs (Purchasing Power Parities) at the level of the BHs. For this reasons we provisionally consider the computed SPI as PPPs.

Indeed, a two-step procedure is adopted (World Bank Group, 2015). In the first step, provincial PPPs are computed at BH level using the Country Product Dummy (CPD) model (which is denoted as a group of similar well-defined goods or services) 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 BH level comparisons to higher level aggregates (food and non-food products) using the GEKS procedure based on Fisher indexes.

Because, the 2018 scanner dataset include all products, identified by the corresponding GTIN, we can obviously include in the comparisons also those products acquired by consumers with reduced quantities. The availability of turnover weights (defined considering sampling weights) allows us to correctly include the corresponding representativeness of these products in terms of the total turnover of the BH at which the products in question belong.

## 3.1 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 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 dividing weighted turnover for the total by 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. With the aim of 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 us 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 item level, which leads to price comparisons at BH level, $p_{ij}$ and $q_{ij}$ represent price and quantity of $i - th$ item in $j - th$ province $i = 1, 2, \ldots, N$; $j = 1, 2, \ldots, Mr$. In order to compute provincial PPPs, we used as already mentioned 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 generalized in order to provide a comprehensive framework for carrying out both international and intra-national. 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:

$$\begin{array}{c} lnp_{_{ij}} = lnPPP_{_j} + lnPPP_{_i} + ln\mu_{_{ijr}} \\ = \pi_j + \gamma_k + \upsilon_{ij} \\ \sum_j^{M_r} \pi_j D^j + \sum_{i=1}^{n} \gamma_i D^i + \upsilon_{ijr} \end{array} \tag{1}$$

where $D^j$ is a provincial-dummy variable that takes value equal to 1 if the price observation is from $j - th$ province; and $D^i$ is a $i - dummy$ variable that takes value equal to 1 if the price observation is for $i - th$ commodity. The random disturbance is assumed to satisfy the standard assumptions of a multiple regression model. In order to estimate parameters of this model we impose normalization $\sum_j^{M_r} \pi_j = 0$ thus treating all provinces in a symmetric manner. If $\widehat{\pi_j} = (1, 2, \ldots M_r)$ are estimated parameters, PPP for the province $j$ in region $r$ is given by $WR\_PPP_j = e^{\widehat{\pi}_j}$ . The CPD method based price comparisons are transitive and base-invariant. With the aim of taking into account the economic importance (representativity) of each product expressed by expenditure weights $w_{ijr}$ based on turnover we used a weighted CPD model:

$$\sqrt{w_{ijr}} lnp_{ijr} = \sum_{j=1}^{M_r} \pi_j \sqrt{w_{ijr}} D^j + \sum_{i=1}^{n} \eta_i \sqrt{w_{ijr}} D^i + \sqrt{w_{ijr}} \tag{2}$$

## 3.2 Aggregation above BHs: Provincial PPPs

The next and final step for compiling provincial price comparisons is to aggregate the results from BH level comparisons to higher level aggregates. Let us assume that there are $L$ basic headings

$(l = 1, \ldots, L)$ and $e_i^r$ 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:

$$P_{rk}^{Fisher} = \sqrt{P_{rk}^{Laspeyres} \cdot P_{rk}^{Paasche}} \tag{3}$$

Where:

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

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

with:

$$s_i^r = \frac{e_i^r}{\sum_{l=1}^{L} e_l^r} = \frac{p_l^r q_l^r}{\sum_{l=1}^{L} p_i^r q_l^r} \tag{6}$$

As the Fisher binary index in eq. 3 is not transitive, it is possible to use the procedure suggested by Gini (1931),Elteto and Koves (1964) and Szulc (1964) referred to as the GEKS index to generate transitive multilateral price comparisons across different regions. The resulting index is given by:

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

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 standardized the GEKS-Fisher based PPPs (S-GEKS).

As these PPPs are now transitive, the ratios between the PPPs for each base are the same. In order to obtain 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.

## 3.3 Results

In order to compile spatial price index for the 103 Italian provinces at BH level, the dataset used includes 2,032,574 annual price quotes concerning 72 BHs for a total of 63,256 products (GTIN code).

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: whole milk, low-fat milk, olive oil, aged cheese, other cheese, lager beer and frozen seafood were excluded since in these cases price data does not exibit spatial chain. Table 1 reports the 65 BHs included in the analysis which corresponds to the ECOICOP 5 digit with number of different products included in each BH.

Table 1: BHs included in the analysis

| Sub-classes | Description | number of products |
|---|---|---|
| 01.1.1.1.0 | Rice | 574 |
| 01.1.1.2.0 | Flour and other cereals | 1056 |
| 01.1.1.3.2 | Bread | 729 |
| 01.1.1.4.2 | Pastry products | 7023 |
| 01.1.1.4.3 | Bakery products | 3227 |
| 01.1.1.6.1 | Dry pasta | 3833 |
| 01.1.1.6.1 | Fresh pasta | 677 |
| 01.1.1.6.2 | Pasta preparations | 1446 |
| 01.1.1.7.0 | Cereal for breakfast | 652 |
| 01.1.2.7.2 | Dried, salted or smoked meat | 1781 |
| 01.1.2.8.2 | Other meat preparation | 546 |
| 01.1.3.2.0 | Frozen fish | 307 |
| 01.1.3.6.0 | Other fish and fruits | 1534 |
| 01.1.4.3.0 | Preserved milk | 692 |
| 01.1.4.4.0 | Yogurt | 2122 |
| 01.1.4.5.2 | Cheese and curd | 1660 |
| 01.1.4.6.0 | Other milk products | 1120 |
| 01.1.4.7.0 | Eggs | 524 |
| 01.1.5.1.0 | Butter | 343 |
| 01.1.5.2.0 | Margarine and other vegetable fats | 32 |
| 01.1.5.4.0 | Other edible oils | 332 |
| 01.1.6.3.0 | Dried fruits and nuts | 2183 |
| 01.1.6.4.0 | Preserved fruits and fruit-based products | 616 |
| 01.1.7.2.0 | Frozen vegetables | 498 |
| 01.1.7.3.2 | Vegetables in packs | 1450 |
| 01.1.7.3.3 | Dried vegetables | 606 |
| 01.1.7.3.4 | Potatoes | 2353 |
| 01.1.7.3.5 | Vegetable-based preparations | 90 |
| 01.1.7.4.0 | Frozen potatoes | 92 |
| 01.1.8.1.0 | Sugar | 163 |
| 01.1.8.2.0 | Jams, marmalades and honey | 1397 |
| 01.1.8.3.0 | Chocolate | 1655 |
| 01.1.8.4.0 | Confectionery produts | 1829 |
| 01.1.8.5.0 | Ice creams | 1537 |
| 01.1.9.1.0 | Sauces and condiments | 2550 |
| 01.1.9.2.0 | Salt, spices and culinary herbs | 2182 |
| 01.1.9.3.0 | Baby food | 571 |
| 01.1.9.4.0 | Ready-made meals | 2820 |
| 01.1.9.9.0 | Yeasts and other preparates | 1851 |
| 01.2.1.1.0 | Coffee | 1076 |
| 01.2.1.2.0 | Tea | 485 |

| | | |
|---|---|---|
| 01.2.1.3.0 | Cocoa and chocolate | 96 |
| 01.2.2.1.0 | Mineral waters | 887 |
| 01.2.2.2.1 | Soft drinks | 835 |
| 01.2.2.2.2 | Other soft drinks | 448 |
| 01.2.2.3.0 | Fruit and vegetable juices | 1396 |
| 02.1.1.1.2 | Spirits | 1038 |
| 02.1.1.2.0 | Alcoholic aperitifs | 127 |
| 02.1.2.1.1 | Table wines | 2373 |
| 02.1.2.1.2 | Quality wines | 4032 |
| 02.1.2.1.3 | Sparkling wines | 819 |
| 02.1.2.3.1 | Fortified wines | 143 |
| 02.1.3.1.0 | Lager beers | 1035 |
| 05.6.1.1.1 | Cleanings and manteinance products | 1587 |
| 05.6.1.1.2 | Dishwashing and detergents | 425 |
| 05.6.1.2.0 | Other non-durable small household products | 3962 |
| 09.3.4.2.1 | Pets and related produts | 2247 |
| 09.3.4.2.2 | Other pets products | 1086 |
| 12.1.3.1.0 | Non-electric appliances | 926 |
| 12.1.3.2.1 | Hair products | 2546 |
| 12.1.3.2.1 | Articles for personal hygiene and wellness | 953 |
| 12.1.3.2.2 | Body products | 2465 |
| 12.1.3.2.3 | Hygienic products | 1905 |

Following the methodology illustrated in the first paragraph, we firstly run a CPD model for each available BH using weighted turnover.

As an example, in Figures 1 - 3 we report the results for Coffee, Fresh Pasta and Eggs BHs.

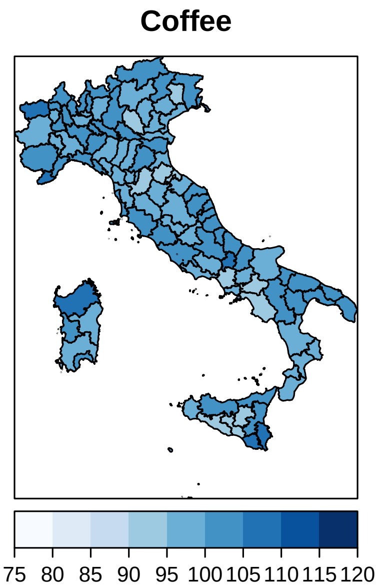


Figure 1: PPPs at provincial level for Coffee BH

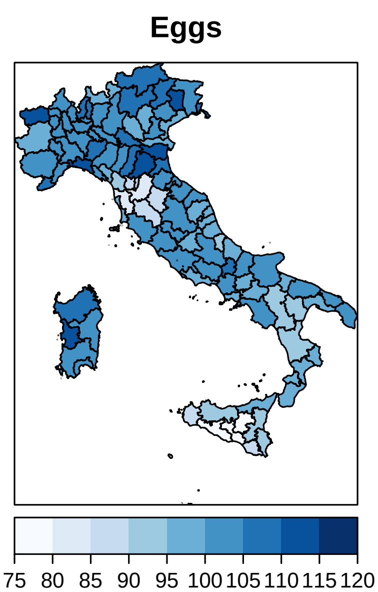


Figure 2: PPPs at provincial level for Eggs BH

Table 2: Descriptive statistics based on provincial PPPs

| | Coffee | Fresh Pasta | Eggs |
|---|---|---|---|
| Min | 92.69 | 83.03 | 75.90 |
| Max | 107.24 | 112.14 | 114.55 |
| Mean | 99.94 | 99.97 | 99.75 |
| Std. Dev | 3.32 | 6.45 | 7.38 |
| CV | 3.32 | 6.45 | 7.40 |

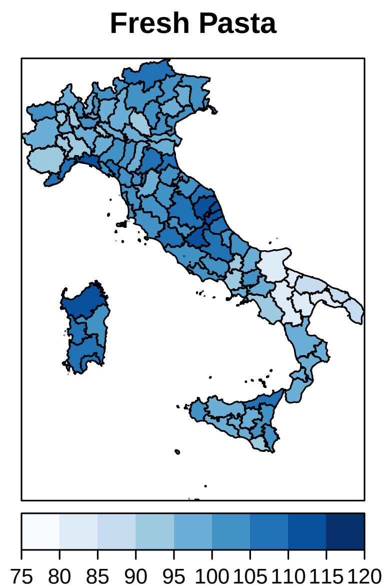


Figure 3: PPPs at provincial level for Fresh Pasta BH

The North-South dualism is confirmed only for some BHs. 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 the Southern Italy. As illustrated in Figure 3, the less expensive provinces are Matera (83.03), Potenza (83.31) located in the Basilicata region and Foggia (83.67) located in Puglia region, while the most expensive provinces are Genova (112.14) located in Liguria region, Ascoli Piceno (111.06) and Fermo (111.02) located in Marche region. On the contrary, as shown in Figure 1, homogeneity in PPPs are observed for the Coffee BH; in this case the coefficient of variation across Italian provinces is equal to 3.32. 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 two aggregates: “Food” and ”Non-food” products using as a weight the weighted turnover for each BH.

