ORIGINAL PAPER

The impact of local cost-of-living differences on relative poverty incidence: an application using retail scanner data and small area estimation models

Stefano Marchetti[1] [ID] · **Caterina Giusti[1]** [ID] · **Francesco Schirripa Spagnolo[1]** [ID] · **Gaia Bertarelli[2]** [ID] · **Luigi Biggeri[3]**

Abstract
Estimating economic poverty indicators at the local level is essential for well-targeted data-driven welfare policies. However, Italy is a country characterized by strong geographical heterogeneity represented by unequal price levels among different areas, and computing poverty indicators with a national monetary poverty threshold can be misleading. This work proposes a novel approach to estimate monetary poverty incidence at the provincial level in Italy considering the different price levels within national boundaries. To account for local price variation, Spatial Price Indices (SPIs) are computed using scanner data on retail prices. The SPIs are estimated in two ways, referring to the mean local prices and using the 20th percentile of the prices. These two kinds of SPIs are used to adjust the national poverty line when computing the poverty incidence at the provincial level using Small Area Estimation (SAE) models. Our findings suggest that adjusting the national poverty line using the SPIs to compute a monetary poverty index can modify the poverty mapping results from the map produced with the traditional national poverty line that ignores price differences.

1 Introduction

In recent decades, reducing poverty has become a much debated and researched issue at the international level. The first of the Sustainable Development Goals adopted by all United Nations member states in 2015 urges a call for action to "end poverty in all its forms everywhere". For statisticians, one of the key ways to support governments in fulfilling this goal is to provide policy-makers with an effective set of indicators that can support their decisions and can be used to evaluate their

---

Extended author information available on the last page of the article

🙋 Springer

policies. To better target these policies where the need for interventions to reduce poverty is higher, local indicators of poverty are an essential tool. However, official poverty measures are often computed using data from national surveys managed by National Statistical Institutes (NSIs), and these data cannot be used to compute reliable poverty indicators at the regional/subregional level or for other unplanned domains of interest. In these cases, small-area estimation methodologies have been shown to be effective by providing a set of models that can compute reliable local indicators of poverty over a vast range of data availability situations and model assumptions (Pratesi 2016; Guadarrama et al. 2016; Tzavidis et al. 2018).

In the international literature, poverty has been recognized as a multidimensional phenomenon that should be measured using a set of indicators covering different dimensions. Under this approach, measuring poverty means going beyond the classical relative monetary indicators, which compute the share of individuals or households with an income/consumption falling below a chosen threshold. Nevertheless, relative measures of poverty incidence are still an important tool to monitor poverty in the territories of interest and are usually included by governments in the set of key poverty indicators together with other measures focusing on nonmonetary dimensions of poverty (World Bank 2021; United Nations Development Programme 2022; United Nations Statistical Commission 2023).

In this paper, we focus on the estimation of the Head Count Ratio (HCR), a measure of poverty incidence, at the local level in Italy. Specifically, we refer to the HCR computed using consumption data from the Household Budget Survey (HBS) managed by Istat, the Italian NSI. By using this indicator, each household is considered poor if its consumption level is below the national poverty threshold, computed as the mean consumption level for households with two members, and then adjusted for households of different sizes by using the Carbonaro equivalence scale (Carbonaro 1985). However, using a national threshold can be misleading in a country historically characterized by strong regional heterogeneity for many socioeconomic phenomena (Istat 2019). It is a matter of fact that for measuring absolute poverty, Istat adjusts the poverty threshold not only by the number of household members but also by other relevant characteristics, such as the macroarea of residence (Istat 2022). Indeed, an experimental study conducted by Istat in 2010 showed that Purchasing Power Parities (PPPs) were strongly heterogeneous among the 20 Italian regional capitals, with Bolzano, located in the north, characterized by the highest costs of living and Napoli, located in the south of the country, characterized by the lowest costs. As suggested by Tonutti et al. (2022) in a context with marked geographical heterogeneity, capturing households' real purchasing power is essential to gain an accurate picture of the incidence of poverty. Indeed, the Italian political debate has recently highlighted that the strong north–south divide should possibly be accounted for in salaries and in national measures such as the guaranteed minimum income known under the name of "Reddito di Cittadinanza" (Saraceno 2021; RaiNews 2022).

Measuring local prices on a regular basis is, however, a difficult task for NSIs, which usually focus on measuring prices for comparisons over time and among countries. Therefore, one possibility to compute local Spatial Price Indices (SPIs)—based on the same idea as PPPs but computed with a different methodology—to adjust the national poverty line is to use survey data and resort, when needed, to

Small Area Estimation (SAE) models (Bertarelli et al. 2021; Pratesi et al. 2021; Marchetti et al. 2020). Moreover, new data sources have recently generated new opportunities in this respect. Since 2014, Istat has introduced scanner data on retail prices in the estimation of inflation, and recent studies have also explored the possibility of using scanner data to measure local PPPs (Biggeri and Pratesi 2022; Laureti and Polidoro 2022).

The paper is organized as follows. In Sect. 2, we describe the data used in the paper, the scanner data, the HBS and other administrative data. In Sect. 3, we describe the proposed methodology to estimate the provincial SPIs, while in Sect. 4, we describe how the SPIs are used in the estimation process of the provincial HCRs using the province-specific poverty lines. The main results are presented in Sect. 5. Finally, in Sect. 6, we discuss the main findings, limitations and future directions of this work.

2 Data description

The scanner data used in this work are those coming from the scanning of bar codes at checkout lines of retail stores made available together with HBS data to the authors of this paper by Istat in the framework of the H2020 research project MAK-SWELL (MAKing Sustainable development and WELL-being frameworks work for policy analysis). We now describe the main characteristics of these data: more information can be found in the project deliverables available at www.makswell.eu.

Starting in 2014, Istat signed an agreement with the market research company ACNielsen and the Association of large-scale retail trade distribution (ADM) to use scanner data of grocery products (excluding fresh food) in the process of estimating inflation. The data are transmitted by ACNielsen on a weekly basis and contain turnover and quantities sold for specific item codes. More precisely, they provide information on sales, expenditures, and quantities with very detailed information on sold products (such as brand, size or type of outlet) at the barcode GTIN (Global Trade Item Number) level. Data concerns 79 aggregates of products belonging to 5 ECOICOP (European Classification of Individual Consumption by Purpose) divisions (01: Food and nonalcoholic beverages, 02: Alcoholic beverages and tobacco, 05: Furniture, home furnishings and items for regular home maintenance, 09: Leisure and culture, 12: Other goods and services). The ECOICOP classification is based on the Classification of Individual Consumption by Purpose (COICOP) developed by the United Nations Statistical Commission (2018). Its aim is to classify and analyse individual consumption expenditures incurred by households, nonprofit institutions serving households, and the general government according to the purpose of each purchased item. This classification has a hierarchical structure with 4 levels of aggregation: (i) divisions, (ii) product groups, (iii) product classes and (iv) product subclasses. The division level is made up of 12 categories, of which we consider the five listed above in this paper.

A probabilistic design is implemented to select the sample of outlets for which Nielsen sends Istat scanner data. In 2018, outlets were stratified according to 107 provinces, 16 market chains and two outlet types, hypermarkets and supermarkets,

for a total of 867 strata. Only strata with at least one outlet were considered. All the provinces are included in the final sample. The inclusion probabilities were assigned to each outlet based on their turnover value. In 2018, 1781 outlets were selected. The number of strata, the number of outlets and the coverage in terms of turnover at the regional and national levels for 2018 are available in the appendix. The number of outlets used at the macroregional level in 2018 is shown in Table 1.

To select items, a cut-off sample of GTINs was selected within each outlet/aggregate of products. This selection covers 40% of turnover, but includes no more than the first 30 GTINs in terms of turnover. The selection was made in December and was maintained as a fixed choice throughout the following year. In total, 1,370,000 price quotes were collected each week. Using these data, for each GTIN, prices are calculated taking into account turnover and total quantities in inventory (weekly price = weekly turnover/weekly quantities). Monthly prices are computed with the arithmetic mean of weekly prices weighted with quantities. We describe the methodology proposed to estimate spatial price indices using these data in Sect. 3.

To estimate the HCR at the provincial level, we use data from the HBS 2017. The HBS focuses on consumption expenditure behaviors of households residing in Italy. It analyzes the evolution of level and composition of household consumption expenditure according to their main social, economic and territorial characteristics. In agreement with Eurostat, the survey is based on the harmonized international classification of expenditure voices to ensure international comparability. The main focus of the HBS is represented by all expenditures incurred by resident households to purchase goods and services exclusively devoted to household consumption (self-consumption, imputed rentals and presents are included); every other expenditure for a different purpose is excluded from the data collection (e. g., payments of fees and business expenditures). The HBS is surveyed yearly, and it is used by Istat to measure the relative and absolute poverty incidence, the structure and the level of household consumption expenditure according to the socioeconomic characteristics of the households. The HBS design is a two-stage, stratified by the type of municipality: metropolitan areas, suburban areas and municipalities with more than 50 thousand inhabitants, and municipalities with fewer than 50 thousand inhabitants. The first stage units are the Italian municipalities (approximately 8 thousand), and the second stage units are the households. The design allows for sound estimates at the regional level and for the type of municipality. Subregional estimates, such as the provincial ones, computed using these data are usually unreliable since the HBS sample size at this level can be rather low. In 2017, the HBS national

Table 1 Number of outlets at the national and macro-regional level. Year 2018

Macro-region	N. Outlets	% Outlets
North	942	52.9
Centre	353	20.4
South	475	26.7
Italy	1781	100

sample was composed of 16,496 households, with provincial sample sizes ranging from 20 to 1036, with a median value of 125, first quartile 59 and third quartile 184. Therefore, to estimate the HCR at the provincial level, we resort to SAE methods. The same data and models are also used to estimate for each province the share of food consumption expenditure of the relative poor, used to properly adjust the national poverty line at the provincial level, as detailed in Sect. 4.

