Preliminary results of scanner data analysis and their use to estimate Italian inflation

Alessandro Brunetti (albrunet@istat.it)
Stefania Fatello (fatello@istat.it)
Federico Polidoro (polidoro@istat.it)

Workshop scanner data. Rome 1-2 October 2015

Istat

Outline of the presentation

1. State of play of Istat project on scanner data (4)
   - Obtaining scanner data for the testing phase: the triangle
   - The build up of ISTAT database
   - Linking EAN codes with COICOP
   - Scanner data and sample design revision
2. Preliminary assessment of different index formulas to aggregate the elementary data (the case study of mineral water) (12)
   - The issue
   - Index formulas compared
   - Description of the dataset
   - Evidences from the case study of mineral water
3. Conclusions and open problems (1)
4. Further steps of Istat project on scanner data (1)

1. State of play of Istat project on scanner data

Obtaining scanner data for the testing phase: the triangle

- Informal agreement with Association of Modern Distribution, representing the main chains of modern retail trade (several meetings in 2013 – 2015 and ongoing consultation)
- Istat sent the requests of the data (signed by the President) to six chains: Coop Italia, Conad, Selex, Esselunga, Auchan, Carrefour (in total almost 57% of the turnover of modern distribution)
- The requests concerned weekly data of turnover and quantities, EAN code by EAN code, outlet by outlet, for a progressively increasing amount of provinces, for the six “big chains”
- The specific request to Nielsen concerned the dictionary (with the descriptions of the EAN and its periodic updates)

1. State of play of Istat project on scanner data

The build up of ISTAT database

From December 2012

1. Torino
2. Ancona
3. Palermo
4. Piacenza
5. Cagliari
6. Roma

From December 2013

7. L'Aquila
8. Potenza
9. Reggio-Calabria
10. Napoli
11. Bologna
12. Trieste
13. Genova
14. Milano
15. Campobasso
16. Bari
17. Firenze
18. Trento
19. Venezia

From December 2014

20. Pescara
21. Catanzaro
22. Ravenna
23. Modena
24. Udine
25. Bergamo
26. Varese
27. Pesaro-Urbino
28. Foggia
29. Sassari
30. Catania
31. Messina
32. Lucca
33. Pisa
34. Perugia
35. Aosta
36. Verona
37. Padova

End 2015: 37 provinces, a coverage of 55% of Italian population

1. State of play of Istat project on scanner data

Linking EAN codes with COICOP

- Istat worked on the mapping from ECR to COICOP linking the ECR groups (at level of "markets" that are about 1600 voices) with Italian COICOP-6 level

- The Nielsen “dictionary” associates to every single EAN code sold in Italy attributes that allow you to identify the product (manufacturer, brand, possible sub-brand, size, packaging, variety) and classifies each EAN within the ECR classification (variation of GPC Global Product Classification applies worldwide)

EAN	EAN description	ECR4 description	COICOP6	COICOP6 description
8020141630005	F.DI VINADIO SANT'ANNA STD TAVOLA MINERALE GAS PLAS 06 150.0 CL	Sparkling water 101-150 cl	01.2.2.1.0	Mineral waters
8020141610007	F.DI VINADIO SANT'ANNA STD TAVOLA MINERALE NO GAS PLAS 06 150.0 CL	Natural water 101-150 cl	01.2.2.1.0	Mineral waters
8020141620006	F.DI VINADIO SANT'ANNA STD TAVOLA MINERALE MED.GAS PLAS 06 150.0 CL	Lightly sparkling water 101-150 cl	01.2.2.1.0	Mineral waters
8003430100557	S.BERNARDO STD TAVOLA MINERALE NO GAS PLAS 08 050.0 CL	Natural water 0-50 cl	01.2.2.1.0	Mineral waters
8008490000021	CO.GE.DI ULIVETO STD TAVOLA MINERALE EFF.NAT PLAS 01 150.0 CL	Natural sparkling water 101-150 cl	01.2.2.1.0	Mineral waters
8001050025663	LEVISSIMA SPORT MINERALE NO GAS PLAS 04 033.0 CL	Natural water 0-50 cl	01.2.2.1.0	Mineral waters
8008490991015	CO.GE.DI ULIVETO STD TAVOLA MINERALE EFF.NAT PLAS 06 050.0 CL FACE03	Natural sparkling water 101-150 cl	01.2.2.1.0	Mineral waters