Results are reported in Table 4 and Figure 4.

Table 3: PPPs for food and non-food aggregates at provincial level (Italy=100)

| Provinces | | PPPs food | PPPs non-food |
|---|---|---|---|
| **North west** | | | |
| Alessandria | AL | 96.49 | 95.28 |
| Aosta | AO | 107.83 | 110.99 |
| Asti | AT | 94.76 | 92.02 |
| Bergamo | BG | 98.63 | 95.69 |
| Biella | BI | 98 | 95.7 |
| Brescia | BS | 100.22 | 98.13 |
| Cuneo | CN | 97.35 | 95.96 |
| Como | CO | 99.92 | 98.35 |
| Cremona | CR | 100.88 | 101.89 |
| Genova | GE | 107.5 | 108.12 |
| Imperia | IM | 104.1 | 106.53 |
| Lecco | LC | 98.44 | 96.74 |
| Lodi | LO | 100.74 | 99.42 |
| Monza e della Brianza | MB | 98.95 | 97.81 |
| Milano | MI | 99.54 | 97.92 |
| Mantova | MN | 100.23 | 99.96 |
| Novara | NO | 100.23 | 98.44 |
| Pavia | PV | 101.43 | 101.22 |
| Sondrio | SO | 99.06 | 97.15 |
| La Spezia | SP | 99.35 | 98.73 |
| Savona | SV | 104.13 | 105.68 |
| Torino | TO | 98.79 | 98.96 |
| Varese | VA | 98.79 | 97.49 |
| Verbano-Cusio-Ossola | VB | 102.07 | 101.08 |
| Vercelli | VC | 102.43 | 102.75 |
| **North west** | | | |
| Belluno | BL | 101.22 | 101.01 |
| Bologna | BO | 101.68 | 102.83 |
| Bolzano | BZ | 101.91 | 102.61 |
| Forli'-Cesena | FC | 96.71 | 95.24 |
| Ferrara | FE | 100.24 | 97.69 |
| Gorizia | GO | 102.36 | 101.47 |
| Modena | MO | 95.93 | 97.71 |
| Piacenza | PC | 100.04 | 99.13 |
| Padova | PD | 98.11 | 93.88 |
| Pordenone | PN | 96.83 | 95.09 |
| Parma | PR | 99.16 | 99.32 |
| Ravenna | RA | 100.38 | 97.98 |
| Reggio nell'Emilia | RE | 100.44 | 102.25 |

| | | | |
|---|---|---|---|
| Rimini | RN | 96.73 | 93.12 |
| Rovigo | RO | 98.54 | 94.44 |
| Trento | TN | 101.99 | 104.47 |
| Trieste | TS | 101.95 | 102.4 |
| Treviso | TV | 96.08 | 94.71 |
| Udine | UD | 102.68 | 100.76 |
| Venezia | VE | 98.39 | 96.52 |
| Vicenza | VI | 97.69 | 95.96 |
| Verona | VR | 93.75 | 92.84 |
| **Centre** | | | |
| Ancona | AN | 103.25 | 103.99 |
| Ascoli Piceno | AP | 105.14 | 106.55 |
| Arezzo | AR | 94.73 | 89.72 |
| Firenze | FI | 92.47 | 88.8 |
| Fermo | FM | 104.85 | 106.55 |
| Frosinone | FR | 101 | 103.04 |
| Grosseto | GR | 98.92 | 100.48 |
| Livorno | LI | 98.94 | 99.55 |
| Latina | LT | 100.05 | 99.62 |
| Lucca | LU | 96.7 | 93.38 |
| Macerata | MC | 103.7 | 105.16 |
| Massa Carrara | MS | 98.56 | 96.69 |
| Perugia | PG | 100.38 | 102.14 |
| Pisa | PI | 93.57 | 90.3 |
| Prato | PO | 94.7 | 88.22 |
| Pistoia | PT | 95.3 | 89.7 |
| Pesaro e Urbino | PU | 97.93 | 96.58 |
| Rieti | RI | 101.35 | 103.65 |
| Roma | RM | 103.15 | 103.09 |
| Siena | SI | 97.5 | 94.36 |
| Terni | TR | 101.27 | 103.05 |
| Viterbo | VT | 102.83 | 106.38 |
| **South** | | | |
| L'Aquila | AQ | 105.4 | 108.2 |
| Avellino | AV | 98.74 | 100.78 |
| Bari | BA | 97.96 | 100.39 |
| Benevento | BN | 101.41 | 104.73 |
| Brindisi | BR | 98.05 | 100.34 |
| Barletta-Andria-Trani | BT | 98.76 | 100.9 |
| Campobasso | CB | 100.91 | 102.01 |
| Caserta | CE | 96.14 | 96.52 |
| Chieti | CH | 102.86 | 103.21 |
| Cosenza | CS | 100.07 | 102.42 |
| Catanzaro | CZ | 98.27 | 100.8 |
| Foggia | FG | 98.1 | 100.28 |
| Isernia | IS | 103.78 | 105.02 |

| | | | |
|---|---|---|---|
| Crotone | KR | 103.78 | 102.42 |
| Lecce | LE | 99.96 | 101.55 |
| Matera | MT | 98.56 | 100.39 |
| Napoli | NA | 97.6 | 99 |
| Pescara | PE | 102.2 | 102.28 |
| Potenza | PZ | 102.83 | 105.55 |
| Reggio di Calabria | RC | 101.44 | 103.63 |
| Salerno | SA | 98.14 | 99.17 |
| Taranto | TA | 97.69 | 100.45 |
| Teramo | TE | 103.42 | 103.38 |
| Vibo Valentia | VV | 102.83 | 101.23 |
| **Islands** | | | |
| Agrigento | AG | 97.28 | 100.78 |
| Cagliari | CA | 100.14 | 100.81 |
| Carbonia-Iglesias | CI | 102.86 | 100.81 |
| Caltanisetta | CL | 102.86 | 106.87 |
| Catania | CT | 102.58 | 106.87 |
| Enna | EN | 98.27 | 100.78 |
| Messina | ME | 105.04 | 107.14 |
| Nuoro | NU | 101.55 | 102.02 |
| Ogliastra | OG | 101.55 | 100.81 |
| Oristano | OR | 105.18 | 107.89 |
| Olbia-Tempio | OT | 105.18 | 100.81 |
| Palermo | PA | 99.18 | 102.79 |
| Ragusa | RG | 104.21 | 103.71 |
| Siracusa | SR | 104.09 | 108.75 |
| Sassari | SS | 101.9 | 100.87 |
| Trapani | TP | 99.49 | 102.3 |
| Medio Campidano | VS | 102.83 | 100.81 |

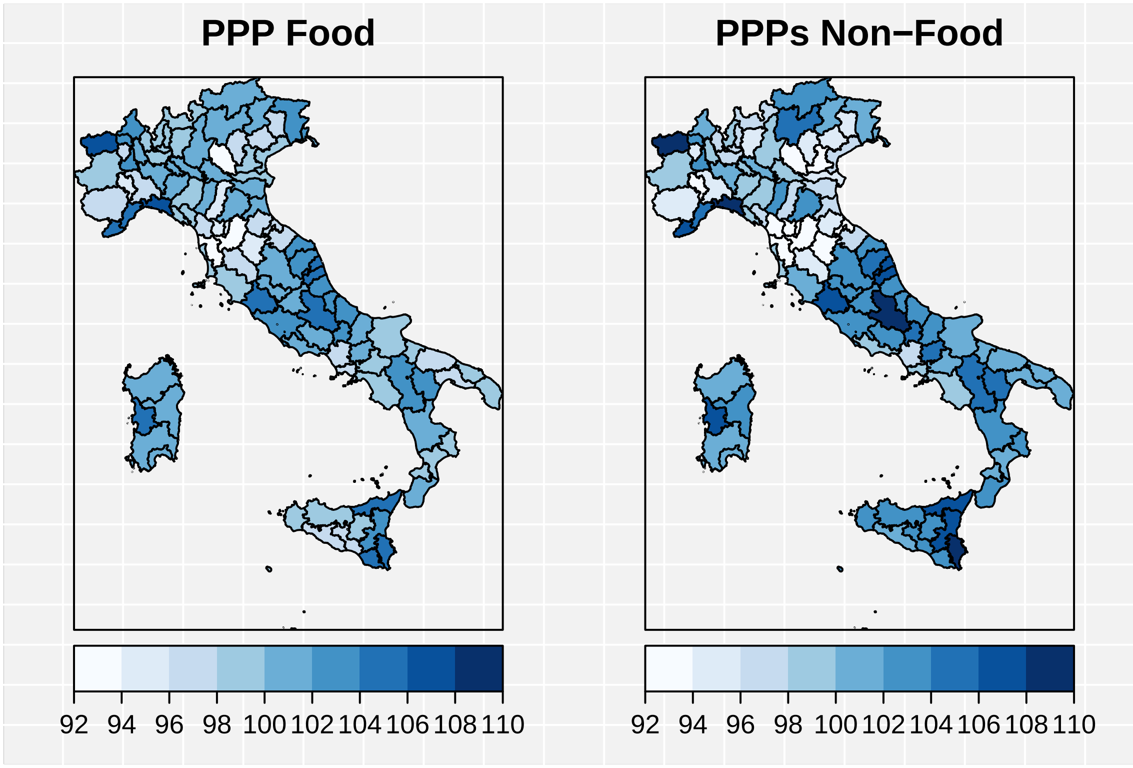


Figure 4: PPPs at provincial level for food and non-food aggregates

From Table 4 and Figure 4 we can observe a high level of price heterogeneity across Italians Provinces both for food and non-food aggregates. For food products the most expensive provinces are Aosta and Genova (with PPPs equal to 107.83 and 107.5 respectively) located in the Northern Italy, while the less expensive provinces are Florence and Pisa (with PPPs equal to 92.47 and 93.57 respectively) located in the Central Italy. For non-food products, the most expensive provinces are Aosta and L'Aquila (with PPPs equal to 110.99 and 108.20 respectively), while the less expensive provinces are Prato and Pistoia (with PPPs equal to 88.22 and 89.70 respectively) located in the Tuscany region, in Central Italy.