Before describing the methodologies used in the paper, we also briefly present here the variables used as covariates in the SAE models. The auxiliary variables were selected from population registers and archives referring to 2017. Therefore there are no missing values for the covariates in the considered provinces. In particular, in this paper, we use data at the provincial level from 1) the Italian Tax Agency: the ratio between taxpayers and inhabitants, per capita taxable income, fraction of payers in different classes of annual income (e.g., 0 to 10,000, 10,000 to 15,000), per capita taxable income by source (labour, pensions, real estate), fraction of payers by source (labour, pensions, real estate), per capita net tax. We also use 2) data from population archives managed and made available by Istat: number of households, mean household size, incidence of foreign people, mean age, structural dependency index (i.e., the ratio between the nonworking age population (0–14 and 65+ years) and the working age population (15–64 years), multiplied by 100), the elderly dependency index (i.e., the ratio between the population aged 65+ and the population of working age (15–64 years), multiplied by 100), and the old-age index (i.e., the ratio between the population aged 65+ and the population aged 0–14 years, multiplied by 100).

It is worth noting that a time lag of one year is present in our data, as HBS data and covariates refer to 2017, while the scanner data refer to 2018. While it is true that a price increase can occur from one year to the next, this usually changes homogeneously throughout the country in such a short period, so we believe that this temporal lag may not influence the final results. Then, concerning the target areas of interest, during 2017 the number of provinces in Italy changed for territorial reasons, with a reduction from 110 to 107. Out of these 107 areas, 2018 scanner data cover a total of 103 provinces. To be able to compare the same territorial units—merging the scanner, HBS and auxiliary data—our work is based on a total of 100 provinces that remained unchanged between 2017 and 2018. As the seven missing provinces are equally distributed in the Italian macroareas, we believe that the 100 provinces included in our analysis are a good representation of the total Italian territory.

3 Estimation of spatial price indices

To compute the consumer spatial price indices (SPIs) for Italian provinces using scanner data, we embrace an innovative approach regarding the definition and satisfaction of the principle of comparability. As already anticipated, this is the reason why we prefer to not call these indices representative of Purchasing Power Parity (PPP). Indeed, in the standard approach, the principle of comparability is applied very tightly by considering the comparisons of like-to-like items (products) for the different subnational

areas (Biggeri and Pratesi 2022; Laureti and Polidoro 2022). In this approach, the lowest level of aggregation of the products is the so-called Basic Heading (BH) level, as defined by the World Bank for the computation of international PPPs (World Bank 2013).

In our approach, we apply the principle of comparability at a different level, i.e., at the level of the ECOICOP-8-digit classification, for which we have 79 available groups (belonging to food and nonalcoholic beverages). We hypothesize that even if the brand, quality, etc., is different, the elementary products (items) within a group are sufficiently similar that consumers are generally indifferent about choices in a product that, in any case, satisfies the same consumer needs (giving him or her the same utility). This approach is justified by the fact that the SPIs are then used to adjust the poverty line when computing the poverty incidence in Italian provinces. In this work, therefore, the comparison is made 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 purchased. Then, the average level of prices of the different groups of products is aggregated to obtain the SPIs for each subnational area (Pratesi et al. 2021; Biggeri and Pratesi 2022).

More specifically, to estimate the SPI for each of the Italian provinces, a two-step procedure is employed. In the first step, we compute the weighted mean price $\bar{p}_{ki}$ for ECOICOP-8-digit k and province i by considering the unit value prices from the consumer side. Let r_{dki} and q_{dki} be the annual turnover and the total quantity sold (which are the expenditure and the quantity purchased by consumers, respectively) of item d belonging to ECOICOP-8-digit k in province i. These quantities are estimated by Istat using the scanner data from 2018 and the sampling weights computed according to the survey design (we refer to Pratesi et al., 2021 for further details). First, we calculate the annual price per gr. or ml. for each item in province i:

$$p_{dki} = \frac{\frac{r_{dki}}{q_{dki}}}{u_{dki}},$$

where r_{dki}/q_{dki} is the turnover for item d and u_{dki} is the quantity sold of the item dki in terms of gr. or ml. The weighted mean price per gr. or ml. for products in ECOICOP-8-digit k and province i is given by:

$$\bar{p}_{ki} = \frac{1}{\sum_{d=1}^{n_{ki}} w_{dki}} \sum_{d=1}^{n_{ki}} p_{dki} w_{dki},$$

where w_{dki} are the relative weights in terms of turnover for each item, calculated as $w_{dki} = r_{dki}/(\sum_{d=1}^{n_{ki}} r_{dki})$, and n_{ki} is the number of items in the kth ECOICOP-8-digit aggregation and the ith province.

In the second step, we aggregate the 79 averages to estimate the SPIs for the 100 provinces. Note that not all the ECOICOP-8-digit aggregates are present in all the provinces. Specifically, we employ a modified Country Product Dummy (CPD) model (Laureti and Rao 2018) where the logarithm of $\bar{p}_{ki}$ is considered a function of the spatial price index of the ith province relative to the other

provinces, and of the provincial average price of the kth group of commodities. The CPD model can be written as follows:

$$\log \bar{p}_{ki} = \alpha_0 + \alpha_i D_i + \beta_k I_k + \varepsilon_{ki}, \quad i = 1, \dots, 100 \quad k = 1, \dots, 79, \tag{1}$$

where D_i is a vector that equals 1 if the mean price is in province i and 0 otherwise, I_k equals 1 if the mean price belongs to the kth ECOICOP-8-digits and 0 otherwise, and the error $\varepsilon_{ki} \sim N(0, \sigma^2)$.

To consider the different levels of turnover between the ECOICOP-8-digit aggregates, we estimate the model (1) using weighted least squares, where the weights are computed as the ratio between the total turnover of one aggregate in one province and the total turnover in the province (where n_i is the number of items in the ith province): $wls_{ki} = (\sum_{d=1}^{n_{ki}} r_{dki})/(\sum_{d=1}^{n_i} r_{dki})$.

The D_is vectors in the model in Eq. (1) are a linear combination of the constant, which introduces an identification problem. To address this issue, we impose the constraint $\alpha_1 = 0$. Once the model is estimated, we obtain the estimates of SPIs at the provincial level by $\exp(\hat{\alpha}_i)$, where $\hat{\alpha}_i$ is the estimate of α_i. The coefficient α_i, adjusted following Suits (1984), is the difference of the fixed effect of province i compared to Italy. In other words, the quantity $\exp(\hat{\alpha}_i)$ represents the SPI for food in province i with respect to Italy, and we indicate it as the SPI of province i (SPI_i).

To obtain poor-specific SPIs, we modify equation (1) using as a target variable the logarithm of the 20th percentile of the prices belonging to ECOICOP-8-digit aggregates. We let $p_{20,ki}$ be the 20th percentile of the price p_{dki}, i.e., the prices of all the items in ECOICOP-8-digit k in province i. The modified CPD model is:

$$\log p_{20,ki} = \alpha_0 + \alpha_i D_i + \beta_k I_k + \varepsilon_{ki}, \quad i = 1, \dots, 100 \quad k = 1, \dots, 79, \tag{2}$$

where covariates and coefficients have the same meaning as in model (1). The SPIs are then obtained following the same procedure described above. These SPIs are poor-specific because we use the lower part of the price distribution, assuming implicitly that (relatively) poor households buy cheaper goods. Further developments of this study could refine the classification of the products to relax the hypothesis of perfect substitutability among items in the same ECOICOP-8-digit k, which is used in models (1) and (2) to control for some characteristics of the items (e.g., the brand).

In what follows, we refer to the SPI_i computed using model (1) as SPI_i^M and to the SPI_i computed using model (2) as SPI_i^{Q20}.

4 Poverty estimates at the provincial level

In this paper, our target is to estimate the HCR based on consumption data at the provincial level in Italy, while also taking into account the territorial price differences as measured by the estimated SPIs.

As already specified, the HCR is a relative poverty index defined by Istat as the percentage of households whose consumption expenditure is below a given threshold, called the poverty line. Relative poverty is estimated using the HBS data, which have been briefly described in Sect. 2. In the computation of the HCR, the poverty line for a household of two members is equal to the mean per capita consumption expenditure computed at the national level. Then, for households with a different number of members, the poverty line is adjusted using the Carbonaro scale (Carbonaro 1985), which gives a weight of 0.6 to households with one member, 1.33 to households with three members, 1.63 to households with four members, 1.9 to households with five members, 2.16 to households with six members and 2.4 to households with seven or more members.

We let y_{ij} be the consumption expenditure for household j in the area (province) i, w_{ij} be the survey weight related to household j in area i such that $\sum_{j=1}^{n_i} w_{ij} = N_i$, where N_i is the population size (persons) in area i and n_i is the sample size in area i. We also let $T(h_{ij})$ be the poverty line at the national level that is a function of the household size h_{ij} of household j in area i, according to the Carbonaro scale described above. The direct estimates—which are estimates that use only survey area-specific units—of the HCRs in area i are

$$\widehat{HCR}_i = \frac{\sum_{j=1}^{n_i} I(y_{ij} < T(h_{ij})) w_{ij}}{N_i}, \tag{3}$$

where $I(\cdot)$ is the indicator function that is equal to 1 when its argument is true and 0 otherwise. The variances of the estimated *HCR*s are estimated using linearization methods.

