A model for measuring the accuracy in spatial price statistics using scanner data

Ilaria Benedetti[1], Federico Crescenzi[1], Tiziana Laureti[1]

Abstract

Given the crucial role of Spatial Price Indices (SPIs) for comparing standard of living across a country, there is a need to assess their accuracy on a regular basis. However, despite the importance attached to SPIs, with few exceptions reliability measures are not computed and published both at an international and national level. Focussing on SPIs, this paper aims at suggesting a measure for assessing SPIs accuracy through the adoption of a new approach based on the Jackknife replication technique. To illustrate the potentialities of the suggested approach an empirical application is provided using the 2018 Italian National Institute of Statistics - Istat scanner data on the ten provinces of Toscana for selected groups of products. Our results demonstrate a large spatial heterogeneity between the SPIs of the Toscana provinces and their standard errors.

Keywords: Scanner data, uncertainty, spatial price indices, price statistics, Jackknife replication method.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_3.2022.01

---

1 Ilaria Benedetti (i.benedetti@unitus.it); Federico Crescenzi (federico.crescenzi@unitus.it); Tiziana Laureti (laureti@unitus.it), Università degli Studi della Tuscia/University of Tuscia - Viterbo, Lazio, Italy.

The authors wish to thank Antonella Simone (Istat) for her support and assistance.

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

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

1. Introduction

The popularity and availability of transaction or scanner data for the compilation of Consumer Price Indices (CPIs) have increased over the past twenty years.

Consensus has emerged on the fact that, besides reducing the administrative burden and cost for both National Statistical Institutes (NSIs) and retailers, scanner data may allow the reduction of both sampling and non-sampling errors thanks to the detailed information available for individual products (product characteristics, quantity sold, etc.), the wide coverage both in terms of product groups and territorial areas, and the opportunity to implement superlative index based on product weights within elementary aggregates.

In addition, the availability of new data sources, such as scanner data and web-scraped data, have stimulated research for adopting more developed statistical techniques for constructing CPIs and for assessing their accuracy (Fenwick and Ball, 2001; Smith, 2021). The number of European NSIs using scanner data for CPI computations is steadily increasing and different choices are made regarding the index formula, which should be able to reduce chain drift bias and substitution bias. In the time domain context, in order to deal with the chain drift problem Ivancic *et al.* (2011) proposed the use of multilateral indices based on scanner data following the work of Balk (1981) who first noted that multilateral index methods, originally developed for price comparisons across countries, can be easily adapted to price comparisons across time. Since these pivotal studies, various multilateral methods have been suggested in literature following different approaches with the common characteristic that price indices are constructed simultaneously for the entire sample period as in the case of spatial price comparisons (Diewert and Fox, 2022).

However, in spite of their potential, to the authors' knowledge, only a few NSIs explored the use of scanner data to measure price level differences across region by computing sub-national spatial price indices (SPIs) which are crucial for assessing regional disparities in the distribution of real incomes and supporting regional policy-making (Rokicki and Hewings, 2019). These indicators, also known as sub-national Purchasing Power Parities (PPPs), are particularly important, especially in the case of EU member states where regional economic analyses have become essential due to the implementation of EU Cohesion Policy promoting more balanced and sustainable territorial development.

Several European NSIs have been using scanner data for replacing on-field collected prices needed for international PPP computations in the framework of the OECD-Eurostat Programme. NSIs use scanner data to identify products and collect prices thus increasing the number of products priced (and assessing their representativity using turnover as weights) and expanding the number of cities where prices are collected (not only national capitals). In this way, it is also easier to compute Spatial Adjustment Factors as required by Eurostat, thus obtaining average national prices that are more representative of the whole country.

Although sub-national price comparisons are not as widespread as international price comparisons through the International Comparison Programme (ICP), several NSIs and individual researchers have conducted interesting research studies on the compilation of sub-national SPIs in various countries (Laureti and Rao, 2018). However, systematic attempts to compile sub-national SPIs on a regular basis have been hindered by the labour-intensive analyses required for processing traditional price data, *e.g.* data used for compiling CPIs, and by the costs involved for carrying out ad-hoc surveys for collecting price data. In this context, the use of scanner data is both a challenging yet feasible solution for solving the difficulties NSIs face when making spatial price comparisons worldwide. The Italian case study by Laureti and Polidoro in the recent Guide to the Compilation of Subnational Purchasing Power Parities (Biggeri and Rao, 2021) illustrates how scanner data can be blended with data from other sources (traditional collected prices, administrative and internet data) in the compilation of subnational SPIs covering the total household consumption expenditure.

The use of scanner data, which are often obtained through a probability sample design, may also allow to provide accuracy measures of point estimates of price differences across space. Uncertainty in the SPIs comes not only from the choice of aggregation procedure but also from the dispersion of relative prices. For countries across which relative prices are very different, the sampling of goods matters, and the PPPs and SPIs are much more uncertain (Deaton, 2012). This source of variation induces substantial uncertainty in the spatial price indices. Measures of statistical errors for CPI and SPI have many uses: to inform on the quality for users, to guide CPI and SPI compilers in allocating resources for compilation in the most efficient way, and to detect possible serious errors in the data when output editing.

Yet, despite their importance, no standard errors or reliability measures are computed and published both at an international and national level (Deaton and Aten, 2017; Rao and Hajargasht, 2016; Smith, 2021). In view of the complexity of price index structures and the common use of non-probability sampling in compiling CPIs and SPIs, an integrated approach to variance estimation appears to be problematic. A single formula for measuring the variance of CPIs and SPIs, which captures all sources of sampling errors, may be impossible to find.

However, it is often possible to develop partial measures, in which only the effect of a specific single source of error is quantified. In recent years, there has been a growing concern for the more explicit use of the concepts and tools of statistical inference to produce estimates of official CPIs/SPIs and, especially, to define the targets of the estimates following a framework typical of statistical survey methods also known as sampling approach in index theory (Balk, 2005).

This paper contributes to the advancement of this literature by exploring the issue of evaluating the uncertainty associated to point estimates of sub-national SPIs, using the Italian scanner data for the year 2018 as the reference source. To the authors' knowledge, the evaluation of uncertainty among SPIs has not been explored yet.

The aim of this paper is twofold: firstly, we demonstrate the feasibility of using scanner data for computing SPIs at a detailed territorial level (NUTS-3) and at elementary aggregates by referring to the recent study carried out by Laureti and Polidoro (2022). Secondly, we suggest a framework for computing variance estimates of sub-regional SPIs that uses Jackknife Repeated Replications (JRR) when scanner data are collected following a probabilistic sample design. We also provide an application using a subset of the 2018 Italian National Institute of Statistics (Istat) scanner data where the sample of large-scale retail outlets is selected according to a probabilistic stratified random design from a universe of more than 9,000 outlets stratified by three variables: province (NUTS-3), distribution chain and outlet type (supermarket or hypermarket). In particular, we consider the outlets of the ten provinces of Toscana (one of the Italian regions) and three group of homogeneous products, called basic headings (BHs), namely, Mineral water, Coffee and Pasta.

The remainder of this paper is structured as follows. Since the aim of this paper is not to provide a detailed literature review on research on spatial price comparisons, Section 2 reports the main methods that have been suggested and case studies in which scanner data have been used. In addition, a brief overview of the issue of measuring the accuracy of a consumer price index is provided. A description of Istat scanner data and our dataset is reported in Section 3, while the methodological approach to point and standard error estimates of SPIs are described in Section 4. Section 5 reports our results for Toscana and for each of the BHs considered. Section 6 draws some concluding remarks.

2. Literature Review

2.1 Scanner data and Spatial Consumer Price Indices

In countries characterised by large territorial differences in prices and quality of products and household characteristics, such as Italy, it is essential to calculate sub-national SPIs in order to assess inequality in the distribution of real incomes and consumption expenditures.

In all spatial price comparisons, the concept of PPP[2] is used to measure the price level in one location compared to that in another location; therefore, PPPs are essentially SPIs[3]. At international level, PPPs facilitate cross-country comparisons of Gross Domestic Product (GDP) and its major aggregates as they can be used in converting aggregates into a common currency. Likewise, sub-national PPPs allow for intra-country spatial comparisons and can serve as inputs and/or improve other inputs for estimating key economic indicators produced by countries, such as real regional price comparisons, real income dimensions and poverty estimates.

The process of compiling PPPs (or SPIs) is quite complex and is carried out in two stages. First, elementary spatial price indices are computed by aggregating, generally without using weights, prices of items belonging to a group of similar well-defined product goods or services (called Basic Headings, BHs). In the second stage, the elementary PPPs (or SPIs) are aggregated using expenditure weights to obtain PPPs (or SPIs) for higher-level aggregates such as consumption, investment and GDP.

