Price Dispersion: The Case of "Pasta"

Isabella Carbonaro, Raffaele Santioni, and Margherita Carlucci

1 Introduction and Problem Definition

The aim of our research is to explore the possibility of utilizing scanner data on pasta purchases to build bilateral and multilateral spatial price indexes, taking a binary approach in the latter.[1]

Pasta plays a major role in the Italian diet. Historically, pasta consumption was mainly concentrated in the Southern regions of the country but today pasta is perhaps the product most representative of the eating habits of the Italians. The range of pasta producers runs from firms of longstanding tradition (some of them mainly directed towards local markets, such as Mastromauro in Puglia) to well known international brands (such as Barilla and De Cecco).

The marked increase in pasta prices over the last two years has aroused great interest, but with little focus on spatial price diversity.

This study stems from the availability of an extremely detailed panel dataset (Nielsen data) on values and quantities of pasta purchased. This data was produced by the use of bar-code scanning at retail outlets and thus includes information which provides weights at an elementary level. The use of scanner data to construct price indexes is not new in literature and there is a widespread consensus on the advantages of this approach in achieving more representative indexes. Average prices (unit values) show a marked spatial price variability: even when only considering the five bestselling products, regional prices vary greatly.

The paper is set out as follows: Sect. 2 provides a description of the pasta scanner dataset and briefly looks for price variability; in Sect. 3 the requirements of comparability and representativity in the case of pasta are discussed; Sect. 4 deals with the methods and formulas chosen to obtain indexes for the regional comparisons of

I. Carbonaro (✉)
DET, University of Rome Tor Vergata, Rome, Italy
e-mail: isabella.carbonaro@uniroma2.it

[1]This research was funded by PRIN 2005. The Authors thank participants in the International Workshop on Price Index Numbers in Time and Space, held in Florence, September 29th 2008, for useful comments. The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy.

L. Biggeri, G. Ferrari (eds.), *Price Indexes in Time and Space*, Contributions to Statistics, DOI 10.1007/978-3-7908-2140-6_8, © Springer-Verlag Berlin Heidelberg 2010

prices; Sect. 5 shows empirical results; in Sect. 6 a brief conclusion and suggestions for future work are given.

2 Pasta Scanner Dataset Description and Price Dispersion

In Italy, A.C. Nielsen has developed a database of scanner data at a regional level for the sales of major supermarket chains. Product groups included are those typically sold in supermarkets (food, beverages and a large number of other commodities for personal care, home cleaning, etc.). The data used results from the aggregation of the information from all the supermarkets examined. Individual product models (hereafter called "items") are identified by a unique product code based on the bar code. If bar codes correspond to a single product configuration, we can be sure that the product match based upon the code is exact.

In the case of pasta, the data covers the entire commodity group, including all types of pasta products: fresh filled pasta, fresh semola pasta, fresh egg pasta, dried filled pasta, dried semola pasta, dried egg pasta and other types. In this paper we concentrate on the sub-group "dried semola pasta/other" (category n. 499 in the Nielsen data). For this sub-group, spatial price variability is analyzed via the construction of spatial price indexes. For the selected category, scanner panel data provide information on 8,071 items of pasta identified by a product code (hereafter CPR). For each item, information on product characteristics (type of pasta as long or short, brand name, weight and variety of pasta e.g. *spaghetti*, *penne*, etc.) is also provided. In Italian, the latter characteristic is called *formato* or *trafila*. The timespan covers a period of 217 weeks (from June 2002 to July 2006). Spatial coverage is 17 regions (all Italian regions with the exception of Basilicata, Valle d'Aosta and Molise).

For each CPR at a temporal and regional level we have weekly sales, quantity and number of packages purchased. Prices are not included: monthly average revenue per unit sold has been used as a proxy. The whole data set consists of about 6 million items. Since the codes differ according to the different weights of packages, we only selected items weighting 500 grams. Some CPR lack a product characteristic description. These CPR were excluded and only 5,618 CPR (or items) were used. Weekly data were aggregated into 50 sets of monthly data. In the following tables, the data for the December of each year (2002, 2003, 2004 and 2005) is used. December 2006 is excluded because of the availability of data only until July 2006.

Scanner data provide highly detailed information which permits the investigation of price differences of the same basket of products sold in two neighboring or distant regions. Binary comparison may successively be inserted in a multilateral context of spatial price index construction. The use of scanner data is not new in the literature of spatial price indexes (Heravi, Heston, & Silver, 2003), but as far as we are aware, analysis of regional price dispersion of pasta at such a detailed level has never been carried out in Italy[2].

[2]Empirical evidence of regional PPPs at a more aggregated level can be found in De Carli (2008).

Tables 1 and 2 provide some database features at regional and aggregate (Italy) level. In Table 1 the number of CPR selected and quantities of purchases of pasta are shown. As can be observed the availability of the selected items varies greatly across regions. In Liguria, Sicilia and Trentino the number of available items is somewhat low, while in Campania, Emilia Romagna, Lazio, Lombardia, Piemonte and Veneto it exceeds 1,300 items. These differences depend mainly upon the number of inhabitants as well as on the tastes of the consumers. The different relative weights of traditional small outlets and modern distribution chains, as well as the selling policy of the latter, might in part account for the difference.

