Measuring inflation through different sampling designs implemented on scanner data

A. Guandalini, A. Bernardini, C. De Vitiis, F. Inglese and M.D. Terribili

ISTAT – ITALY

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
Summary

1. Aims of the presentation
2. The current sample strategy of the CPI survey
3. The new general sampling design for CPI survey
4. Different sampling strategies from Scanner Data
5. Results
6. Open issues and conclusions
1. Aims of the presentation

Comparing properties
of weighted and un-weighted higher level elementary price indices
when different sampling scheme are implemented
on scanner data (SD)

- Preliminary results:
  - Probabilistic versus Non-probabilistic

- New results:
  - Different probability sampling strategies
    - Allocation
    - Selection scheme
    - Aggregation price index formula
2. The current sample strategy of the CPI survey

1st stage
Municipalities

Forced by law
2. The current sample strategy of the CPI survey

1st stage
Municipalities

2nd stage
Outlets

Forced by law
2. The current sample strategy of the CPI survey

1st stage
Municipalities

2nd stage
Outlets

3rd stage
EAN

FIXED BASKET
about 1,500 products

FIXED BASKET
about 1,500 products

FIXED BASKET
about 1,500 products

Forced by law

Purposive sample

Purposive sample

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
3. The new general sampling design for CPI survey

ASIA-LU

Register of
Local Unit Active Enterprises
3. The new general sampling design for CPI survey

ASIA-LU
Register of Local Unit Active Enterprises

Modern distribution

Traditional distribution
3. The new general sampling design for CPI survey

ASIA-LU
Register of Local Unit Active Enterprises

Modern distribution

Food & Grocery
Other retail sector

Traditional distribution

Probability sampling
3. The new general sampling design for CPI survey

ASIA-LU

Register of Local Unit Active Enterprises

Modern distribution

Food & Grocery

Probability sampling

Other retail sector

Probability sampling

Traditional distribution
4. Different sampling strategies from Scanner Data

Operational framework

- Turin province
- 2014
- Six consumption segments (COICOP-6)
- Relevant week (one of the first three full weeks in a month)
- Cleaned dataset
  - Formal check (completeness of data)
  - Quality check (quantities sold and turnover)
  - “Mobile trimming” (inadmissible unit prices)
- Permanent series (only series always observed from December 2013 to December 2014)
PRELIMINARY EXPERIMENTS
Probability versus non-probability selection schemes

Comparison among:

- Cut-off selection of series based on thresholds of covered sales:
  60-80% coverage of all expenditure of the consumption segment in previous year, 2013

- Reference method
  most sold elementary items for each representative product (current fixed basket approach)

- Probability sampling – independent samples in consumption segments
PRELIMINARY EXPERIMENTS
Probability versus non-probability selection schemes

Comparison among:

- Cut-off selection of series based on thresholds of covered sales:
  60-80% coverage of all expenditure of the consumption segment in previous year, 2013

- Reference method
  most sold elementary items for each representative product (current fixed basket approach)

- Probability sampling – independent samples in consumption segments

MAIN RESULTS

- Probability sampling always brings more efficient estimates than non-probability selection scheme (i.e. most sold and cut-off)

- Sample scheme is not neutral with respect to the choice of aggregation formulas

- Sampling error varies among consumption segments
FURTHER EXPERIMENTS
Probability selection schemes

- Elementary price indices computed for 13 months with three aggregation formulas:
  - Jevons (unweighted),
  - Fisher
  - Lowe
- Montecarlo simulation (500 replications)
- Comparison among aggregation formulas performances under different sampling designs in terms of:
  - Bias $ B(\hat{\theta}) = E[\hat{\theta}] - \theta $
  - Variance – Relative sampling error $ RE(\hat{\theta}) = \frac{\sqrt{Var(\hat{\theta})}}{\hat{\theta}} $
## FURTHER EXPERIMENTS
Probability selection schemes

### Population size
- 50,000 EAN
- 121 outlets

### Sample size
- 7,400 EAN

<table>
<thead>
<tr>
<th>Sampling design</th>
<th>Level</th>
<th>Stratification</th>
<th>Allocation</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stratified</td>
<td>EAN</td>
<td>Market (ECR4)</td>
<td>Neyman</td>
<td>SRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPS (turnover)</td>
</tr>
<tr>
<td>Stratified 1 stage</td>
<td>Outlet</td>
<td>Chain * type</td>
<td>Neyman</td>
<td>SRS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPS (turnover)</td>
</tr>
<tr>
<td>Stratified 2 stages</td>
<td>Outlet</td>
<td>Chain * type</td>
<td>Proportional</td>
<td>SRS</td>
</tr>
<tr>
<td></td>
<td>EAN</td>
<td>Market</td>
<td></td>
<td>PPS (turnover)</td>
</tr>
</tbody>
</table>
Stratified sampling

- Stratification of EAN by market (ECR4)
- Neyman allocation with respect to price relative variability in 2013 in each stratum

\[
\text{SEGMENT} = \text{COFFEE}
\]
Stratified sampling

- Stratification of EAN by market (ECR4)
- Neyman allocation with respect to price relative variability in 2013 in each stratum

**SEGMENT = COFFEE**
Stratified sampling

Comparison of RE% under SRS and PPS (size=turnover) for different consumption segments (COICOP-6)
Stratified sampling

Relative bias of monthly index estimates when EAN are selected with SRS or with PPS
## Stratified 1 stage

- Stratification of outlets by chain and type

<table>
<thead>
<tr>
<th>Chain</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPER MARKET</td>
<td></td>
</tr>
<tr>
<td>SUPER MARKET</td>
<td></td>
</tr>
</tbody>
</table>

... ... ...

... ... ...

... ... ...

... ... ...

---

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
Stratified 1 stage

- Stratification of outlets by chain and type

<table>
<thead>
<tr>
<th>Chain</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPER MARKET</td>
<td></td>
</tr>
<tr>
<td>SUPER MARKET</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>IPER MARKET</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
Stratified 1 stage

- Stratification of outlets by chain and type

<table>
<thead>
<tr>
<th>Chain</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPER MARKET</td>
<td></td>
</tr>
<tr>
<td>SUPER MARKET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>IPER MARKET</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
Stratified 1 stage sampling

Comparison of RE% for PPS selection of outlets allocated with Neyman and Proportional allocation

Jevons

Lowe

Fisher
Comparison between the relative bias in Stratified sampling and in Stratified 1 stage sampling

<table>
<thead>
<tr>
<th>Stratified sampling</th>
<th>Stratified 1 stage sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jevons</td>
<td>Lowe</td>
</tr>
</tbody>
</table>

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)
Stratified 2 stages sampling

- Stratification of outlets by chain and type
- Stratification of EAN by market (ECR4)
Stratified 2 stages sampling

Comparison of RE% when Second Stage Units (EAN) are selected with SRS or PPS (size=turnover)
Stratified 1 stage vs. Stratified 2 stages (Turin, Coffee 2014)

Meeting of the Group of Experts on Consumer Price Indices
2 - 4 May 2016
Geneva, Switzerland
7. Conclusions

Future developments

- Evaluation of the impact of attrition and substitution of elementary items during the yearly data collection
- Implementation of monthly chained matched-item index

Open issues

- Integration between scanner data and traditional data for index compilation
- Different hypothesis are under evaluation:
  - Combining indices obtained with different approaches
  - Gradually stop manual collection, at least for food and grocery, considering the high expenditure coverage of modern distribution
  - Aim at define and realise a probability sampling also for traditional distribution