Some NSIs and many researchers have conducted computations of subnational spatial price indices for household consumption. Indeed, given the limited resources, the goal of producing sub-national PPPs or spatial CPIs at a subnational level is most feasibly achieved using information from the national CPIs.

2 PPP for a given country represents the number of currency units required to buy a similar basket of goods and services in the given country in relation to a reference or base country). At international level, PPPs for countries are compiled by the ICP, administered by the World Bank with the collaboration of the OECD, EUROSTAT and other organisations (World Bank, 2013; Biggeri and Rao, 2021). A description of the framework adopted by the International Comparison Programme (ICP) is presented in Rao (2013) and the full set of ICP procedures are discussed in various chapters of World Bank (2013), "Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Programme (ICP)".

3 In this paper, we will use the term sub-national SPIs instead of PPPs as we refer to spatial comparison using scanner data for grocery products.

Early research on sub-national price comparisons was mostly conducted in the United States in the early 1990s (Kokoski *et al.*, 1999). These pioneering efforts at the Bureau of Labour Statistics and at the Bureau of Economic Analysis, were later continued by Aten (2008) finally leading to the regular compilation of spatial price differences in the United States through the computation of Regional Price Parities (RPPs). The Australian Bureau of Statistics started a research project for compiling SPIs. Similarly, Statistics New Zealand has been evaluating the possibility of carrying out spatial price comparisons of prices since 2005 and two experts have been assigned to develop a methodology for constructing spatial cost of living indices (Melser and Hill, 2007). However, to the authors' knowledge, estimates of sub-national SPIs have never been disseminated (Tam and Clarke 2015).

The General Statistical Office (GSO) of Vietnam started a pilot research project in 2010 to compute subnational PPPs in terms of the SCOLI (Spatial COst of LIving) index, supported by the World Bank, based on CPI data available. The GSO produced and published SCOLI indices for the period 2010 to 2017 (GSO, 2019).

At the beginning of the 2000s, the Italian National Institute of Statistics (Istat) conducted experiments on the use of CPI data to calculate regional consumer price level indices and disseminated results on two occasions: in 2008 with reference to price data from 2006; and in 2010 with reference to 2009 data (Istat, 2008 and 2010).

In France, the National Institute of Statistics and Economic Studies (INSEE) conducted *ad hoc* price surveys in 1985, 1992, 2010, and 2015 and published analyses based on these surveys (INSEE, 2019).

As far as the estimation of sub-national consumer spatial price index numbers using CPI data and/or COICOP classification, numerous attempts have been made by researchers in various countries (Janský and Kolcunová, 2017; Biggeri *et al*, 2017*a*; Chen *et al*, 2019; Rokicki and Hewings, 2019; Weinand and von Auer, 2020; Biggeri *et al*, 2017*b*; Montero *et al.* 2019; Fenwick and O'Donoghue, 2003; Roos, 2006, Kocourek *et al.*, 2016).

Recently, several NSIs around the world have been integrating price data collected from traditional sources in the process of compiling official CPIs with other new sources of data. Based on the experienced gained, for example in

Italy, scanner source of data makes it possible to identify representative and comparable products across subnational areas, resolving the important issue of balancing the two competing requirements of representativity and comparability. Indeed, when compiling SPIs, the first step is the definition of a product list consisting of goods and services that are to be priced which should adequately cover goods and services purchased in different regions of the country.

To this respect, scanner data allow for the construction of fine level SPIs as the volume of information contained in such data makes it more likely for spatial price comparisons to be made on the same product across geographical areas and it allows using information on sales for obtaining weights at individual product level. Indeed, GTIN codes uniquely distinguish products, and they are generally the same for each item at national level. Thanks to the high territorial coverage which characterises scanner data, it is possible to compare price levels at different territorial levels within a country (NUTS-3, NUTS-2 and NUTS-1). In this way, the issue of comparability can be solved.

Moreover, scanner data may allow to provide information on quality characteristics that may influence the price of a product, such as the chain or the type of outlet in which the product is sold. It is also possible to add a time dimension to multilateral spatial price comparisons since detailed data are usually available at the point of sale and usually on a weekly basis. In addition, using transaction data it is possible to account for the economic importance of each item in its market by using data on turnover, thus providing a reliable indicator of the importance of individual products. Finally, as already mentioned, using scanner data to carry out spatial comparisons will result in cost efficiencies since price data collection can then be limited to traditional outlets thus lowering data collection costs for the NSIs.

When it comes to the usage of scanner data in computing SPIs however we notice that scanner data have been mainly used for making international price comparisons (Heravi *et al.*, 2003; Feenstra *et al.*, 2017) leaving its usage for sub-national SPIs unexplored.

A first attempt in this context is the study by Laureti and Polidoro (2018) who estimated Italian SPIs for 2017 using a scanner dataset constructed for experimental CPI computation. Later, Laureti and Polidoro (2022) demonstrated the feasibility of using scanner data for comparing consumer prices at different territorial levels, which are representative of local consumption patterns and comparable based on a set of prices determining characteristics. The encouraging results obtained stimulated further research by Istat to achieve the aim of producing sub-national SPIs for Italy on a regular basis through the adoption of a multi-source approach to cover all the retail trade channels and product baskets.

2.2 Addressing the issue of accuracy in Spatial Consumer Price Indices

The issue of measuring the accuracy of temporal and spatial consumer price indices has a long history and it is still controversial. Indeed, it has not been clear if it is feasible to make such assessments. The underlying population price index, for even the smallest of countries, involves so many transactions on so many individual products in so many places as to be inaccessible. In addition, usually, the universe of individual products on the market is dynamic, thus introducing additional difficulty in the definition of an appropriate sampling design. As a result, most official price index numbers are published without an explicit statement about their accuracy. Therefore, the situation has not changes since Morgenstern's remark on this issue "In spite of the widespread use of government price indices, there has been little done in attempting to determine the error inherent in these indices" (Morgenstern, 1963).

However, during the last years, there has been a growing interest in a more explicit evaluation of the accuracy of CPIs thanks to the availability of new data sources (De Haan *et al.*, 1999; Fenwick *et al*, 2001). However, since CPI and SPI are not in general obtained from a single survey, the sampling and non-sampling errors, being related to all the surveys used for the construction of the index, cannot be easily specified by a single complex model.

The pioneering works on this subject date back to Banerjee (1956) and Adelman (1958), while systematic research on the sampling variance of CPI indices has been developed since the mid-eighties (Balk and Kersten, 1986; Balk, 1989). In the same framework, Biggeri and Giommi (1987) suggested a structure for the classification of errors within a price index based on a breakdown of the mean squared error.

Several approaches have been used to estimate the variance of CPIs (Smith, 2021). Model-based and design-based approaches have both been applied for CPI variance calculation.

More specifically, design-based approaches with Taylor linearisation have been adopted by several authors in the literature. Andersson *et al.* (1987a) used this approach to assess the variance due to sampling outlets in the Swedish CPI, Leaver and Valliant (1995) used this approach in the US context in order to give an approximation of the variance of the Laspeyres index between two time periods. Despite several applications, there is a suggestion that Taylor linearisation may underestimate the empirical variances for smaller sample sizes (see, *e.g.* Andersson *et al.*, 1987b). Moreover, the computation of Taylor linearisation estimator may be extremely complicated, and may require several layers of approximation, whose aggregate effect is not clear without detailed investigation (Smith, 2021).

A replication-based approach is the longest-standing approach in the computation of standard errors for the US CPI (Leaver *et al.*, 1991). The drawback of this method is related to the resampling procedure. Indeed, if the sampling procedure generates small samples, this approach can produce large variances

The Jackknife and the Bootstrap methods have become increasingly popular for CPI variance estimation since they can be used in complex designs. The Jackknife is already in use in the US for special item categories. Leaver and Cage (1997) discussed the implementation of the Jackknife in the US CPI by comparing variance estimates for a series of alternatively aggregated prices. Klick and Shoemaker (2019) used the Jackknife to evaluate significant differences between urban population and other populations (wage earner and clerical worker population, and elderly population) in some cities in the US.

Some authors have used a model-based approach to estimate the price index formula variance. For example, Kott (1984) modelled the variance of the Laspeyres type index where the prices are assumed to be “nearly homogeneous”. The key advantage of this approach is that sampling weights are not required to obtain the estimates, on the contrary, it is more difficult to justify a particular model and to claim objectivity for the variance estimates.

