Measuring geographical and population coverage in CPI internet price collection: An application with groceries web scraping in Italy

Tiziana Laureti[a], Luigi Palumbo[ab*]
[a]Università degli Studi della Tuscia, Viterbo, Italy
[b]Banca d'Italia, Roma, Italy *

Abstract

Consumer price indices (CPIs) are instrumental in the development of monetary policy and in monitoring economic developments. Prices collection for CPI compilation has come a long way in the past 20 years. However, while ideally, the index should include expenditure made by all households, urban and rural, throughout the country, CPIs in various countries have limited geographic coverage both for price collection and consumption expenditures. The introduction of new data sources, such as web scraping and scanner data, have contributed to reduce price collection costs and increase the reach across national territories, thus allowing to enhance the accuracy and quality of the CPI. The aim of this paper is to suggest a finer measurement CPI geographical coverage based on geostatistical fuzzy indices that would be particularly useful in cases where prices vary substantially across space, as it is proven that consumers only travel within limited extents for their purchases and a sparse network of outlets may lead to biased measurements. To explore the potential of the suggested measure we estimate relative price levels across regions for a time period and price changes over the period for each region region-time-dummy method. This analyses is further validated by referring to structural breaks in our coverage metric and in spatio-temporal CPIs. Using a dataset deriving from geo-localized groceries web scraping in Italy, we provide a practical application calculating coverage at a regional level adopting different functional forms. Our findings corroborate the robustness of the proposed coverage metric and allow to embed information on geographic coverage in price statistics.

Keywords: geographical coverage; geostatistics; fuzzy logic; prices; web scraping.

1 Introduction

Consumer price indices (CPIs) measure price changes of the goods and services purchased by households in their role as consumers. CPIs are instrumental in the development of monetary policy and in monitoring economic developments. As a result, many policy debates have arisen surrounding the accuracy and reliability of price indices. Over the last decades, substantial progress has been made in developing new data sources, price collection methods, and related index calculation methods with the aim of reducing CPIs biases and errors (Smith, 2021). Price collection is becoming increasingly multimodal with prices being web scraped from the internet or obtained from scanner data, as well as being traditionally collected by collectors visiting individual outlets for several goods and services. Due to the fact that it is impossible to regularly record all the prices of the universe, CPIs are a sample statistics that represent the change in prices over the target universe in the two periods (International Monetary Fund, 2020).

Consequently, sampling techniques are used to select a subset of prices that enter the CPI compilation. The sampling process occurs on geographical location, outlet type, products and time dimensions. Within each of the different sampling levels, the sampling approach can differ from country to country, reflecting different administrative arrangements and practical reasons. Either probability or non-probability sampling methods can be adopted in each dimension. The geographic or spatial dimension is a key component in assessing the methodological soundness of the CPI (Berry, Graf, Stanger, & Ylä-Jarkko, 2019). concerning both product price collection and elementary price aggregation (average prices calculated for all item transactions in the country, province or state, city, neighbourhood). Diewert (2021) underlined that there is not a clear consensus on what the optimal degree of spatial disaggregation should be. Therefore, each NSI can make its own judgements on this matter, taking into account the costs of data collection and the demands of users for a spatial dimension for the CPI.

*The views expressed herein are those of the authors and do not necessarily represent the views of the Bank of Italy and/or the Eurosystem.

Several CPI sampling operations are effectively cut-off samples with parts of the population of interest excluded thus producing a resulting in coverage error (Smith, 2021). Unless it is possible to sample outlets directly from a national sampling frame such as a business register (which often cannot identify small outlets or the precise range of products available in them), the sampling of outlets generally needs to be done in two stages. In the first stage, a sample of locations such as regions, cities or shopping areas is drawn/selected throughout the country, and in the second stage outlets are sampled.

When sampling locations, two major requirements must be considered: representativity and cost effectiveness. Areas where the bulk of consumer purchases take place need to be covered with certainty or by a probability sample to make the sample representative. Location samples are generally fixed for a long period of time, as they determine the whole organisation of work for the statistical office. When a large country is divided into administrative areas (state, region, province, etc.), all areas are often included with certainty, after which there may be sampling of locations within each of them, thus leading to increased representativity where price movements may differ due to different climates and/or transport costs. It is also a necessary requirement if regional CPIs are disseminated. In this case expenditure weights should be used to aggregate the regional indexes into the national 'all-items' index. If regional indexes are not disseminated, a representative sample of geographic areas can be selected for price collection, but the index weights should be based on the expenditure of all households in the country. In small countries it is common to select a few of the larger cities for price collection. This leaves out smaller towns and rural areas, but as consumers living in areas close to city will go there for some of their shopping the effect of their exclusion will be smaller than might be inferred from population numbers, and a sufficient coverage may still be achieved. It is then important that the selected cities are such that their outlets are used by a large part of the population and that they are situated in different parts of the country for maximum coverage. Car-friendly shopping centres situated immediately outside a city should be included if they are significant. While geographic coverage of CPIs is an indicator of quality, as ideally the index should include expenditure made by all households - urban and rural - throughout the country, little attention has been devoted to this issue in literature (Diewert, 2021; Guerreiro, Baer, & Silungwe, 2022; Hawkes & Piotrowski, 2003).

Many countries have CPIs with limited geographic coverage — capital city, including few of the largest areas (such as large and medium-sized cities) and prices are collected in urban areas only because their movements are considered to be representative of the price movements in rural areas. The geographical dimension, which is related to the scope of the index, becomes more important the smaller the region to which the index relates. The aim of this paper is to suggest a finer measurement of CPI geographical coverage to provide price statisticians with better insights on the actual reach of data collection. Since CPI compilation is becomes more important in economic planning and inflation monitoring, efforts should be made to expand the CPI to cover more geographic areas including all urban and rural areas. This would be particularly useful in cases where prices vary substantially across space, as it is proven that consumers only travel within limited extents for their purchases and a sparse network of outlets may lead to biased measurements.

The popularity and availability of new data sources for the compilation of the CPI, such as web-scraping and scanner data, has increased over the past twenty years and have contributed to reduce price collection costs and increased the reach across national territories, thus allowing to enhance the accuracy and quality of the CPI. Scanner data offer an opportunity to examine the effect of population exclusions, and Brunetti, Fatello, Polidoro, and Simone (2018) make such calculations for Italy, where sampling in the main CPI is restricted to the main provincial towns and uses only a sample of the most-sold products. They find only some differences, mainly due to sampling towns only, and concentrated in the south of Italy. Web-scraped data collection has been increasingly used by NSIs recent years and many countries are developing web-scraping tools tailored to specific CPIs requirements that allow to quickly collect large numbers of prices for a wide variety of online products and cover new consumpetion segments (Eurostat, 2020). Although the compilation of price indices from such large datasets is not straightforward, these new sources of data have proved to be of benefit to CPIs thanks to the detailed information available for individual products and the wide coverage both in terms of product groups and territorial areas.