1. State of play of Istat project on scanner data

Scanner data and sample design revision

- Scanner data one of the sides of the coin of Multipurpose Price Statistics Project concerning modernization of data sources (scanner data) and of data collection techniques (web scraping)
- Scanner data and use of web scraping techniques raised the need to overcome the purposive sample design used until now for Italian survey on consumer prices
- Test and analysis on scanner data carried out for both the aims:

  A. improving the use of elementary data in terms of more precise aggregation schemes

  B. Testing different sampling design

2. Preliminary assessment of different index formulas

The issue

- Until now, elementary indices aggregated by Jevons formula, essentially for the lack of information about weights
- Scanner data bring detailed information about quantity and turnover
- Weekly prices calculated as the ratio between turnover and quantity
- Monthly prices calculated as weighted average of the prices of the three central weeks of the months
- Going on to use Jevons formula to aggregate elementary indices referred to single EAN code appears no longer suitable
- **Which index has to be adopted ?**

2. Preliminary assessment of different index formulas

Index formulas compared

- Focus on indices with fixed base (exclusion, for the time being, of monthly-chained indices)

Laspeyres: $$P_{Lasp}^{m,t} = \frac{\sum_{n=1}^{N} p_n^{m,t} \cdot q_n^{0,t}}{\sum_{n=1}^{N} p_n^{0,t} \cdot q_n^{0,t}}$$

$m = 1, \dots, 12 \quad t = 2014$

$p_n^{0,t} = p_n^{12,t-1}$

$q_n^{0,t} = q_n^{12,t-1}$

Paasche: $$P_{Paas}^{m,t} = \frac{\sum_{n=1}^{N} p_n^{m,t} \cdot q_n^{m,t}}{\sum_{n=1}^{N} p_n^{0,t} \cdot q_n^{m,t}}$$

Jevons (unweighted): $$P_{Jev}^{m,t} = \left(\prod_{n=1}^{N} \frac{p_n^{m,t}}{p_n^{0,t}} \right)^{1/N}$$

Lowe (Laspeyres type): $$P_{Lowe}^{m,t} = \frac{\sum_{n=1}^{N} p_n^{m,t} \cdot \bar{q}_n^{t-1}}{\sum_{n=1}^{N} p_n^{0,t} \cdot \bar{q}_n^{t-1}}$$

$\bar{q}_n^{t-1} = \sum_{m=1}^{12} q_n^{m,t-1}$

- Comparison with Fisher (ideal) index, considered as a benchmark.

$$P_{Fish}^{m,t} = \left(P_{Lasp}^{m,t} \cdot P_{Paas}^{m,t} \right)^{1/2}$$

2. Preliminary assessment of different index formulas

Description of the dataset

- Period considered: December 2013 – December 2014
- Panel of EAN codes present seamlessly in the period considered. No sample selection
- Focus on Mineral water (ECOICOP 02.1.1.2)
- All the supermarkets and hypermarkets of the province of Turin

ECR 4	N° Ean	N° series 2014	% series 2014	turnover 2014	% turnover 2014
Natural water 101-150 cl	78	1.796	25,7	9.179.240	54,5
Sparkling water 101-150 cl	53	1.107	15,8	2.444.845	14,5
Natural water 0-50 cl	53	1.256	18,0	1.397.556	8,3
Natural sparkling water 101-150 cl	13	492	7,0	1.173.462	7,0
Other markets (12)	125	2.339	33,5	2.656.456	15,8
Mineral water	**322**	**6.990**	**100,0**	**16.851.559**	**100,0**

Type of outlet	N° outlet	N° Ean	N° series 2014	% series 2014	turnover 2014	% turnover 2014
Hypermarket	28	252	2.294	32,8	9.963.236	59,1
Supermarket	99	232	4.696	67,2	6.888.323	40,9
Total	**127**	**322**	**6.990**	**100,0**	**16.851.559**	**100,0**