For each province, we estimate three different HCR values: 1) the HCR with the poverty line adjusted at the provincial level using the SPI_i^M; 2) the HCR with the poverty line adjusted at the provincial level using the SPI_i^{Q20}; and 3) the HCR with the 'traditional' national poverty line. Practically, in cases 1) and 2), the poverty line is a threshold that is a function of the household size—as usual in case 3)—but also of the SPI in area i. Therefore, we define three different poverty lines: $T(h_{ij}) = T^{na}$, the national poverty line (equal for all provinces) and not adjusted for local prices; $T(h_{ij}, \text{SPI}_i^M) = T_i^M$, the threshold adjusted using SPI_i^M; and $T(h_{ij}, \text{SPI}_i^{Q20}) = T_i^{Q20}$, the threshold adjusted using SPI_i^{Q20}.

As explained in Sect. 3, the SPIs - SPI_i^M and SPI_i^{Q20} - are obtained using only food and nonalcoholic beverage prices based on scanner data. For this reason, when using them, we adjust the poverty thresholds only partially. Specifically, we use the share of food consumption expenditure of the poor households, λ_i, for each province, estimated using HBS data, as a weight in the adjustment of the poverty line. Therefore, the price-adjusted provincial poverty lines T_i^M and T_i^{Q20} are equal to:

$$T_i^M = T^{na}(\lambda_i \text{SPI}_i^M + 1 - \lambda_i); \quad T_i^{Q20} = T^{na}(\lambda_i \text{SPI}_i^{Q20} + 1 - \lambda_i). \tag{4}$$

Since our goal is to adjust the poverty line, we estimate λ_i conditionally for poor people, as described at the end of this section.

The three types of direct estimates of the HCR at the province level have been computed by plugging T^{na}, T_i^M or T_i^{Q20} into equation (3):

$$\widehat{HCR}_i^{na} = \frac{\sum_{j=1}^{n_i} I(y_{ij} < T^{na})w_{ij}}{N_i}$$

$$\widehat{HCR}_i^{M} = \frac{\sum_{j=1}^{n_i} I(y_{ij} < T_i^M)w_{ij}}{N_i}$$

$$\widehat{HCR}_i^{Q20} = \frac{\sum_{j=1}^{n_i} I(y_{ij} < T_i^{Q20})w_{ij}}{N_i}$$

To obtain reliable estimates at the provincial level, we need SAE methods because provinces are unplanned domains in the HBS. Therefore, their sample size (n_i) is often small, resulting in unreliable direct estimates.

Modern SAE techniques are model-based and can be divided into two main approaches: unit-level approaches that relate the unit values of a study variable to unit-specific auxiliary variables and area-level approaches that relate small area direct estimates to area-specific auxiliary variables. More details can be found in Rao and Molina (2015). As in this work we have auxiliary variables available only at the provincial level, we use the area-level approach.

The most common area-level model is the Fay-Herriot (FH) model introduced by Fay and Herriot (1979), which is composed of two levels. The first is the sampling model:

$$\hat{\theta}_i = \theta_i + e_i, \tag{5}$$

where $\hat{\theta}_i$ is an unbiased direct estimator for the population indicator of interest θ_i and e_i are the sampling errors in the ith area, normally distributed with $E(e_i|\theta_i) = 0$ and $V(e_i|\theta_i) = \sigma_{e_i}^2$, which is assumed to be known for all areas.

In the second level, the parameter of interest, θ_i, is linked to the vector of p auxiliary variables, $\mathbf{x}_i = (x_{i1}, x_{i1}, ...x_{ip})^T$, via the following model:

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i, \quad i = 1, ...m, \tag{6}$$

where $\boldsymbol{\beta}$ is the vector of the regression coefficients and γ_i's are area-specific random effects assumed to be independent and identically normally distributed with mean 0 and variance σ_γ^2.

Combining Eq. 6 and Eq. 5 we obtain the FH linear mixed model

$$\hat{\theta}_i = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i + e_i, \tag{7}$$

where γ_i is independent of e_i for all i.

The Empirical Best Linear Unbiased Predictor (EBLUP) of θ_i under the model in equation (7) is given by:

Table 2 Arc-sin area-level model covariates for the not adjusted, mean adjusted and Q20 adjusted poverty index

	Not adjusted		Mean adjusted		Q20 adjusted	
	Coeff.	p-val	Coeff.	p-val	Coeff.	p-val
(Intercept)	−0.107	0.869	−0.091	0.889	0.047	0.945
Structural dependency index	0.013	0.011	0.011	0.035	0.010	0.076
Ratio between taxpayers and inhabitant	−5.219	0.000	−4.658	0.000	−4.958	0.000
Per capita taxable income	0.000	0.110	0.000	0.105	0.000	0.156
Fraction of employed taxpayers	2.920	0.003	2.344	0.017	2.580	0.012
Fraction of taxpayer annual income < 10000 euro	4.540	0.008	4.312	0.012	4.299	0.017
Per capita net tax	0.000	0.105	0.000	0.104	0.000	0.164
Fraction of pensioned taxpayers	1.658	0.104	1.466	0.151	2.001	0.059

$$\hat{\theta}_i^{FH} = \hat{\phi}_i \hat{\theta}_i + (1 - \hat{\phi}_i)(\mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\gamma}_i),$$

where $\hat{\boldsymbol{\beta}}$ is the weighted least square estimator of $\boldsymbol{\beta}$, $\hat{\sigma}_\gamma^2$ is a consistent estimator of σ_γ^2, and $\hat{\phi}_i = \hat{\sigma}_\gamma^2 / (\hat{\sigma}_\gamma^2 + \sigma_{e_i}^2)$, called *shrinkage factor*, measures the model variance with respect to the total variance. Consequently, the EBLUP places more weight on the direct estimator when the sampling variance is small and vice versa. Moreover, the shrinkage factor approximately measures the reduction of the EBLUP mean squared error with respect to the variance of the direct estimator, given that the magnitude of the mean squared error of the EBLUP is given by $\phi_i \sigma_{e_i}^2$.

Since our target parameter is a proportion and the estimate in an area should be within the interval [0, 1], we use a popular arc-sin transformation (Casas-Cordero Valencia et al. 2016; Schmid et al. 2017; Jiang et al. 2001). In this case, the direct estimates are transformed as follows:

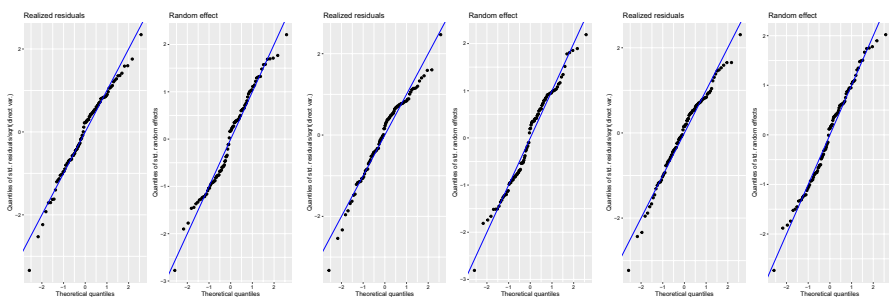$$\theta_i^{arcsin} = \sin^{-1}\left(\sqrt{\theta_i}\right),$$

Fig. 1 QQplot for model residuals (sampling errors and random effects) for arc-sin transformed FH models for the poverty index with poverty line not adjusted (first two on the left), mean adjusted (the two on the centre) and Q20 adjusted (last two on the right)
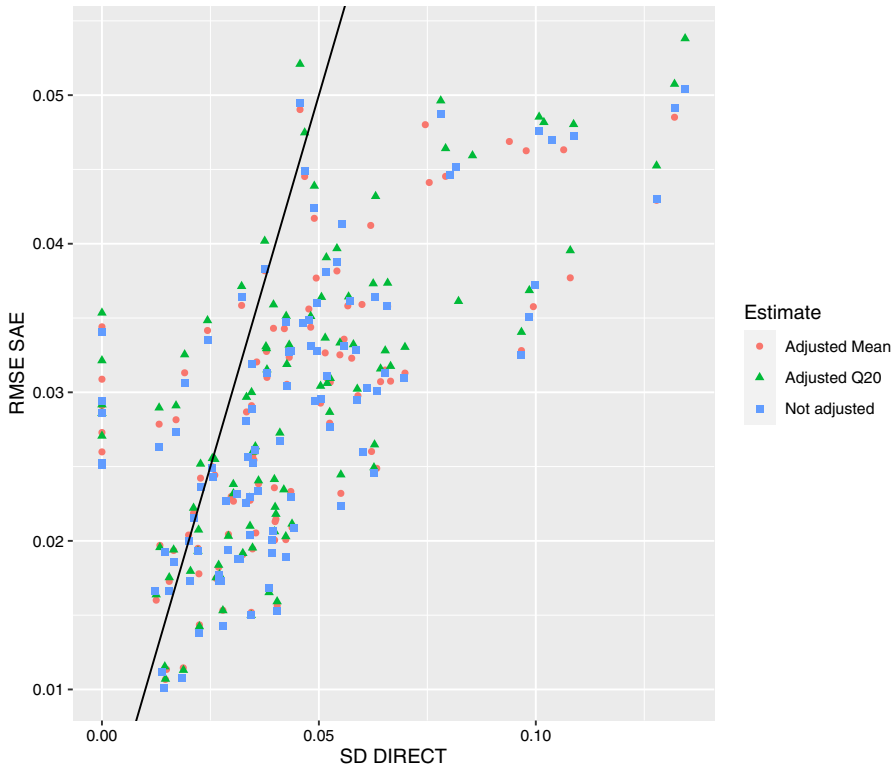
Fig. 2 Estimated standard deviation of direct estimates (SD Direct) vs. estimated root mean squared error of small area estimates (RMSE SAE), by type of adjustment. The solid line is the equality line

and the sampling variance is approximated by $\sigma^2_{e_i} = 1/(4\tilde{n})$, where $\tilde{n}$ is the effective sample size, which is derived by dividing the sample size by an estimate of the design effect (Jiang et al. 2001). The FH model is then estimated using equation (7) with results truncated to the interval $[0, \pi/2]$ if needed. To obtain final estimates on the original scale, the final estimation results must be subjected to a back-transformation. We use the back-transformation with the bias correction proposed by Sugasawa and Kubokawa (2017); Hadam et al. (2020):

$$\theta_i^{FH,bc} = \int\limits_{-\infty}^{+\infty} \sin^2(t) \frac{1}{2\pi \frac{\hat{\sigma}_u^2 \sigma_{e_i}^2}{\hat{\sigma}_u^2 + \sigma_{e_i}^2}} \exp\left\{ \frac{\left(t - \theta_i^{FH,arcsin}\right)^2}{2\pi \frac{\hat{\sigma}_u^2 \sigma_{e_i}^2}{\hat{\sigma}_u^2 + \sigma_{e_i}^2}} \right\} dt, \tag{8}$$

where $\theta_i^{FH,arcsin}$ denotes the FH estimator obtained by using the arc-sin transformed scale. The R package "emdi" (Kreutzmann et al. 2019) was used to compute the estimates.