We suggest a geostatistical fuzzy index (Zadeh, 1977; Zimmermann, 2011) to measure the reach of data collection in terms of geographical and population coverage of outlets where prices are collected. This index may be used to evaluate the degree of coverage for price data collection, both in the context of probabilistic and non-probabilistic outlet selection. An advantage of the fuzzy set theory approach is to overcome the limits of discrete classifications of data, preserving a higher degree of information for analysis. A properly designed membership functions may enable us to achieve a better classification of the data, smoothing distortions caused by outliers while still including them into the analysis. Using a fuzzy membership function it is possible to calculate the coverage value for each municipality, inversely proportional the driving distance in minutes from the closest outlet where prices have been collected. Total coverage value for a given territory is calculated as an average over municipalities coverage values, either using unweighted or weighted formulae.

Using a dataset deriving from geo-localized groceries web scraping in Italy, we provide a practical application calculating coverage at a regional level and comparing results from two different functional forms – linear and non-linear – as well as a set of different parameters for spatial decay of coverage. Our findings corroborate the robustness of the coverage index, as rank correlations amongst different parameters and functional values are close to 1 and statistically significant. In addition, with the aim of emphasising the importance of measuring and

monitoring the degree of coverage, we carry out spatial-temporal panel comparisons using the Time-interaction-Region Product Dummy (TiRPD) method to assess what happens to consumer price indices when there is an abrupt change in coverage. TiRPD is a natural extension of the well known time-dummy and region-dummy methods which have widely been used in literature for constructing consumer price indices (Corrado & Ukhaneva, 2016). We conclude by illustrating our methodology using web scraped data from 616 online supermarkets belonging to 23 different retail chains in Italy from November 2020 to February 2023. Abrupt change in spatio-temporal CPIs and coverage are identified using a Bayesian estimator, in order to validate the link amongst the two phenomena.

The remainder of this paper is structured as follows. Section 2 illustrates the methodological approach. Section 3 describes the data and reports descriptive statistics. Section 4 presents the main results of the empirical analysis. Finally, Section 5 draws some conclusions.

2 Methodology

2.1 Fuzzy Coverage Index

The fundamental concept behind our proposed measure of coverage for price collection is that price information decays with space and travel time. Consumers may travel for certain distances and time to make purchases, thus providing an incentive for sellers to maintain competitive prices in different municipalities (see for example Kerr et al., 2012). However, consumers' inclination to commit time and money for purchasing trips is directly connected to the expected economic benefit in terms of savings.

Given the average basket value for groceries shopping, it is reasonable to affirm that there are limits to shopping trips distances, even if those may vary between consumers because of different travel costs, cost-opportunity of travel time and other individual characteristics.

Empirical evidence of spatial effects underlying consumer price differences among geographical areas have been observed both at country and sub-national level (Aten, 1996; Biggeri, Laureti, & Polidoro, 2017; Montero, Laureti, Mínguez, & Fernández-Avilés, 2020; Rao, 2001). Therefore, we need to conclude that prices may be different between municipalities situated at a certain distance, and the information value of collected prices in a certain location will decay with space and travel time.

In order to appropriately model this decay we resort to Fuzzy Set theory, as it seems inappropriate to specify hard boundaries regarding the validity of price information in binary terms. We then propose two different membership functions to calculate the coverage value for each municipality. The first one is a simple linear function, where coverage is inversely proportional to travel distance.

$$lc(x) = max(1 - \frac{x}{D}, 0) \tag{1}$$

Where x is the travel time by car in minutes between a municipality and the closest municipality where prices have been collected, and D is a parameter indicating at which travel time level the price information is considered no longer valid.

The second type of membership function is based on an inverse sigmoid modeling of price information decay. In fact, it is reasonable to assume that consumer willingness to travel for purchases is not linear, therefore price persistence in space is relatively stronger at short distances and weaker at longer ones. We can then propose a different membership function as follow:

$$c(x) = 1 - \frac{1}{1 + e^{-k(x - \frac{D}{2})}} \tag{2}$$

Where x is again the travel time by car in minutes between a municipality and the closest municipality where prices have been collected, D is a parameter indicating at which travel time level the price information is considered no longer valid (and $\frac{D}{2}$ is the midpoint of the inverse sigmoid), and k is a parameter indicating the steepness of the inverse sigmoid curve.

Once we calculate coverage values for all municipalities in a region, we need to synthesize a metric to indicate the overall coverage for the region. In order to do so, we can aggregate individual municipalities as units or by weighting them according to their population.

If we chose to treat municipalities as individual units, the coverage for a given region could be expressed as a simple arithmetic mean as in (3).

$$C_{mun} = \frac{\sum_{i=1}^{n} c_i}{n} \tag{3}$$

On the other side, if we chose to weight coverage in each municipality by its population the overall Region coverage would be:

$$C_{pop} = \frac{\sum_{i=1}^{n} c_i * pop_i}{\sum_{i=1}^{n} pop_i} \tag{4}$$

Formulas for coverage are applied at a regional level, since the main purpose is to provide a coverage metric for sub-national spatio-temporal price indices over space and time.

2.2 Spatio-temporal Price Indices

While there are several methodologies that could be applied to calculate regional SPIs (Laureti & Rao, 2018), for this work we selected a Time-interaction-Region Product Dummy (TiRPD) model which helps us reconcile aggregate SPIs across space and time and it has already been applied to price data from web scraping (Benedetti, Laureti, Palumbo, & Rose, 2022). This model was first proposed by Aizcorbe and Aten (2004), who referred to it as the Time-interaction-Country Product Dummy method. This model was designed as combination between the Country Product Dummy (CPD) model (Summers, 1973), which focuses on spatial price variation, and the Time Product Dummy model (TPD) (Aizcorbe, Corrado, & Doms, 2000; De Haan & Krsinich, 2014), which focuses on price variation over time. The specification of the TiRPD model is:

$$lnP_{ijt} = \sum_{i=1}^{N} \beta_i D_{ijt} + \sum_{t=1}^{T} \sum_{j=1}^{M} \delta_{jt} R_{ij} T_{jt} + \eta_{ijt} \tag{5}$$

Where lnP_{ijt} and D_{ijt} are respectively the log-price and the dummy for product i in area j at time t ($i = 1, 2, ..., N$; $j = 1, 2, ..., M$; $t = 1, 2, ..., T$). R_{ij} and T_{jt} are dummy variables for each combination of area and time period. A price index for each region-period jt is obtained directly from the parameter of the dummy by exponentiation of the δ_{jt} coefficient, and it is possible to perform direct comparisons across regions or between time as TiRPD is a multilateral method. The TiCPD provides the same answers as separate CPD or TPD models, with the advantage that it normalizes the relationships on a single region and time period.

2.3 Structural breaks

In order to further validate the importance of measuring and monitoring coverage, we assess what happens when there are structural breaks. For this purpose, we use the Bayesian Estimator of Abrupt change, Seasonal change, and Trend (BEAST) proposed by Zhao et al. (2019), and implemented in the R package Rbeast.

The BEAST model, a Bayesian statistical model that performs time series decomposition into multiple trend and seasonal signals, provides us with the probability for each of the time series points to be a trend change point. The general form of the model is:

$$y_i = S(t_i; \Theta_s) + T(t_i; \Theta_t) + \varepsilon_i \tag{6}$$

where y_i is the observed value at time t_i, Θ_s and Θ_t are respectively the season and trend signals, and ε_i is noise with an assumed Gaussian distribution. Given the relative short length of our time series, we removed the seasonal component from the model, which is then formalized as:

$$y_i = T(t_i; \Theta_t) + \varepsilon_i \tag{7}$$

Trend change points are implicitly encoded in Θ_t, and the trend function is modeled as a piecewise linear function with m knots and $m + 1$ segments. In each segment, the trend is built as:

$$T(t) = a_j + b_j t \; for \; \tau_j \leq t < \tau_{j+1}, j = 0, ..., m \tag{8}$$

where a_j and b_j are parameters for the linear trend in the j segment, which spans from τ_j to τ_{j+1}.