2. Preliminary assessment of different index formulas

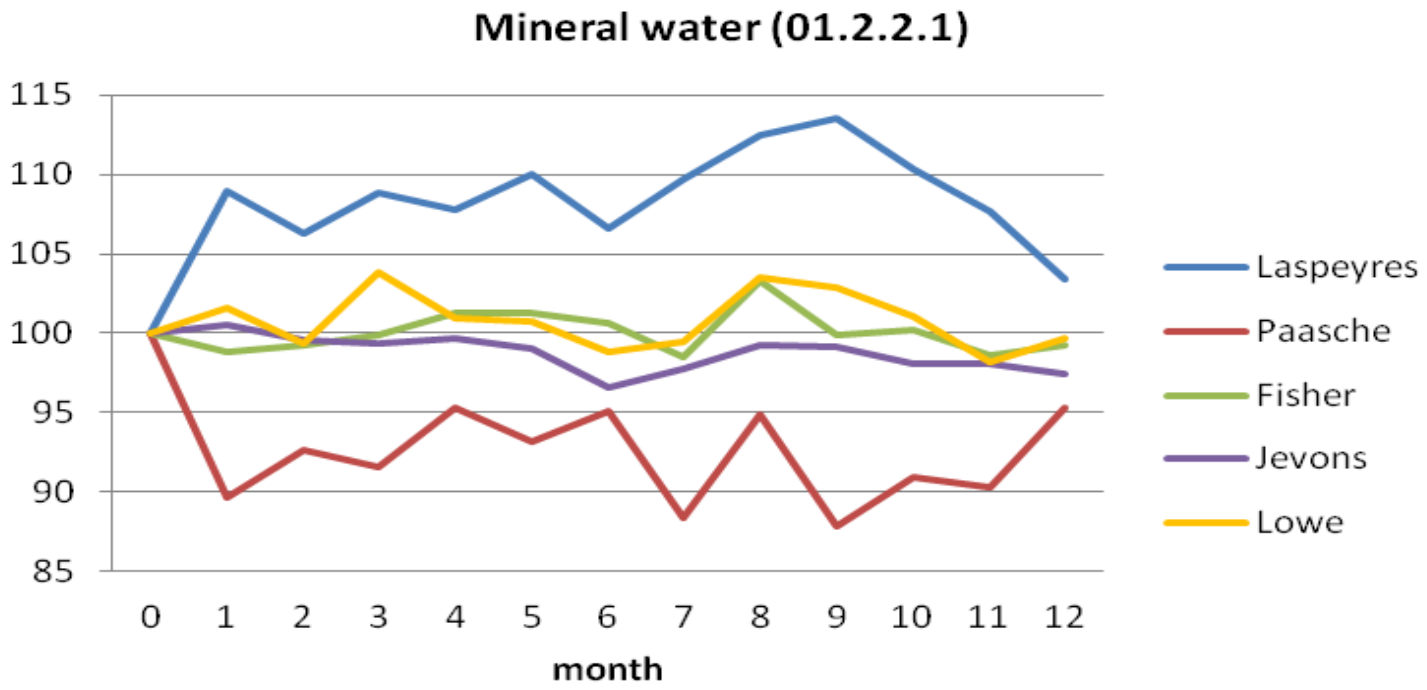
Description of the dataset

Mineral water	N° Ean	N° outlet	turnover 2014
Panel	322	127	16.851.559
Total dataset	468	130	18.506.760

% N° series panel													
ECR 4	**Dec-13**	**Jan-14**	**Feb-14**	**Mar-14**	**Apr-14**	**May-14**	**Jun-14**	**Jul-14**	**Aug-14**	**Sep-14**	**Oct-14**	**Nov-14**	**Dec-14**
Natural water 101-150 cl	71,0	70,1	69,2	70,0	70,9	69,5	69,8	68,5	70,5	69,6	69,7	68,7	69,5
Sparkling water 101-150 cl	72,9	73,9	73,1	71,4	72,4	70,5	70,3	69,8	71,1	68,3	69,0	67,1	67,8
Natural water 0-50 cl	77,0	76,8	77,2	76,6	75,8	78,5	77,1	75,3	75,4	75,1	76,4	77,0	76,9
Natural sparkling water 101-150 cl	80,8	81,2	82,3	81,7	81,1	85,0	84,7	81,2	81,9	81,3	82,7	81,9	81,9
Mineral water	**73,8**	**73,8**	**73,4**	**72,4**	**73,0**	**72,9**	**72,1**	**71,7**	**72,9**	**71,5**	**72,4**	**71,6**	**71,5**
% Turnover panel													
ECR 4	**Dec-13**	**Jan-14**	**Feb-14**	**Mar-14**	**Apr-14**	**May-14**	**Jun-14**	**Jul-14**	**Aug-14**	**Sep-14**	**Oct-14**	**Nov-14**	**Dec-14**
Natural water 101-150 cl	93,5	95,3	92,6	91,7	93,1	93,8	95,2	91,3	92,9	91,5	93,5	92,1	91,3
Sparkling water 101-150 cl	92,7	94,4	92,8	91,9	93,3	92,2	92,3	89,5	91,9	90,3	91,5	89,1	89,5
Natural water 0-50 cl	91,3	89,3	91,0	92,6	92,3	93,1	93,0	91,8	93,2	92,2	92,7	92,6	90,5
Natural sparkling water 101-150 cl	90,6	90,6	91,4	91,2	93,7	94,8	94,9	91,8	90,8	90,6	92,9	90,8	90,0
Mineral water	**91,8**	**92,8**	**91,5**	**90,1**	**92,7**	**92,2**	**93,0**	**90,4**	**91,5**	**88,3**	**91,5**	**88,7**	**89,6**