During the last years, various research studies have been carried out for adopting probability sampling design and assessing CPI accuracy by using simulation data and new sources of data (see, for example, De Gregorio, 2012; De Vitiis *et al.*, 2017). De Gregorio (2012) analysed the sample sizes needed to estimate Laspeyres consumer price sub-indices under a combination of alternative sample designs, aggregation methods, and temporal targets using simulated data. The author found that the optimal sample size depends crucially on the degree of relative variability and skewness of items and underlined the crucial role of stratification in saving sample size. De Vitiis *et al.* (2017) studied the properties of alternative aggregation formulas of the elementary price index in different sampling schemes implemented on scanner data. Bias and efficiency of the estimated indices are evaluated through a Monte Carlo simulation.

At official level, only the Swedish NSI provides official variance estimates for CPIs. Official statistics published by Statistics Sweden are disseminated with a quality declaration per survey year and the sampling uncertainty measures are assessed annually for monthly change, annual change (inflation rate), and monthly change in the inflation rate (Norberg and Tongur, 2022).

In this context, scanner data may play a crucial role in stimulating research on sampling errors by improving sampling methods, checking the representativity of the achieved sample, and controlling initial sample selection. Recently, Tongur (2019) focussed on the case of scanner data for daily consumer products and their inclusion in the Sweden CPI, particularly regarding the issue of the trade-off between item related variance and the bias from disregarding explicit quality adjustments. Results show that the contribution to the variance from a randomly sampled item in the daily products survey is rather small and would tend to decrease with appropriate sampling, given that the samples are based on size-proportional sampling strategies. The sample size related variance is estimated through a Jackknife method.

It is worth noting that scanner data may also reduce non-sampling errors. Among these types of errors, it is possible to distinguish between measurement errors, representativeness of items and coverage errors. The firsts are mitigated thanks to the increased number of products priced and to the improved territorial and population coverage. Indeed, prices may be collected in each city across the province and not only in the provincial capital. In contrast, the traditional basket is a relatively small subset of the complete universe of goods while quantities sold are not available. Moreover, the use of unit value (calculated as the total expenditure for that item code divided by the total quantities sold) instead of price, represents a more accurate measure of the actual price of an individual product than an isolated price quotation (Balk,1995).

In addition, by using scanner data, it is possible to consider a wide range of methods for calculating spatial price indices due to the availability of quantities and expenditure information (Heravi *et al.*, 2003) which is usually not collected in traditional surveys. In fact, information on expenditure and quantities allows to calculate indices based on a variety of "superlative" index number formulae, including the Fisher ideal index (see Imai *et al.*, 2015; Laureti and Polidoro, 2018). This may reduce the so-called formula error (Dorfman *et al.*, 2006).

According to the previous literature review, this paper offers an advancement to the literature on spatial price index variance estimation by applying the Jackknife as a variance estimation technique to scanner data. To the authors' knowledge, the use of the Jackknife technique for evaluating uncertainty among spatial price indices computed for geographical areas within a country has not yet been explored.

Although in the context of CPI both design-based and model-based estimators have been used, the Jackknife method has been selected since it can be easily applied to the scanner data sampling design by setting BHs as strata and market classification as Primary Sampling Units (PSU). As a result, Jackknife estimates have an interpretation in terms of the error arising from the sampling processes for prices, and the proposed stratification variables isolate possibly homogeneous product groups and clusters of pricing policies.

3. Italian scanner data for computing Spatial Price Index

In this paper, we use a scanner dataset provided by Istat for the year 2018[4]. Since January 2018, the Italian NSI has introduced in its consumer price production process the use of scanner data from large-scale retail trade/modern distribution chains (hypermarkets and supermarkets) for grocery products (packaged food, household, and personal care goods).

Istat acquires data, through a market research company, for individual outlets of 16 large-scale retail groups in Italy for all 107 provinces of the national territory by type of outlets (hypermarkets and supermarkets).

The sample of large-scale retail trade outlets is representative of the entire universe of large-scale retail trade and includes 1,781 outlets, of which 510 hypermarkets and 1,271 supermarkets distributed throughout the country.

The 16 chains collaborating with Istat represent, at national level, more than 90% of the total turnover of hypermarkets and supermarkets, with a high coverage also at regional level (the highest coverage value is recorded in Toscana with 99.9% and the lowest in Basilicata with 67.9%).

Italian scanner data cover all grocery products for a total of 79 product aggregates, belonging to five ECOICOP[5] divisions (01, 02, 05, 09, 12) and substitute on field price collection. The individual products included in the index calculation are identified by codes (GTINs), which uniquely identify products (items) throughout the country.

The sample of large-scale retail outlets is selected according to a probabilistic stratified random design. The universe, made up of over 9,000 outlets, is stratified by considering three variables: the province (all 107 provinces), the chain or group to which it belongs (16 large-scale retail trade distribution) and the outlet type (supermarket or hypermarket). The sampled outlets are extracted within each of the 888 strata of the universe with probability proportional to the sales turnover of the previous year. Since scanner data do not provide the "shelf price" of the product but quantity sold the purchasing price should be defined as unit value of average weekly price.

4 The scanner data set was provided by Istat and treated in order to respect the statistical confidentiality. The authors worked at Istat to perform the various analyses.

5 European Classification of Individual Consumption according to Purpose (ECOICOP).

In our analysis, the unit value price for each individual product/GTIN is calculated using turnover and quantities sold (unit value = turnover/quantity) of each province obtained as a weighted sum of outlet sampling weights. All the products sold in each sampled outlet are selected within homogeneous groupings of products corresponding to the markets. Probabilities of selection were assigned to each outlet based on the corresponding turnover value. The classification of homogeneous products within markets represents an objective and detectable identification of commodity products shared by industrial and distribution companies. This classification is provided by the Efficient Consumer Response (ECR) community, used for category management by both industrial and distribution companies. In the Italian case, the market classification is made available by the Nielsen company, but in general they could be retrieved through statistical registers of enterprises and local units.

The ERC classification is based on a hierarchical structure. Each of the defined categories is linked to a reference sheet that contains the definition of the category and the criteria for exclusion and inclusion of products. As an example, for the Water BH one of the possible markets is "sparkling water". Classification includes the following description: "Natural water from mineral water sources, with various chemical and therapeutic properties. They have carbon dioxide added and are labelled 'carbonated' or 'sparkling'". Only mineral waters with added carbon dioxide called carbonated or sparkling are included. On the contrary non-carbonated mineral waters (natural or still), natural sparkling waters and lightly or slightly carbonated waters are excluded. The classification reports also some examples of inclusion, *e.g. "Boario gassata", "San Pellegrino frizzante"*, and example of exclusion: *e.g. "Ferrarelle effervescente naturale"*.

Provincial prices are calculated as weighted arithmetic mean according to the probability proportional to size (PPS) sampling design within each stratum.

In our paper, we refer to a portion of this big dataset since we used 2018 data for all the outlets sampled for Toscana region, assembled after the stratification process. To illustrate the potential of the suggested methodology, in this analysis we considered three BHs namely: Mineral water, Coffee, and Pasta. The dataset consists of 8,856 annual price quotes from the ten Toscana provinces, which make up Toscana: Arezzo (AR), Firenze (FI), Grosseto (GR), Livorno (LI), Lucca (LU), Massa-Carrara (MS), Pisa (PI), Prato (PO), Pistoia (PS) and Siena (SI).

4. Methods

4.1 Computation of Spatial Price Index among provinces

We estimate provincial SPIs at BH level for the 10 Toscana provinces. As a first step, we compute within-province SPIs using the Fisher and the Törnqvist formulas, as they are known to be superlative (Diewert, 1976). In particular, the Fisher type formula is used for computing international SPIs also by the Eurostat-OECD (2012).

The Fisher index is an "almost ideal" index since it satisfies all the standard properties and tests except the transitivity test and approximates a cost-of-living index (Diewert, 1976). Instead, the Törnqvist index satisfies the country reversal test, and it is superlative and exact. The two indices are very close to each another due to the similarity in their definitions (Pilat and Rao, 1996). In the second step, we compute a multilateral index by combining bilateral comparisons between pairs of provinces.

When constructing multilateral SPIs is important to satisfy two basic properties: transitivity and base invariance (province invariance). Transitivity requires that the SPI computed between two provinces should be the same whether it is computed directly or indirectly through a third region. Base invariance means that all provinces be treated equally in deriving the matrix of SPIs that satisfy transitivity.

Let us assume that we are attempting to make a spatial comparison of prices between M provinces. At BH level, p_{ij} and q_{ij} represent the price and quantity of the *i-th* item in the *j-th* province with $i=1,2,\ldots,N$ and $j=1,2,\ldots, k,\ldots, M$. In most cases, the price indices for elementary aggregates are calculated without the use of explicit weights. Contrastingly, our scanner data set contains detailed quantity and sales information within an elementary aggregate so that there are no constraints on the type of index number that may be used.