To extend the analysis of the spatial price variability to a comparable basket of items consumed in the various regions, further study is needed. Nielsen data give information on brands, types (long or short) and variety (spaghetti, penne, etc.) of pasta. As regards the construction of regional price indexes, brand appears to be the most relevant issue. Grouping the CPR by brand, two important results may be achieved. Firstly, the spatial indexes can be assigned to the types of pasta that correspond to the choices of Italians consumers, as in general these choices are influenced by the brand. The second result is that we may obtain a preliminary data set from which a subset can be derived which has the requisites of comparability[3] and representativity[4] required by spatial indexes.

Following these criteria, the CPR have been grouped according to brands: the results are shown in Table 2, where we see that the number of brands available to Italian consumers grew over the period from 2002 to 2006 (212 brands in December 2002 against 247 in the last month of July 2006). As can be seen in Table 2 there are also differences in the number of brands existing in each region, only partially dependent on the different sizes of the Italian regions.

Analysis of the five bestselling brands purchased at a regional level showed that on a monthly basis within the same region there are few changes, while there is much diversity between the brands purchased in the various regions.

At the aggregate level (Italy) the top five brands are the same in all the selected months, with a market share of approximately 71%. The consumption shares of pasta (quantity) goes on average from 67.6% in Campania to 90.5% in Trentino. Price dispersion is also high: a kilogram of pasta of the same brand has different average prices in Italian regions, but there is also a great deal of variability in the prices of the available brands of pasta within each region.

It is worth noting that regional coverage is achieved differently by the five bestselling brands. The lower results for Campania show that brands sold only in the local market attract higher preference in terms of quantity consumed. In the following we refer to this type of brand as local or typical brands. Typically, in each region

[3]Comparability depends on the way of defining each product, as 'product' can cover a large variety of types depending on various characteristics, such as raw materials used, weight, and packaging, all of which affect price. A product must have the same characteristics in order to be strictly comparable over the different areas (Biggeri, Brunetti, & Laureti, 2008).

[4]Representative products are defined here as products that are purchased in relatively large quantities in a country (Hill, 2008).

Table 1 Number of Cpr and quantity of pasta (thousand kilograms) sold in Italian regions

	December 2002		December 2003		December 2004		December 2005		July 2006	
Italian regions	N. CPR	Quantity	N. CPR	Quantity	N. CPR	Quantity	N. CPR	Quantity	N. CPR	Quantity
Abruzzo	1,195	839	1,335	964	1,465	942	1,296	796	1,446	886
Calabria	1,021	1,649	1,102	1,923	1,178	2,000	970	1,547	932	1,664
Campania	1,582	3,249	1,680	4,018	1,674	3,812	1,532	3,044	1,610	2,335
Emilia Romagna	1,371	2,203	1,582	2,629	1,709	2,451	1,415	2,094	1,526	2,053
Friuli V.G.	938	637	1,098	837	1,165	736	1,015	662	1,116	640
Lazio	1,554	3,783	1,700	4,469	1,805	4,051	1,598	3,389	1,773	2,986
Liguria	679	709	772	895	860	838	926	606	982	657
Lombardia	1,538	5,126	1,759	6,211	1,983	5,587	1,772	4,499	1,852	4,305
Marche	1,048	933	1,171	1,110	1,326	1,082	1,177	955	1,321	1,004
Piemonte	1,364	1,860	1,534	2,355	1,684	2,410	1,455	1,996	1,544	1,754
Puglia	1,413	2,397	1,498	2,692	1,497	2,307	1,456	2,007	1,496	2,073
Sardegna	913	1,250	1,093	1,414	1,051	1,457	718	929	802	919
Sicilia	582	491	620	558	648	505	665	418	847	388
Toscana	1,114	2,265	1,287	2,726	1,312	2,728	1,196	2,461	1,222	2,440
Trentino A.A.	423	412	590	592	645	545	573	416	611	406
Umbria	1,065	505	1,170	644	1,097	609	1,014	619	1,049	561
Veneto	1,329	2,424	1,525	2,993	1,655	2,667	1,450	2,340	1,593	2,251
Italia	3,806	30,728	4,031	37,031	4,218	34,726	4,009	28,779	4,163	27,323

Table 2 Number of brands sold in Italian regions

Italian regions	December 2002	December 2003	December 2004	December 2005	July 2006
Abruzzo	61	72	84	86	86
Calabria	52	54	67	57	52
Campania	70	77	78	71	76
Emilia Romagna	98	115	121	113	111
Friuli V.G.	66	85	89	82	79
Lazio	89	99	104	95	97
Liguria	51	61	54	65	72
Lombardia	109	123	138	133	134
Marche	68	80	91	75	82
Piemonte	88	100	113	110	116
Puglia	63	69	74	72	68
Sardegna	64	72	74	50	54
Sicilia	43	49	59	58	59
Toscana	73	85	93	85	90
Trentino A.A.	36	54	52	44	44
Umbria	60	70	69	69	65
Veneto	102	116	131	129	127
Italia	212	226	244	237	247

we see a large number of local brands, but only a few of these have considerable market shares[5].