For each of the three poverty indices, we adapt the arc-sin area-level model, where the auxiliary variables have been selected among those already presented in

Table 3 Distribution among provinces of the estimated root mean squared error of direct and small area estimates of the poverty index

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\widehat{HCR}_i^{na}$	0.000	0.027	0.039	0.044	0.055	0.134
$HCR_i^{FH,na}$	0.010	0.021	0.028	0.029	0.035	0.050
$\widehat{HCR}_i^{M}$	0.000	0.027	0.040	0.044	0.055	0.134
$HCR_i^{FH,M}$	0.011	0.021	0.029	0.029	0.034	0.050
$\widehat{HCR}_i^{Q20}$	0.000	0.027	0.040	0.045	0.056	0.134
$HCR_i^{FH,Q20}$	0.011	0.022	0.029	0.030	0.035	0.054

Sect. 2 by using a stepwise approach modified for the FH model as in Marhuenda et al. (2014). The sample size adjusted for the design effect, which is needed for the arc-sin transformation, has been computed following Kish (1965).

Finally, the root mean squared error for the arc-sin small area model presented in this section has been obtained by a bootstrap technique as in Kreutzmann et al. (2019).

In Table 2, we show the selected variables for the three arc-sin FH models and their estimated coefficients. The interpretation of the coefficients is not straight-forward because of the arc-sin transformation. However, the coefficients are similar to each other among the models, indicating that the auxiliary variables act similarly on the responses. We decided to use in the model auxiliary variables with a significance at most around 10%.

Model diagnostics about the distribution of residuals in the three models are shown in Fig. 1. For each of the three models, a quantile-quantile plot (QQplot) for the sampling errors and for the random effects is shown. The normality assumption can be reasonable for the sampling errors, while it is not for the random effects. However, using the arc-sin transformation, we obtain a better approximation to normality than that we would have obtained using a linear FH model. Moreover, the arc-sin transformation has the advantage of stabilizing the variance in those areas where the sample size is very small.

In Fig. 2, we plot the estimated standard deviation (SD) of direct estimates against the estimated root mean squared error (RMSE) of HCR small area estimates computed with the three different poverty lines. We can see that in most of the provinces, the small area estimates are more efficient than the direct estimates; more precisely, for the 89% of the provinces, the RMSE of small area estimates is lower than the SD of direct estimates.

We refer to the small area arc-sin bias-corrected HCR estimate in province i obtained without any adjustment as $HCR_i^{FH,na}$, to that obtained using T_i^M thresholds as $HCR_i^{FH,M}$ and to that obtained using the T_i^{Q20} thresholds as $HCR_i^{FH,Q20}$.

Table 3 shows the distribution among provinces of the estimated root mean squared error (RMSE) of direct and small-area estimates of the poverty index. We compare the root mean square errors given that the Handbook on Precision Requirements and Variance Estimation for ESS Households (Eurostat 2013) discourages the use of the coefficient of variation to compare the efficiency of
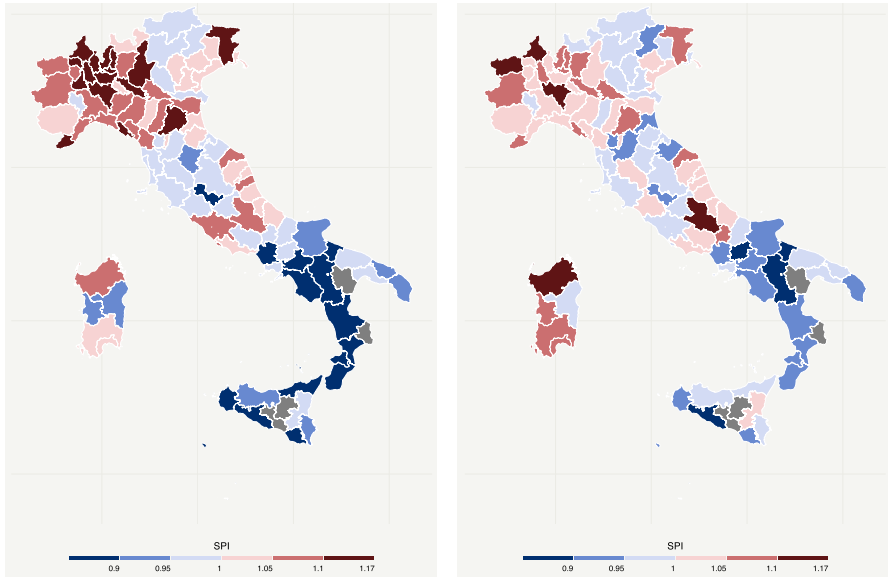
Fig. 3 Spatial price indices: SPI_i^M (left) and SPI_i^{Q20} (right)

proportion estimates as the poverty index. We can see from the table that the gain in efficiency for all the model-based estimates is evident over the direct estimates. In particular, the reduction of the RMSE for small-area estimates from the direct estimates increases as the RMSE of the direct estimates increases. From Table 3, we can see that the RMSE of some direct estimates is equal to 0 (min value of 0). When the RMSE of a direct estimate is equal to 0, then no poor households have been surveyed in that province. However, it is unlikely that a province has no poor households, and this tendency is corrected by the small area estimates, which results in a poverty index different from 0 and thus a RMSE different from 0.

As already specified, in each province, the weight λ_i—the share of food consumption—used to adjust the poverty threshold is area-specific and conditional to poor people. As these weights were estimated using HBS data, direct estimation often led to unreliable estimates, even if we observed a very low variability among the share of food consumption expenditure values. In any case, for these values, we also used a small-area arc-sin model to improve the efficiency of the direct estimates, selecting

Table 4 Small area estimates of the share of food consumption expenditure at province level summarised by Italian partitions	Partition	Min	Max	Median
	North	0.151	0.341	0.220
	Centre	0.203	0.414	0.262
	South	0.209	0.336	0.273

Fig. 4 HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates mean adjusted ($HCR_i^{FH,M}$) (**a**) and HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates Q20 adjusted ($HCR_i^{FH,Q20}$) (**b**) - northern Italy
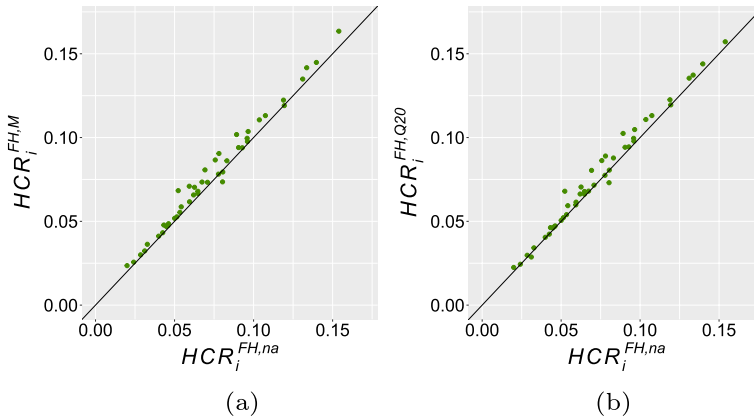

Fig. 5 HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates mean adjusted ($HCR_i^{FH,M}$) (**a**) and HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates Q20 adjusted ($HCR_i^{FH,Q20}$) (**b**) - Central Italy

the auxiliary variables using the same stepwise selection method and the same data sources used for HCR_i. Additionally, in this case, the root mean square errors were computed by applying a bootstrap technique as in Kreutzmann et al. (2019). More details about the application of the small-area arcsin transformed model for the estimation of HCR_i and λ_i are available from the authors upon request.
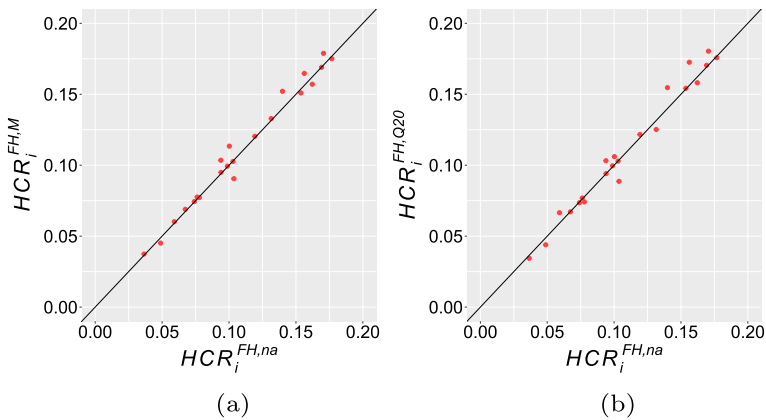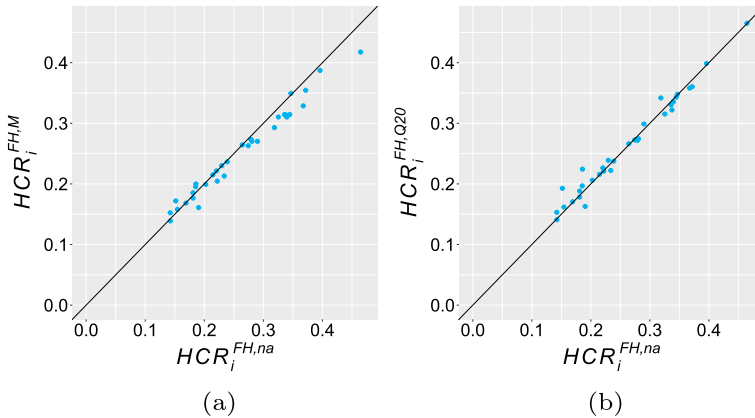
Fig. 6 HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates mean adjusted ($HCR_i^{FH,M}$) (**a**) and HCR SAE estimates not adjusted ($HCR_i^{FH,na}$) vs HCR SAE estimates Q20 adjusted ($HCR_i^{FH,Q20}$) (**b**) - southern Italy