We calculate matched-model indices, *i.e* Laspeyres and Paasche indices, between areas j and k. Since not all items may be priced in all areas included in the comparison, N_{jk} represents the number of products that are priced in both areas j and k. Therefore, N_{jk} is usually smaller than the number of commodities N in the basic heading. If a commodity is not priced in one of the two provinces, that item cannot be included in the SPI computation. Because SPI_{jk}^{P} and SPI_{jk}^{L} use only information on areas j and k from the price tableau, the resulting indices are not transitive.

In a spatial framework, the Laspeyres index measures the change in the fixed basket cost, taking a base province as a reference, where substitution is not considered when there is a change in relative prices (Paredes Araya and Iturra Rivera, 2013).

As known, the Laspeyres and Paasche indices can be calculated in two ways: either as the ratio of two value aggregates or as an arithmetic weighted average of the price ratios for the individual products using the hybrid expenditure shares as weight:

$$SPI_{jk}^{L} = \frac{\sum_{i\in N_{jk}} p_{ik}\cdot q_{ij}}{\sum_{i\in N_{jk}} p_{ij}\cdot q_{ij}} \equiv \sum_{i\in N_{jk}} \left(p_{ik}/p_{ij}\right) s_{ij} \tag{1}$$

$$SPI_{jk}^{P} = \frac{\sum_{i\in N_{jk}} p_{ik}\cdot q_{ik}}{\sum_{i\in N_{jk}} p_{ij}\cdot q_{ik}} \equiv \sum_{i\in N_{jk}} \left[\left(p_{ik}/p_{ij}\right) s_{ik}{}^{-1}\right]^{-1} \tag{2}$$

For each pair of provinces, we calculate the two bilateral SPIs, that is SPI_{jk}^{P} and SPI_{jk}^{L}, using expenditure shares equal to s_{ij} and s_{ik} respectively. These expenditure shares are computed using only the common goods in each pair of provinces. Therefore s_{ij} represents the expenditure share of *i-th* item in the province *j-th*, while s_{ik} represents the expenditure share of *i-th* item in the province *k-th*.

By following this procedure each BH is provided with a matrix of Fisher SPIs. The Fisher price index, which has good axiomatic and economic properties (Balk, 1995), is a geometric average of the Laspeyres and Paasche indices given by:

$$SPI_{jk}^{F} = \sqrt{SPI_{jk}^{L} \cdot SPI_{jk}^{P}} \tag{3}$$

From an economic approach, the Fisher index is preferred as it uses quantities at different times and allows for substitution effects. Since the Fisher SPIs, SPI_{jk}^{F}, are not transitive, the Gini-Éltető-Köves-Szulc (GEKS) methodology is used to obtain a transitive index that deviates the least from a given matrix of binary comparisons. The GEKS SPIs can be obtained as an unweighted geometric average of the linked (or chained) comparison between provinces j and *k* using each province *l* in the comparisons as a link:

$$SPI_{jk}^{GEKS-F} = \prod_{j=1}^{M} \left[SPI_{jl}^{F} \cdot SPI_{lk}^{F} \right]^{1/M} \tag{4}$$

The second method that has been used for computing sub-national SPIs is the Törnqvist spatial price index, that is defined as a geometric average of the price relatives weighted by the average expenditure shares in the two provinces *j* and *k*:

$$SPI_{jk}^{T} = \prod_{j=1}^{M} \left(\frac{p_{ik}}{p_{ij}} \right)^{\frac{s_{ij}+s_{ik}}{2}} \tag{5}$$

Where $\frac{s_{ij}+s_{ik}}{2}$ is the arithmetic average of the share of expenditure on item i in two provinces. Fisher and Törnqvist are superlative indices and show up as being "best" in all the approaches to the index number theory as they satisfied all the axiomatic properties expected of a price index number formula with the exception of the circularity test (Hill, 2004; ILO, 2020). More specifically, they satisfy the base invariance test and commensurability test; they are symmetric indices given that they make equal use of prices and quantities in both the areas compared and treat them in a symmetric manner. Within the economic approach, Diewert (1976) obtained a characterisation of the Törnqvist price index, as being the economic price index that corresponds to a linearly homogeneous translog unit cost or revenue function.

Unfortunately, similarly to Fisher, the Törnqvist SPIs express in equation (5) are not transitive. To obtain a set of transitive SPIs, the bilateral SPIs must be transitivised using the GEKS procedure:

$$SPI_{jk}^{GEKS-T} = \prod_{j=1}^{M} \left[SPI_{jl}^{T} \cdot SPI_{lk}^{T} \right]^{1/M} \tag{6}$$

The Törnqvist version of GEKS is often referred to as the CCD method (Caves *et al.*, 1982). One attractive feature of the CCD method is that it can also be represented as a star method with an artificial country at the centre of the star (Hill and Timmer, 2006).

4.2 Addressing the issue of accuracy in Spatial Price Indices

Originally introduced as a technique of bias reduction, the Jackknife Repeated Replication (JRR) method has by now been widely tested and used for variance estimation (Durbin, 1959). Like other resampling procedures, the JRR method estimates the sampling error from comparisons among sample replications which are generated through repeated resampling of the same parent sample. Each replication needs to be a self-representative sample and to reflect the full complexity of the parent sample. The JRR variance estimates consider the effect on variance of aspects of the estimation process which are allowed to vary from one replication to another. In principle, these can include complex effects such as those of imputation and weighting. The basic JRR model which shall be adopted in this work can be summarised as follows. Consider a design in which two or more primary units have been selected independently from each stratum in the population. As in the case of the linearisation approach, sub-sampling of any complexity may be involved within each PSU, this does not affect the variance computation formulae. In the standard delete one-PSU at a time Jackknife version (Leaver and Cage, 1997), each replication is formed by eliminating one sample PSU from a particular stratum at a time and increasing the weight of the remaining sample PSUs in that stratum appropriately to obtain an alternative but equally valid estimate to that obtained from the full sample. This procedure involves creating as many replications as the number of primary units in the sample.

Let r be a subscript to indicate a sample PSU and let h indicate its stratum; moreover, let $a_h \geq 2$ be the number of PSUs in stratum h, assumed to be selected independently. Let λ be a full sample estimate of any complexity, and $\lambda_{(hr)}$ the estimate obtained after eliminating primary unit r in stratum h and increasing the weight of the remaining $a_h > 1$ units in that stratum. Also, let $\lambda_{(h)}$ be the simple average of the $\lambda_{(hr)}$ over the a_h values of r in h. The variance of λ is then estimated as (Betti *et al.*, 2018):

$$var(\lambda) = \sum_h [(1 - f_h)\left(\frac{a_h - 1}{a_h}\right)\sum_r\left(\lambda_{(hr)} - \lambda_{(h)}\right)^2] \qquad (7)$$

Where $1 - f_h$ is the finite population correction which in typical social surveys is approximately equal to 1.

Under quite general conditions for the application of the procedure, the same and relatively simple variance estimation in Formula (7) holds for λ of any complexity. This in fact is the major attraction of the JRR method for practical application.

De Gregorio (2012) showed that the choice of the strata is of paramount importance, since it involves theoretical and microeconomic issues, including for example market criteria, to isolate possibly homogeneous product groups and clusters of pricing policies. Our dataset considers 8,856 observations, subdivided in 3 Strata (BHs) and 46 PSUs (the markets).

Therefore, by conditioning on each BH, we use the JRR to estimate the uncertainty in SPIs due to the allocation of products in Markets. This is done by substituting the λ-s in Formula (7) with their counterparts illustrated in Formulas (4) and (6) respectively.

5. Results

In our analysis, we consider a total of 8,856 price observations referring to a total of 1,726 unique products. Table 5.1 reports the number of observations, number of unique products and number of markets in each BH.

In Figure 5.1, we compare price variability observed across Toscana provinces in the three BHs included in our analysis. The highest price variability value is reported for Coffee products where the coefficient of variation (CV) is equal to 96%. Contrastingly, Pasta products show similar prices as the value of the CV is equal to 59%.

Considering the variability within each product group, it is worth noting that, among the various provinces, for the Mineral water product group the highest heterogeneity in price levels is observed for Pistoia and Lucca provinces (CV equal to 98% and 97% respectively). The highest values of price variability for Pasta products are observed in Lucca and Firenze (CV equal to 62% and 59% respectively), while for the Coffee BH, Firenze and Pisa show the highest levels of price variability (CV equal to 64% and 63% respectively).