Further details about the Bayesian formulation of BEAST, its Markov Chain Monte Carlo inference and posterior inference of change points, seasonality, and trends can be found in Zhao et al. (2019). [1]

For our purposes, once we obtain the probability of trend change for each time point in each Region for the coverage and spatio-temporal price index level, we first check the stationarity of both time series using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992), and then we calculate the Pearson correlation between the two series. A positive significant correlation would signal a potential effect on the estimated price level from the abrupt change in the coverage.

3 Data

Data used for the empirical validation of our proposed methodology has been scraped from 616 online supermarkets belonging to 23 different chains in 19 Italian Regions from November 2020 to February 2023[2]. The portfolio of online supermarkets changed over time, as new sources were added and other became unavailable, due to failure in the scraping routines or anti-scraping measures implemented by the source website.

[1] By construction, the probability of being a trend change point is additive over time. In other words, the total probability of encountering a trend change point between time t and s equals the sum of all probabilities for time points between t and s.

[2] No data has been collected for the Trentino-Alto Adige region.

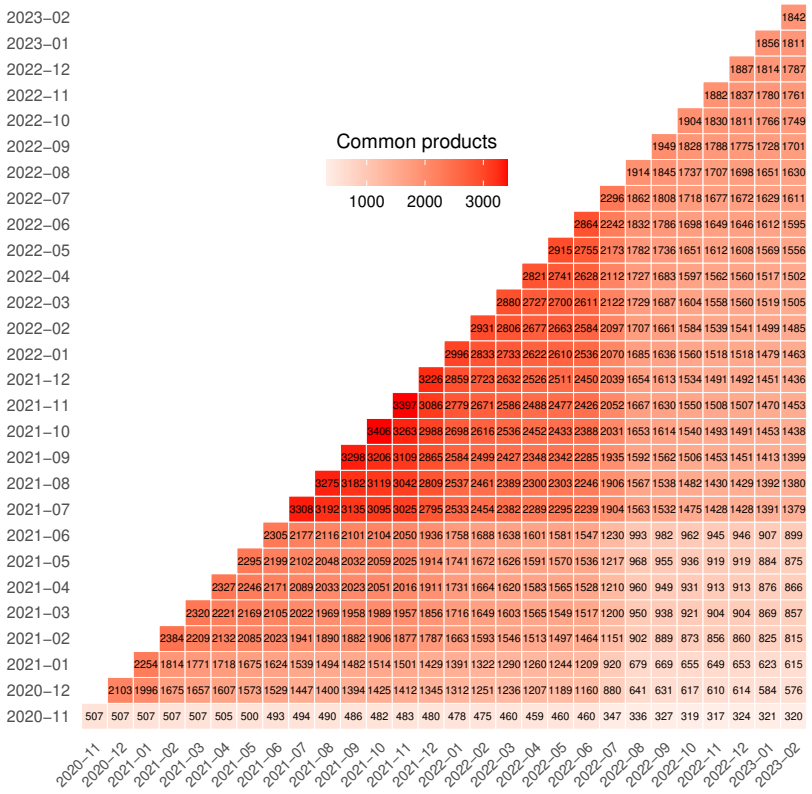
Each online supermarket has been located with GPS coordinates and placed in a specific municipality using geographical merging functions. We collected prices for each supermarket using the "pick up" option for purchase delivery. Therefore, validity of price information is considered linked to its geographical position.

For exemplification purposes, we selected the Coffee category (ECOICOP code 01.2.1.1), which in 2021 accounted for an average Italian household monthly expenditure of 11.91 EUR. Weights for the 5 digit subclass Coffee ranges from 0.38% in 2020 to 0.43% in 2023. Our data collection for this category amounted to 5338 unique products and 1221755 total observations. In 2056 cases we were able to identify products using their Global Trade Item Number (GTIN), and therefore we could accurately match them across different retail chains. In the other cases, instead, each unique product was identified according to retailer-specific attributes such as product code or product name and could not be automatically matched across different retail chains.

We classified products according to their commercial category, which is not standard across retailer, and with filters based on inclusion (or exclusion) of specific terms in the product name when the commercial category was not sufficiently specific. It should be noted that product naming could vary substantially across retailer for the same product - for instance: the inclusion or exclusion of the category name, brand and size or the use of abbreviations - but at the same time different products in the same category may be quite similarly named.

Our products are overlapping across the different time periods and regions, as showed in Figures 1 and 2. The large number of common products is reassuring for the validity of our results, as if there is little overlap across time and space price levels are inherently difficult to compare (Hill & Timmer, 2006).

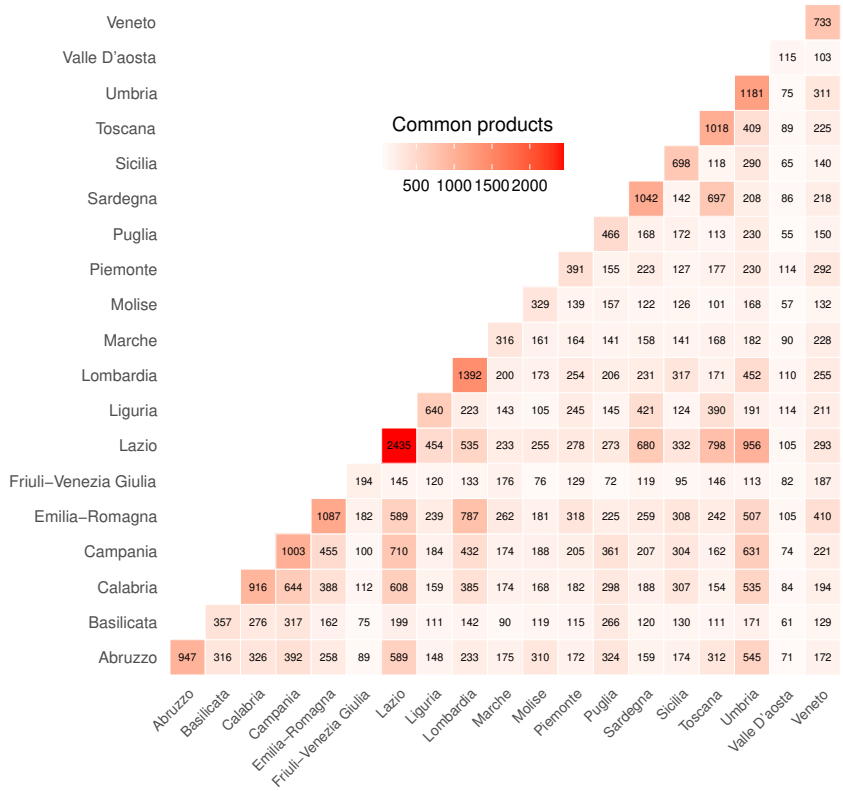
Figure 1: Common products across time periods.