5 Results and discussion

The SPIs computed according to the methodology described in Sect. 3 reveal different price levels for food and nonalcoholic beverages within Italy. Specifically, for both the SPIs—the poverty-specific SPI_i^{Q20} and the general ones SPI_i^M—provinces from southern Italy show lower values than provinces in northern and central Italy, with exceptions. The largest change between the SPI_i^M and SPI_i^{Q20} is observed for the provinces of Vibo Valentia in Calabria, Trapani in Sicilia and Oristano in Sardegna; these three provinces are located in southern Italy and the main islands.

The territorial differences of the SPIs across provinces are easily appreciated by examining Fig. 3, where we map the SPI_i^M (on the left) and SPI_i^{Q20} (on the right). Indeed, as stated before, it is evident that for the SPI_i^M, the higher values are in the provinces of Lombardia and Piemonte in the northwest, and the lower values are in the provinces of Calabria, Sicilia and Campania (south of Italy). For the poverty-specific SPI_i^{Q20}, the territorial differences seem to be attenuated. Even if the highest values are in northern Italy, some exceptions are evident, such as the provinces of L'Aquila in Abruzzo and Sassari and Cagliari in Sardegna (south Italy and islands). It is worth emphasizing that, as described in Sect. 2, the scanner data in this work come from supermarkets and hypermarkets, while discount stores are excluded. This could result in a smoothed effect on the computed SPIs, especially for SPI_i^{Q20}. The SPI values at the provincial level together with their statistical significance are shown in Table 7 in Appendix A.

Using the SPIs, we then adjusted the national poverty thresholds, as specified in Equation (4), so that the HCR estimates take into account the different price levels within the country. The adjustment has been weighted according to the estimates of the share of food consumption expenditure of poor people at the provincial level, as described in Sect. 4. These estimates are summarized in Table 4, where we can

Table 5 Summary of the HCR small area estimates by partitions

Partition	Min	'1st Qu'	Median	Mean	'3rd Qu'	Max
$HCR_i^{FH,na}$						
North	0.020	0.050	0.067	0.073	0.093	0.154
Centre	0.037	0.077	0.102	0.110	0.150	0.177
South	0.142	0.186	0.239	0.260	0.336	0.465
$HCR_i^{FH,M}$						
North	0.024	0.052	0.073	0.077	0.098	0.163
Centre	0.037	0.077	0.103	0.111	0.152	0.179
South	0.139	0.195	0.236	0.250	0.310	0.418
$HCR_i^{FH,Q20}$						
North	0.023	0.050	0.069	0.076	0.098	0.157
Centre	0.033	0.074	0.103	0.111	0.155	0.180
South	0.141	0.210	0.242	0.264	0.331	0.467

see that the share of food consumption expenditure for poor people varies between 0.151 and 0.414. However, the median is 0.22 in the north, 0.262 in the centre and 0.273 in the south. Overall, we adjust 24.5% of the poverty line (T^{na}).

Figures 4, 5 and 6 show the comparison of the SAE HCR_i values obtained using the unadjusted (national) poverty line (T^{na}) and the price-adjusted poverty lines—according to Equation (4), T_i^M and T_i^{Q20}—referring to Italian provinces broken down by partition: north (both northwest and northeast), centre and south.

Starting from northern Italy (Fig. 4), we can see that in most of the provinces, the HCR increases when we use the adjusted poverty lines, T_i^Ms and T_i^{Q20}s. The unadjusted HCR ($HCR_i^{FH,na}$) for the provinces in the north varies between 0.02 and 0.154, with a median HCR of 0.0674. Looking at the HCR obtained with the mean-adjusted poverty lines T_i^M, i.e., $HCR_i^{FH,M}$, the HCRs vary between 0.0236 and 0.163 with a median of 0.0733, with very similar values for the HCRs obtained using the T_i^{Q20}s poverty-specific poverty lines, $HCR_i^{FH,Q20}$.

Regarding central Italy, Fig. 5, the HCRs do not show an identifiable behaviour as in the north. Indeed, in 14 provinces, $HCR_i^{FH,M}$ increases with respect to HCR_i^{na}, while in 8 provinces, it decreases. The $HCR_i^{FH,na}$ among the central provinces varies between 0.037 and 0.177 with a median of 0.102. $HCR_i^{FH,M}$ and $HCR_i^{FH,Q20}$ show very similar values. We conclude that the HCR values in the central provinces are very slightly affected by the price levels. Indeed, only 3 provinces out of 22 in central Italy have an SPI_i^M significantly different from 1, namely, Lucca (Toscana, greater than 1), Terni (Umbria, lesser than 1) and Roma (Lazio, greater than 1), and only 7 of 22 provinces have an SPI_i^{Q20} significantly different from 1.

Looking at the southern provinces, the HCR is higher than in the central and northern provinces. The HCR_i^{na} varies between 0.142 and 0.465 with a median value of 0.239. As we can see from Fig. 6a, when we adjust the poverty line using mean prices, T_i^M, the HCRs decrease in most of the provinces. In particular, $HCR_i^{FH,M}$ is
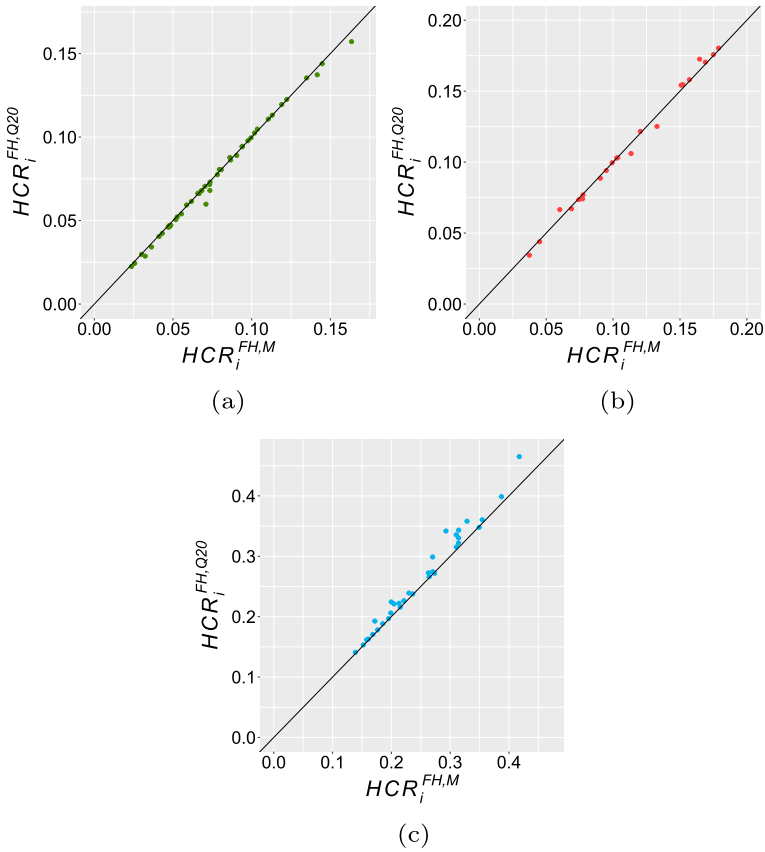
Fig. 7 HCR SAE estimates mean adjusted ($HCR_i^{FH,M}$) vs Q20 adjusted ($HCR_i^{FH,Q20}$). Northern provinces (**a**); central provinces (**b**); southern provinces (**c**)

smaller than $HCR_i^{FH,na}$ in 22 out of 33 provinces. This means that considering the mean price levels, we have a decrease in relative poverty values in southern Italy. When we take into account the 20th percentile of the prices for adjusting the poverty line, T_i^{Q20}, then the effect on the HCR changes, as shown in Fig. 6b. Indeed, the $HCR_i^{FM,Q20}$ values are smaller than the $HCR_i^{FH,na}$ values in only 14 provinces out of 33, while they are higher in the remaining 19 provinces. This means that considering the lower tail of the prices, there is a swinging effect on the HCRs in the southern provinces, with a null effect on average.

Table 5 shows the distribution of the estimated HCRs over provinces by partitions (northern, central and southern Italy), grouped by the type of adjustment. Overall, the HCRs are lower in the northern provinces, increase slightly in the central provinces and are quite higher in the southern provinces. These results are expected given the economic north–south divide characterizing Italy. Adjusting the poverty threshold using the mean prices, the HCRs increase in the northern provinces, swing in the central provinces and decrease in the southern provinces. When we adjust the

poverty threshold using the lower tail of the price distribution, the HCRs increase in the northern provinces, while they swing in the central and southern provinces. Nevertheless, in both cases, the adjustments are small, particularly for the $HCR_i^{FM,Q20}$. The effects on the HCRs are also small because the poverty threshold is adjusted partially, according to the incidence of food consumption expenditure on total consumption expenditure (24.5% on average, for poor people).