Table 5.1 - Number of observations, GTIN and markets in each stratum

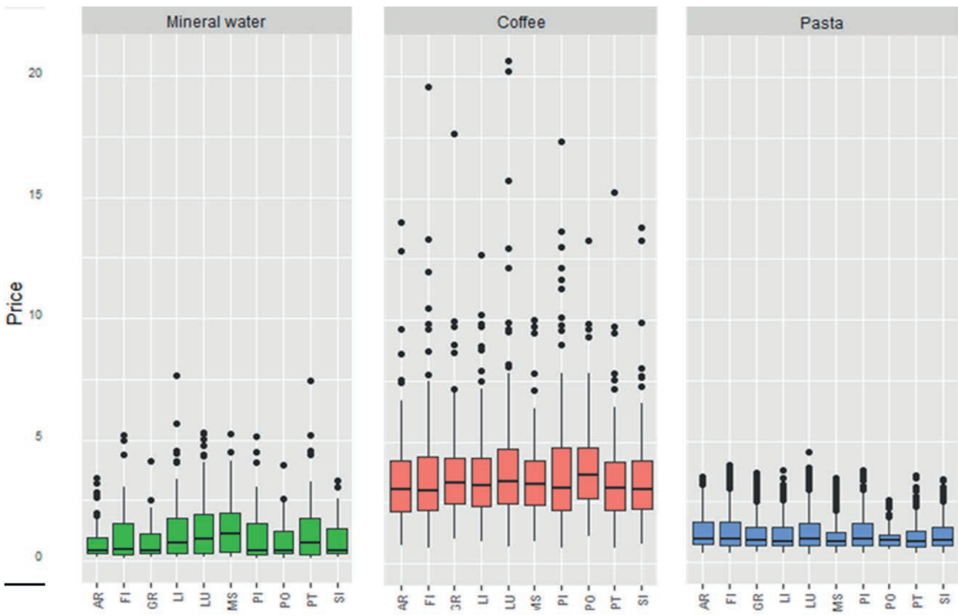
BHs	N. of Observations	N. of GTIN	N. of Markets
Mineral water	2,010	298	13
Coffee	1,184	337	12
Pasta	5,662	1,091	21

Source: Authors' elaboration from Italian Scanner Data

To ensure comparability among products sold in two provinces, individual products are matched by using GTIN codes, which contain elementary information and identify unique products.

For price comparisons at the product level, we identified 1,091 different individual products for Pasta, 337 different products for Coffee and 298 different products for Mineral water that are priced in at least two different provinces. Table 5.2 reports the number of different individual products (GTINs) sold in each province of Toscana by group of products (BH) and outlet type. This Table shows that there is a large variety of products over provinces that can be explained by the fact that the large-scale retail trade distribution is not uniformly distributed across the Toscana territory in terms of types of outlets, retail chains and market share. Unfortunately, our data do not allow us to analyse the distribution of retail chains among the provinces in question due to confidentiality constraints. In addition it is worth noting that consumers may purchase different individual products due to different consumption behaviours strictly related to the city in which they live. While well-known brands are sold in all the Toscana provinces (*e.g.* “De Cecco”, “Barilla”, etc.) other local-produced pasta are sold in only few provinces (*e.g.* “Antichi Poderi Toscani”).

Figure 5.1 - Price distributions in Toscana provinces by groups of products: Mineral water (left), Coffee (centre), Pasta (right)



Source: Authors’ elaboration from Italian Scanner Data

Table 5.2 - Number of GTINs priced in each province for each product groups

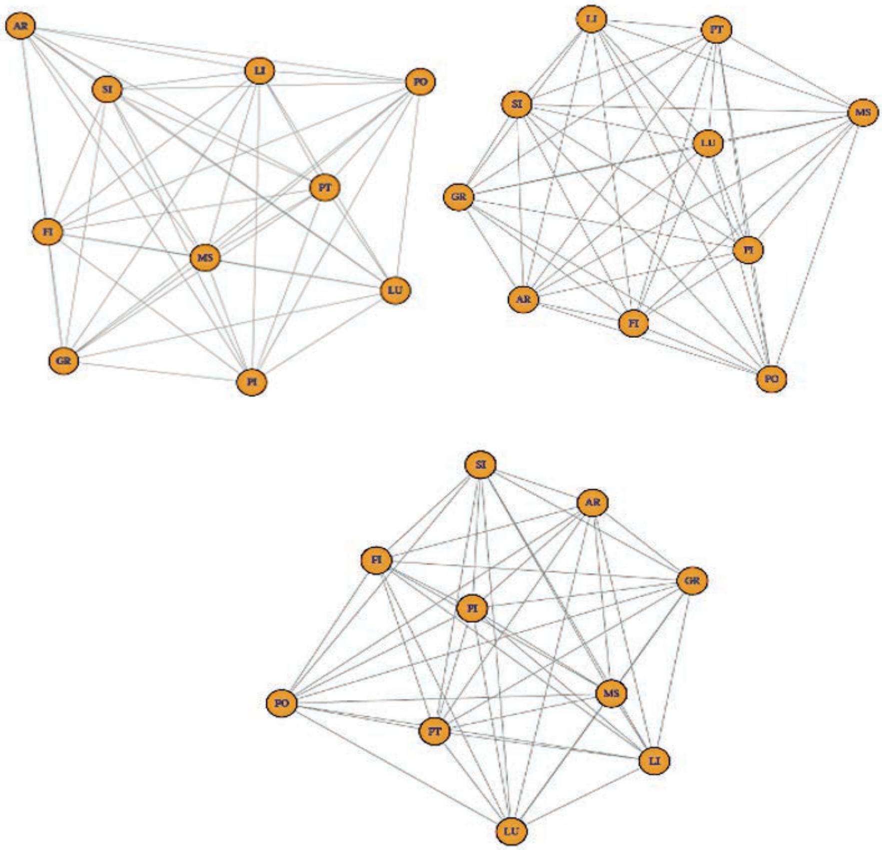
Province	Basic Heading		
	Mineral water	Coffee	Pasta
Arezzo	103	131	500
Firenze	159	176	470
Grosseto	112	110	354
Livorno	210	148	538
Lucca	193	194	611
Massa-Carrara	168	103	327
Pisa	142	172	437
Prato	78	76	178
Pistoia	177	141	485
Siena	87	108	352
Total	298	337	1091

Source: Authors' elaboration from Istat scanner data 2018

Given that not all the GTINs are priced in all the provinces, our data lead to an incomplete tableau of prices. In this case, it is important that the collected price data are connected to allow price comparison between all the areas involved (World Bank, 2013). For binary SPIs, little overlap in the products priced by the two provinces implies that the two geographical areas are very different and, by implication, inherently difficult to compare (Hill and Timmer, 2006).

Figure 5.2 demonstrates that reliable price comparisons can be carried out by illustrating the existence of the links among all the provinces. Indeed, we can state that our price data for the three product groups in question are connected among provinces and therefore it is not possible to place the provinces in two groups in which no GTINs sold by any province in one group is sold by any other province in the second group. We note multiple links between the same two nodes, even if not all the provinces are directly linked. Left panel of Figure 5.2 reports an interesting case in which the Pistoia (PT) province is not directly linked with Grosseto (GR) province. However, four indirect links exist. For example, one of the links compares prices across Pistoia and Firenze (FI) and then across Firenze and Grosseto. Therefore, in this case, through indirect links it is possible to make reliable price comparisons across provinces (World Bank, 2013).

Figure 5.2 - Figure 5.2 Product links among provinces. Mineral water (left), Coffee (right), Pasta (bottom)



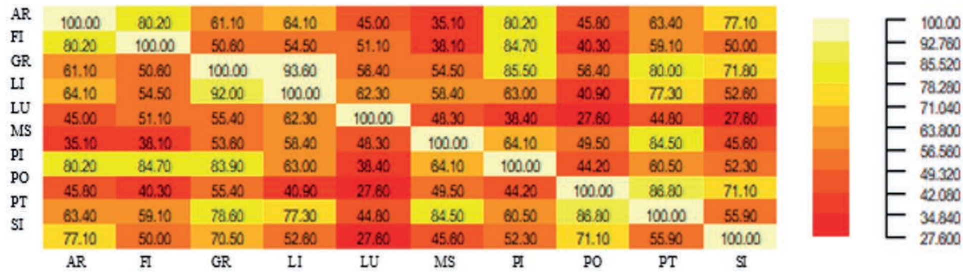
Source: Authors' elaboration from Istat scanner data 2018

In order to evaluate the strength of the overlaps among provinces, Figures 5.3, 5.4 and 5.5 show the product-matched matrices calculated for the 10 provinces for the aggregates included in our analysis. Each value in the matrix reports the matching products sold in two provinces. As an example, Figure 5.3, which reports the matrix for Mineral water, illustrates that 71.40% of products sold in Firenze are also purchased in the province of Arezzo. Considering Pasta products, the highest overlap is observed between Prato and Pistoia (97.75%), while the minimum overlap is observed between Pistoia and Lucca (25.90%).