## 3.4 PPPs for the first quintile of the price distribution at regional level

In order to calculate PPPs for the lowest price in the price distribution, in particular for the first quintile of the price distribution, we used data for the 20 regions (NUTS 2 level), by applying a two-step procedure for the food BHs. Results are reported in Figure 5. In this case, we provide PPPs at regional level since at Provincial level there was not a reliable level of overlap. Results in Table **??** provide PPPs estimates for food aggregates.

As reported in Figure 5, for food aggregates calculated for the first quintile of the distribution, the most expensive regions are Valle d'Aosta and Liguria (with PPPs equal to 107.39 and 101.78 respectively), while the less expensive regions are Toscana and Campania (with PPPs equal to 96.23

and 96.90 respectively). Similar results are obtained if we considered the total distribution of the price. Indeed the most expensive regions are Valle d'Aosta and Liguria (with PPPs equal to 105.86 and 102.14 respectively), while the less expensive regions are Toscana and Veneto (with PPPs equal to 95.40 and 96.94 respectively).

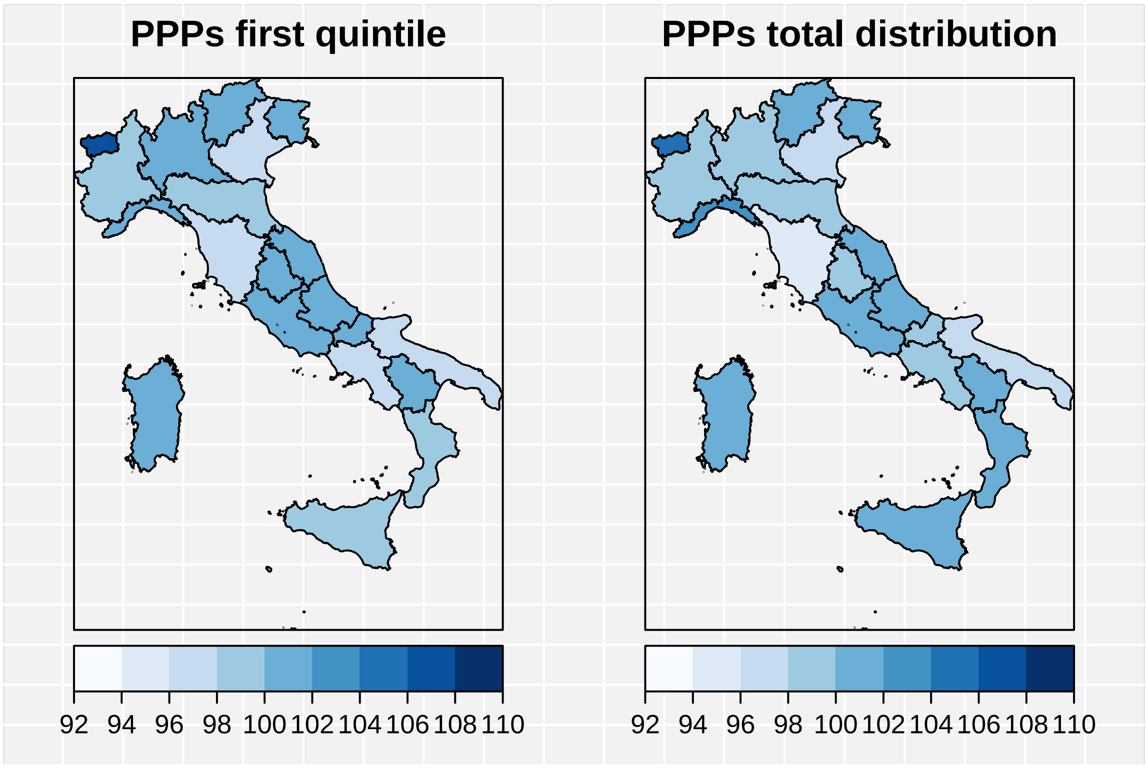


Figure 5: PPPs at regional level for first quintile and total distribution of the price

Table 4: PPPs for food aggregates for first quintile and total distribution of price (Italy=100)

| REGIONS | PPPs FIRST QUINTILE | PPPS TOTAL DISTRIBUTION |
|---|---|---|
| **North-west** | | |
| Liguria | 101.78 | 102.14 |
| Lombardia | 100.04 | 99.28 |
| Piemonte | 99.34 | 98.75 |
| Valle d'Aosta | 107.39 | 105.86 |
| **North-east** | | |
| Emilia-Romagna | 99.10 | 99.08 |
| Friuli Venezia-Giulia | 100.96 | 100.28 |

| | | |
|---|---|---|
| Trentino Alto Adige | 100.78 | 100.20 |
| Veneto | 97.82 | 96.94 |
| **Centre** | | |
| Lazio | 100.56 | 101.04 |
| Marche | 100.97 | 101.11 |
| Toscana | 96.23 | 95.40 |
| Umbria | 100.02 | 98.90 |
| **South** | | |
| Abruzzo | 100.95 | 101.96 |
| Basilicata | 100.43 | 101.79 |
| Calabria | 98.32 | 100.26 |
| Campania | 96.90 | 98.00 |
| Molise | 101.29 | 99.73 |
| Puglia | 97.70 | 97.53 |
| **Islands** | | |
| Sardegna | 100.17 | 100.34 |
| Sicilia | 99.74 | 101.90 |

As already pointed out, there are various advantages in using scanner data to compute SPIs and our PPPs. On the other hand, the like to like approach may have some limitations. Indeed, in order to use strictly comparable products, it is possible that some products are excluded since they are produced and consumed at 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. From Table 5 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 channel may be considered. In Southern regions consumers tend to buy food and non-food products in open markets and traditional shops more frequently than consumers in Northern regions.

# 4 Spatial Price Indexes: ASESD approach

## 4.1 Estimation of Spatial Consumer Price Indexes for the Italian Provinces

The results we present in this section concern the computation of spatial consumer price indices (SN-SCPI) 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 milliliter, 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 (prod-

Table 5: Scanner data: % market shares (hypermarket + supermarket) – year 2016. Image taken from a computation made by Istat on the scanner dataset.

| RETAIL CHAINS | NORTH - W | | | | NORTH - E | | | | CENTER | | | | SOUTH AND ISLANDS | | | | | | | | ITALIA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PIEMONTE | VALLE D'AOSTA | LIGURIA | LOMBARDIA | TRENTINO-ALTO ADIGE | VENETO | FRIULI-VENEZIA GIULIA | EMILIA-ROMAGNA | TOSCANA | UMBRIA | MARCHE | LAZIO | ABRUZZO | MOLISE | CAMPANIA | PUGLIA | BASILICATA | CALABRIA | SICILIA | SARDEGNA | |
| COOP ITALIA | 18,2 | - | 42,2 | 7,9 | 18,0 | 9,1 | 21,3 | 41,2 | 51,2 | 30,8 | 18,5 | 14,3 | 10,0 | - | 4,4 | 18,6 | 6,9 | - | 6,3 | - | 18,5 |
| CONAD | 4,3 | 22,3 | 17,0 | 3,3 | 13,8 | 3,6 | 7,7 | 26,5 | 14,8 | 29,9 | 12,6 | 24,5 | 29,8 | 30,9 | 20,5 | 9,6 | 10,3 | 30,2 | 19,5 | 30,6 | 13,3 |
| ESSELUNGA | 12,4 | - | 3,9 | 31,3 | - | 1,2 | - | 9,9 | 22,1 | - | - | 0,9 | - | - | - | - | - | - | - | - | 12,1 |
| SELEX COMMERCIALE | 17,9 | 8,6 | 4,8 | 9,9 | - | 32,3 | 9,4 | 6,6 | 1,1 | 22,1 | 18,2 | 3,4 | 2,7 | 23,4 | 7,6 | 29,1 | 6,0 | 3,3 | 4,4 | 12,8 | 11,1 |
| GRUPPO AUCHAN | 7,0 | - | 0,7 | 8,2 | - | 6,3 | 1,1 | 1,5 | 1,9 | 2,7 | 25,8 | 10,7 | 11,1 | - | 8,1 | 17,2 | 10,4 | 17,3 | 20,1 | 12,6 | 7,8 |
| GRUPPO CARREFOUR ITALIA SPA | 16,4 | 45,1 | 8,8 | 9,9 | - | 2,1 | 4,2 | 1,8 | 2,8 | 0,7 | 0,9 | 13,3 | 5,7 | 1,6 | 9,2 | - | 0,9 | 8,9 | 1,5 | 5,6 | 7,1 |
| FINIPER | 1,5 | - | - | 6,4 | - | 1,6 | 2,9 | 1,4 | - | - | 4,1 | - | 8,3 | - | - | - | - | - | - | - | 2,3 |
| GRUPPO VEGE | - | - | 1,5 | 1,1 | - | 6,2 | - | 0,2 | 0,1 | 0,2 | - | 0,7 | 2,6 | 5,7 | 20,7 | 1,2 | 5,0 | 4,0 | 19,8 | 13,8 | 3,2 |
| GRUPPO SUN | 1,4 | - | 3,2 | 2,6 | - | 2,0 | 1,2 | 0,3 | - | 2,4 | 9,8 | 14,4 | 18,2 | 27,6 | - | - | - | - | - | - | 3,1 |
| AGORA' NETWORK SCARL | 2,5 | - | 13,5 | 6,1 | 34,4 | 0,4 | - | 0,2 | 0,2 | - | - | - | - | - | - | - | - | - | - | - | 2,8 |
| GRUPPO PAM | 3,7 | - | 2,7 | 0,9 | 0,6 | 3,1 | 8,0 | 1,8 | 5,4 | 3,1 | - | 8,5 | 0,7 | - | 0,2 | 1,4 | - | - | - | 3,8 | 2,7 |
| ASPIAG | - | - | - | - | 32,4 | 12,7 | 29,9 | 1,8 | - | - | - | - | - | - | - | - | - | - | - | - | 2,7 |
| BENNET SPA | 8,7 | - | 1,3 | 5,2 | - | 1,2 | 4,0 | 1,9 | - | - | - | - | - | - | - | - | - | - | - | - | 2,5 |
| SIGMA | 0,1 | - | - | 1,1 | - | 2,8 | 2,6 | 3,0 | 0,3 | 0,3 | 7,0 | 0,8 | 3,2 | 6,4 | 2,8 | 6,9 | 5,3 | 1,6 | 1,1 | 5,0 | 1,8 |
| CRAI | 1,6 | - | 0,3 | 0,2 | - | 2,6 | 2,1 | 0,5 | 0,0 | - | 0,4 | 1,7 | 0,7 | 0,9 | 2,3 | 0,2 | 5,4 | 3,5 | 7,5 | 9,7 | 1,4 |
| DESPAR SERVIZI | - | - | - | 0,6 | - | - | - | - | - | - | - | 0,0 | - | - | 1,8 | 7,1 | 17,6 | 18,4 | 6,2 | 4,3 | 1,2 |
| **TOTAL** | **95,9** | **76,0** | **99,8** | **94,8** | **99,1** | **87,0** | **94,3** | **98,5** | **99,9** | **92,2** | **97,4** | **93,2** | **92,9** | **96,6** | **77,5** | **91,3** | **67,9** | **87,2** | **86,4** | **98,0** | **93,7** |

ucts). 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[2].