Driving distances between municipalities have been obtained from a distance matrix published by the Italian National Institute of Statistics (Istat). Istat calculated the driving distance between centroids for all Italian municipalities in 2013 using a commercial road graph (Istat, 2019). We performed a basic elaboration in order to adjust for merging between small municipalities in the 2013-2021 period, also excluding minor islands and municipalities disconnected from the road graph[3]. The total amount of population living in excluding municipalities is marginal when compared to the relative Region population.

[3]Municipalities of Monte Isola (BS) and Campione d'Italia (CO) do not have any connection with the road graph used for distance calculation. Istat only provides distance from the closest municipality for them. For minor islands Istat provides a travel time by ferry to the closest port.

Figure 2: Common products across regions.



4 Results

In table 1 we report the results for coverage in December 2021 by Region calculated according to the different membership functions presented in (1) and (2), using as weight both individual municipalities, according to (3), and their population according to (4) in order to calculate the overall regional coverage and utilizing different D parameters for distance. Figures 3 to 6 are graphical representations of municipalities' coverage values according to the above mentioned membership functions at selected values of D in December 2021.

In order to evaluate the stability and consistence of our coverage metrics, we performed a series of measurement leveraging the Spearman Rank Correlation non-parametric test on the coverage values calculated for each region in December 2021, as presented in Table 1. Results are presented in Table 2, and we can appreciate that rank correlations are very strong and significant in all cases, indicating that our proposed indicator can deliver robust and consistent results irrespective of the parameters chosen. The link is somehow less strong between population-weighted indexes and municipalities-weighted ones, but within each methodology seems fairly stable and consistent. Results for other months deliver a substantially identical picture and are available under request.

We complete the illustration in Table 3 pairing our spatio-temporal price indices for Coffee, computed with the TiPRD equation as in (5) taking as normalization prices in Lazio region in December 2021 with coverage values for each month and region. We can note a marked upward trend in prices starting in 2022, clearly showed in Figure 7.

The spatio-temporal indexes calculated in our exercise reflect the specific composition of retailers in our web scraping operations, as well as the addition and termination of data sources. In our specific example, different retailers may have very different positioning and geographical presence. The main purpose of the illustration in Table 3 is to show the pattern of structural breaks in each Region when there is a structural break in the coverage index

Finally, we calculated the correlation between the trend change point probability for the coverage and spatio-temporal price index level time series for each region. The KPSS test failed to reject the null hypothesis of stationarity in all cases. Results from the Pearson correlation test are presented in Table 5.

Table 1: Coverage value in December 2021 by Region according to different specifications and parameters.

	Municipalities								Population							
	Sigmoid				Linear				Sigmoid				Linear			
Region	20	30	40	50	20	30	40	50	20	30	40	50	20	30	40	50
Abruzzo	0.705	0.568	0.416	0.262	0.633	0.546	0.418	0.250	0.910	0.845	0.754	0.648	0.843	0.805	0.744	0.641
Basilicata	0.166	0.119	0.080	0.050	0.188	0.130	0.083	0.046	0.355	0.297	0.249	0.207	0.376	0.307	0.253	0.203
Calabria	0.482	0.364	0.252	0.149	0.466	0.375	0.258	0.141	0.631	0.549	0.470	0.397	0.619	0.555	0.474	0.392
Campania	0.590	0.492	0.383	0.264	0.557	0.475	0.373	0.261	0.861	0.796	0.703	0.580	0.796	0.750	0.681	0.580
Emilia-Romagna	0.585	0.477	0.364	0.252	0.553	0.468	0.365	0.244	0.751	0.674	0.588	0.503	0.724	0.664	0.585	0.497
Friuli-Venezia Giulia	0.748	0.629	0.484	0.322	0.666	0.588	0.476	0.314	0.698	0.601	0.481	0.336	0.637	0.551	0.468	0.331
Lazio	0.526	0.394	0.271	0.166	0.520	0.413	0.281	0.156	0.874	0.831	0.785	0.741	0.860	0.828	0.786	0.738
Liguria	0.431	0.312	0.204	0.113	0.443	0.333	0.212	0.104	0.736	0.657	0.584	0.522	0.730	0.667	0.593	0.515
Lombardia	0.600	0.479	0.348	0.217	0.562	0.466	0.346	0.208	0.813	0.713	0.586	0.445	0.736	0.673	0.579	0.437
Marche	0.516	0.387	0.261	0.147	0.496	0.393	0.269	0.137	0.728	0.604	0.465	0.334	0.658	0.583	0.474	0.316
Molise	0.265	0.188	0.119	0.062	0.275	0.202	0.124	0.057	0.331	0.285	0.243	0.204	0.343	0.289	0.244	0.202
Piemonte	0.207	0.148	0.094	0.047	0.216	0.156	0.097	0.043	0.472	0.419	0.360	0.299	0.458	0.409	0.357	0.295
Puglia	0.318	0.235	0.162	0.102	0.337	0.250	0.168	0.097	0.536	0.457	0.385	0.323	0.543	0.471	0.387	0.318
Sardegna	0.442	0.356	0.267	0.182	0.437	0.352	0.265	0.176	0.659	0.602	0.538	0.473	0.646	0.592	0.534	0.468
Sicilia	0.268	0.192	0.125	0.067	0.278	0.206	0.131	0.064	0.503	0.448	0.392	0.339	0.491	0.448	0.395	0.335
Toscana	0.445	0.336	0.233	0.145	0.446	0.351	0.239	0.137	0.718	0.637	0.551	0.470	0.699	0.636	0.552	0.464
Umbria	0.776	0.657	0.514	0.364	0.698	0.623	0.508	0.352	0.946	0.907	0.851	0.788	0.906	0.882	0.845	0.782
Valle d'Aosta	0.453	0.337	0.230	0.134	0.466	0.353	0.233	0.127	0.496	0.347	0.229	0.123	0.510	0.394	0.230	0.120
Veneto	0.700	0.590	0.455	0.309	0.629	0.550	0.450	0.298	0.743	0.622	0.480	0.328	0.669	0.589	0.474	0.316

Figure 3: Coverage representation in December 2021 - Linear membership function - D: 20 min.

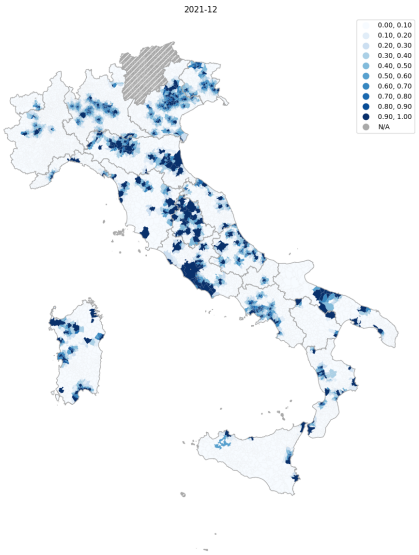


Figure 4: Coverage representation in December 2021 - Linear membership function - D: 50 min.

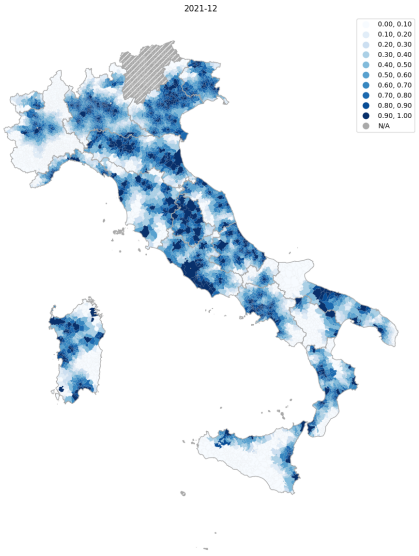


Figure 5: Coverage representation in December 2021 - Sigmoid membership function - D: 20 min.