Figure 5.3 - Product-matched matrix for Mineral water BH across provinces

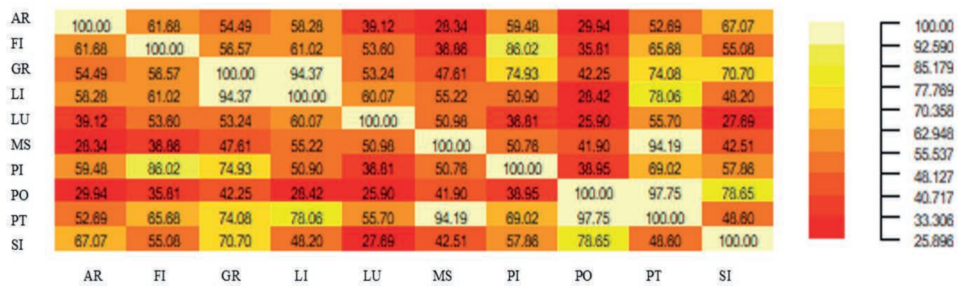
Source: Authors' elaboration from Istat scanner data 2018

Figure 5.4 - Product-matched matrix for Coffee BH across provinces



Source: Authors' elaboration from Istat scanner data 2018

Figure 5.5 - Product-matched matrix for Pasta BH across provinces



Source: Authors' elaboration from Istat scanner data 2018

SPIs point estimates obtained with GEKS-Fisher and GEKS-Törnqvist methods are reported in Table 5.3, SPIs are computed by considering Firenze province as reference (Firenze=100). As already mentioned, Fisher and Törnqvist price indices are symmetric and allow for substitution effect. Fisher price index measures the change in the expenditure function described by the Cobb-Douglas function, while Törnqvist price index measures the change in the expenditure function in a Translog utility function. Results reported in Table 5.3 reveal that Fisher and Törnqvist indices are very close to each other as shown by Diewert (1978). In addition, the Törnqvist index is approximately bounded by the Laspeyres and Paasche indices (Reinsdorf *et al.*, 2002).

The territorial heterogeneity among Toscana SPIs provinces is highlighted in Figure 5.6. From our results, it is possible to observe that Firenze is the less expensive province for Pasta and Coffee BHs, while for the Mineral water product group the less expensive area is Siena, followed by Arezzo and Prato.

It is interesting to note that concerning the Mineral water product group, Massa-Carrara is found to be the most expensive province (SPIs equal to 104.09 estimated with GEKS-Fisher and 104.07 estimated with Törnqvist-GEKS), followed by Grosseto (SPIs equal to 101.95 estimated with GEKS-Fisher and 101.92 estimated with Törnqvist-GEKS). For Coffee product group, the provinces of Grosseto and Siena proved to be the most expensive provinces according to the two procedures. The SPIs for Grosseto are equal to 105.35 with GEKS-Fisher and 105.20 with Törnqvist-GEKS, while the SPIs for Siena are equal to 105.13 with GEKS- Fisher and 105.10 with Törnqvist-GEKS. Interestingly, for the Pasta product group, the most expensive province is Siena (SPIs equal to 104.09 estimated with GEKS-Fisher and 104.07 estimated with Törnqvist-GEKS).

Table 5.3 reports the standard errors (in italics) obtained using JRR for the GEKS-Fisher and the Törnqvist-GEKS SPIs for each BH[6.] These are similar to each other although the GEKS-Fisher procedure has standard errors slightly greater than the Törnqvist-GEKS.

The highest standard errors are observed for Mineral water product group estimates obtained with the Fisher-GEKS procedure: the standard errors range from 0.003 in Pisa province to 0.049 in Grosseto province. By focussing on

6 These correspond to the inner summation in Formula (3).

Coffee product group, the highest standard error computed with Fisher-GEKS procedure, is observed for Grosseto (0.029) while the lowest standard error is observed for Pistoia (0.003). The Pasta product group has the lowest standard errors among the groups considered either for the Fisher-GEKS and the Törnqvist-GEKS. This is reasonably related to the high number of products and markets included in the analysed BHs. High standard errors may be due to the different territorial distribution of types of outlets, retail chains and market share in the Toscana territory.

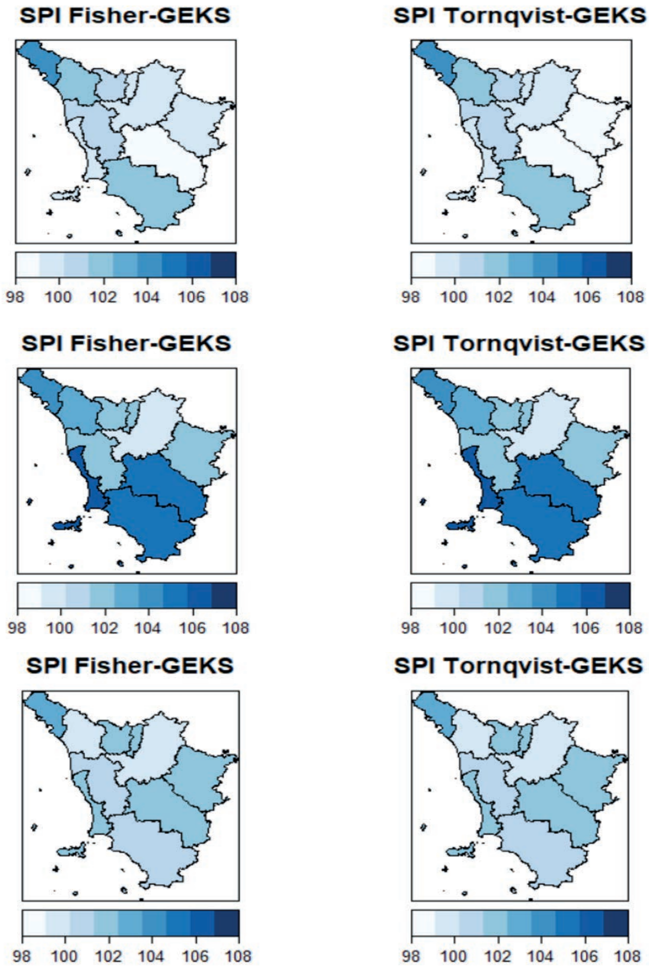
On the basis of the JRR replications, we derived 95% confidence intervals for each Provincial SPI. Confidence intervals overlap for the provinces of Livorno and Lucca for the three product groups analysed in this paper, thus suggesting that the SPIs in these two provinces are not significantly different from each other. As an example, for Mineral water product group, the confidence interval for Livorno province at 95% becomes [101.52; 101.68] and for Lucca province, the confidence interval becomes [101.59; 101.62].

Table 5.3 - Fisher-GEKS and Törnqvist-GEKS SPIs estimates. Standard errors in italics (Firenze=100)

Province	Fisher-GEKS			Törnqvist-GEKS		
	Basic Heading			Basic Heading		
	Mineral water	Coffee	Pasta	Mineral water	Coffee	Pasta
Arezzo	99.12	101.58	102.25	99.09	101.53	102.23
	0.010	*0.014*	*0.002*	*0.009*	*0.013*	*0.002*
Grosseto	101.95	105.35	100.95	101.92	105.20	100.94
	0.049	*0.029*	*0.020*	*0.048*	*0.028*	*0.020*
Livorno	101.61	102.65	100.18	101.55	102.66	100.21
	0.045	*0.022*	*0.018*	*0.044*	*0.021*	*0.018*
Lucca	101.61	102.65	100.18	101.55	102.66	100.21
	0.008	*0.014*	*0.007*	*0.008*	*0.014*	*0.007*
Massa-Carrara	104.09	104.50	102.80	104.07	104.40	102.82
	0.038	*0.018*	*0.014*	*0.038*	*0.018*	*0.014*
Pisa	100.45	102.02	100.70	100.46	101.99	100.70
	0.003	*0.008*	*0.005*	*0.003*	*0.008*	*0.005*
Prato	99.17	101.81	101.37	99.16	101.76	101.35
	0.010	*0.007*	*0.008*	*0.009*	*0.007*	*0.008*
Pistoia	100.63	102.41	102.21	100.61	102.34	102.19
	0.004	*0.003*	*0.006*	*0.004*	*0.003*	*0.006*
Siena	98.72	105.13	105.31	98.67	105.01	105.24
	0.028	*0.017*	*0.007*	*0.028*	*0.017*	*0.007*

Source: Authors' elaboration from Istat scanner data 2018.