3 Comparability and Representativity: List of Common Brands and List of Common Products

As is repeatedly emphasized in the literature, the most difficult steps in spatial price index construction are: (i) the preparation of the basket, i.e. a common list of CPR, (ii) compliance with the two important requirements of comparability and representativity which are usually in conflict with one another, (iii) avoiding a severe loss of characteristicity and (iv) the achievement of transitivity. The latter two refer to multilateral spatial price comparisons.

To achieve comparability a common list of types of pasta purchased in all regions must be compiled. Given that, as noted above, pasta consumption patterns differ between Italian regions, the risk of selecting marginally representative kinds of pasta for some regions must be avoided.

[5]If ten brands are involved, as in a study we subsequently developed, the consumption coverage grows on average to 92 percent, for example at December 2002, for all Italian regions. However, the increases show considerable differences between regions: the average increase in coverage passing from five to ten brands is about 11 percentage points, with a maximum of about 21 percentage points in Campania and a minimum in Trentino A.A. of about 5 percent.

As regards representativity, even if the constraint of an equally representative product list for all regions has been relaxed, the issue remains unresolved. Another difficulty stems from the varying meanings that the word "common" can assume in the case of pasta: it can refer to brands, items (CPR) or even to a particular variety of pasta (i.e. spaghetti, penne, etc.).

In our study we chose to select those brands, and the products (CPR) of these common brands, which are common to all regions.

Table 3 summarizes spatial common brands at an aggregate level. The number of common brands varies from 11 to 15. In every period, the bestselling brand is Barilla, followed by Agnesi, De Cecco, Divella, PL (Private Label) and Voiello. These brands, along with Colussi, are present in all the years considered. The numbers of CPR included in common brands are 1,253, 1,429, 1,507, 1,233 (respectively at December 2002, 2003, 2004 and 2005) and 1,251 at July 2006.

At regional level, in all periods common brands cover a substantial share of the aggregate sales of pasta. For example, in July 2006 this share exceeded 90% in 4 regions (Friuli V.G., Liguria, Piemonte and Trentino A.A.), 80% in 8 regions (Abruzzo, Calabria, Emilia Romagna, Lazio, Lombardia, Marche, Sicilia and Veneto) about 75% in Sardegna and 79% in Umbria, attaining minimum values in Puglia (62%) and Campania (63%). In these regions, particularly in Puglia and Campania, the importance of local brands reduces the market shares of common brands.

Table 4 shows the number of common CPR for each region.

Table 3 The common brands list

	Number of brands	Total expenditure (current euro)	Quantity purchased (kilograms)	Share of total quantity sold (%)
December 2002	12	27,236,346	23,700,733	77.1
December 2003	14	33,562,586	28,553,866	77.1
December 2004	15	31,252,208	27,546,265	79.3
December 2005	11	25,999,402	23,562,368	81.9
July 2006	11	24,863,682	22,416,679	82.0

Table 4 The common brands list. Number of CPR purchased

	Abr	Cal	Cam	Emi	Fri	Laz	Lig	Lom	Mar	Pie	Pug	Sar	Sic	Tos	Tre	Umb	Ven
December 2002	711	536	772	788	567	799	472	827	588	854	694	563	416	671	320	694	730
December 2003	879	588	863	940	699	918	541	1003	738	993	769	678	464	790	400	771	908
December 2004	930	673	828	960	679	980	611	1056	792	1019	769	657	458	770	416	743	899
December 2005	723	621	741	771	576	753	613	866	655	813	727	500	441	631	365	561	718
July 2006	770	592	767	839	643	801	638	911	711	860	734	549	578	655	396	605	812

A properly constructed spatial price index would require complete regional series of equal CPR within the common brands. This is not the case, because, even though the analysis is restricted only to products of common brands, observations are incomplete for many CPR, as there are many items that are not purchased (or do not exist) in a given region in the chosen month.

To construct a basket of comparable CPR, two lists of common products produced by common brands have been compiled.

The first list was compiled with the CPR present in the two regions being compared. These binary matches allow the measurement of the spatial variability between every pair of regions. The second list includes the CPR that are present in all regions in the selected months (December of every year from 2002 until 2005 and July 2006). The CPR included in these two lists ensure bilateral (first list) and multilateral (second list) comparability.

Table 5 shows the number of common CPR, in binary-matched cases, for every pair of regions. For the multilateral comparison, the number of selected CPR, common to all regions, is 128, 146, 145, 148 and 179 (at December 2002, 2003, 2004, 2005 and July 2006, respectively).