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[3] 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

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

[2] 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

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

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 $j$th ECOICOP-8-digit aggregation and the $i$th 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:

$$\log \bar{p}_{ij} = \alpha_0 + \alpha_i D_i + \beta_j I_j + \varepsilon_{ij}, \quad i = 1, \ldots, 103 \quad j = 1, \ldots, 102, \tag{8}$$

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 $j$th 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 $\varepsilon_{ij} \sim N(0, \sigma^2)$.

To take into account the different level of the turnover between the ECOICOP-8-digit aggregates we estimate the model equation (8) 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 $i$th province).

Model equation (8) – as it is specified – is not identified, because the $D_i$s 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 $\alpha_i$. The coefficient $\alpha_i$ 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, $\alpha_i$ 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 equation (8) 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 equation (8) is modified as follows:

$$\log Q(\tau,p)_{ij} = \gamma_0 + \gamma_i D_i + \beta_j I_j + \varepsilon_{ij}, \quad i = 1,\dots,103 \quad j = 1,\dots,102, \tag{9}$$

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

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.

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

The estimates of the SPIs and SPI($Q_{0.2}$)'s we obtained are shown in table 6. Moreover, we show a choropleth map of estimated SPIs (based on model equation (8)) for the Italian provinces in figure 6 on the left.

Table 6: Estimates of SPI's based on the mean prices and on quantile 0.2.

| Prov. | SPI(Mean) | SPI($Q_{0.2}$) | Region |
|---|---|---|---|
| TO | 1.123 | 1.043 | PIEMONTE |
| VC | 1.044 | 1.058 | PIEMONTE |
| NO | 1.102 | 1.094 | PIEMONTE |
| CN | 1.035 | 1.012 | PIEMONTE |
| AT | 0.914 | 0.960 | PIEMONTE |
| AL | 1.087 | 1.057 | PIEMONTE |
| BI | 0.973 | 1.021 | PIEMONTE |
| VB | 1.108 | 1.134 | PIEMONTE |
| AO | 1.047 | 1.143 | VALLE D'AOSTA |
| IM | 1.057 | 1.084 | LIGURIA |
| SV | 1.072 | 1.044 | LIGURIA |
| GE | 1.069 | 1.021 | LIGURIA |
| SP | 1.087 | 1.090 | LIGURIA |
| VA | 1.100 | 1.032 | LOMBARDIA |
| CO | 1.134 | 1.106 | LOMBARDIA |
| SO | 0.996 | 0.938 | LOMBARDIA |
| MI | 1.098 | 1.027 | LOMBARDIA |
| BG | 1.110 | 1.055 | LOMBARDIA |

| | | | |
|---|---|---|---|
| BS | 1.097 | 1.047 | LOMBARDIA |
| PV | 1.132 | 1.101 | LOMBARDIA |
| CR | 1.105 | 1.074 | LOMBARDIA |
| MN | 1.035 | 1.035 | LOMBARDIA |
| LC | 1.100 | 1.078 | LOMBARDIA |
| LO | 1.005 | 0.992 | LOMBARDIA |
| MB | 1.119 | 1.086 | LOMBARDIA |
| BZ | 0.974 | 0.970 | TRENTINO-ALTO ADIGE |
| TN | 1.023 | 0.964 | TRENTINO-ALTO ADIGE |
| VR | 1.024 | 0.979 | VENETO |
| VI | 1.017 | 1.000 | VENETO |
| BL | 0.930 | 0.956 | VENETO |
| TV | 1.039 | 1.016 | VENETO |
| VE | 1.051 | 1.010 | VENETO |
| PD | 1.010 | 0.985 | VENETO |
| RO | 1.000 | 1.012 | VENETO |
| UD | 1.077 | 1.062 | FRIULI-VENEZIA GIULIA |
| GO | 0.998 | 1.003 | FRIULI-VENEZIA GIULIA |
| TS | 1.012 | 1.018 | FRIULI-VENEZIA GIULIA |
| PN | 0.972 | 0.958 | FRIULI-VENEZIA GIULIA |
| PC | 1.042 | 1.058 | EMILIA-ROMAGNA |
| PR | 1.064 | 1.038 | EMILIA-ROMAGNA |
| RE | 1.064 | 1.006 | EMILIA-ROMAGNA |
| MO | 1.060 | 1.002 | EMILIA-ROMAGNA |
| BO | 1.085 | 1.055 | EMILIA-ROMAGNA |
| FE | 1.026 | 1.043 | EMILIA-ROMAGNA |
| RA | 0.985 | 0.927 | EMILIA-ROMAGNA |
| FC | 1.025 | 0.956 | EMILIA-ROMAGNA |
| RN | 1.008 | 0.971 | EMILIA-ROMAGNA |
| PU | 0.981 | 0.952 | MARCHE |
| AN | 1.075 | 1.068 | MARCHE |
| MC | 1.032 | 1.033 | MARCHE |
| AP | 1.000 | 1.020 | MARCHE |
| FM | 0.997 | 1.045 | MARCHE |
| MS | 0.984 | 1.015 | TOSCANA |
| LU | 1.028 | 0.999 | TOSCANA |
| PT | 0.938 | 0.925 | TOSCANA |
| FI | 1.005 | 0.939 | TOSCANA |
| LI | 1.040 | 0.980 | TOSCANA |
| PI | 1.028 | 0.975 | TOSCANA |
| AR | 0.974 | 0.975 | TOSCANA |
| SI | 0.974 | 0.991 | TOSCANA |
| GR | 1.010 | 0.979 | TOSCANA |
| PO | 1.000 | 1.034 | TOSCANA |
| PG | 0.999 | 0.958 | UMBRIA |
| TR | 0.937 | 0.949 | UMBRIA |

| | | | |
|---|---|---|---|
| VT | 1.024 | 1.000 | LAZIO |
| RI | 0.982 | 0.983 | LAZIO |
| RM | 1.062 | 0.986 | LAZIO |
| LT | 1.024 | 1.013 | LAZIO |
| FR | 1.016 | 1.018 | LAZIO |
| CE | 0.931 | 0.946 | CAMPANIA |
| BN | 0.828 | 0.865 | CAMPANIA |
| NA | 0.964 | 0.963 | CAMPANIA |
| AV | 0.898 | 0.913 | CAMPANIA |
| SA | 0.904 | 0.900 | CAMPANIA |
| AQ | 1.115 | 1.125 | ABRUZZO |
| TE | 1.022 | 1.032 | ABRUZZO |
| PE | 1.012 | 1.026 | ABRUZZO |
| CH | 1.061 | 1.033 | ABRUZZO |
| CB | 0.989 | 1.022 | MOLISE |
| IS | 0.957 | 1.068 | MOLISE |
| FG | 0.940 | 0.946 | PUGLIA |
| BA | 0.974 | 0.959 | PUGLIA |
| TA | 0.947 | 0.954 | PUGLIA |
| BR | 0.929 | 0.943 | PUGLIA |
| LE | 0.915 | 0.909 | PUGLIA |
| BT | 0.884 | 0.888 | PUGLIA |
| PZ | 0.827 | 0.905 | BASILICATA |
| CS | 0.908 | 0.913 | CALABRIA |
| CZ | 0.893 | 0.908 | CALABRIA |
| RC | 0.882 | 0.937 | CALABRIA |
| VV | 0.845 | 0.956 | CALABRIA |
| TP | 0.889 | 0.920 | SICILIA |
| PA | 0.928 | 0.968 | SICILIA |
| ME | 0.956 | 0.981 | SICILIA |
| AG | 0.707 | 0.820 | SICILIA |
| CT | 0.975 | 0.995 | SICILIA |
| RG | 0.915 | 0.943 | SICILIA |
| SR | 0.931 | 0.995 | SICILIA |
| SS | 1.062 | 1.091 | SARDEGNA |
| NU | 0.927 | 0.972 | SARDEGNA |
| CA | 1.041 | 1.081 | SARDEGNA |
| OR | 0.978 | 1.081 | SARDEGNA |
| SU | 1.024 | 1.077 | SARDEGNA |

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

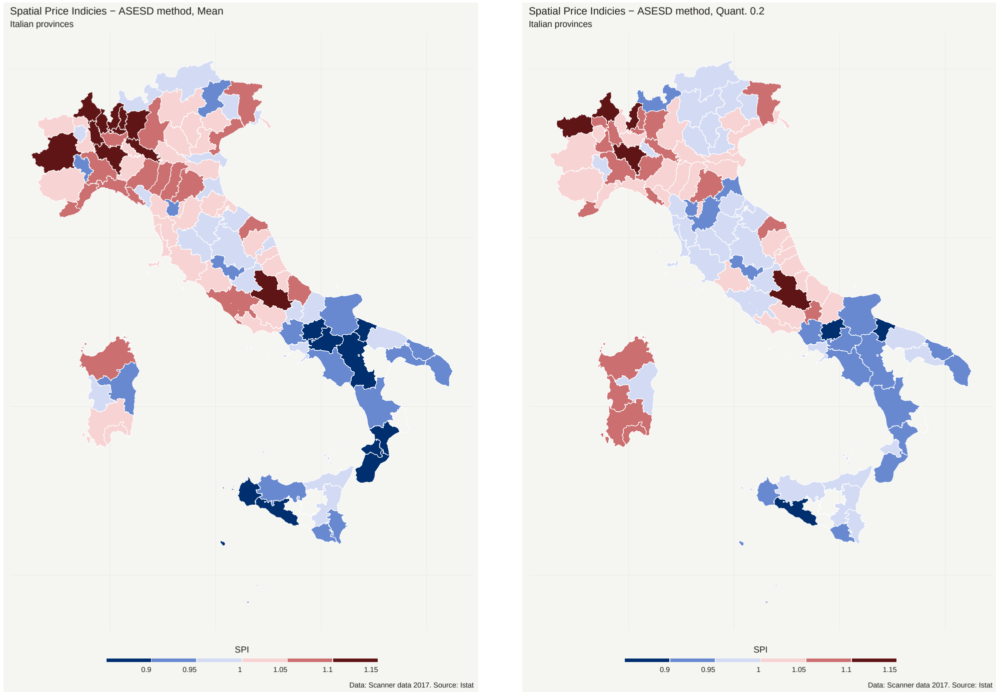


Figure 6: Choropleth map of SPI computed according to ASESD method. SPI obtained using mean unit prices (left) and quantile 0.2 of unit prices (right).

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[4] of Italy.

As it concerns the estimates of SPI($Q_{0.2}$) mapped in figure 6 on the right, we recall that they are obtained modifying model equation (8). 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.

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

# 5 A general analysis of the results of the experiments: some concluding remarks

The results obtained with the two experiments are undoubtedly interesting. Actually the estimations of the PPPs and SPIs (according the ASEDS approach) at provincial level are quite different as we see from Figure 8 that reports the indexes computed with the two approaches using the same scale.