Figure 5.6 - Fisher-GEKS and Törnqvist-GEKS SPIs estimates Mineral water, Coffee, Pasta. Firenze=100



Source: Authors' elaboration from Istat scanner data 2018

6 Conclusions

Scanner data has received considerable attention from statistical agencies during recent years since they proved to have significantly improved the efficiency of traditional price collection techniques in that they contain transactions on the goods sold, the prices paid by consumers, and the quantities sold for each item code or GTIN (Laureti and Polidoro, 2018). Scanner data may help to overcome the issue of price data availability in the various areas involved in spatial price comparisons thus fulfilling the requirements of representativeness and comparability that emerge when compiling sub-national SPIs. Due to the high territorial coverage, which characterises scanner data, it is possible to compare price levels at different territorial levels within a country (NUTS-3, NUTS-2, and NUTS-1). In this paper, we provided point estimates of SPIs of three basic headings for the provinces of Toscana in Italy using scanner data provided by Istat. Our results reveal that Fisher and Törnqvist-based GEKS numerically approximate each other and the two index methods tend to coincide, even if in some cases the SPIs estimated with Fisher-GEKS procedure are slightly greater than those obtained with Törnqvist-GEKS. This means that the unit elasticity of substitution implied by the geometric formula seems to overestimate the extent to which consumer responds to process changes relative to the market level in each product group based on ECR classification.

Along with point estimates, we used Jackknife Repeated Replications to provide an estimate of the associated standard errors with. In the introduction of the paper, we outlined that the measurement of price level differences across regions within a country is essential for assessing inequalities among populations residing in different parts of a country, for example in the distribution of real incomes (Laureti and Rao, 2018). Consequently, assessing the uncertainty of point estimates of price differences is essential for a better understanding of the phenomenon that we are investigating. In this paper we estimated the variance using the Jackknife estimator. However, the Jackknife depends on the sampling design and therefore to have more accurate estimates of SPIs it may be of crucial importance reconsidering the sampling design. On the basis of the JRR replications, we derived 95% confidence intervals for each Provincial SPI. Confidence intervals overlap for the provinces of Livorno and Lucca for the three product groups analysed in this paper, thus suggesting that the SPIs in these two provinces are not significantly different from each other. Moreover, our results showed that sometimes the uncertainty due to the reference selection obtained stratifying the GTINs by market (ECR group) is such that for some area we are not able to say whether the SPI obtained is significantly higher or lower than that of the base area. Of course, the Jackknife is not the only estimator of variance that has been proposed in the literature (*e.g.* bootstrap, linearisation) but no comparison has been addressed in the literature so far (at least, to the authors best knowledge). This paper is a stepping-stone to the development of variability associated to SPIs. This is an important topic that will constitute a further line of research focussed on estimating uncertainty and comparing the results from different models.

References

Adelman, I. 1958. "A New Approach to the Construction of Index Numbers". *The Review of Economics and Statistics*, Volume 40, N. 3: 240-249.

Andersson, C., G. Forsman, and J. Wretman. 1987a. "On the Measurement of Errors in the Swedish Consumer Price Index". *Bulletin of the International Statistical Institute*, N. 52, Book 3: 155–171.

Andersson, C., G. Forsman, and J. Wretman. 1987b. "Estimating the Variance of a Complex Statistic: A Monte Carlo Study of Some Approximate Techniques". *Journal of Official Statistics - JOS*, Volume 3, N. 3: 251–265.

Aten, B.H. 2017. "Regional Price Parities and Real Regional Income for the United States" "Regional price parities and real regional income for the United States". *Social Indicators Research*, Volume 131, Issue 1: 123-143. https://doi.org/10.1007/s11205-015-1216-y.

Aten, B.H. 2008. "Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005". *Papers*, N. 05. Suitland, MD, U.S.: Bureau of Economic Analysis – BEA, U.S. Department of Commerce.

Balk, B.M. 2005. "Price Indexes for Elementary Aggregates: The Sampling Approach". *Journal of Official Statistics* - JOS, Volume 21, N. 4: 675–699.

Balk, B.M. 1995. "Axiomatic Price Index Theory: A Survey". *International Statistical Review*, Volume 63, N. 1: 69–93.

Balk, B.M. 1989. "On calculating the precision of consumer price indices". In Proceedings of the *47th Session of ISI World Statistics Congress* – WSC, Paris: contributed papers, Book 1.

Balk, B.M. 1981. "A simple method for constructing price indices for seasonal commodities". *Statistische Hefte*, Volume 22, Issue 1: 72-78.

Balk, B.M., and H.M.P. Kersten. 1986. "On the Precision of Consumer Price Indices Caused by the Sampling Variability of Budget Surveys". *Journal of Economic and Social Measurement*, Volume 14, N. 1: 19-35.

Banerjee, K.S. 1956. "A Note on the Optimal Allocation of Consumption Items in the construction of a Cost-of-Living Index". Econometrica, Volume 24, N. 3: 294-295.

Betti, G., F. Gagliardi, and V. Verma. 2018. "Simplified Jackknife Variance Estimates for Fuzzy Measures of Multidimensional Poverty". *International Statistical Review*, Volume 86, Issue 1: 68-86.

Biggeri, L., G. Ferrari, and Y. Zhao. 2017*a*. "Estimating Cross Province and Municipal City Price Level Differences in China: Some Experiments and Results". *Social Indicators Research*, Volume 131, Issue 1: 169-187.

Biggeri, L., and A. Giommi. 1987. "On the accuracy and precision of the consumer price indices. Methods and applications to evaluate the influence of the sampling of households". *Bulletin of the International Statistical Institute*, N. 52, Book 3: 155–171.

Biggeri, L., T. Laureti, and F. Polidoro. 2017*b*. "Computing Sub-national PPPs with CPI Data: An Empirical Analysis on Italian Data Using Country Product Dummy Models". *Social Indicators Research*, Volume 131, Issue 1: 93-121. https://doi.org/10.1007/s11205-015-1217-x.

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

Caves, D.W., L.R. Christensen, and W.E. Diewert. 1982. "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity". *Econometrica*, Volume 50, N. 6: 1393-1414.

Chen, M., Y. Wang, and D.S. Prasada Rao. 2019. "Measuring the Spatial Price Differences in China with Regional Price Parity Methods" *The World Economy*, Volume 43, Issue 4: 1103-1146. https://doi.org/10.1111/twec.12899.

De Gregorio, C. 2012. "Sample size for the estimate of consumer price sub-indices with alternative statistical designs". *Rivista di statistica ufficiale/Review of official statistics*, N. 1/2012: 19-47. Roma, Italy: Istat. https://www.istat.it/it/archivio/72183.

De Haan, J., E. Operdoes, and C. Schut. 1999. "Item selection in the Consumer Price Index: Cut-off versus probability sampling". *Survey Methodology*, Volume 25, N. 1: 31-41.

De Vitiis, C., A. Guandalini, F. Inglese, and M.D. Terribili. 2017. "Sampling Schemes Using Scanner Data for the Consumer Price Index". In Petrucci, A., F. Racioppi, and R. Verde (Eds.). *New Statistical Developments in Data Science. Atti Convegno della Società Italiana di Statistica - SIS 2017*. Springer Proceedings in Mathematics & Statistics, Volume 288: 203-2017. Cham, Switzerland: Springer.

Deaton, A. 2012. "Calibrating measurement uncertainty in purchasing power parity exchange rates". International Comparison Program – ICP, *7th Technical Advisory Group Meeting*, Washington, D.C., U.S., 17-18 September 2012.

Deaton, A., and B.H. Aten. 2017. "Trying to Understand the PPPs in ICP 2011: Why Are the Results so Different?". *American Economic Journal: Macroeconomics*, Volume 9, N. 1: 243-264.

Diewert, W.E., and K.J. Fox. 2022. "Substitution Bias in Multilateral Methods for CPI Construction". *Journal of Business & Economic Statistics*, Volume 40, Issue 1: 355-369.

Diewert, W.E. 1978. "Superlative Index Numbers and Consistency in Aggregation". *Econometrica*, Volume 46, N. 4: 883-900.

Diewert, W.E. 1976. "Exact and superlative index numbers". *Journal of Econometrics*, Volume 4, Issue 2: 115-145.

Dorfman, A.H., J. Lent, S.G. Leaver, and E. Wegman. 2006. "On Sample Survey Designs for Consumer Price Indexes". *Survey Methodology*, Volume 32, N. 2: 197-216.

Durbin, J. 1959. "A Note on the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios". *Biometrika*, Volume 46, N. 3/4: 477-480.