To comply with the representativity requirements, a slightly modified Sergeev approach (2003) was used. As in Sergeev, the CPR included in the lists are grouped in four subsets: the first includes CPR that are representative in both the regions involved in the comparison; the second contains the CPR representative for one region (the base region) but not for the other (the reference region); the third is formed by the CPR representative for the reference region but not for the base

Table 5 Common brands list. Binary-matched CPR at July 2006

Italian regions	Abr	Cal	Cam	Emi	Fri	Laz	Lig	Lom	Mar	Pie	Pug	Sard	Sic	Tos	Tre	Umb	Ven
Abr	–																
Cal	460	–															
Cam	602	540	–														
Emi	605	423	565	–													
Fri	515	349	478	614	–												
Laz	665	481	621	659	566	–											
Lig	514	321	461	591	497	533	–										
Lom	654	455	610	790	614	676	617	–									
Mar	624	439	571	617	534	642	500	646	–								
Pie	606	408	561	795	629	662	622	807	611	–							
Pug	599	485	628	578	506	625	476	597	589	576	–						
Sar	437	364	436	491	471	498	405	506	478	503	465	–					
Sic	485	413	521	482	438	516	404	518	491	489	510	407	–				
Tos	547	379	500	634	535	592	537	630	566	623	519	455	452	–			
Tre	370	228	306	385	370	376	367	396	354	379	329	275	285	371	–		
Umb	559	342	472	540	473	573	497	554	541	543	486	400	403	515	354	–	
Ven	604	411	551	718	642	654	571	758	615	731	562	515	490	622	397	533	–

region; the fourth subset includes the CPR which are not representative for either region and consequently, because irrelevant for the construction of the bilateral index, were deleted.

The CPR in each region have been considered representative on the basis of their share of total quantity of common CPR sales. Let us illustrate the procedure with an example: in July 2006 respectively 770 and 911 items were purchased in Abruzzo and Lombardia, but only 654 common CPR (see Table 5). Within this subset, a given CPR was considered representative if the quantity of it sold in each of the two regions is more than $1/654 = 0.001529$ (i.e. more than 0.1529%) of the aggregate quantity of common CPR sold in the same region. This threshold corresponds to the hypothesis of uniform distribution among the common CPR, in each region, of the purchased quantity of pasta.

On the basis of this procedure the 654 CPR that are common for Abruzzo and Lombardia have been grouped, as shown in Table 6, part (a), in the three subsets cited above: the first includes 72 representative CPR of both regions, that is the CPR whose "quantity bought" shares exceed the threshold in both regions; the second subset includes 39 representative CPR of Abruzzo (base region) but not for Lombardia; the third subset includes 62 representative CPR of Lombardia but not of Abruzzo. These three subsets (173 CPR) constitute the basket. The remaining 481 CPR are irrelevant for the construction of the spatial indexes, as they are not representative of the two regions considered in our example.

The procedure is the same for the two lists of common products of common brands: CPR with a quantity share above (below) the threshold are representative (unrepresentative) and define the four subsets.

It must be noted that in the first list the threshold changes for each pair of regions compared because the number of common CPR is different in the regions; in the second list the threshold changes only over time because the common list was based on the matched CPR for all regions (the number at July 2006 is always 179 and the threshold is 0.559%) as shown in Table 6, part (b).

Table 6 The representativity subsets. Coverage, number of CPR and threshold (%)

	Abruzzo (base)	Lombardia	No. of CPR	Abruzzo (base)	Lombardia	No. of CPR
**	64.48	64.30	72	65.80	57.48	31
–*	17.04	1.99	39	12.01	2.12	9
*–	4.26	17.38	62	8.77	20.38	26
Total	85.78	83.66	173	86.58	79.98	66
Threshold		0.1529 Part (a)			0.5587 Part (b)	

Note: ** representative items of both regions; –* items representative of base region; *– items representative for comparison region

4 Notation and Methodological Issues

The notation below refers only to the common brands in the selected month. Moreover, notation and methodological issues are mainly offered with respect to the first list, but with minor changes can also be adapted to the second, since the procedure is the same.

Let N_t denote the number of products that are included in the common brand in period t ($t = 1, 2, ..., T$). Let N_{tj} denote the number of products purchased in region j($j = 1, 2, \ldots, m$) in period t. Generally $N_{tj}<N_t$. Let N_{tjk} denote the number of products purchased both in region j and k on the same date. Generally $N_{tjk}<N_{tj}$ and $N_{tjk}<N_{tk}$. N_{tjk} is the number of products forming the first list in the comparison between region j and k. We denote with p^i_{tj} and q^i_{tj} respectively, the average price and quantity of i-th CPR in j-th region ($i = 1,2 \ldots, N_{tj}$; $j = 1,2, \ldots, m$ and $t = 1,2, \ldots,$ T). Only with respect to the items belonging to the binary-matched N_{tjk} we do define the quantity share (%) for each product, $s^i_{tj} = \frac{q^i_{tj}}{\sum_{i=1}^{N_{tjk}} q^i_{tj}}$ in j-th region and $s^i_{tk} = \frac{q^i_{tk}}{\sum_{i=1}^{N_{tjk}} q^i_{tk}}$ k-th region. The threshold corresponding to the hypothesis of a uniform distribution of the purchased quantity of pasta among the common CPR is $thr_{tjk} = \frac{100}{N_{tjk}}$. Obviously, the threshold is the same for each pair of regions compared in the binary-matched case. Each item is representative of j-th region if $s^i_{tj} \geq thr_{tjk}$, the same is true for k-th region if $s^i_{tk} \geq thr_{tjk}$.

We will now define the following subsets of products:

- products that are representative in either region j and k. Their number is denoted by n^{**}_{tjk};
- products representative in base region j, but not in comparison region k: The number is: n^{-*}_{tjk};
- products representative in comparison region k, but not in base region j: number n^{*-}_{tjk}

There are also products that are not representative at all (neither in region j nor in region k) and denote the number of the products of this last subset with n^{--}_{tjk}. These are excluded from the analysis.