2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water



- The patterns of indices are quite different (as expected)
- Major bias concern Laspeyres and Paasche indices, while Jevons and Lowe remain – relatively - close to Fisher

2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water

- The different temporal evolution of the different elementary indices reflects the underlying movements of prices and quantities

- In order to get a closer insight into the relationship between prices and quantities, it is useful to breakdown mineral water into more homogeneous markets, according to ECR4 classification

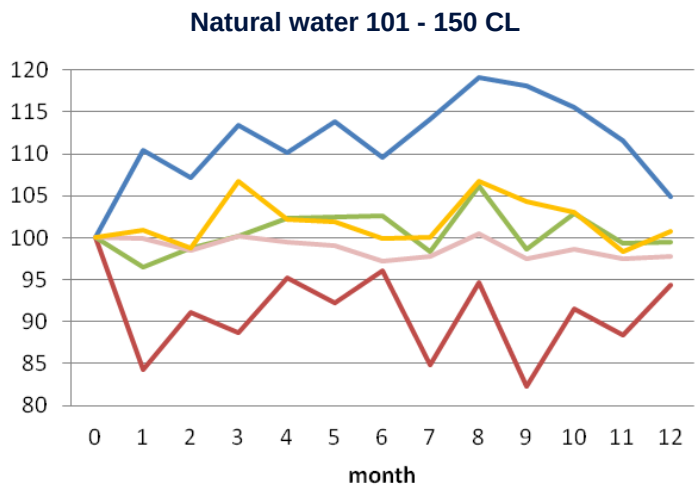
- Focus on the market (ECR4 level) of "*Natural water, 101 -150 cl* " and of "*Natural sparkling water, 0 – 50 cl* "

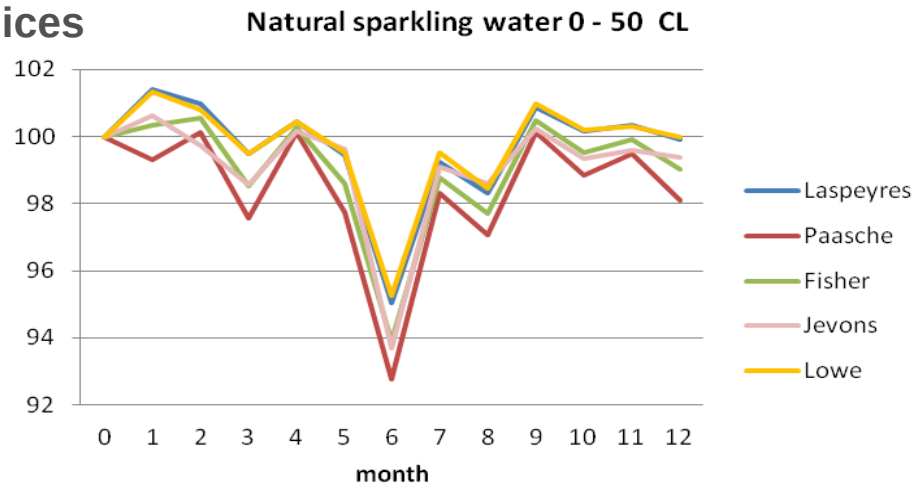
ECR 4	N° Ean	N° series 2014	% series 2014	turnover 2014	% turnover 2014
Natural water 101-150 cl	78	1.796	25,7	9.179.240	54,5
Natural sparkling water 0-50 cl	6	325	4,6	176.172	1,0