As illustrated in the previous sections, the PPPs obtained with the WB method are smoother and indicate that the prices are lower in the provinces of Tuscany, located in Central Italy, and in some northern provinces. However, also some provinces in the south of the Country show values below 1. On the opposite, the SPIs computed using the ASESD methodology reported in the section, show a general north/south divide, although with some exceptions, as already commented in the previous section. Another main difference is that the range of the ASESD SPIs values is higher, as it covers the lower and the higher classes of values represented in the Figure (values lower than 0.9 and higher than 1.1).

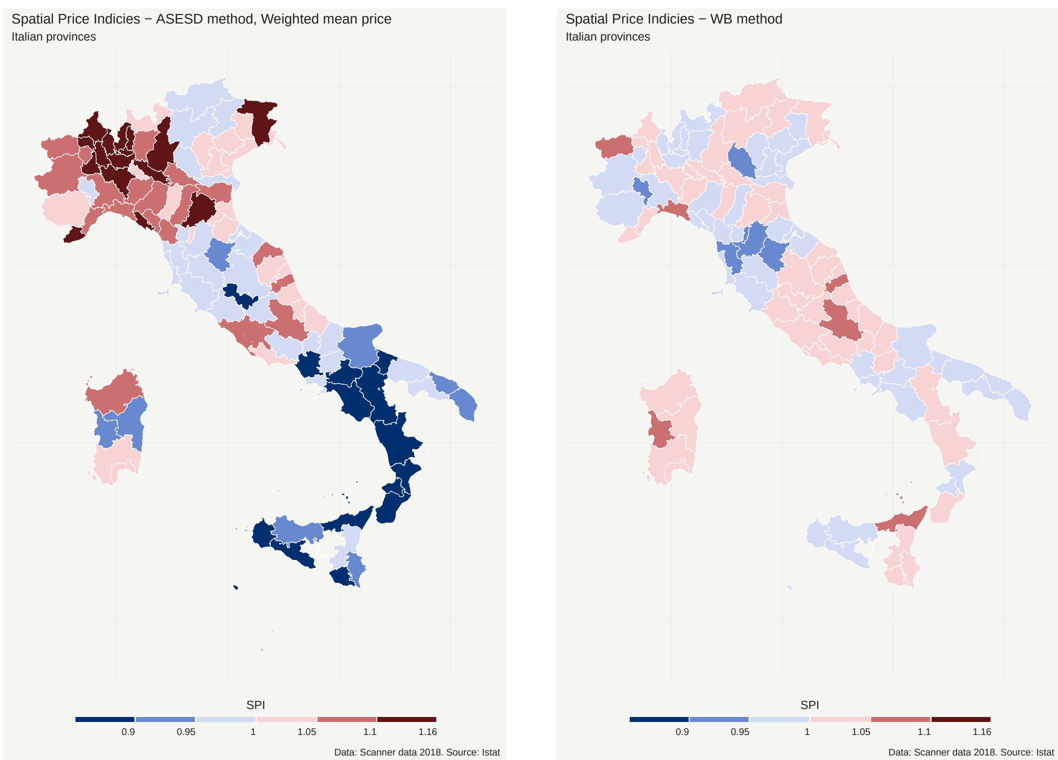


Figure 7: Choropleth map of SPI computed according to ASESD and the WB method.

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

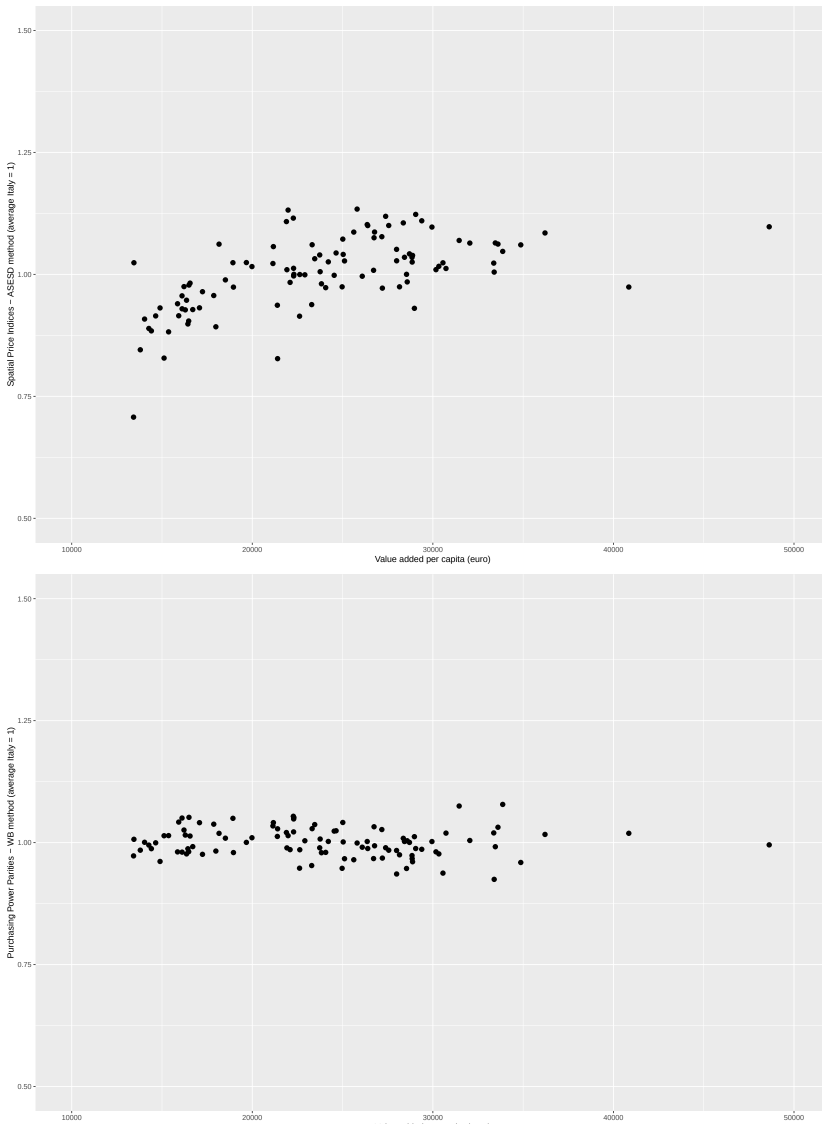


Figure 8: Scatterplot of the SPIs and PPPs versus the value added per capita, Italian provinces.

compute the within-country PPPs one would expect some internal consistency. Price level in poor areas should be generally lower than those in richer areas and showed a similar patter cross the basic headings. To check this hypothesis, we have done a comparison between the two computed indexes and the value added per capita by Italian province as reported in the figure 8. It is evident that the Indexes computed with ASESD method satisfy, in some way the previous 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.)

As explained in the previous subsections, the two sets of SPIs values have been computed with different methodologies and with slightly different sets of data. We think that the difference in the results mainly depends on the different methodologies that have been used: the WB approach is based on like-to-like product comparisons, and we expect that the range of the price for the same product cannot vary so highly in the 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 analyzed and to have a more clear picture of the reasons of the differences, the analyses should be done at disaggregated level. Unfortunately, a finer comparison at Basic heading level is not possible in this moment, but we plan to better investigate in the future. In any case the results obtained are interesting and useful from scientific and official statistics point of view. We think that the unit of research has to continue the experiments, as well as the same experiments should be conducted also by the units of research of other European countries.

Finally, we would like to inform the readers, that, outside of the computation of the SN-SCPIs by using scanner data, the unit of research has done the computation of sub-national SPIs for Housing costs which acquires particular importance since housing costs can be a substantial financial burden to household, especially for low-income families. Such latest indexes have surely an autonomous relevance in assessing and comparing poverty and in designing housing policies for poor at a local level, but they can also be used as proxies of the general sub-national household consumption SPIs or in combination with them, for adjusting poverty thresholds. We computed the SPIs for Housing Rents (SPIHRs) by using two different sources of data coming from: (i) Archives on the evaluation of rents, for all the 7,914 Italian municipalities, made by the Revenue and Tax Agency; (ii) Rents and house information collected through the Household Expenditure Survey (HES), with a sample of about 25,000 units. In both cases, the hedonic regression models, by including location and house characteristics, have been used to obtain the SPIHRs for the Italian regions. The results obtained, making the Italian average=100, show significant level differences across the various Italian regions: (i) by using the revenue and tax agency data for all type of dwellings the max is 152,73 and the min is 53,47 and for the economic apartments the max is 156,87 and the min 48,48; (ii) by using the HES data for the rented dwellings the max is 156, 80 and the min 50,60. This results point out the importance and the possibility of calculating the SPIHR in Italy on a regular basis.

# 6 Where people in condition of absolute poverty purchase some large consumption products

In the two-weeks Diary (aimed at investigating the more frequent expenditures referred to large consumption products) of Italian Households Budget Survey, HES, since 2015 Istat asks to households to indicate, in a dedicated section, the type of outlets where they have purchased a list of 25 products. The products are the most frequently purchased and the types of outlets listed are seven: traditional shop, open market and street vendors, hard discount, hypermarkets and supermarkets, department stores and outlet chains, farm or direct producer, internet. Households are asked to indicate the two places where more frequently they have purchased that specific good or to select the option ‘not purchased’.. On 2019 HES data an analysis has been carried out with the objective of detecting possible differences in terms of choices of types of outlets between households over (non poor) and households on or below the thresholds of absolute poverty (poor) calculated and updated each year by Istat. In terms of frequency of purchase, fresh vegetables and fruit, bread, meat and cheeses are the products purchased most frequently (figure 7) by the entire population: more than 85% buys at least once in two weeks. At the opposite end of this ranking we find toys and video games, frozen fish, wine and olive oil that are purchased by 40%, and even less, of the population considering the time span of two weeks and thus less frequently. If we limit the observation only to households in conditions of absolute poverty, differences emerge in purchasing attitudes. In particular, the purchase of some essential products such as bread, milk and eggs is (relatively) more important for poor families. On the contrary, the purchase of cured meats loses importance, (it is at the 6th place in the overall ranking of products purchased at least once and falls to 11th place in the case of absolute poor households).

Particular attention should be paid to the purchase of medicines whose general frequency is relatively low, about 60% of the population makes at least one purchase in the period under review (16th place in the ranking). In the case of the poor, the percentage drops to 40% (20th in the ranking). The reasons for this difference are probably mainly the use of drugs only in situations of particular gravity and a greater possibility of obtaining drugs for free from the National Health Service. In red, in Table 7, we find those products that show a wider difference between poor and non-poor in terms frequency of purchase. In addition to the case of cured meats and medicines, that of fresh fish is highlighted. The products with the smallest differences are highlighted in gray.