Let $w^i_{tj} = \frac{p^i_{tj}q^i_{tj}}{\sum_{i=1}^{N_{tjk}} p^i_{tj}q^i_{tj}}$ and $w^i_{tk} = \frac{p^i_{tk}q^i_{tk}}{\sum_{i=1}^{N_{tjk}} p^i_{tk}q^i_{tk}}$ be the expenditure share of the i-th item, respectively in the j-th and k-th region

The binary elementary indexes are calculated using unit value ratios of all the CPRs belonging to the three subsets identified among the matched CPR whose prices are available in both regions. Since expenditure share weights are available, the binary purchasing power parity between the same currency (euro) in the j-th region (base) vs k-th comparison of regions are calculated by the Törnqvist formula.

$$PPP^{T}_{j,k} = \prod_{i=1}^{n_{tjk}(**)} \left[\left(\frac{p^{i}_{tk}}{p^{i}_{tj}} \right)^{\frac{w^{i}_{tj}+w^{i}_{tk}}{2}} \right] \prod_{i=1}^{n_{tjk}(-*)} \left[\left(\frac{p^{i}_{tk}}{p^{i}_{tj}} \right)^{\frac{w^{i}_{tj}+w^{i}_{tk}}{2}} \right] \prod_{i=1}^{n_{tjk}(*-)} \left[\left(\frac{p^{i}_{tk}}{p^{i}_{tj}} \right)^{\frac{w^{i}_{tj}+w^{i}_{tk}}{2}} \right] \tag{1}$$

In the second list, N_{tjk} is equal for every comparison, because it includes those items which are purchased simultaneously in all regions. In this case we denote the set of matched items as ${}^{AR}N_t$ where "AR" means "all regions". The same is also true for the threshold ${}^{AR}thr_t = \frac{100}{{}^{AR}N_t}$ because the denominator is constant for every pair of regions. With some modification it is possible to rearrange the notations and formula and derive Törnqvist-type binary PPPs, denoted with ${}^{AR}PPP^{T}$. A binary approach to multilateral comparison has subsequently been selected applying the modified GEKS[6] formula (Gini, 1924, 1931; Eltetö & Köves, 1964; Szulc, 1964) to binary Törnqvist PPPs to construct transitive multilateral PPPs (only for the second list):

$$GEKS_{jk} = \left[\prod_{l=1}^{m} {}^{AR}PPP^{T}_{j,l} * {}^{AR}PPP^{T}_{l,k} \right]^{\frac{1}{m}} \tag{2}$$

The GEKS formula is widely used in most countries and by international organizations (Balk, 1996; Hill, 1997; ILO et al., 2004). GEKS is recognized as an attractive method for various reasons, in particular because its property of transitivity[7] is imposed by construction. The GEKS method for multilateral comparisons also has its origins in the property of characteristicity[8]. The modified GEKS refers (i) to the use of Törnqvist binary indexes (Caves, Christensen and Diewert, 1982) instead of Fisher-type formula as originally proposed and (ii) to the choice of the Sergeev procedure to satisfy the requirements of representativity.

5 Results and Discussion

Tables 7, 8 and 9 show the results[9] regarding spatial price diversity of the CPR included in the baskets of the two common lists: bilateral PPPs for the first and modified GEKS for the second. The binary parities of Table 7 (the indexes are base

[6]Gini proposed this method long before the others, so EKS should in fact be termed GEKS and we therefore adopt this acronym.

[7]Transitivity requires that the application of a formula to make a direct comparison between j and k countries should result in the same numerical measure as an indirect comparison between j and k through a link country m (ILO et al., 2004).

[8]This property requires that any set of multilateral comparisons satisfying the transitivity property should retain the essential features of the binary comparisons constructed without the transitivity requirement (ILO et al., 2004).

[9]Due to space restraints, only results for July 2006 are given.