2. Preliminary assessment of different index formulas

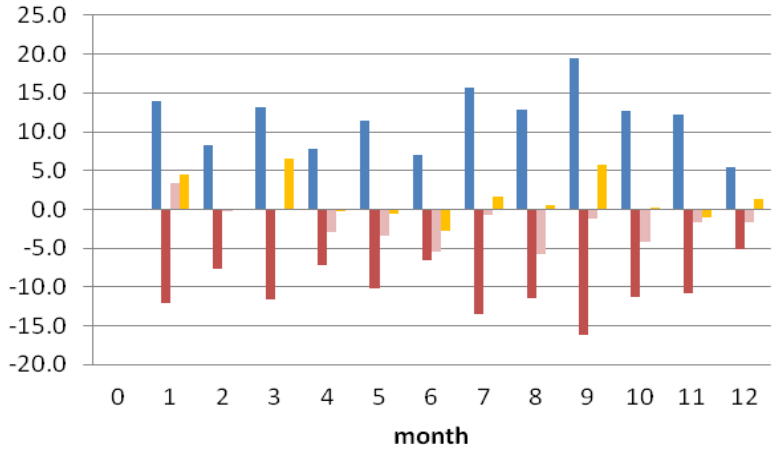
Evidences from the case study of mineral water

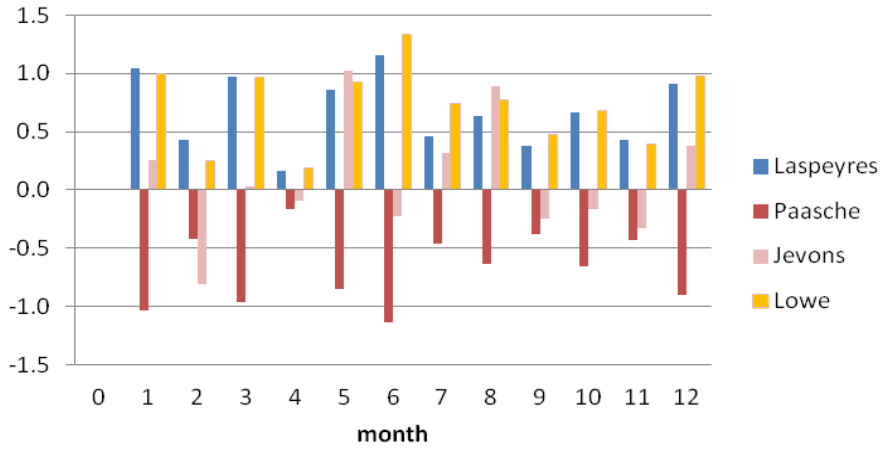
Indices





Absolute differences with respect to Fisher index





2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water

Indicators from relative differences with respect to Fisher index

Natural water, 101 -150 cl

Variable	Obs	Mean	Std. Dev.	Min	Max
dif_Laspey~r	12	11.61	4.10	5.44	19.77
dif_Paasch~r	12	-10.29	3.27	-16.51	-5.16
dif_Jevons~r	12	-1.93	2.52	-5.37	3.56
dif_Lowe_F~r	12	1.34	2.85	-2.72	6.51

Natural sparkling water, 0 – 50 cl

Variable	Obs	Mean	Std. Dev.	Min	Max
dif_Laspey~r	12	0.68	0.32	0.16	1.23
dif_Paasch~r	12	-0.68	0.32	-1.21	-0.16
dif_Jevons~r	12	0.09	0.53	-0.81	1.04
dif_Lowe_F~r	12	0.74	0.36	0.19	1.42