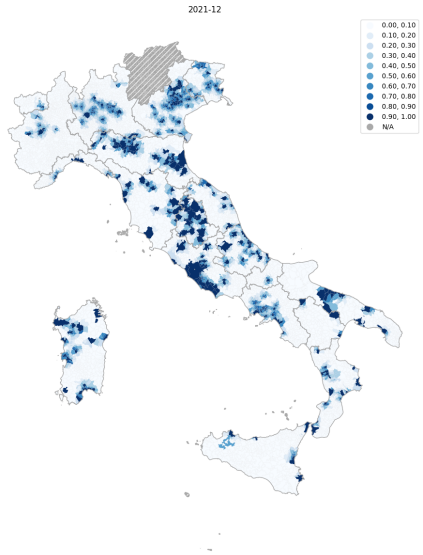


Figure 6: Coverage representation in December 2021 - Sigmoid membership function - D: 50 min.

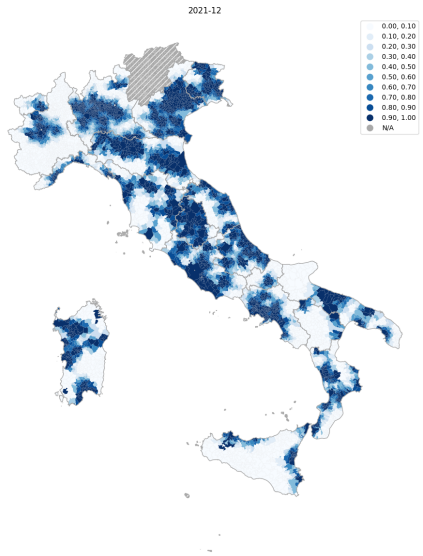


Table 2: Spearman Rank Correlation Test between pairs of membership functions.

W			Municipalities								Population							
	Fun		Sigmoid				Linear				Sigmoid				Linear			
		Min	20	30	40	50	20	30	40	50	20	30	40	50	20	30	40	50
Mun	Sig	20	1.000	0.993	0.981	0.956	0.998	0.988	0.974	0.946	0.604	0.730	0.737	0.795	0.589	0.670	0.698	0.732
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.003)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.004)*	*(0.001)*	*(0.000)*	*(0.000)*
		30	0.993	1.000	0.991	0.974	0.989	0.998	0.986	0.965	0.611	0.744	0.763	0.823	0.593	0.688	0.718	0.756
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.003)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.004)*	*(0.001)*	*(0.000)*	*(0.000)*
		40	0.981	0.991	1.000	0.991	0.974	0.995	0.996	0.986	0.570	0.711	0.747	0.818	0.549	0.658	0.696	0.746
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.005)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.007)*	*(0.001)*	*(0.000)*	*(0.000)*
		50	0.956	0.974	0.991	1.000	0.951	0.981	0.991	0.998	0.558	0.705	0.753	0.821	0.537	0.649	0.695	0.753
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.007)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.009)*	*(0.001)*	*(0.000)*	*(0.000)*
	Lin	20	0.998	0.989	0.974	0.951	1.000	0.982	0.967	0.940	0.612	0.739	0.740	0.793	0.602	0.675	0.704	0.735
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.003)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.003)*	*(0.001)*	*(0.000)*	*(0.000)*
		30	0.988	0.998	0.995	0.981	0.982	1.000	0.991	0.974	0.600	0.737	0.765	0.828	0.579	0.684	0.714	0.758
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.003)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.005)*	*(0.001)*	*(0.000)*	*(0.000)*
		40	0.974	0.986	0.996	0.991	0.967	0.991	1.000	0.988	0.554	0.695	0.735	0.809	0.533	0.642	0.679	0.735
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.007)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.009)*	*(0.002)*	*(0.001)*	*(0.000)*
		50	0.946	0.965	0.986	0.998	0.940	0.974	0.988	1.000	0.561	0.709	0.760	0.828	0.540	0.656	0.700	0.761
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.006)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.008)*	*(0.001)*	*(0.000)*	*(0.000)*
Pop	Sig	20	0.604	0.611	0.570	0.558	0.612	0.600	0.554	0.561	1.000	0.961	0.909	0.860	0.993	0.963	0.939	0.893
			(0.003)	*(0.003)*	*(0.005)*	*(0.007)*	*(0.003)*	*(0.003)*	*(0.007)*	*(0.006)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		30	0.730	0.744	0.711	0.705	0.739	0.737	0.695	0.709	0.961	1.000	0.970	0.942	0.954	0.982	0.975	0.954
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		40	0.737	0.763	0.747	0.753	0.740	0.765	0.735	0.760	0.909	0.970	1.000	0.988	0.895	0.977	0.989	0.991
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		50	0.795	0.823	0.818	0.821	0.793	0.828	0.809	0.828	0.860	0.942	0.988	1.000	0.842	0.947	0.965	0.982
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
	Lin	20	0.589	0.593	0.549	0.537	0.602	0.579	0.533	0.540	0.993	0.954	0.895	0.842	1.000	0.949	0.928	0.879
			(0.004)	*(0.004)*	*(0.007)*	*(0.009)*	*(0.003)*	*(0.005)*	*(0.009)*	*(0.008)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		30	0.670	0.688	0.658	0.649	0.675	0.684	0.642	0.656	0.963	0.982	0.977	0.947	0.949	1.000	0.988	0.967
			(0.001)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.001)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		40	0.698	0.718	0.696	0.695	0.704	0.714	0.679	0.700	0.939	0.975	0.989	0.965	0.928	0.988	1.000	0.988
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.001)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*
		50	0.732	0.756	0.746	0.753	0.735	0.758	0.735	0.761	0.893	0.954	0.991	0.982	0.879	0.967	0.988	1.000
			(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*	*(0.000)*

Table 3: Spatio-temporal price index levels for Coffee (ECOICOP 01.2.1.1).
In parenthesis coverage values calculated with the Sigmoid function and weighted by pouplation (D= 30 min)