Table 7 The binary Törnqvist PPPs

Italian regions	Abruzzo	Calabria	Campania	Emilia Romagna	Friuli V.G.	Lazio	Liguria	Lombardia	Marche	Piemonte	Puglia	Sardegna	Sicilia	Toscana	Trentino A.A.	Umbria	Veneto
Abruzzo	100	97.22	99.87	98.07	97.89	97.98	101.86	99.17	100.85	100.80	96.40	104.61	109.94	92.36	98.63	92.63	97.67
Calabria	102.86	100	102.93	106.32	103.93	104.34	110.98	106.23	107.37	109.54	98.70	117.88	115.39	100.32	106.16	100.40	103.69
Campania	100.13	97.15	100	100.12	97.38	97.89	102.73	100.32	101.96	101.14	98.12	105.89	109.66	94.82	97.54	93.77	97.90
Emilia Romagna	101.97	94.06	99.88	100	99.86	99.20	102.74	100.79	102.36	101.34	96.45	103.09	111.16	95.23	100.49	94.16	98.67
Friuli V.G.	102.15	96.22	102.69	100.14	100	99.65	103.58	101.49	102.88	101.79	98.97	105.41	111.63	95.08	100.20	94.14	99.32
Lazio	102.06	95.84	102.16	100.80	100.35	100	104.97	101.84	103.20	102.52	98.15	107.51	111.95	95.30	100.58	95.06	99.88
Liguria	98.17	90.11	97.35	97.33	96.54	95.27	100	97.41	98.57	97.27	91.37	99.57	107.76	92.70	97.60	90.51	95.80
Lombardia	100.84	94.13	99.68	99.21	98.53	98.19	102.66	100	101.14	100.58	97.08	104.38	110.73	94.98	99.13	93.09	98.20
Marche	99.15	93.14	98.08	97.70	97.20	96.90	101.45	98.87	100	99.91	95.25	102.63	109.25	92.35	97.81	92.19	96.66
Piemonte	99.21	91.29	98.87	98.67	98.24	97.55	102.80	99.42	100.09	100	93.63	102.78	108.96	94.12	99.20	92.20	97.69
Puglia	103.74	101.31	101.91	103.68	101.04	101.88	109.44	103.01	104.99	106.81	100	112.72	115.23	99.90	101.65	100.29	102.53
Sardegna	95.59	84.83	94.44	97.00	94.87	93.01	100.44	95.80	97.43	97.30	88.72	100	105.68	89.96	93.25	88.84	93.86
Sicilia	90.96	86.66	91.19	89.96	89.58	89.32	92.80	90.31	91.54	91.78	86.79	94.62	100	84.06	90.61	84.71	88.86
Toscana	108.28	99.68	105.47	105.01	105.18	104.93	107.88	105.29	108.29	106.25	100.10	111.16	118.96	100	106.01	99.22	103.99
Trentino A.A.	101.39	94.20	102.52	99.52	99.80	99.42	102.46	100.87	102.24	100.80	98.38	107.24	110.37	94.33	100	92.92	99.30
Umbria	107.96	99.60	106.65	106.21	106.23	105.19	110.49	107.42	108.48	108.45	99.71	112.56	118.05	100.78	107.62	100	105.73
Veneto	102.38	96.44	102.15	101.35	100.69	100.12	104.38	101.83	103.46	102.36	97.53	106.55	112.54	96.16	100.71	94.58	100

Table 8 Ranking of Italian regions due to GEKS PPPs

(1) Toscana	100	(10) Lombardia	107.42
(2) Umbria	100.25	(11) Abruzzo	107.44
(3) Calabria	100.25	(12) Campania	108.7
(4) Puglia	103.11	(13) Piemonte	108.98
(5) Lazio	105.15	(14) Marche	109.09
(6) Veneto	105.53	(15) Liguria	110.8
(7) Friuli Venezia Giulia	105.83	(16) Sardegna	114.61
(8) Trentino	106.11	(17) Sicilia	119.28
(9) Emilia Romagna	106.81		

reversible, so the figures above the principal diagonal are the inverse of the corresponding figures below the same diagonal) show considerable price differences between the pairs of regions. Reading the table by row, the main results are the following:

- Calabria, Puglia, Umbria and Toscana are the regions where prices are lower relative to other regions (it must be stressed that the comparison of Table 7 regards pairs of regions; consequently the figures of the rows cannot be used to build rankings, as can be done with the figures of Table 9). For Calabria and Puglia, 15 indexes of the respective rows are above 100. In Toscana and Umbria more than 14 indexes are over 100.
- As regards the scale of the differences between Calabria and other regions, in 7 of these (Abruzzo, Campania, Friuli-Venezia Giulia, Lazio, Toscana, Umbria and Veneto) prices are higher by less than 5%. In 8 other regions (Emilia-Romagna, Liguria, Lombardia, Marche, Piemonte, Sardegna, Sicilia and Trentino) the difference exceeds 5%, with Piemonte (+9.54%), Liguria (+10.98%), Sicilia (+15.39%) and Sardegna (+17.88%) at the top of the list.
- The results are different for Toscana and Umbria, In many of the other regions, prices are higher than in Toscana by less than 6%; in Abruzzo, Liguria, Marche and Sardegna the difference is between 6 and 11%; in Sicilia 19%. Price differences are greater in Umbria. In 13 regions common basket prices exceed those of Umbria by more than 5% and in 5 of these (Liguria, Marche, Piemonte, Sardegna and Sicilia) by more than 8% (in Sicilia +18%).
- The results for Sicilia and Sardegna must be interpreted with caution, as they may be due in part to two factors: transport costs for the common brands and the exclusion from our analysis, as not being common, of local brands (for instance Tomasello, Piatti and Gallo in Sicilia, Pastificio Cellino Sardegna, Casa del Grano, Pastificio Cellino Sant'Alberto and Pastifici Cagliaritani in Sardegna), which offer the same products as the common brands at lower prices. In July 2006 these local brands accounted for substantial consumption shares (about 13% in Sicilia and 20% in Sardegna).The problem of products that are typical in one region and are not purchased or do not exist in the other regions is common in binary