In addition to bread, milk and eggs already mentioned, there is a frequent consumption, even by the poor, of fresh meat and vegetables. With reference to the entire population, the most suitable place for shopping is, in all cases, the super-hypermarket with only two exceptions: medicines and toys (Table 8). The other two main shopping places are the same in almost all the cases: the traditional shop and the hard discount. For the other types of outlets, different situations arise which, however, are in line with what expected in relation to the nature of the asset (i.e. fresh products - fruit, vegetables and fish - are the most purchased in open markets). The exception is fresh meat for which traditional shop are more frequently visited as well as supermarkets and hypermarkets. Purchases 'via the internet' are actually sporadic excluding toys and, in part, coffee (especially in pods).

Table 9 shows the types of outlet chosen by families in absolute poverty who have made at least one purchase in the two weeks of observation. Any comparison between these data and those reported in Table 8 have to take into account the fact that the universe of poor households is small (6.4% in 2019) and that for many of the products in the table 'no purchase' is selected. Nevertheless, the main comments that it is possible to sketch are the following:

- On average, for the 25 products considered, only 10.6% of non-poor households made a purchase in a hard discount. This percentage reaches 27.2% (+ 16.6%) in the case of families in conditions of absolute poverty.
- The difference in the case of hypermarkets / supermarkets is almost identical but, of course, of the opposite sign (from 65.5% to 48.8%; - 16.7%).
- The difference is relevant for all products even if in the case of the products bought more frequently it tends to decrease. In the case of bread, for example, the use of the hard discount goes from 7.2% to 19.1% (+ 11.9%) and similarly occurs in the case of meat and vegetables.
- In general, the use of the traditional shop and open market is very similar for the two categories of families. On average, for the 25 products considered, 18.3% of non-poor households and 19.9% of poor ones make a purchase at a traditional store. For the open markets and street vendors the share is 2.3% and 2.9% respectively.

Table 7: Families who did not purchase in the last two weeks (%) by product - Year 2019. Source: elaboration on 2019 Istat HES data.

| Products | No purchase - percentages of Households | | |
|---|---|---|---|
| | No poor | Poor | Total |
| Bread | 7.1 | 12.5 | 7.5 |
| Pasta | 18.8 | 32.1 | 19.7 |
| Biscuits, rusks, snacks | 17.7 | 31.2 | 18.6 |
| Fresh meat | 9.9 | 20.0 | 10.5 |
| Frozen meat | 25.6 | 39.9 | 26.5 |
| Cured meats | 16.6 | 41.8 | 18.2 |
| Fresh fish | 54.2 | 81.0 | 55.9 |
| Frozen fish | 63.8 | 80.6 | 64.9 |
| Milk | 19.6 | 31.1 | 20.3 |
| Chesees | 10.3 | 26.8 | 11.4 |
| Yogurt | 44.2 | 65.7 | 45.6 |
| Eggs | 27.8 | 38.8 | 28.5 |
| Fres fruit | 6.0 | 19.1 | 6.9 |
| Fresh vegetables, potatoes and legumes | 4.9 | 15.5 | 5.6 |
| Dried or frozen vegetables, potatoes and legumes | 44.0 | 63.3 | 45.3 |
| Olive oil | 58.5 | 75.9 | 59.6 |
| Mineral water | 29.0 | 48.1 | 30.3 |
| Soft drinks | 37.9 | 52.8 | 38.8 |
| Wine | 59.1 | 85.6 | 60.8 |
| Coffee | 42.7 | 65.2 | 44.2 |
| Medicines | 38.4 | 68.7 | 40.4 |
| Personal hygiene products (soaps, deodorant, baby diapers, etc.) | 25.1 | 48.7 | 26.6 |
| Cleaning products | 22.8 | 44.5 | 24.2 |
| Disposable items for the kitchen (napkins, dishes, etc.) | 41.1 | 63.1 | 42.5 |

- The last three categories of businesses (department stores, farms and Internet) are almost completely unused by poor families (an expected result). If on average 3.3% of non-poor families buy from one of these types of outlets, in the case of the poor, the share drops to 1.2%.

The results obtained from these preliminary analyses of 2019 HES data show some interesting differences between non poor and absolutely poor households in terms of choice of the type of outlet where purchasing a list of 25 large consumption products. 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 the poor families in different Italian geographical areas by taking into account their different behavior 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 8: Types of outlet where households make purchases (% distribution) - Year 2019. Source: elaboration on 2019 Istat HES data.

| Products | Traditional shop | Open market and street vendors | Hard discount | Hypermarkets and supermarkets | Department stores and outlet chains | Farm or direct producer | Internet |
|---|---|---|---|---|---|---|---|
| Bread | 44,9 | 1,1 | 7,9 | 45,4 | 0,3 | 0,2 | 0,1 |
| Pasta | 9,8 | 0,5 | 13,3 | 75,5 | 0,7 | 0.1 | 0,1 |
| Biscuits, rusks, snacks | 8,7 | 0,8 | 13,9 | 75,8 | 0,7 | 0.1 | 0,1 |
| Fresh meat | 31,7 | 0,9 | 9,3 | 56,7 | 0,6 | 0,8 | 0,1 |
| Frozen meat | 9,0 | 1,7 | 12,1 | 75,2 | 1,6 | 0.3 | 0,2 |
| Cured meats | 14,8 | 1,0 | 12,2 | 70,9 | 0,6 | 0.3 | 0,1 |
| Fresh fish | 35,0 | 9,8 | 4,3 | 49,8 | 0,6 | 0.4 | 0,0 |
| Frozen fish | 9,1 | 1,3 | 13,0 | 74,8 | 1,5 | 0.2 | 0,1 |
| Milk | 10,5 | 0,5 | 13,6 | 74,4 | 0,7 | 0.3 | 0,1 |
| Chesees | 11,7 | 1,6 | 13,1 | 72,2 | 0,7 | 0,7 | 0,1 |
| Yogurt | 6,1 | 0,5 | 13,0 | 79,3 | 0,7 | 0.3 | 0,2 |
| Eggs | 11,7 | 3,2 | 13,7 | 67,9 | 0,8 | 2,6 | 0,1 |
| Fres fruit | 22,0 | 11,7 | 9,5 | 54,9 | 0,5 | 1.1 | 0,1 |
| Fresh vegetables, potatoes and legumes | 20,7 | 11,3 | 9,9 | 56,1 | 0,6 | 1,4 | 0,1 |
| Dried or frozen vegetables, potatoes and legumes | 10,3 | 4,3 | 13,4 | 69,8 | 1,3 | 0,7 | 0,1 |
| Olive oil | 8,0 | 0,8 | 12,7 | 71,8 | 1,1 | 5.3 | 0,2 |
| Mineral water | 8,0 | 0,9 | 13,7 | 75,4 | 0,9 | 1,0 | 0,2 |
| Soft drinks | 6,4 | 0,7 | 14,9 | 76,8 | 0,9 | 0.1 | 0,1 |
| Wine | 10,8 | 0,7 | 11,1 | 70,1 | 1,1 | 5,8 | 0,3 |
| Coffee | 11,9 | 0,6 | 12,3 | 71,7 | 1,4 | 0,5 | 1,5 |
| Medicines | 93,4 | 0,1 | 0,8 | 5,0 | 0,3 | 0,2 | 0,2 |
| Personal hygiene products (soaps, deodorant, baby diapers, etc.) | 10,9 | 0,6 | 11,9 | 71,1 | 4,9 | 0,2 | 0,3 |
| Cleaning products | 9,1 | 0,8 | 13,4 | 71,4 | 5,0 | 0.1 | 0,2 |
| Disposable items for the kitchen (napkins, dishes, etc.) | 9,2 | 0,8 | 14,9 | 70,6 | 4,3 | 0,0 | 0,2 |
| Toys and videogames | 36,4 | 2,0 | 5,3 | 36,1 | 14,0 | 0,1 | 6,0 |

# 7 The impact of the local cost-of-living differences on the measure of the poverty incidence

Intra-country comparisons of poverty indicators are important for many reasons. For example, when measuring the poverty incidence, the use of a national poverty line allows to establish a general scheme of how local areas (e.g. regions or provinces) compare with national standards. However, considering the same poverty line for each area implies an equity concept in which individuals with equal income are assumed to have similar wellbeing regardless of the area where they live. The use of local poverty lines allows to gauge intra-country poverty, which can be important for planning local policies.

A possibile approach to compute local poverty lines is by taking into account the different price levels within the country. Under this approach the national poverty line can be modified using area-specific Purchasing Power Parities (PPPs), following the methodology currently applied at international level for international comparisons among different countries (see section 3). In this section we show an application where we compute the Head Count Ratio – a measure of poverty incidence – using Household Expenditure Survey data in Italy, adjusting the national poverty line

Table 9: Types of outlet where poor households make purchases (% distribution) - Year 2019. Source: elaboration on 2019 Istat HES data.

| **Products** | Traditional shop | Open market and street vendors | Hard discount | Hypermarkets and supermarkets | Department stores and outlet chains | Farm or direct producer | Internet |
|---|---|---|---|---|---|---|---|
| Bread | 41,5 | 1,7 | 19,1 | 37,6 | 0,1 | 0,0 | 0,0 |
| Pasta | 13,9 | 1,0 | 31,3 | 53,3 | 0,5 | 0,0 | 0,0 |
| Biscuits, rusks, snacks | 10,9 | 1,6 | 32,3 | 55,0 | 0,2 | 0,0 | 0,0 |
| Fresh meat | 30,2 | 1,3 | 23,9 | 43,9 | 0,3 | 0,3 | 0,0 |
| Frozen meat | 10,9 | 1,6 | 16,4 | 69,3 | 0,8 | 0,8 | 0,1 |
| Cured meats | 15,7 | 1,1 | 29,6 | 53,1 | 0,2 | 0,3 | 0,0 |
| Fresh fish | 41,8 | 13,6 | 14,0 | 30,3 | 0,0 | 0,3 | 0,0 |
| Frozen fish | 10,3 | 1,7 | 32,6 | 54,4 | 1,0 | 0,0 | 0,0 |
| Milk | 13,7 | 1,1 | 32,5 | 52,5 | 0,2 | 0,0 | 0,0 |
| Chesees | 14,7 | 0,8 | 30,4 | 53,7 | 0,2 | 0,2 | 0,0 |
| Yogurt | 10,7 | 0,9 | 32,5 | 55,6 | 0,3 | 0,0 | 0,0 |
| Eggs | 14,3 | 2,4 | 32,6 | 49,7 | 0,2 | 0,7 | 0,0 |
| Fres fruit | 25,1 | 10,8 | 24,9 | 38,8 | 0,1 | 0,2 | 0,0 |
| Fresh vegetables, potatoes and legumes | 25,1 | 9,6 | 24,1 | 40,7 | 0,2 | 0,3 | 0,0 |
| Dried or frozen vegetables, potatoes and legumes | 10,3 | 5,0 | 30,6 | 53,4 | 0,4 | 0,4 | 0,0 |
| Olive oil | 8,5 | 1,3 | 33,2 | 55,4 | 0,3 | 1,4 | 0,0 |
| Mineral water | 11,5 | 0,7 | 29,1 | 58,1 | 0,4 | 0,1 | 0,0 |
| Soft drinks | 8,1 | 2,0 | 31,2 | 58,2 | 0,5 | 0,0 | 0,0 |
| Wine | 13,4 | 0,0 | 32,3 | 53,5 | 0,0 | 0,9 | 0,0 |
| Coffee | 10,1 | 1,1 | 31,4 | 56,5 | 0,1 | 0,3 | 0,5 |
| Medicines | 94,0 | 0,0 | 0,7 | 5,4 | 0,0 | 0,0 | 0,0 |
| Personal hygiene products (soaps, deodorant, baby diapers, etc.) | 10,3 | 0,9 | 31,7 | 55,1 | 1,7 | 0,1 | 0,2 |
| Cleaning products | 12,8 | 2,3 | 31,1 | 51,7 | 2,1 | 0,0 | 0,0 |
| Disposable items for the kitchen (napkins, dishes, etc.) | 13,5 | 1,5 | 34,1 | 49,6 | 1,3 | 0,0 | 0,0 |
| Toys and videogames | 27,2 | 9,1 | 17,3 | 34,4 | 11,9 | 0,0 | 0,0 |

using the SPI($Q_{0.2}$) values computed using the methodology presented in section 4.