Eurostat, and Organisation for Economic Co-operation and Development - OECD. 2012. "Eurostat-OECD Methodological Manual on Purchasing Power Parities. 2012 edition". *Methodologies and Working papers*. Luxembourg: Publications Office of the European Union.

Feenstra, R.C., M. Xu, and A. Antoniades. 2017. "What is the Price of Tea in China? Towards the Relative Cost of Living in Chinese and US Cities". NBER, *Working Paper*, N. 23161. Cambridge, MA, U.S.: National Bureau of Economic Research - NBER.

Fenwick, D., A. Ball, P. Morgan, and M. Silver. 2001. "Sampling in Consumer Price Indices: what role for scanner data?". Paper presented at the *Sixth Meeting of the International Working Group on Price Indices*, Canberra, Australia, 2-6 April 2001.

Fenwick, D., and J. O'Donoghue. 2003. "Developing estimates of relative regional consumer price levels". *Economic Trends*, N. 599: 72-83, Office for National Statistics – ONS. London, UK: Palgrave Macmillan.

General Statistical Office of Vietnam - GSO. 2019. "Subnational PPPs in Vietnam". Paper presented at the *4th Meeting of the ICP Task Force on the Country Operational Guidelines and Procedure*, Paris, France, 2-3 May 2019.

Heravi, S., A. Heston, and M. Silver. 2003. "Using scanner data to estimate country price parities: A hedonic regression approach". *Review of Income and Wealth*, Volume 49, Issue 1: 1-21. https://doi.org/10.1111/1475-4991.00071.

Hill, R.J. 2004. "Constructing Price Indexes across Space and Time: The Case of the European Union". *American Economic Review*, Volume 94, N. 5: 1379-1410.

Hill, R.J., and M.P. Timmer. 2006. "Standard Errors as Weights in Multilateral Price Indexes". *Journal of Business & Economic Statistics*, Volume 24, N. 3: 366-377.

Imai, S., E. Diewert, and C. Shimizu. 2015. "Consumer Price Index Biases. Elementary Index Biases vs. Sampling Biases". Paper presented at the *Fourteenth Meeting of the Ottawa Group – the International Working Group on Price Indices*, Tokyo, Japan, 20-22 May 2015.

Institut National de la Statistique et des Études Économiques – INSEE, Direction des statistiques démographiques et sociales. 2019. "Enquete de comparaison spatiale des niveaux de prix à la consommation entre territoires francais". *Sources statistiques et indicateurs*. Paris, France: INSEE.

International Monetary Fund – IMF, International Labour Organization – ILO, Eurostat, United Nations Economic Commission for Europe – UNECE, Organisation for Economic Co-operation and Development – OECD, and The World Bank. 2020. *Consumer Price Index Manual. Concepts and Methods* - 2020. Washington, D.C., U.S.: IMF.

Istituto Nazionale di Statistica – Istat, Unioncamere e Istituto Guglielmo Tagliacarne. 2010. "Le differenze nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane. Anno 2009". *Comunicato Stampa*. Roma, Italy: Istat. https://www.istat.it/it/archivio/6279.

Ivancic, L., W.E. Diewert, and K.J. Fox. 2011. "Scanner data, time aggregation and the construction of price indexes". *Journal of Econometrics*, Volume 161, Issue 1: 24-35.

Janský, P., and D. Kolcunová. 2017. "Regional differences in price levels across the European Union and their implications for its regional policy". *The Annals of Regional Science*, Volume 58, Issue 3: 641-660.

Klick, J., and O. Shoemaker. 2019. "Measures of Variance Across CPI Populations". *Research papers*. Washington, D.C., U.S.: U.S. Bureau of Labor Statistics, Office of Survey Methods Research.

Kocourek, A., J. Šimanová, and L. Šmída. 2016. "Modelling of Regional Price Levels in the Districts of the Czech Republic". Technical University of Liberec, 2018-09-25T11:50:33Z.

Kokoski, M.R., B.R. Moulton, and K.D. Zieschang. 1999. "Interarea Price Comparisons for Heterogenous Goods and Several Levels of Commodity Aggregation". In Heston, A., and and R.E. Lipsey (Eds.). "International and Interarea Comparisons of Income, Output and Prices": 123-166. NBER, *Studies in Income and Wealth*, Volume 61. Chicago, IL, U.S.: The University of Chicago Press.

Kott, P.S. 1984. "A Superpopulation Theory Approach to the Design of Price Index Estimators with Small Sampling Biases". *Journal of Business & Economic Statistics*, Volume 2, N. 1: 83-90.

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

Laureti, T., and F. Polidoro. 2018. "Big data and spatial price comparisons of consumer prices". Presented at the *49th Scientific meeting of the Italian Statistical Society*, 20-22 June 2018, Palermo, Italy.

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

Leaver, S.G., and R.A. Cage. 1997. *Estimating the Sampling Variance for Alternative Estimators of the U.S. Consumer Price Index*. Washington, D.C., U.S.: U.S. Bureau of Labor Statistics, Office of Survey Methods Research.

Leaver, S.G., J.E. Johnstone, and K.P. Archer. 1991. "Estimating unconditional variances for the U.S. consumer price index for 1978–1986". In American Statistical Association - ASA. *Proceedings of the Survey Research Methods Section:* 614-619. Alexandria, VA, U.S.: ASA.

Leaver, S.G., and R. Valliant. 1995. "Statistical Problems in Estimating the U.S. Consumer Price Index". In Cox, B.G., D.A. Binder, B. Nanjamma Chinnappa, A. Christianson, M.J. Colledge, and P.S Kott (Eds.). *Business Survey Methods*, Chapter 28: 543–566. Hoboken, NJ, U.S.: John Wiley & Sons, *Wiley Series in Probability and Statistics*.

Melser, D., and R.J. Hill. 2007. "Methods for Constructing Spatial Cost of Living Indexes". *Statisphere - Official Statistics Research*, Series 1, Article 3: 1-113.

Montero, J.-M., T. Laureti, R. Mínguez, and G. Fernández-Avilés. 2019. "A Stochastic Model with Penalized Coefficients for Spatial Price Comparisons: An Application to Regional Price Indexes in Italy". *Review of Income and Wealth*, Volume 66, Issue 3: 512-533. https://doi.org/10.1111/roiw.12422.

Morgenstern, O. 1963. *On the Accuracy of Economic Observations*. Princeton, NJ, U.S.: Princeton University Press.

Norberg, A., and C. Tongur. 2022. "Balancing the Swedish CPI". Paper presented at the *Seventeenth Meeting of the Ottawa Group – the International Working Group on Price Indices*, Roma, Italy 7-10 June 2022.

Paredes Araya, D., and V. Iturra Rivera. 2013. "Substitution bias and the construction of a spatial cost of living index". *Regional Science*, Volume 92, Issue 1: 103-117.

Pilat, D., and D.S. Prasada Rao. 1996. "Multilateral Comparisons Of Output, Productivity, And Purchasing Power Parities In Manufacturing". *Review of Income and Wealth*, Volume 42, Issue 2: 113-130.

Rao, D.S. Prasada. 2013. "The framework for the international comparison program (ICP)". In World Bank (Ed.). *Measuring the Real Size of the World Economy*, Chapter 1: 13-45. Washington, D.C., U.S.: World Bank.

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

Reinsdorf, M.B., W.E. Diewert, and C. Ehemann. 2002. "Additive Decompositions for Fisher, Törnqvist and Geometric Mean Indexes". J*ournal of Economic and Social Measurement*, Volume 28, N. 1-2: 51-61.

Rokicki, B., and G.J.D. Hewings. 2019. "Regional price deflators in Poland: Evidence from NUTS-2 and NUTS-3 regions". *Spatial Economic Analysis*, Volume 14, Issue 1: 88-105. https://doi.org/10.1080/17421772.2018.1503705.

Roos, M. 2006. "Regional price levels in Germany". *Applied Economics*, Volume 38, N. 13: 1553-1566. https://doi.org/10.1080/00036840500407207.

Smith, P.A. 2021. "Estimating Sampling Errors in Consumer Price Indices". *International Statistical Review,* Volume 89, Issue 3: 481-504. https://doi.org/10.1111/insr.12438.

Tam, S.-M., and F. Clarke. 2015. "Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics". *International Statistical Review*, Volume 83, Issue 3: 436-448. https://doi.org/10.1111/insr.12105.

Tongur, C. 2019. "Inflation Measurement with Scanner Data and an Ever-Changing Fixed Basket". *Economie et Statistique*, N. 509: 31-47.

Weinand, S., and L. von Auer. 2020. "Anatomy of regional price differentials: evidence from micro-price data". *Spatial Economic Analysis*, Volume 15, Issue 4: 1-28. https://doi.org/10.1080/17421772.2020.1729998.

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