In Fig. 7, we compare the $HCR_i^{FH,M}$ and $HCR_i^{FH,Q20}$ values. The HCRs in the southern provinces show higher values when obtained with the T_i^M thresholds than when obtained with the T_i^{Q20} thresholds, except for two provinces. Provinces located in northern and central Italy show swinging HCR values when obtained using T_i^{Q20} or T_i^M thresholds. Therefore, the effect of using the lower tail of the price distribution instead of its mean affects mainly the provinces in southern Italy, smoothing the decreasing effect observed between $HCR_i^{FH,na}$ and $HCR_i^{FH,M}$.

6 Conclusions, limitations and future directions

In this work, we presented a novel approach that included the territorial price differences to estimate the relative poverty incidence in Italian provinces. This approach, based on the use of scanner data from the retail distribution and survey data from the Household Budget Survey (HBS), and resorting to small-area estimation (SAE) models, highlighted the following main findings.

When using the traditional national poverty line, the HCR is generally higher in southern provinces than in central and northern provinces. This is a well-known result. However, it does not take into account the fact that price levels can be different within a country. Indeed, we showed that prices computed using scanner data referring to grocery products (excluding fresh food and alcoholics) are usually lower in southern Italy. This result was obtained by relaxing the classical like-to-like product comparison approach under the hypothesis that people are prepared to change their choice of a given product (e.g., of a specific brand) if a similar product is available.

When including the proposed Spatial Price Indices (SPIs) in the computation of the HCRs, the national poverty line is substituted by province-specific poverty lines that include the different price levels in the estimation of the poverty incidence. In this work, we proposed to adjust the national poverty line only partially, considering the average share of food consumption expenditure over the total consumption for (relatively) poor people in each province.

Our results show that when adjusting the national poverty line using the SPIs referring to the mean provincial prices, $SPIs^M$, the HCR values are higher in the northern provinces and lower in the southern provinces, while no main differences are observed in central provinces. We also considered an alternative set of SPIs based on the 20th percentile of the price distribution, $SPIs^{Q20}$. The idea under this proposal is to adjust the national poverty line using the prices corresponding to the purchasing power of poor households. In this case, the modifications of the HCR

Table 6 Sample size: number of strata, number of outlets and coverage in terms of turnover at NUTS2 level - scanner data year 2018. Regional codes: 01 Piemonte, 02 Valle d'Aosta, 03 Lombardia, 04 Trentino A.A., 05 Veneto, 06 Friuli Venezia Giulia, 07 Liguria, 08 Emilia Romagna, 09 Toscana, 10 Umbria, 11 Marche, 12 Lazio, 13 Abruzzi, 14 Molise, 15 Campania, 16 Puglie, 17 Basilicata, 18 Calabria, 19 Sicilia, 20 Sardegna

Region	N. strata	N. outlets	% markets
01	78	152	95.9
02	4	6	76
03	148	286	99.8
04	12	35	94.8
05	85	161	99.1
06	45	77	87
07	31	62	94.3
08	84	163	98.5
09	65	142	99.9
10	16	32	92.2
11	43	83	97.4
12	38	106	95.3
13	34	38	92.9
14	11	18	96.6
15	35	88	77.5
16	34	85	91.3
17	6	10	67.9
18	25	51	87.2
19	44	111	86.4
20	29	55	96
IT	867	1781	93.7

values are less relevant: the HCR values are slightly higher in the northern provinces, while the values swing in the central and southern provinces.

These results support the idea that for given categories of products, such as grocery products, the price levels within the Italian territory are indeed different. Therefore, including these differentials in the computation of a monetary poverty index, such as the HCR based on HBS consumption data, can modify the poverty mapping results from the traditional national poverty line that ignores price differences. This is an important result that could be considered by policy-makers, especially for policies that are implemented at the national level.

From a modelling perspective, the HCRs and shares of food consumption expenditures (for poor people) were computed at the provincial level using SAE models for proportions. By using auxiliary variables from population registers and archives, the models were able to reduce the direct estimate variability computed using only HBS data. Therefore, the current application has shown that SAE models represent a powerful tool to drive local policy decisions.

In any case, it is important to emphasize that the present work also has some limitations. The main limitation is represented by the partial correction of the national poverty line, which is limited to the share of food consumption expenditure (approximately equal to 25% of total expenditures for poor people at the national level). It would be important to extend the current study by including the correction for all the expenditure categories. Preliminary work based on HBS data (Bertarelli et al.

Table 7 Spatial price indices: SPI_i^M, SPI_i^{Q20}

Province	SPI_i^M	p-val.	SPI_i^{Q20}	p-val.	Region	Partition
Torino	1.08	0.0125	1.05	0.0064	Piemonte	Northwest
Novara	1.14	0.0001	1.09	0.0000	Piemonte	Northwest
Cuneo	1.01	0.6622	1.01	0.4436	Piemonte	Northwest
Asti	0.96	0.2410	0.96	0.0267	Piemonte	Northwest
Alessandria	1.08	0.0151	1.05	0.0197	Piemonte	Northwest
Biella	1.09	0.0081	1.01	0.5296	Piemonte	Northwest
Valle d'Aosta	1.10	0.0049	1.16	0.0000	Valle d'Aosta	Northwest
Imperia	1.12	0.0003	1.08	0.0001	Liguria	Northwest
Savona	1.06	0.0806	1.03	0.0747	Liguria	Northwest
Genova	1.06	0.0758	1.02	0.4156	Liguria	Northwest
La Spezia	1.12	0.0008	1.10	0.0000	Liguria	Northwest
Varese	1.13	0.0001	1.04	0.0426	Lombardia	Northwest
Como	1.15	0.0000	1.10	0.0000	Lombardia	Northwest
Sondrio	1.01	0.8708	0.97	0.0799	Lombardia	Northwest
Milano	1.12	0.0005	1.04	0.0437	Lombardia	Northwest
Bergamo	1.07	0.0275	1.05	0.0068	Lombardia	Northwest
Brescia	1.13	0.0002	1.05	0.0118	Lombardia	Northwest
Pavia	1.15	0.0000	1.10	0.0000	Lombardia	Northwest
Cremona	1.10	0.0029	1.08	0.0001	Lombardia	Northwest
Mantova	1.07	0.0375	1.05	0.0071	Lombardia	Northwest
Lecco	1.11	0.0021	1.08	0.0001	Lombardia	Northwest
Lodi	1.04	0.1839	1.01	0.7235	Lombardia	Northwest
Monza Brianza	1.13	0.0001	1.08	0.0001	Lombardia	Northwest
Bolzano	0.97	0.4257	0.95	0.0100	Trentino - A.A	Northeast
Trento	0.97	0.3553	0.95	0.0142	Trentino - A.A	Northeast
Verona	0.97	0.2755	0.97	0.1352	Veneto	Northeast
Vicenza	1.00	0.9368	0.99	0.7374	Veneto	Northeast
Belluno	0.99	0.8580	0.94	0.0016	Veneto	Northeast
Treviso	1.05	0.1613	1.02	0.2287	Veneto	Northeast
Venezia	1.05	0.1680	1.01	0.5866	Veneto	Northeast
Padova	1.01	0.8720	0.97	0.1629	Veneto	Northeast
Rovigo	0.98	0.6192	1.00	0.8364	Veneto	Northeast
Udine	1.12	0.0004	1.07	0.0005	Friuli - V.G	Northeast
Gorizia	1.04	0.2167	1.00	0.9914	Friuli - V.G	Northeast
Trieste	1.02	0.6102	1.01	0.7323	Friuli - V.G	Northeast
Pordenone	1.02	0.5334	0.96	0.0249	Friuli - V.G	Northeast
Piacenza	1.06	0.0962	1.05	0.0162	Emilia Romagna	Northeast
Parma	1.07	0.0430	1.04	0.0219	Emilia Romagna	Northeast
Reggio Emilia	1.04	0.2311	1.00	0.9562	Emilia Romagna	Northeast
Modena	1.10	0.0051	1.01	0.6072	Emilia Romagna	Northeast
Bologna	1.12	0.0004	1.05	0.0096	Emilia Romagna	Northeast
Ferrara	1.06	0.0991	1.04	0.0226	Emilia Romagna	Northeast

Table 7 (continued)