In this application we use HES data as they are internationally used – together with EU-SILC data – to compute monetary poverty indicators, such as the HCR. It is important to underline that, therefore, the current application could be extended to other countries and/or datasets, as it present a general methodology to compute local poverty lines.

According to ISTAT, the Head Count Ratio (HCR), a relative measure of poverty incidence, is computed using HES consumption data by defining for each household an indicator variable which takes value 1 if the Monthly Consumption Expenditure (MCE) of the household is less or equal the poverty line, value 0 otherwise. The values are then averaged by using the sample weights. To compute the HCR values, it is thus necessary to first compute the poverty line. At national level, the poverty line for households of two components is set equal to the per-capita mean MCE at country level:

$$nPL = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_j} CE_{ij}w_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n_j} a_{ij}w_{ij}} \tag{10}$$

where $CE_{ij}$ represent the Consumption Expenditure, $w_{ij}$ the survey weight and $a_{ij}$ the household size of household $j$ living in area $i$, with $i = 1, \dots, m$ and $j = 1, \dots, n_j$. To take into account the existence of economies of scale in consumption within households, the poverty line is then adjusted by using the Carbonaro scale (Istat, 2010). In this way, household expenditures can be directly compared with those of households composed of two members. The value of the $HCR_{ij}$ is thus computed for each household as

$$HCR_{ij} = I(CE_{ij} \leq PL \cdot s_{ij}) \tag{11}$$

where $s_ij$ represents the values of the Carbonaro scale, a specific coefficient depending on the household size. Specifically, according to the Carbonaro scale $s_ij = 0.66$ for households with $a_{ij=1}$, $s_{ij} = 1.33$ for a household with $a_{ij} = 3$, $s_{ij} = 1.63$ when $a_{ij} = 4$, $s_{ij} = 1.90$ when $a_{ij} = 5$, $s_{ij} = 2.16$ when $a_{ij} = 6$ and $s_{ij} = 2.40$ for households with 7 members or more. The *HCR* of a given area $i$ computed by using the national poverty line PL is then computed as

$$HCR_i = \frac{\sum_{j=1}^{n_j} HCR_{ij} w_{ij}}{\sum_{j=1}^{n_j} w_{ij}}. \tag{12}$$

A corresponding measure of variability can be computed to derive the coefficient of variation and the confidence intervals for the *HCR* estimates. We computed direct estimates using the *sae* package that is available in *R* (Molina and Marhuenda, 2015).

To allow intra-country comparisons, local poverty lines can be computed and used in the HCR definition. A possibility to compute local poverty lines is by taking into account the different price levels in each area, modifying the national poverty line using area-specific Purchasing Power Parities (PPPs), so that the poverty lines represent approximately the same standard of living across the different areas. By considering the provincial SPI(s values computed using the ASESD methodology and using mean prices (see section 4), the national poverty line can be adjusted for each province using the SPI($Q_{0.2}$) values opportunely weighted (adapting the idea in Renwick et al. (2014)):

$$nPL_i^* = nPL \times (\lambda_i SPI_i + 1 - \lambda_i) \tag{13}$$

where $nPL_i^*$ is the adjusted poverty line for province $i$, $\lambda_i$ is the estimated share of food consumption in province $i$. The quantities $\lambda_i$'s are estimated from the HES 2017 as the provincial mean of the ratios between the rent expenditure and the total consumption expenditure:

$$\lambda_i = \frac{1}{\sum_{j=1}^{n_i} w_{ij}} \sum_{j=1}^{n_i} \frac{p_{ij}}{t_{ij}} w_{ij}, \tag{14}$$

where $n_i$ is the sample size in province $i$, $w_{ij}$ is the survey weight of household $j$ in area $i$, $p_{ij}$ is the food expenditure of household $j$ in area $i$ and $t_{ij}$ is the total consumption expenditure of household $j$ in area $i$. The survey weights have been calibrated to sum to the total households at provincial level.

Although the $\lambda_i$'s are estimated at the provincial level – thus possibly unreliable because of small sample size – we judge the direct estimates suitable for our purpose. Indeed, about half of the provinces have a 95% confidence interval for $\lambda_i$%'s direct estimates that is less than 4% and it is less than 5% for about 75% of the provinces[5]. In table 10 we show the distribution over provinces of the $\lambda_i$'s grouped by the main Italian geographic areas, which is similar among provinces in the north, center and south of Italy.

From Table 10 we can see that distribution of the share of expenditure for food is higher in the southern provinces than in the central and northern provinces, with a mean value equal approximately to 20%.

[5] Standard error of $\lambda_i$'s are obtained ignoring the design effect at the province level.

Table 10: Distribution over provinces of the $\lambda_i$%'s grouped by Italian geographic repartitions

| Repartition | Min | 1st Q. | Median | Mean | 3rd Q. | Max |
|---|---|---|---|---|---|---|
| North | 12.89 | 16.36 | 17.74 | 18.04 | 18.94 | 26.83 |
| Centre | 13.87 | 18.05 | 19.11 | 19.38 | 20.29 | 25.47 |
| South | 18.96 | 21.34 | 23.71 | 23.34 | 25.27 | 27.74 |

Having computed the adjusted nPLs, we then calculated the corresponding direct estimates of the poverty rates. We computed the direct estimates using the `direct` function of the R (R Core Team, 2019) package `sae` (Molina and Marhuenda, 2015). We judged the variability of the direct estimates to be too high, in particular to carry out comparisons among provinces. Indeed, about half of the provinces had a 95% confidence interval length grater than 6%, and about one third of them greater than 9%. Looking at the coefficient of variation (CV) of the direct estimates, we obtained for approximately half of the provinces a CV greater than 30% and for about 25% of the provinces a CV greater than 45%. Therefore, we decided to resort to small area estimation methodologies to try to improve the efficiency of the poverty incidence estimates.

We used a Fay-Herriot (FH) model for the HCR at provincial level in Italy using the adjusted poverty lines (used to compute direct estimates). As auxiliary variables we used the ratio between number of taxed persons over the population, and the ratios between the number of persons with *i*. income coming from salary, *ii*. income coming from pensions and *iii*. income lower than 10,000 euros per year, over the number of taxed persons. These data come from the Italian tax agency database 2017.

The EBLUPs (Empirical Best Linear Unbiased Predictors) obtained with the FH model showed a gain in efficiency with respect to direct estimates. We obtained a CV smaller than 16% in 37 provinces, while half of the provinces had a CV smaller than 20%. The gain in term of variability is shown in figure 9 where we can see that the EBLUP is more efficient than the Direct estimator in all the provinces and the gain in efficiency is greater in those areas where the sample size is smaller, as expected.

Figure 10 maps the HCR computed using the price-adjusted poverty lines referring to the Italian provinces[6]. As we can see, the results confirm the well-known north/south divide, with HCR values that are generally higher in the south of the country, lower in the north.

[6] The HCR value has not been estimated for out of sample provinces in the HES data. Specific predictions could be made for this provinces.

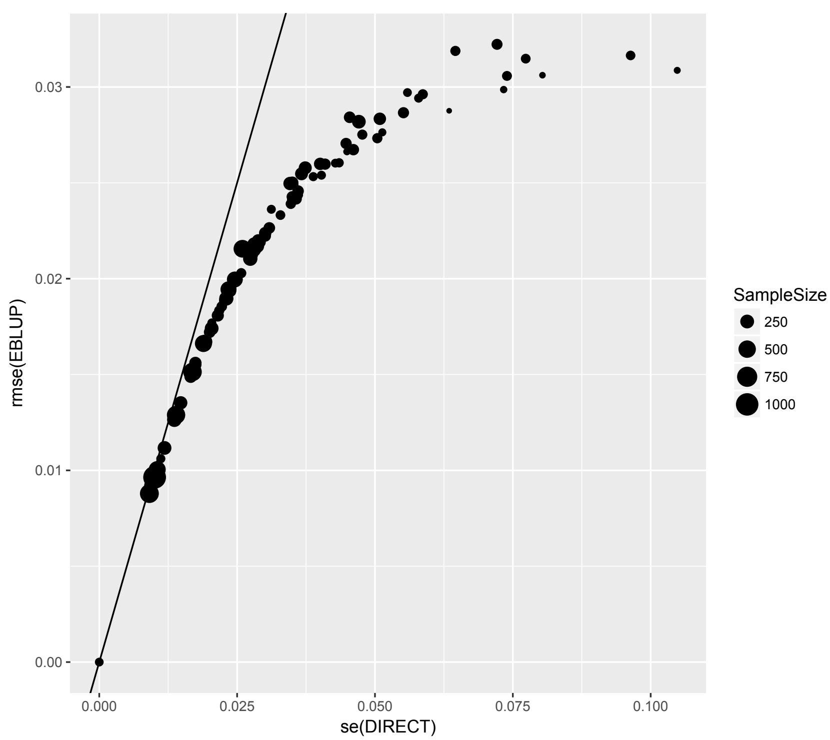


Figure 9: Comparison of the Root Mean Squared Error of EBLUP and direct estimates.