Month	Abruzzo	Basilicata	Calabria	Campania	Emilia-Romagna	Friuli-Venezia Giulia	Lazio	Liguria	Lombardia	Marche	Molise
2020-11	104.55						102.88			102.00	101.79
	(0.668)						*(0.562)*			*(0.207)*	*(0.243)*
2020-12	105.60		97.00	97.29	101.43		101.10	102.52	99.48	103.56	105.03
	(0.695)		*(0.272)*	*(0.477)*	*(0.122)*		*(0.690)*	*(0.498)*	*(0.359)*	*(0.207)*	*(0.243)*
2021-01	105.49		101.48	100.30	101.92		101.75	101.64	99.88	103.46	104.56
	(0.700)		*(0.272)*	*(0.512)*	*(0.122)*		*(0.690)*	*(0.498)*	*(0.409)*	*(0.207)*	*(0.243)*
2021-02	104.99	101.86	100.70	99.68	101.17	99.44	101.00	103.34	100.08	101.11	104.77
	(0.700)	*(0.177)*	*(0.369)*	*(0.556)*	*(0.510)*	*(0.497)*	*(0.725)*	*(0.567)*	*(0.470)*	*(0.574)*	*(0.243)*
2021-03	104.98	102.21	100.98	99.80	100.57	99.79	100.29	103.11	99.67	101.57	104.31
	(0.700)	*(0.177)*	*(0.369)*	*(0.561)*	*(0.510)*	*(0.497)*	*(0.725)*	*(0.567)*	*(0.470)*	*(0.574)*	*(0.243)*
2021-04	104.73	101.62	101.33	99.73	101.12	100.00	101.15	103.25	100.24	101.75	104.84
	(0.700)	*(0.177)*	*(0.369)*	*(0.561)*	*(0.522)*	*(0.497)*	*(0.739)*	*(0.567)*	*(0.470)*	*(0.574)*	*(0.243)*
2021-05	104.96	101.86	101.79	100.08	101.19	99.72	100.96	102.47	100.28	101.90	103.73
	(0.674)	*(0.177)*	*(0.369)*	*(0.521)*	*(0.523)*	*(0.497)*	*(0.739)*	*(0.584)*	*(0.470)*	*(0.574)*	*(0.243)*
2021-06	104.87	102.33	101.20	100.72	101.67	100.30	100.80	102.80	99.93	101.91	103.02
	(0.674)	*(0.307)*	*(0.369)*	*(0.525)*	*(0.488)*	*(0.509)*	*(0.739)*	*(0.584)*	*(0.465)*	*(0.574)*	*(0.243)*
2021-07	102.45	100.31	100.52	97.72	98.50	99.80	100.10	101.49	99.49	101.87	102.97
	(0.758)	*(0.425)*	*(0.470)*	*(0.703)*	*(0.678)*	*(0.509)*	*(0.801)*	*(0.584)*	*(0.617)*	*(0.574)*	*(0.243)*
2021-08	102.52	100.76	100.16	97.57	98.49	99.29	100.03	101.88	99.91	101.41	102.97
	(0.758)	*(0.425)*	*(0.470)*	*(0.703)*	*(0.678)*	*(0.509)*	*(0.791)*	*(0.584)*	*(0.617)*	*(0.574)*	*(0.243)*
2021-09	102.72	100.59	100.46	96.86	97.73	100.14	99.86	102.62	99.27	101.76	102.97
	(0.743)	*(0.425)*	*(0.470)*	*(0.703)*	*(0.678)*	*(0.509)*	*(0.791)*	*(0.584)*	*(0.617)*	*(0.574)*	*(0.243)*
2021-10	102.67	101.08	99.33	97.26	97.61	98.98	99.87	102.26	99.73	101.20	102.97
	(0.754)	*(0.425)*	*(0.470)*	*(0.703)*	*(0.678)*	*(0.509)*	*(0.791)*	*(0.584)*	*(0.617)*	*(0.574)*	*(0.243)*
2021-11	102.99	100.58	100.01	97.55	97.04	100.36	100.26	103.02	98.78	101.91	103.12
	(0.754)	*(0.263)*	*(0.470)*	*(0.703)*	*(0.678)*	*(0.509)*	*(0.791)*	*(0.584)*	*(0.617)*	*(0.574)*	*(0.243)*
2021-12	102.86	100.46	99.42	96.91	101.18	100.55	100.00	103.37	99.68	102.44	104.11
	(0.754)	*(0.249)*	*(0.470)*	*(0.703)*	*(0.588)*	*(0.481)*	*(0.785)*	*(0.584)*	*(0.586)*	*(0.465)*	*(0.243)*
2022-01	103.46	100.79	100.55	97.99	100.59		101.33	101.56	99.81		105.09
	(0.728)	*(0.351)*	*(0.414)*	*(0.660)*	*(0.275)*		*(0.768)*	*(0.498)*	*(0.586)*		*(0.243)*
2022-02	103.68	102.02	100.69	99.04	100.74		101.87	101.74	100.13		103.52
	(0.662)	*(0.369)*	*(0.414)*	*(0.687)*	*(0.275)*		*(0.768)*	*(0.498)*	*(0.586)*		*(0.243)*
2022-03	103.78	102.41	100.71	99.12	100.81		102.39	102.70	100.98		103.52
	(0.567)	*(0.430)*	*(0.414)*	*(0.710)*	*(0.275)*		*(0.768)*	*(0.498)*	*(0.586)*		*(0.243)*
2022-04	103.76	102.20	101.33	100.18	101.29		103.16	102.64	101.18		
	(0.458)	*(0.430)*	*(0.414)*	*(0.710)*	*(0.275)*		*(0.768)*	*(0.498)*	*(0.586)*		
2022-05	104.52	103.26	101.62	100.73	101.61		103.59	102.55	101.85		
	(0.458)	*(0.312)*	*(0.414)*	*(0.710)*	*(0.275)*		*(0.768)*	*(0.498)*	*(0.586)*		
2022-06	104.71	102.79	101.82	100.47	102.00		103.95	101.57	102.59		
	(0.458)	*(0.430)*	*(0.414)*	*(0.702)*	*(0.275)*		*(0.768)*	*(0.395)*	*(0.586)*		
2022-07	103.66	103.58	101.69	100.59	102.15		103.10	102.27	102.62		
	(0.387)	*(0.188)*	*(0.414)*	*(0.660)*	*(0.275)*		*(0.749)*	*(0.395)*	*(0.586)*		
2022-08	105.43		101.05	102.44	103.66		104.37	103.08	103.39		
	(0.158)		*(0.378)*	*(0.635)*	*(0.275)*		*(0.742)*	*(0.395)*	*(0.586)*		
2022-09	108.56		102.43	103.01	103.25		105.25	103.77	103.17		
	(0.158)		*(0.378)*	*(0.635)*	*(0.275)*		*(0.742)*	*(0.395)*	*(0.586)*		
2022-10	104.02		101.07	102.38	105.36		105.12	104.05	105.23		
	(0.158)		*(0.378)*	*(0.635)*	*(0.275)*		*(0.742)*	*(0.395)*	*(0.586)*		
2022-11	109.41		103.45	104.56	106.13		106.43	104.47	105.90		
	(0.158)		*(0.378)*	*(0.635)*	*(0.275)*		*(0.742)*	*(0.395)*	*(0.586)*		
2022-12	109.34		102.05	103.33	105.28		106.33	104.18	105.44		
	(0.158)		*(0.378)*	*(0.635)*	*(0.275)*		*(0.742)*	*(0.395)*	*(0.586)*		
2023-01	108.23		103.33	104.47	106.67		107.47	106.19	106.52		
	(0.158)		*(0.372)*	*(0.617)*	*(0.187)*		*(0.728)*	*(0.395)*	*(0.560)*		
2023-02	108.66		103.37	104.82	104.61		107.77	107.04	104.45		
	(0.158)		*(0.372)*	*(0.617)*	*(0.187)*		*(0.728)*	*(0.395)*	*(0.560)*		

Table 4: Continue from Table 3.