Province	SPI_i^M	p-val.	SPI_i^{Q20}	p-val.	Region	Partition
Ravenna	1.03	0.4141	0.93	0.0001	Emilia Romagna	Northeast
Forli Cesena	1.02	0.5957	0.95	0.0150	Emilia Romagna	Northeast
Rimini	0.99	0.6954	0.95	0.0077	Emilia Romagna	Northeast
Pesaro Urbino	0.98	0.6070	0.95	0.0052	Marche	Centre
Ancona	1.06	0.0998	1.08	0.0001	Marche	Centre
Macerata	1.00	0.9091	1.04	0.0695	Marche	Centre
Ascoli Piceno	1.06	0.0565	1.02	0.3950	Marche	Centre
Fermo	1.02	0.5891	1.03	0.0746	Marche	Centre
Massa Carrara	1.06	0.0906	1.03	0.1197	Toscana	Centre
Lucca	1.08	0.0204	1.01	0.5015	Toscana	Centre
Pistoia	0.99	0.7081	0.94	0.0021	Toscana	Centre
Firenze	0.98	0.4663	0.95	0.0047	Toscana	Centre
Livorno	1.00	0.8783	0.99	0.5163	Toscana	Centre
Pisa	0.98	0.4799	0.99	0.5951	Toscana	Centre
Arezzo	0.95	0.1144	1.00	0.8299	Toscana	Centre
Siena	0.98	0.4612	1.01	0.7227	Toscana	Centre
Grosseto	0.99	0.7854	0.98	0.4089	Toscana	Centre
Prato	1.05	0.1662	1.05	0.0166	Toscana	Centre
Perugia	0.96	0.2622	0.96	0.0332	Umbria	Centre
Terni	0.90	0.0010	0.95	0.0036	Umbria	Centre
Viterbo	0.99	0.8431	1.02	0.3768	Lazio	Centre
Rieti	0.95	0.1186	0.98	0.3884	Lazio	Centre
Roma	1.08	0.0175	0.99	0.5332	Lazio	Centre
Latina	1.01	0.8407	1.02	0.2182	Lazio	Centre
Frosinone	0.98	0.6090	1.01	0.6958	Lazio	Centre
Caserta	0.86	0.0000	0.93	0.0002	Campania	South
Benevento	0.95	0.1277	0.87	0.0000	Campania	South
Napoli	0.97	0.3273	0.97	0.0693	Campania	South
Avellino	0.88	0.0001	0.90	0.0000	Campania	South
Salerno	0.83	0.0000	0.90	0.0000	Campania	South
L'Aquila	1.10	0.0038	1.12	0.0000	Abruzzo	South
Teramo	1.02	0.5273	1.03	0.1028	Abruzzo	South
Pescara	1.03	0.3396	1.02	0.2477	Abruzzo	South
Chieti	1.03	0.3014	1.03	0.1211	Abruzzo	South
Campobasso	0.99	0.6430	1.00	0.9686	Molise	South
Isernia	0.98	0.5220	1.06	0.0046	Molise	South
Foggia	0.90	0.0019	0.94	0.0021	Puglia	South
Bari	0.96	0.2290	0.96	0.0336	Puglia	South
Taranto	0.96	0.1776	0.96	0.0170	Puglia	South
Brindisi	0.94	0.0790	0.95	0.0072	Puglia	South
Lecce	0.92	0.0125	0.90	0.0000	Puglia	South
Barletta Andria Trani	0.88	0.0001	0.89	0.0000	Puglia	South

Table 7 (continued)

Province	SPI$_i^M$	p-val.	SPI$_i^{Q20}$	p-val.	Region	Partition
Potenza	0.83	0.0000	0.88	0.0000	Basilicata	South
Cosenza	0.86	0.0000	0.91	0.0000	Calabria	South
Catanzaro	0.82	0.0000	0.91	0.0000	Calabria	South
Reggio Calabria	0.85	0.0000	0.93	0.0001	Calabria	South
Vibo Valentia	0.74	0.0000	0.94	0.0004	Calabria	South
Trapani	0.79	0.0000	0.93	0.0004	Sicilia	South
Palermo	0.90	0.0016	0.98	0.2634	Sicilia	South
Messina	0.90	0.0011	0.97	0.1140	Sicilia	South
Agrigento	0.73	0.0000	0.81	0.0000	Sicilia	South
Catania	0.98	0.6112	1.00	0.8296	Sicilia	South
Ragusa	0.88	0.0001	0.93	0.0001	Sicilia	South
Siracusa	0.92	0.0147	1.00	0.8795	Sicilia	South
Sassari	1.09	0.0063	1.12	0.0000	Sardegna	South
Nuoro	0.92	0.0163	0.99	0.5898	Sardegna	South
Cagliari	1.05	0.1499	1.10	0.0000	Sardegna	South
Oristano	0.95	0.0996	1.09	0.0000	Sardegna	South

2021; Marchetti et al. 2020) shows that the cost of housing represents approximately another 20% of total household expenditures in Italy. Therefore, provincial spatial price indices referring to this expenditure category could be estimated and used together with the SPIs referring to food expenditure to further adjust the national poverty line. However, estimation of the spatial housing price indices using survey data (e.g., HBS) is still an open issue because for households that own their house, imputed rents are available, and these are already estimated by Istat using hedonic regression (Bertarelli et al. 2021). Therefore, alternative data sources, such as the archive of real estate quotations maintained by the Italian Tax Agency, should be considered.

Moreover, it should be noted that in our approach, we assume perfect substitutability between goods in the same ECOICOP-8-digit category when applying the CPD model. Further developments could relax this assumption and/or include a measure of the degree of substitutability between goods.

Another limitation is represented by the fact that in 2018, the scanner data provided by Istat did not take into account discount stores, which usually sell products at a lower price. This, together with the fact that scanner data also exclude local street markets, could lead to improper measurements of the spending levels of the poorest individuals in the population. Therefore, future developments of the current work could be devoted to a refinement in the computation of the local SPIs, provided that new scanner data–which are not public but collected by a private company–will be made available.

Table 8 Small area estimates of the Head Count Ratio at the province level in Italy, by threshold: national ($HCR_i^{FH,na}$), adjusted using SPI_i^M ($HCR_i^{FH,M}$) and adjusted using SPI_i^{Q20} ($HCR_i^{FH,Q20}$). In parentheses are the estimated root mean squared errors

Province	$HCR_i^{FH,na}$	$HCR_i^{FH,M}$	$HCR_i^{FH,Q20}$
Torino	0.107(0.014)	0.113(0.014)	0.113(0.014)
Novara	0.052(0.023)	0.068(0.024)	0.068(0.024)
Cuneo	0.09(0.019)	0.094(0.02)	0.094(0.02)
Asti	0.053(0.031)	0.055(0.031)	0.054(0.032)
Alessandria	0.065(0.028)	0.067(0.029)	0.066(0.029)
Biella	0.02(0.025)	0.024(0.027)	0.022(0.028)
Valle d'Aosta	0.062(0.011)	0.066(0.011)	0.066(0.011)
Imperia	0.134(0.035)	0.142(0.036)	0.137(0.036)
Savona	0.078(0.029)	0.09(0.031)	0.089(0.031)
Genova	0.14(0.015)	0.145(0.015)	0.144(0.015)
La Spezia	0.119(0.029)	0.122(0.03)	0.123(0.03)
Varese	0.059(0.017)	0.071(0.018)	0.06(0.018)
Como	0.076(0.02)	0.087(0.021)	0.086(0.021)
Sondrio	0.033(0.029)	0.036(0.031)	0.034(0.031)
Milano	0.065(0.01)	0.068(0.011)	0.068(0.011)
Bergamo	0.096(0.017)	0.1(0.018)	0.1(0.018)
Brescia	0.154(0.025)	0.163(0.025)	0.157(0.025)
Pavia	0.097(0.019)	0.104(0.02)	0.105(0.02)
Cremona	0.045(0.027)	0.047(0.028)	0.046(0.029)
Mantova	0.089(0.03)	0.102(0.031)	0.102(0.031)
Lecco	0.043(0.026)	0.048(0.028)	0.046(0.028)
Lodi	0.071(0.033)	0.073(0.034)	0.072(0.035)
Monza Brianza	0.046(0.019)	0.049(0.02)	0.047(0.02)
Bolzano	0.029(0.017)	0.03(0.016)	0.03(0.016)
Trento	0.119(0.017)	0.119(0.017)	0.119(0.017)
Verona	0.08(0.017)	0.074(0.017)	0.073(0.017)
Vicenza	0.069(0.019)	0.081(0.019)	0.08(0.019)
Belluno	0.024(0.025)	0.026(0.026)	0.024(0.027)
Treviso	0.104(0.019)	0.111(0.019)	0.111(0.019)
Venezia	0.078(0.019)	0.078(0.019)	0.077(0.019)
Padova	0.05(0.019)	0.052(0.019)	0.051(0.019)
Rovigo	0.083(0.022)	0.086(0.023)	0.088(0.024)
Udine	0.131(0.015)	0.135(0.016)	0.135(0.016)
Gorizia	0.081(0.03)	0.079(0.029)	0.081(0.03)
Trieste	0.063(0.023)	0.07(0.023)	0.07(0.023)
Pordenone	0.093(0.028)	0.094(0.028)	0.094(0.029)
Piacenza	0.046(0.024)	0.047(0.024)	0.047(0.025)
Parma	0.04(0.017)	0.041(0.017)	0.04(0.017)
Reggio Emilia	0.096(0.023)	0.098(0.024)	0.098(0.024)
Modena	0.051(0.018)	0.053(0.018)	0.052(0.018)
Bologna	0.059(0.014)	0.062(0.015)	0.061(0.015)
Ferrara	0.054(0.021)	0.059(0.021)	0.059(0.022)
Ravenna	0.067(0.02)	0.073(0.021)	0.068(0.021)

Table 8 (continued)