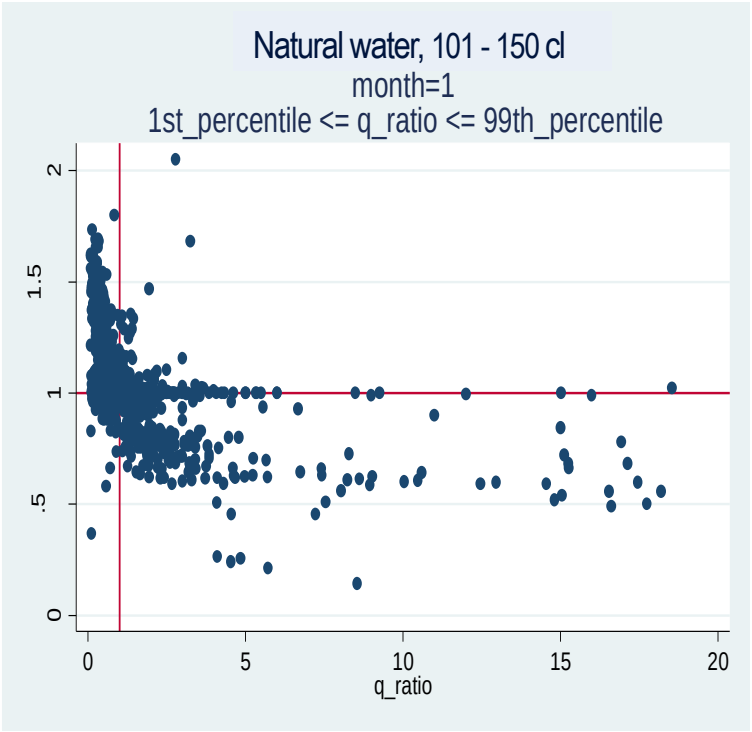
- As expected Laspeyres always overestimates inflation (very low performance in the first market).
- As expected Paasche is always negative biased (very low performance in the first market).
- Jevons seems to have a small negative bias in the first market. It performs relatively well in second.
- Lowe: good performance in the first market; comparatively low in the second market but in absolute acceptable

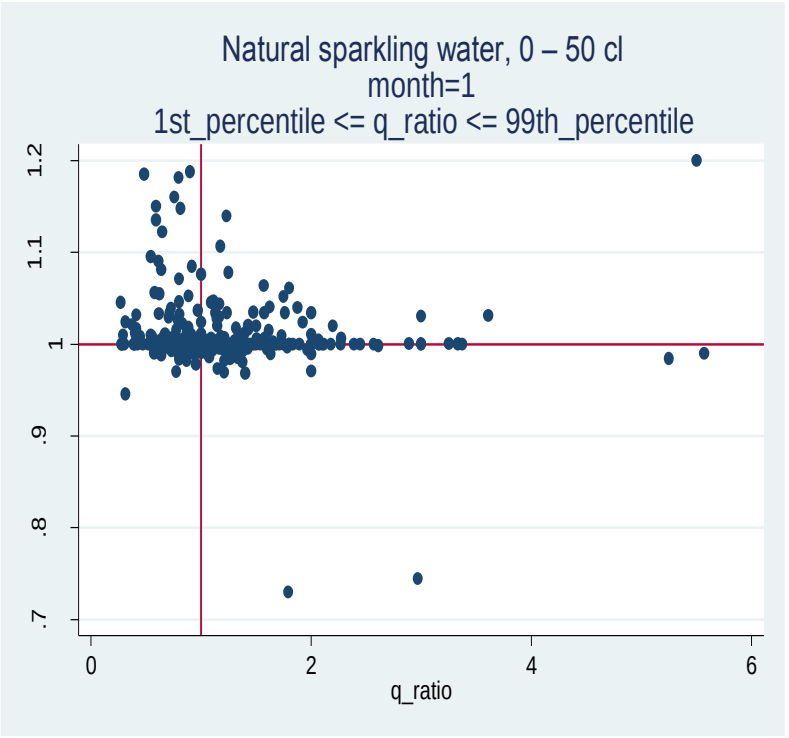
2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water

- The bias of different indices depends on the characteristics of the market in term of price elasticity of demand

- Prices and quantities movements in the two markets: evidences from January data