Month	Piemonte	Puglia	Sardegna	Sicilia	Toscana	Umbria	Valle D'Aosta	Veneto
2020-11		94.83						
		(0.106)						
2020-12		90.19	104.02	102.80	100.39	98.09		107.55
		(0.120)	*(0.406)*	*(0.164)*	*(0.523)*	*(0.283)*		*(0.122)*
2021-01		91.36	103.42	104.58	100.06	100.79		
		(0.120)	*(0.423)*	*(0.164)*	*(0.523)*	*(0.283)*		*(0.122)*
2021-02	105.33	94.83	102.14	102.73	100.11	99.47	110.70	98.13
	(0.403)	*(0.120)*	*(0.542)*	*(0.397)*	*(0.563)*	*(0.283)*	*(0.653)*	*(0.499)*
2021-03	105.47	95.45	102.12	103.07	100.52	98.47	110.76	98.70
	(0.403)	*(0.120)*	*(0.542)*	*(0.397)*	*(0.563)*	*(0.283)*	*(0.653)*	*(0.499)*
2021-04	105.58	96.39	102.75	102.65	100.57	99.87	110.20	98.83
	(0.403)	*(0.120)*	*(0.542)*	*(0.397)*	*(0.563)*	*(0.283)*	*(0.653)*	*(0.499)*
2021-05	104.92	96.45	102.69	101.76	100.47	99.38	109.80	98.65
	(0.403)	*(0.120)*	*(0.542)*	*(0.397)*	*(0.563)*	*(0.283)*	*(0.653)*	*(0.499)*
2021-06	105.21	95.29	102.56	102.27	100.43	99.10	111.18	99.66
	(0.403)	*(0.106)*	*(0.542)*	*(0.361)*	*(0.563)*	*(0.283)*	*(0.653)*	*(0.487)*
2021-07	99.79	94.62	102.17	102.38	100.02	98.52	109.80	94.19
	(0.551)	*(0.398)*	*(0.572)*	*(0.460)*	*(0.591)*	*(0.851)*	*(0.653)*	*(0.565)*
2021-08	98.31	94.33	102.28	102.86	99.72	98.42	110.31	97.10
	(0.554)	*(0.362)*	*(0.572)*	*(0.460)*	*(0.591)*	*(0.851)*	*(0.653)*	*(0.541)*
2021-09	105.76	94.26	101.39	102.23	99.88	98.67	110.43	96.82
	(0.403)	*(0.362)*	*(0.561)*	*(0.460)*	*(0.591)*	*(0.851)*	*(0.653)*	*(0.541)*
2021-10	104.96	93.72	102.31	102.36	100.00	98.96	109.46	96.45
	(0.403)	*(0.362)*	*(0.572)*	*(0.460)*	*(0.591)*	*(0.851)*	*(0.653)*	*(0.541)*
2021-11	105.24	95.03	102.72	103.23	100.04	99.69	110.26	97.21
	(0.403)	*(0.385)*	*(0.572)*	*(0.460)*	*(0.591)*	*(0.851)*	*(0.653)*	*(0.541)*
2021-12	105.19	95.53	103.22	102.23	99.96	99.77	107.21	99.91
	(0.360)	*(0.385)*	*(0.538)*	*(0.392)*	*(0.551)*	*(0.851)*	*(0.229)*	*(0.480)*
2022-01		95.55	104.31	104.97	101.19	101.77		
		(0.292)	*(0.423)*	*(0.399)*	*(0.544)*	*(0.839)*		
2022-02		96.53	103.96	106.24	100.53	101.75		
		(0.269)	*(0.423)*	*(0.428)*	*(0.523)*	*(0.839)*		
2022-03		97.43	105.40	107.65	102.09	102.06		
		(0.269)	*(0.423)*	*(0.428)*	*(0.523)*	*(0.839)*		
2022-04		97.90	106.80	108.82	103.94	102.30		
		(0.269)	*(0.423)*	*(0.428)*	*(0.523)*	*(0.839)*		
2022-05		98.42	107.59	109.29	103.11	101.92		
		(0.269)	*(0.423)*	*(0.428)*	*(0.404)*	*(0.839)*		
2022-06		98.34	106.73	109.19	100.68	101.57		
		(0.269)	*(0.423)*	*(0.428)*	*(0.138)*	*(0.839)*		
2022-07		101.22		109.40		101.62		
		(0.254)		*(0.428)*		*(0.839)*		
2022-08				109.27		103.54		
				(0.428)		*(0.839)*		
2022-09				109.35		105.23		
				(0.428)		*(0.839)*		
2022-10				109.98		104.79		
				(0.428)		*(0.839)*		
2022-11				110.80		106.34		
				(0.428)		*(0.839)*		
2022-12				110.07		105.68		
				(0.428)		*(0.839)*		
2023-01				111.60		107.20		
				(0.428)		*(0.839)*		
2023-02				113.02				
				(0.428)				

Figure 7: Spatio-temporal price index levels for Coffee (ECOICOP 01.2.1.1).

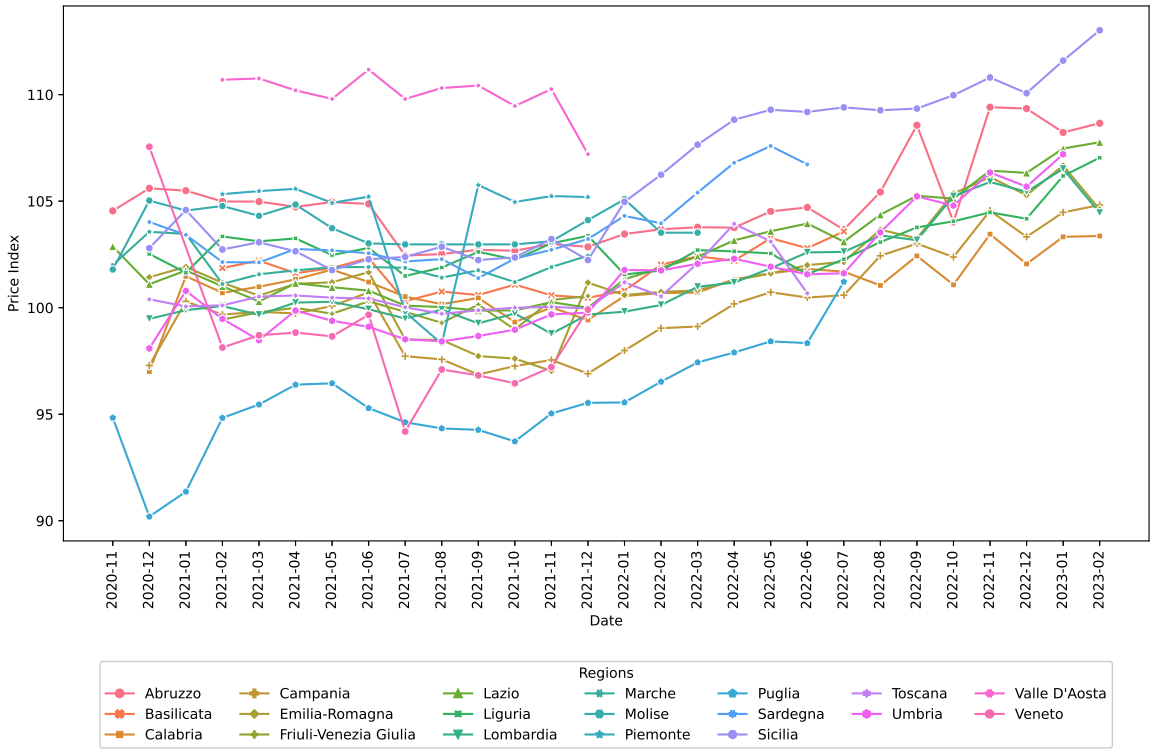


Table 5: Correlation between structural breaks in coverage and spatio-temporal price index level time series.