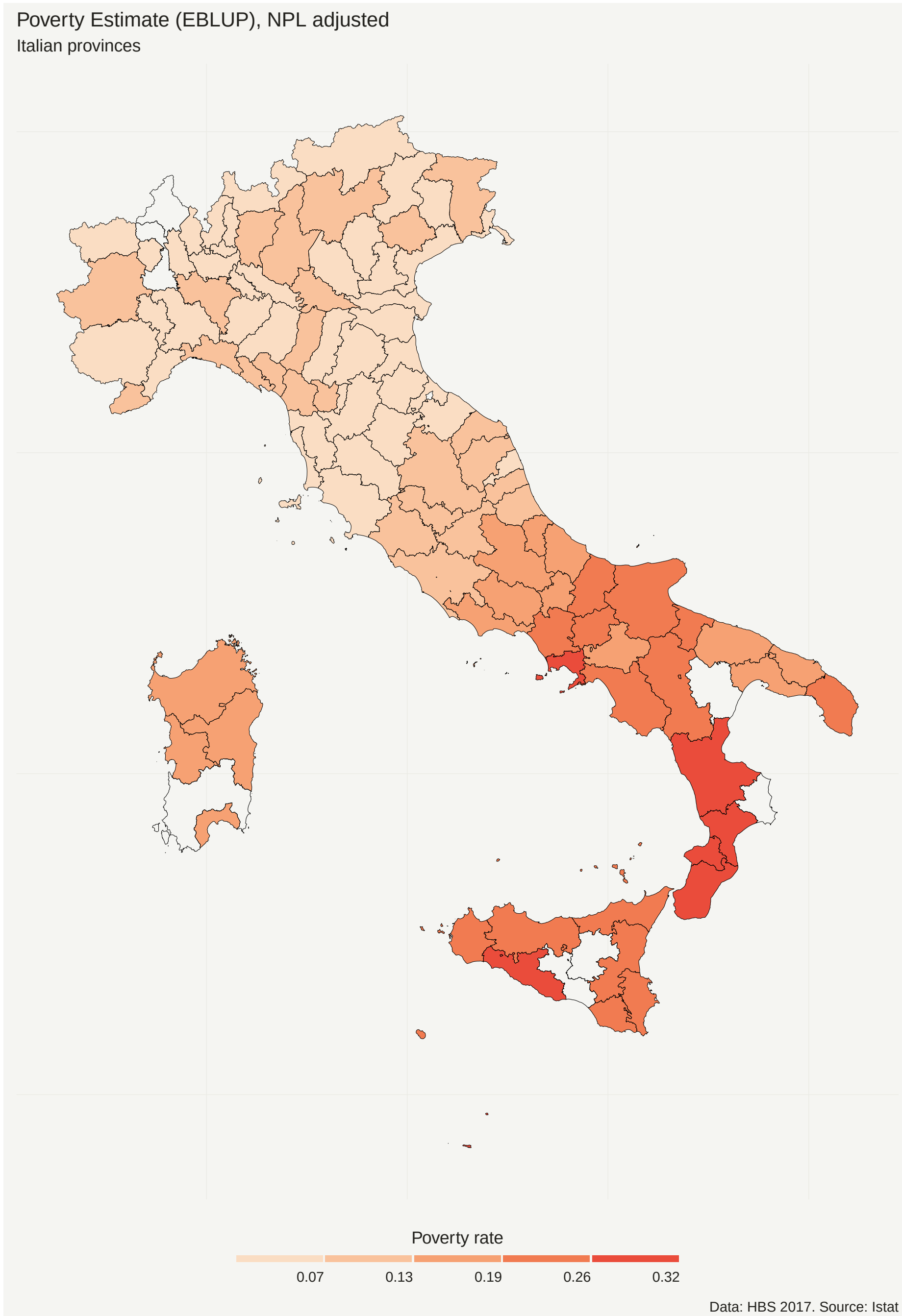


Figure 10: Poverty rate at provincial level in Italy: EBLUPs computed using the adjusted national poverty line.

We also computed the EBLUPs without any adjustment of the national poverty line, using the same small area model as for adjusted EBLUPs. In this case, however, the map of the HCR using the national poverty line is essentially the same as the one using the price-adjusted poverty lines, as the SPI($Q_{0.2}$) are applied only to approximately the 20% of the poverty line, as explained above. A finer comparison is represented in Figure 11, where we can see that using the SPI($Q_{0.2}$) to adjust the poverty lines, the HCRs in northern and central provinces slightly decrease.

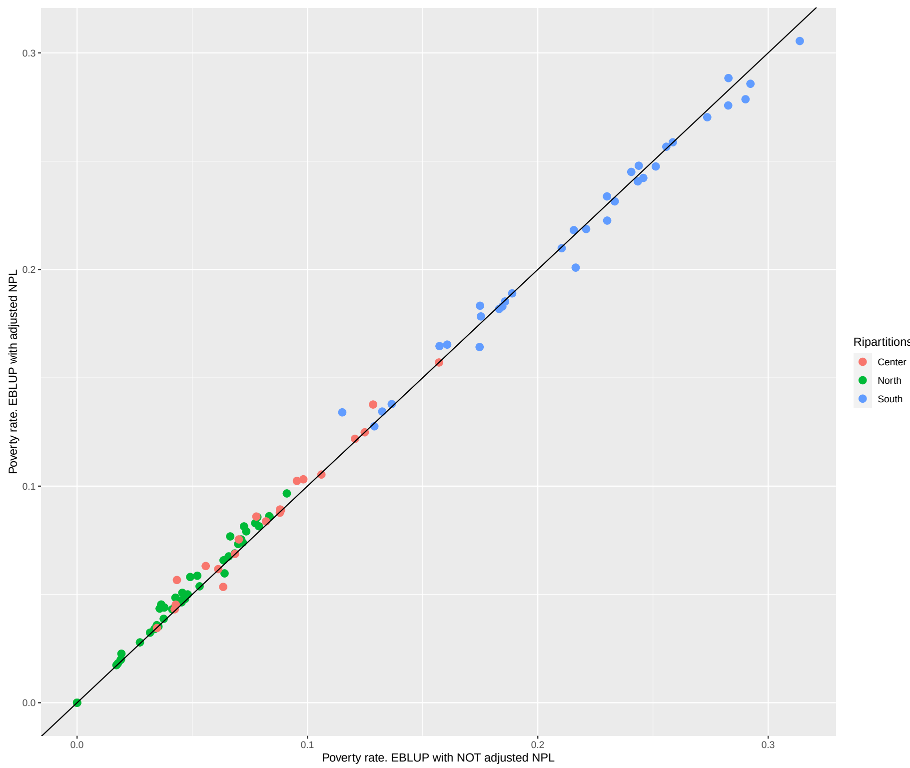


Figure 11: Poverty rate at provincial level in Italy: provincial EBLUPs estimates using the SPI($Q_{0.2}$) adjusted vs not adjusted national poverty line.

The results obtained here suggest that the methodology can be extended to include other Spatial Price Indexes, therefore adjusting the national poverty line with other components of households' consumption expenditure. Indeed, our results suggest the products included in the scanner data represent a relevant but still limited share of the total household consumption expenditure, approximately equal to the 20%. Therefore, by including other consumption expenditure components, such as for example the expenditure for the rent, the national poverty line could be adjusted in a more complete manner. Figure 12 reports the same analysis as figure 11 but also using, in addition to the SPI($Q_{0.2}$), a Spatial Rent Index (SRI) to further adjust the national poverty line. The SRI has been estimated using HES data and small area estimation methodologies. It is important to underline that the cost for the rent covers, in mean, another 20% of the total household consumption expenditure. From figure 12 we can see that the effect of the adjustment in this case is more pronounced.

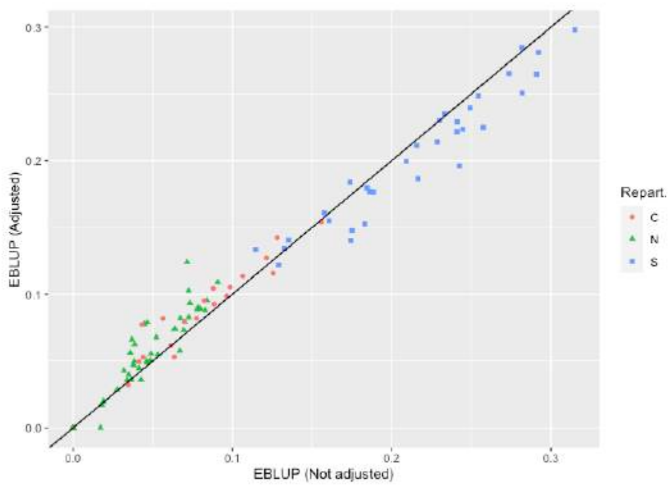


Figure 12: Poverty rate at provincial level in Italy: provincial EBLUPs estimates using the national poverty line adjusted with the SPI($Q_{0.2}$) and SRIs vs EBLUPs estimates using the not adjusted national poverty line.

Finally, as already stressed, it is important to underline that, being based on EU–SILC and HES data, the proposed analyses - applied here only to Italian data - could be extended to other European countries provided that survey data at subnational level are available.

# 8 Final remarks

Integration of traditional data with Big data sources is the red 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. The approach is model– based and uses tools which are now in the current tool–box of the majority of NSIs in European Statistical System for the analysis of price data (the CPD models) and of survey data (SAE models).

Integrating data sources is generally a complex process, for this it can be useful a list of recommendations which stem from our analysis:

- price distributions for groups of items (BHs or other groupings) obtained by scanner data are a valuable information. The first quantiles (quintiles, or deciles) of these distributions are mimicking the prices paid by the poor. This is the evidence from Section 6 where we see that with reference to the poor, a suitable place for shopping is the hard discount.
- the SPIs (f.i. SPI($Q_{0.2}$)) alone are not enough to appreciate the subregional variations of the cost of living, they should be accompanied by the SPIs for housing Rents (SPI Rs) obtained by HES or by archives of Revenue and Tax Agencies.
- the prices from scanner data may be affected by the price policy of the RTCs, this can have a uniform effect at country level, smoothing the subregional variability of the prices. In this Distinguishing between the so–called first– price products and the others can facilitate the analysis. The first–price products group all the products with the lowest price that exists for each product category in a supermarket assortment.
- integration of different data sources as done in this study requires for a deep reflection on the sources of uncertainty affecting the whole process. They come from the design of the data sources and from the application of models as CPD and SAE models. Scanner data base can be the results of probability or not probability sampling of items or observations (as in the design of the Istat data base of prices from scanner data). While the accuracy of the model-based estimates has already been studied in the literature, the effect of the design of the data sources, especially in case of Big data sources, has not yet completely studied. This last topic in combination with the accuracy of model-based estimates necessitate further research.

# 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.

Diewert, W. E. (1995). Axiomatic and economic approaches to elementary price indexes. Technical report, National Bureau of Economic Research.

Elteto, O. and P. Koves (1964). On a problem of index number computation relating to international comparison. *Statisztikai Szemle 42*, 507–518.

Gini, C. (1931). On the circular test of index numbers. *Metron 9*(9), 3–24.

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

Istat (2010). *La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane.* Roma, Italy: Istat, 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.

Molina, I. and Y. Marhuenda (2015). sae: An R package for small area estimation. *The R Journal 7*(1), 81–98.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rao, D. P. and G. Hajargasht (2016). Stochastic approach to computation of purchasing power parities in the international comparison program (icp). *Journal of econometrics 191*(2), 414–425.

Renwick, T., B. Aten, E. Figueroa, and T. Martin (2014). Supplemental poverty measure: A comparison of geographic adjustments with regional price parities vs. median rents from the american community survey. Technical report, Bureau of Economic Analysis.

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

Szulc, B. (1964). Indices for multiregional comparisons. *Przeglad statystyczny 3*, 239–254.

World Bank (2013). *Measuring the Real Size of the World Economy, The Framework, Methodology, and results of the International Comparison Pogram-ICP.* Washington, DC: World Bank.

World Bank Group (2015). *Operational Guidelines and Procedures for Measuring the Real Size of the World Economy.* Washington, DC: 2011 International Comparison Program.