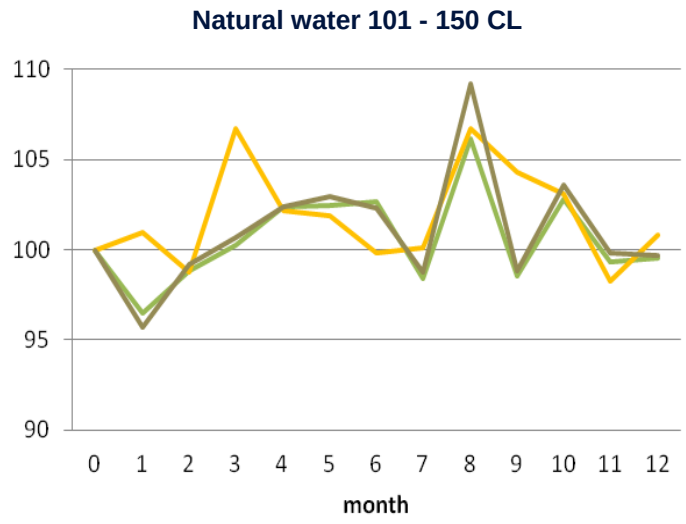


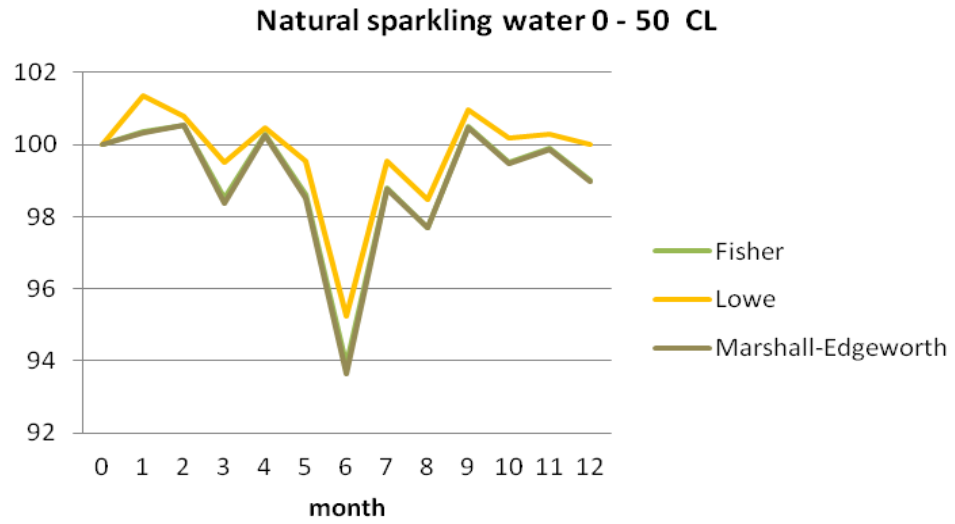
2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water

- Analysis of the factors behind the performance of the Lowe index in the two markets, through the comparison with a symmetric index (Marshall-Edgeworth) besides Fisher index

$$P_{ME}^{m,t} = \frac{\sum_{n=1}^{N} p_n^{m,t} \cdot \left(q_n^{m,t} + q_n^{0,t}\right)/2}{\sum_{n=1}^{N} p_n^{0,t} \cdot \left(q_n^{m,t} + q_n^{0,t}\right)/2}$$