Table 9 The modified GEKS PPPs

Italian regions	Abruzzo	Calabria	Campania	Emilia Romagna	Friuli V.G.	Lazio	Liguria	Lombardia	Marche	Piemonte	Puglia	Sardegna	Sicilia	Toscana	Trentino A.A.	Umbria	Veneto
Abruzzo	100	93.31	101.17	99.42	98.50	97.87	103.13	99.98	101.53	101.44	95.97	106.67	111.02	93.07	98.76	93.30	98.22
Calabria	107.17	100	108.43	106.55	105.57	104.89	110.53	107.15	108.82	108.71	102.86	114.33	118.99	99.75	105.85	100.00	105.27
Campania	98.84	92.23	100	98.27	97.37	96.73	101.94	98.82	100.36	100.26	94.86	105.44	109.74	92.00	97.62	92.23	97.09
Emilia Romagna	100.59	93.85	101.76	100	99.08	98.44	103.73	100.56	102.13	102.03	96.53	107.30	111.67	93.62	99.34	93.85	98.80
Friuli V.G.	101.52	94.72	102.71	100.93	100	99.35	104.69	101.49	103.07	102.98	97.43	108.29	112.71	94.49	100.26	94.72	99.71
Lazio	102.18	95.34	103.38	101.59	100.65	100	105.38	102.16	103.75	103.65	98.07	109.00	113.44	95.10	100.91	95.34	100.36
Liguria	96.97	90.48	98.10	96.40	95.52	94.90	100	96.95	98.45	98.36	93.06	103.44	107.66	90.25	95.77	90.47	95.24
Lombardia	100.02	93.33	101.19	99.44	98.53	97.89	103.15	100	101.56	101.46	95.99	106.70	111.05	93.10	98.78	93.32	98.24
Marche	98.49	91.90	99.64	97.92	97.02	96.39	101.57	98.47	100	99.90	94.52	105.06	109.35	91.67	97.27	91.89	96.74
Piemonte	98.59	91.99	99.74	98.01	97.11	96.48	101.67	98.56	100.10	100	94.61	105.17	109.45	91.76	97.36	91.98	96.83
Puglia	104.20	97.22	105.42	103.59	102.64	101.97	107.46	104.17	105.79	105.69	100	111.15	115.68	96.98	102.90	97.22	102.34
Sardegna	93.74	87.47	94.84	93.20	92.34	91.74	96.67	93.72	95.18	95.09	89.97	100	104.08	87.25	92.58	87.47	92.07
Sicilia	90.07	84.04	91.12	89.55	88.72	88.15	92.89	90.05	91.45	91.36	86.44	96.08	100	83.83	88.95	84.04	88.47
Toscana	107.44	100.25	108.70	106.81	105.83	105.15	110.80	107.42	109.09	108.98	103.11	114.61	119.28	100	106.11	100.25	105.53
Trentino A.A.	101.26	94.48	102.44	100.67	99.74	99.09	104.42	101.23	102.81	102.71	97.18	108.01	112.42	94.24	100	94.47	99.45
Umbria	107.18	100.00	108.43	106.55	105.57	104.89	110.53	107.15	108.82	108.72	102.86	114.33	118.99	99.75	105.85	100	105.27
Veneto	101.81	95.00	103.00	101.22	100.29	99.64	105.00	101.79	103.37	103.27	97.71	108.61	113.03	94.76	100.55	94.99	100

comparisons. The debate on how to include products that are characteristic in a single region in the analysis (or area) continues (Biggeri et al., 2008).
– The differences of prices in Puglia with other regions are small, being generally less than 5%. Only in Liguria, Piemonte, Sardegna and Sicilia are prices higher than in Puglia by more than 5%, being from 7 to 15%.
– Liguria, Piemonte, Marche, Sardegna and Sicilia are the regions where prices, in binary comparisons, are highest relative to other regions.

Turning to multilateral comparisons, Table 8 shows the non-decreasing ranking of the 17 regions according to GEKS PPPs (Table 9).

The differences are notable. For Sicilian consumers prices are 19% higher than for the inhabitants of Toscana, while for Sardinians they are 14% higher. In 11 regions prices exceed prices in Toscana by more than 5%; in Campania, Liguria, Marche and Piemonte the difference is greater than 8%.

As regards Sardegna and Sicilia, we refer to the considerations above. From the above ranking it appears that in the Northern and Central regions, pasta prices are lower than in the Southern regions.

It may be observed that, within the baskets utilized to calculate GEKS, we have implicitly defined a minimum basket. This is formed by those items which are perfectly comparable (matched in all the regions) and also representative in all the regions (after the applied procedure to construct binary PPPs). This subset is composed of 13 varieties of pasta belonging to 2 brands.

6 Conclusion and Future Work

Pasta is perhaps the product most representative of the eating habits of Italians. Using the extremely rich information provided by AC Nielsen (about 6 million pieces of elementary data) we have calculated binary and multilateral PPPs for 17 Italian regions.

One result emerges. There are vast differences in regional prices: living in Toscana but also in Umbria, Calabria and Puglia entails a remarkable saving in the expenditure for pasta. This is confirmed by both Törnqvist binary parities and GEKS indexes.

In our analysis it was necessary to make many choices, in some cases necessitated by the data available and in others inevitably discretionary.

Due to the characteristics of the available data, we had to use unit values such as average prices and unit value ratios as elementary binary indexes. Other choices relate to (i) the brands, given the preference of the Italian consumers for certain brands, (ii) the products of each brand, (iii) the procedure for representativity, (iv) the thresholds and (v) the index numbers methodology.