Region	Correlation	p-value
Abruzzo	0.359	*(0.061)*
Basilicata	0.399	*(0.101)*
Calabria	0.142	*(0.481)*
Campania	0.735	*(0.000)*
Emilia-Romagna	0.010	*(0.961)*
Friuli-Venezia Giulia	1.000	*(0.000)*
Lazio	-0.078	*(0.695)*
Liguria	0.410	*(0.034)*
Lombardia	-0.048	*(0.813)*
Marche	0.815	*(0.000)*
Molise	0.993	*(0.000)*
Piemonte	0.994	*(0.000)*
Puglia	0.300	*(0.186)*
Sardegna	-0.047	*(0.848)*
Sicilia	-0.033	*(0.868)*
Toscana	0.954	*(0.000)*
Umbria	0.554	*(0.003)*
Valle d'Aosta	0.999	*(0.000)*
Veneto	0.919	*(0.000)*

We can see that in 10 cases out of 19 we are able to identify a significant and strong positive correlation between the two time series. Therefore, we can maintain that structural breaks in coverage may effectively impact the stability of the measured price index level.

5 Conclusions and future research

We believe coverage information is a relevant metric for price statistics. Modelling accurately where price collection takes place considering consumer purchasing habits and travel distance. Embedding this information in CPIs can provide tremendous insights at several levels in the price statistics compilation and utilization process.

During selection and sampling of outlets for price collection it would be important to have an accurate view of geographical and population coverage in order to make sure that no dark spot is left systematically in price surveys and there is continuity and consistent overlap over time for the covered area. As demonstrated, substantial changes in coverage between different period may effectively impact the stability of price index measurement.

When using price statistics this coverage view would be equally important. Local dynamics in economic and social measures are object of a growing number of studies, and granular coverage information could help to better integrate price statistics in this stream of research.

One of the main points for future improvement is the ability to correctly identify and match product across multiple retailers when GTINs are not available. In our case, less than half of the unique products had a GTIN associated. It is quite likely that amongst the others there will be matching products, but the large number of unique products - even in the very limited perimeter we selected for this exercise - combined with the sheer similarity in naming across different products makes manual vetting a complex and time consuming task on the one side, and tricky for automatic matching algorithms on the other. In this area we plan to explore the use of Large Language Models, as those tools have extended language capabilities and the potential to leverage context information and prompting specific for the task.

Finally, other services may be explored for obtaining updated travel time calculations in the future, such as Google Distance Matrix API or TravelTime API. Furthermore, we foresee additional application for a measure of information decay over space or travel time as presented in this work for other fields beyond price statistics.

References

Aizcorbe, A., & Aten, B. (2004). An approach to pooled time and space comparisons..

Aizcorbe, A., Corrado, C., & Doms, M. (2000). Constructing price and quantity indexes for high technology goods. *Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System, July*, *19*.

Aten, B. (1996). Evidence of spatial autocorreilation in international prices. *Review of Income and Wealth*, *42*(2), 149–163.

Benedetti, I., Laureti, T., Palumbo, L., & Rose, B. M. (2022). Computation of high-frequency sub-national spatial consumer price indexes using web scraping techniques. *Economies*, *10*(4), 95.

Berry, F., Graf, B., Stanger, M. M., & Ylä-Jarkko, M. (2019). Price statistics compilation in 196 economies: The relevance for policy analysis. *International Monetary Fund Working Papers*, *2019*. DOI: https://doi.org/10.5089/9781513508313.001

Biggeri, L., Laureti, T., & Polidoro, F. (2017). Computing sub-national PPPs with CPI data: an empirical analysis on Italian data using country product dummy models. *Social Indicators Research*, *131*(1), 93–121.

Brunetti, A., Fatello, S., Polidoro, F., & Simone, A. (2018). Improvements in Italian CPI/HICP deriving from the use of scanner data. In *50th scientific meeting of the italian statistical society.* Retrieved from http://meetings3.sis-statistica.org/index.php/sis2018/50th/paper/viewFile/1484/32

Corrado, C., & Ukhaneva, O. (2016). Hedonic prices for fixed broadband services: estimation across oecd countries. *OECD Science, Technology and Industry Working Papers*(2016/07). DOI: https://doi.org/https://doi.org/10.1787/5jlpl4sgc9hj-en"

De Haan, J., & Krsinich, F. (2014). Scanner data and the treatment of quality change in nonrevisable price indexes. *Journal of Business & Economic Statistics*, *32*(3), 341–358.

Diewert, W. (2021). Elementary indexes. *Consumer Price Index Theory*.

Eurostat. (2020). *Practical Guidelines on Web Scraping for the HICP.* Retrieved 2022-04-20, from https://ec.europa.eu/eurostat/documents/272892/12032198/Guidelines-web-scraping-HICP-11-2020.pdf

Guerreiro, V., Baer, M. A., & Silungwe, A. (2022). *The availability, methodological soundness, and scope of consumer price statistics in 2020.* International Monetary Fund.

Hawkes, W. J., & Piotrowski, F. W. (2003). Using scanner data to improve the quality of measurement in the consumer price index. In *Scanner data and price indexes* (pp. 17–38). University of Chicago Press.

Hill, R. J., & Timmer, M. P. (2006). Standard errors as weights in multilateral price indexes. *Journal of Business & Economic Statistics*, *24*(3), 366–377.

International Monetary Fund. (2020). *Consumer price index manual.* London, England: International Monetary Fund.

Istat. (2019). *Matrici di contiguità, distanza e pendolarismo.* Retrieved 2022-04-20, from https://www.istat.it/it/archivio/157423

Kerr, J., Lawrence, F., Sallis, J. F., Saelens, B., Glanz, K., & Chapman, J. (2012). Predictors of trips to food destinations. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(58). DOI: 10.1186/1479-5868-9-58

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159-178.

Laureti, T., & Rao, D. P. (2018). Measuring spatial price level differences within a country: Current status and future developments. *Studies of Applied Economics*, *36*(1), 119–148.

Montero, J.-M., Laureti, T., Mínguez, R., & Fernández-Avilés, G. (2020). A stochastic model with penalized coefficients for spatial price comparisons: An application to regional price indexes in Italy. *Review of Income and Wealth*, *66*(3), 512–533.

Rao, D. S. P. (2001). Weighted EKS and generalised CPD methods for aggregation at basic heading level and above basic heading level. In *Joint World Bank-OECD seminar on purchasing power parities, recent advances in methods and applications.* Washington DC.

Smith, P. A. (2021). Estimating sampling errors in consumer price indices. *International Statistical Review*, *89*(3), 481–504.

Summers, R. (1973). International price comparisons based upon incomplete data. *Review of Income and Wealth*, *19*(1), 1–16.

Zadeh, L. A. (1977). Fuzzy sets and their application to pattern classification and clustering analysis. In *Classification and clustering* (pp. 251–299). Elsevier.

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., ... Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment*, *232*, 111181.

Zimmermann, H.-J. (2011). *Fuzzy set theory—and its applications.* Springer Science & Business Media.