Province	$HCR_i^{FH,na}$	$HCR_i^{FH,M}$	$HCR_i^{FH,Q20}$
Forli Cesena	0.043(0.022)	0.043(0.022)	0.042(0.022)
Rimini	0.031(0.029)	0.032(0.029)	0.029(0.029)
Pesaro Urbino	0.074(0.025)	0.074(0.025)	0.074(0.025)
Ancona	0.156(0.026)	0.165(0.026)	0.173(0.026)
Macerata	0.119(0.031)	0.12(0.031)	0.122(0.032)
Ascoli Piceno	0.103(0.038)	0.103(0.038)	0.103(0.04)
Fermo	0.099(0.03)	0.099(0.031)	0.1(0.031)
Massa Carrara	0.14(0.037)	0.152(0.038)	0.155(0.039)
Lucca	0.094(0.03)	0.103(0.031)	0.103(0.031)
Pistoia	0.132(0.023)	0.133(0.023)	0.125(0.023)
Firenze	0.049(0.019)	0.045(0.02)	0.044(0.02)
Livorno	0.104(0.033)	0.09(0.033)	0.089(0.033)
Pisa	0.067(0.025)	0.069(0.026)	0.067(0.026)
Arezzo	0.059(0.02)	0.06(0.02)	0.067(0.021)
Siena	0.076(0.032)	0.078(0.033)	0.077(0.034)
Grosseto	0.037(0.034)	0.037(0.034)	0.034(0.035)
Prato	0.078(0.031)	0.077(0.031)	0.074(0.031)
Perugia	0.162(0.021)	0.157(0.021)	0.158(0.021)
Terni	0.169(0.027)	0.169(0.027)	0.17(0.027)
Viterbo	0.094(0.029)	0.095(0.029)	0.094(0.03)
Rieti	0.154(0.049)	0.151(0.049)	0.154(0.051)
Roma	0.1(0.011)	0.113(0.011)	0.106(0.011)
Latina	0.177(0.033)	0.175(0.032)	0.176(0.033)
Frosinone	0.171(0.033)	0.179(0.033)	0.18(0.033)
Caserta	0.274(0.035)	0.263(0.034)	0.273(0.035)
Benevento	0.203(0.045)	0.199(0.045)	0.206(0.046)
Napoli	0.347(0.023)	0.349(0.023)	0.348(0.023)
Avellino	0.19(0.035)	0.161(0.034)	0.163(0.035)
Salerno	0.222(0.038)	0.205(0.038)	0.221(0.039)
L'Aquila	0.154(0.031)	0.158(0.031)	0.162(0.032)
Teramo	0.142(0.042)	0.139(0.042)	0.141(0.043)
Pescara	0.169(0.043)	0.168(0.043)	0.17(0.044)
Chieti	0.185(0.026)	0.195(0.025)	0.197(0.026)
Campobasso	0.181(0.024)	0.185(0.024)	0.188(0.025)
Isernia	0.152(0.033)	0.172(0.034)	0.193(0.035)
Foggia	0.325(0.033)	0.31(0.033)	0.315(0.033)
Bari	0.215(0.026)	0.215(0.026)	0.216(0.026)
Taranto	0.239(0.031)	0.236(0.031)	0.238(0.032)
Brindisi	0.181(0.036)	0.177(0.036)	0.178(0.037)
Lecce	0.279(0.039)	0.273(0.038)	0.272(0.039)
Barletta Andria Trani	0.281(0.036)	0.27(0.036)	0.274(0.036)
Potenza	0.234(0.023)	0.213(0.023)	0.222(0.024)
Cosenza	0.372(0.033)	0.355(0.032)	0.36(0.033)

Table 8 (continued)

Province	$HCR_i^{FH,na}$	$HCR_i^{FH,M}$	$HCR_i^{FH,Q20}$
Catanzaro	0.345(0.036)	0.315(0.036)	0.343(0.037)
Reggio Calabria	0.465(0.048)	0.418(0.047)	0.465(0.048)
Vibo Valentia	0.34(0.049)	0.31(0.048)	0.335(0.049)
Trapani	0.368(0.047)	0.329(0.046)	0.358(0.048)
Palermo	0.29(0.036)	0.27(0.036)	0.299(0.036)
Messina	0.319(0.045)	0.293(0.044)	0.342(0.046)
Agrigento	0.337(0.045)	0.314(0.045)	0.322(0.046)
Catania	0.264(0.035)	0.264(0.034)	0.266(0.035)
Ragusa	0.336(0.047)	0.314(0.046)	0.33(0.048)
Siracusa	0.396(0.049)	0.387(0.049)	0.399(0.05)
Sassari	0.22(0.036)	0.221(0.036)	0.226(0.037)
Nuoro	0.23(0.05)	0.23(0.05)	0.239(0.053)
Cagliari	0.142(0.032)	0.152(0.032)	0.153(0.033)
Oristano	0.186(0.041)	0.2(0.041)	0.224(0.043)

Appendix 1

See Tables 6, , 7 and 8.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Bertarelli G, Biggeri L, Giusti C, et al (2021) Methodological paper on intra-country comparisons of poverty rates. Tech. rep., InGRID-2 EU-Project, Deliverable 9.9., [Online. Available at: https://www.inclusivegrowth.eu/project-output. Accessed 30 June 2023.]

Biggeri L, Pratesi M (2022) Estimation of local spatial price indices using scanner data: methods and experiments applied also for assessing poor-specific indices. Rev Off Stat 2:9–42

Carbonaro G (1985) Commissione di indagine sulla povertà e sull'emarginazione, primo rapporto sulla povertà in Italia, Istituto Poligrafico e Zecca dello Stato, chap Nota sulle scale di equivalenza

Casas-Cordero Valencia C, Encina J, Lahiri P (2016) Analysis of Poverty Data by Small Area Estimation, Wiley and Sons, chap Poverty Mapping for the Chilean Comunas

Eurostat (2013) Handbook on precision requirements and variance estimation for ESS households surveys. Eurostat, 2013th edn

Fay R, Herriot R (1979) Estimation of icome from small places: an application of james-stein procedures to census data. J Am Stat Assoc 74:269–77

Guadarrama M, Molina I, Rao J (2016) A comparison of small area estimation methods for poverty mapping. Stat Transit New Ser 17(1):41–66

Hadam S, Würz N, Kreutzmann AK (2020) Estimating regional unemployment with mobile network data for Functional Urban Areas in Germany. Refubium Freie Universitat Berlin Repository. https://doi.org/10.17169/refubium-26791

Istat (2019) Le statistiche dell'Istat sulla povertà: anno 2018. Istat

Istat (2022) Calcolo della soglia di povertà assoluta. https://www.istat.it/it/dati-analisi-e-prodotti/contenuti-interattivi/soglia-di-povert%C3%A0, Accessed 30 June 2023

Jiang J, Lahiri P, Wan SM, et al (2001) Jackknifing in the fay-herriot model with an example. In: Proceedings of the Sem Funding Opportunity in Survey Research pp 75–97

Kish L (1965) Survey Sampling. Wiley, New Jersey

Kreutzmann AK, Pannier S, Rojas-Perilla N, et al (2019) The r package emdi for estimating and mapping regionally disaggregated indicators. J Stat Softw 91

Laureti T, Polidoro F (2022) Using scanner data for computing consumer spatial price indexes at regional level: an empirical application for grocery products in italy. J Off Stat 38(1):23–56

Laureti T, Rao DP (2018) Measuring spatial price level differences within a country: current status and future developments. Stud Appl Econ 36(1):119–148

Marchetti S, Giusti C, Biggeri L et al (2020) Small area poverty indicators adjusted using local spatial price indices. Book of Short Papers SIS 2020. Pearson, London, pp 341–346

Pratesi M (2016) Analysis of poverty data by small area estimation. Wiley, New Jersey

Pratesi M, Giusti C, Marchetti S, et al (2021) Estimations of local spatial price indices using scanner data and their impact on the measure of poverty incidence. Tech. rep., MAKSWELL Project, Deliverable 3.2

RaiNews (2022) Scuola, Valditara: "Differenziare gli stipendi e aprire ai finanziamenti privati". https://www.rainews.it/articoli/2023/01/scuola-valditara-ministro-istruzione-stipendi-b8b3810e-3f3f-4d5f-9a61-f6c245b78cb0.html, Accessed on 30 June 2023

Rao J, Molina I (2015) Small Area Estimation. Wiley Series in Survey Methodology. Wiley, New Jersey

Saraceno C (2021) Reddito di cittadinanza: quale soluzione per lo squilibrio nord-sud. https://lavoce.info/archives/91163/reddito-di-cittadinanza-quale-soluzione-per-lo-squilibrio-nord-sud/, Accessed on 30 June 2023

Schmid T, Bruckschen F, Salvati N, et al (2017) Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in senegal. J. R Stat Soc Ser A (Statistics in Society) pp 1163–1190

Sugasawa S, Kubokawa T (2017) Transforming response values in small area prediction. Comput Stat Data Anal 114:47–60

Suits DB (1984) Dummy variables: Mechanics v. interpretation. Rev Econ Stat 66(1):177–180

Tonutti G, Bertarelli G, Giusti C et al (2022) Disaggregation of poverty indicators by small area methods for assessing the targeting of the "reddito di cittadinanza" national policy in italy. Socio-Econ Plan Sci 82:101327

Tzvidis N, Zhang LC, Luna A et al (2018) From start to finish: a framework for the production of small area official statistics. J R Stat Soc Ser A Stat Soc 181:927–979

United Nations Development Programme (2022) 2022 Global Multidimensional Poverty Index (MPI): Unpacking deprivation bundles to reduce multidimensional poverty. Tech. rep., United Nations, https://hdr.undp.org/content/2022-global-multidimensional-poverty-index-mpi#/indicies/MPI

United Nations Statistical Commission (2018) Classification of individual consumption according to purpose (coicop) 2018. Tech. rep., United Nations, https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018_-_pre-edited_white_cover_version_-_2018-12-26.pdf, [Online. Accessed 12 October 2023.]

United Nations Statistical Commission (2023) Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development. Tech. rep., United Nations, https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202023%20refinement_Eng.pdf

World Bank (2013) Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP). World Bank

World Bank (2021) Global monitoring database, multidimensional poverty measure. Tech. rep., World Bank, https://www.worldbank.org/en/topic/poverty/brief/multidimensional-poverty-measure

Authors and Affiliations

Stefano Marchetti[1] · **Caterina Giusti[1]** · **Francesco Schirripa Spagnolo[1]** · **Gaia Bertarelli[2]** · **Luigi Biggeri[3]**

✉ Caterina Giusti
caterina.giusti@unipi.it

Stefano Marchetti
stefano.marchetti@unipi.it

Francesco Schirripa Spagnolo
francesco.schirripa@unipi.it

Gaia Bertarelli
gaia.bertarelli@unive.it

Luigi Biggeri
luigi.biggeri@unifi.it

[1] Dipartimento di Economia e Management, Università di Pisa, Via Cosimo Ridolfi, 10, Pisa, Italy

[2] Dipartimento di Economia, Università Ca' Foscari, Campus San Giobbe Cannaregio 873, Venezia, Italy

[3] Dipartimento di Statistica, Informatica, Applicazioni "Giuseppe Parenti" (DiSIA), Università degli Studi di Firenze, Viale Morgagni 59, Firenze, Italy