The choice regarding the brands (only those present in all the regions have been selected) was suggested by the need to ensure comparability of results and by the high percentage of purchases of pasta produced by the brands present in all regions.

A major problem is the exclusion of local brands. Including them would have extended the control for comparability, now limited to some features of the product (in our case brand, type, variety, weight etc.), to the percentage of purchases. We are currently at work on developing solutions to this problem.

Products were selected according to the various aims of the binary and the multilateral indexes. With the former, the aim was to analyze price differences between pairs of regions in order to explore diversities between contiguous regions or regions of different macro areas. The aim of multilateral indexes is to rank the regions on the basis of prices: this requires that the products considered for the comparison are the same and are present in all regions. Consequently, the basket used is smaller than the one used for binary comparisons.

The procedure for evaluating product representativity follows the Sergeev approach. We modified the procedure both by using the expenditure shares of each CPR, information available from the Nielsen data set, and by determining the thresholds for representativity.

The thresholds were set on the basis of the number of the products and of the quantities purchased in the regions matched. Obviously, these could have been defined on the basis of various criteria (the first 10 or more products, or those representing a given percentage of overall sales, for instance) and more complex methods. In any case, some arbitrariness is unavoidable. The soundness of our choice was confirmed by the fact that the baskets developed by our method proved to be well balanced: the number of representative products for one region and not-representative for the other, and vice versa, do not differ greatly from the total expenditure shares achieved in most cases.

For the construction of multilateral indexes we selected the binary approach, which is widely used in many countries and by international organizations. The use of the Törnqvist index for the calculation of binary parities was suggested, in addition to its characteristic of ideal index, by the Sergeev procedure and by the availability of the expenditure shares. Moreover, the use of the Törnqvist index instead of the Fisher formula for the construction of GEKS indexes is widely accepted in the relevant literature. The need to weight GEKS indexes has often been discussed in the literature (Ferrari & Riani, 1998; Rao, 2001), chiefly when consumption habits or when the general economic conditions of the countries or areas compared are very different. We chose not to use weighting, as we are analysing regions in the same country.

In conclusion, a methodological suggestion derives from the consideration that the selected baskets can be viewed as "nested baskets": (i) CPR perfectly comparable but not always representative (the CPRs inserted in the second list); (ii) representative but no longer perfectly comparable baskets (first list) and iii) a final basket (minimum basket) where the products are both fully comparable and representative. Comparing all these baskets (even with less regions, products or periods) might be useful in order to better evaluate the trade–off between comparability and representativity.

Acknowledgments Data have been kindly provided by CEIS of Tor Vergata University of Rome, whose Chief manager, Prof. Giovanni Tria, we would like to warmly thank.

References

Balk, B. M. (1996). A comparison of ten methods for multilateral international price comparisons. *Journal of Official Statistics, 12*(2), 199–222.

Biggeri, L., Brunetti, A., & Laureti, T. (2008). *The Interpretation of the Divergences between CPIs at Territorial Level: Evidence from Italy*. Paper presented at the Joint UNECE/ILO meeting on Consumer Price Indexes, May 8–9, Geneva.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers. *Economic Journal, 92*, 73–86.

De Carli, R. (2008). An Experiment to Calculate PPPs at Regional Level in Italy: Procedure Adopted and Analyses of the Results. Paper presented at the Joint UNECE/ILO meeting on Consumer Price Indexes, May 8–9, Geneva.

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

Ferrari, G., & Riani, M. (1998). On purchasing power prities calculation at the basic heading level. *Statistica, 58*, 91–108.

Gini, C. (1924). Quelques Considérations au Sujet de la Construction des Nombres Indexes des Prix et des Questions Analogues. *Metron, 4*, 3–162.

Gini, C. (1931). On the circular test of index numbers. *International Review of Statistics, 9*(2), 3–25.

Heravi, S., Heston, A., & Silver, M. (2003). Using scanner data to estimate country price parities: A hedonic regression approach. *Review of Income and Wealth, 49*(1), 1–21.

Hill, R. (1997). A Taxonomy of multilateral methods for making international comparisons of prices and quantities. *Review of Income and Wealth, 43*(1), 49–69.

Hill, R. (2008). *Elementary Indexes for Purchasing Power Parities (PPPs)*. Paper presented at the Joint UNECE/ILO meeting on Consumer Price Indexes, May 8–9, Geneva.

ILO et al. (2004). Consumer price index manual: Theory and practice, Geneva. An electronic updated version of the manual can be found at the web site of ILO.

Rao, D. S. Prasada (2001). *Weighted EKS and Generalized Country Product Dummy Methods for Aggregation at Basic Heading Level and Above Basic Heading Level*. Seminar on PPPs, World Bank Washington, Jan/Feb 2001.

Sergeev, S. (2003). *Equi-representativity and Some Modifications of the EKS Method at the Basic Heading Level*. Paper presented at the Joint Consultation on the European Comparison Programme, ECE, Geneva 31 March-2 April 2003.

Szulc, B. J. (1964). Indexes for multiregional comparisons. *Przeglad Statystyczny*(*Statistical Review*), *3*, 239–254.