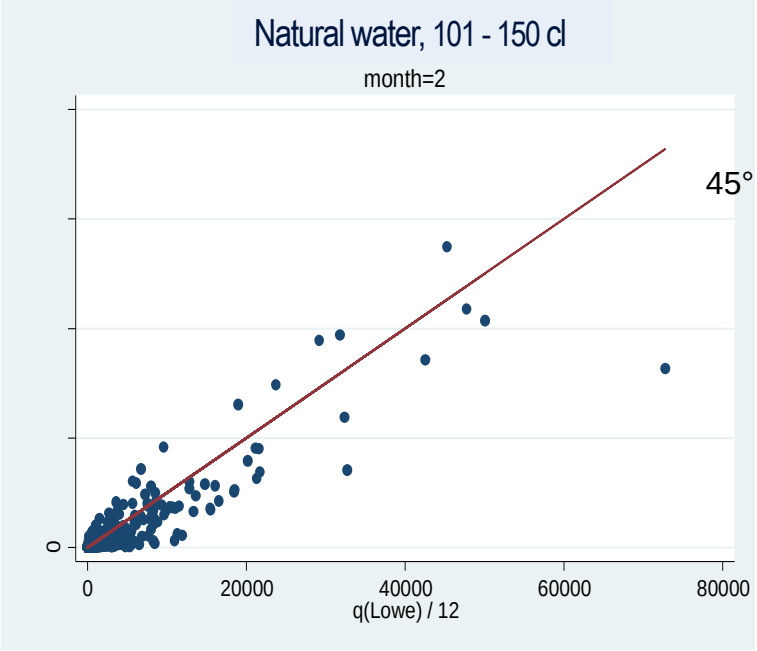


2. Preliminary assessment of different index formulas

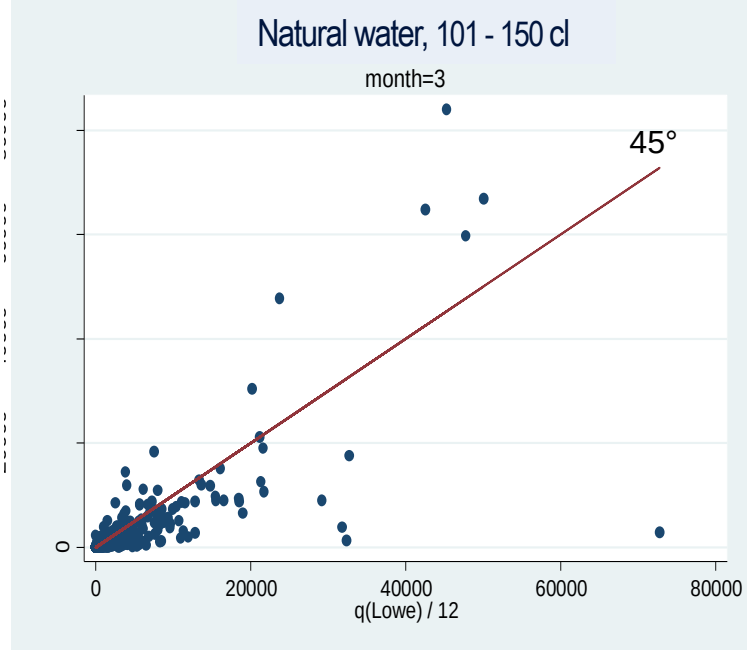
Evidences from the case study of mineral water

- Lowe approximates Marshall-Edgeworth if, in response to prices changes, the quantities sold of the different items fluctuate around values which are relatively stable and close to the monthly average quantities sold in the previous year. That is, if: $\left(q_n^{m,t} + q_n^{0,t}\right)/2 \approx \bar{q}_n^{t-1}/12$

February:
PLOWE = 98.74
PME = 99.20



March:
PLOWE = 106.75
PME = 100.68



2. Preliminary assessment of different index formulas

Evidences from the case study of mineral water

In formal terms, if we define:

$$w_{Lowe}^{t}(n) = \frac{p_n^{0,t} \cdot \bar{q}_n^{t-1}}{\sum_{n=1}^{N} p_n^{0,t} \cdot \bar{q}_n^{t-1}} \qquad w_{ME}^{t}(n) = \frac{p_n^{0,t} \cdot \left(q_n^{m,t} + q_n^{0,t}\right)}{\sum_{n=1}^{N} p_n^{0,t} \cdot \left(q_n^{m,t} + q_n^{0,t}\right)}$$

and

$$d_n^{m,t} = w_{Lowe}^{t}(n) - w_{ME}^{t}(n)$$

It can be shown that:

the smaller the linear correlation between $d_n^{m,t}$ and $p_n^{m,t}/p_n^{0,t}$

the closer $P_{Lowe}^{m,t}$ to $P_{ME}^{m,t}$

3. Conclusions and open problems

- The choice of the formula to aggregate microindices is not neutral with respect to the inflation measurement
- The availability of data on quantities sold and turnover in the scanner data make possibile to overcome an unweighted approach to the aggregation of micro indices
- If we consider Fisher as a benchmark, and given for granted that we cannot compile currently Fisher indices, Jevons and Lowe (and the Laspeyres type) indices emerge as the best proxies of Fisher ideal index
- But the elasticity of the demand in each specific market plays a relevant role to make a formula of aggregation better than another one
- Taken for granted that HICPs are Laspeyres type indices, **may we move towards a case by case approach for the aggregation at the lower level ?**

4. Further steps of Istat project on scanner data

- Testing the possibility to reduce the triangle to a straight line for obtaining scanner data
- In 2016 working on scanner data referred to the entire national territory
- The general idea of abandoning Jevons as formula to aggregate the elementary indices
- Extending the analysis and tests carried out for mineral water in Turin, to other groups of products and other geographical areas
- Extending the analysis to the issues of monthly-chained indices, attrition and replacement issues
- Compiling a parallel index in 2016 using scanner data and comparing different results obtained using different index formulas and probabilistic sample design (look at the presentation tomorrow)
- Evaluating after 2016 summer if and how to start the use of scanner data in current production process of HICP

Thank you for your attention