Balancing the Swedish CPI

Meeting of the Ottawa Group in Rome, Italy 7-10 June 2022

Anders Norberg
Can Tongur

Statistics Sweden
Outline

• The Swedish CPI (Konsumentprisindex)

• Variance estimation

• Divisions of CPI in this study

• Methodology

• Results

• Conclusions
The Konsumentprisindex

- Sampling of companies, outlets, products, product offers, time
  - Rotated by PRN (20%)
  - Multistage: outlets – product offers
  - Two-phase: areas – outlets (many price collectors, historically)
  - Order πps, srs, judgemental
  - Two-dimensional; outlet/product in daily necessities (sort of)

- "Old" and new data sources
  - Conventional price collection; field, online, telephone
  - Regional price collectors, central CPI-staff
  - New collection methods in broad usage and ongoing transition: internet, web scraping, transaction data
  - Mixed and adapted methodology ↔ Swedish CPI Board
Variance estimation for the CPI

• Complex statistical measure with "derivatives"
  Central for the analysis: short-term link, inflation, change in inflation
  One/two baskets, different weights, within-year patterns & business cycles

• Variance assessments
  1) Design based
  2) Model based
  3) Randomisation

Notable work by Dalén & Ohlsson (1995) and Norberg (2004) for KPI
Model approaches by Shoemaker (1999,+,+)+ and Bialek (2020)
The Quality Declaration (2022)

- Official statistics provided with QD
- Relies partly on previous studies, partly on simple variance estimates
- Many components in CPI surveys contribute to the final estimate, and the variance
- Does QD improve interpretation for important users?

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Length of 95% confidence interval</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly change</td>
<td>±0.14</td>
<td>Somewhat shorter for April, May, June and November</td>
</tr>
<tr>
<td>Annual change (inflation rate)</td>
<td>±0.23</td>
<td>Somewhat shorter for December*</td>
</tr>
<tr>
<td>Monthly change in inflation rate</td>
<td>±0.20</td>
<td>Somewhat shorter for April, May, June, November and December, somewhat longer for other months</td>
</tr>
</tbody>
</table>

*The change from December to December is based on one and the same sample.
CPI divisions in the study

- COICOP 01 and 02.2 (Food, non-alcoholic beverages, tobacco + non-food)
- COICOP 03  Clothing and footwear (=Apparel)
- COICOP 05.1 Furniture
- COICOP 07.1 Domestic and international air travel
- COICOP 11  Restaurants

- Domains compare with previous assessments
  Varying data sources and quality adjustment methods
Bootstrap resampling method

- Select $B$ samples from the original data by srs-wr, thus obtaining the same distributions in all samples. Compute the estimates $\hat{\theta}^*(b)$.

$$\hat{V}(\hat{\theta}) = \sum_{b=1}^{B} (\hat{\theta}^*(b) - \hat{\theta}^*)^2 / (B - 1)$$

- The formulation allows for setting up most estimators. Specifically, the twelve-month change from two baskets can be explicitly estimated.

- Challenge to mimic the underlying design, many moving parts also between years (=not so simple in practice)
Two-dimensional sampling

Transaction Data
- Outlets cheap
- Products expensive due to market analysis, replacements and manual quantity adjustments

Correlations between outlets (in general)
- Very high in multi-store companies
- High for chains

Very high design effect, simple variance estimator not appropriate
# Food/daily necessities

## Table 5.1 01 & 02.2 Food and non-alcoholic beverages, Tobacco, and non-food

<table>
<thead>
<tr>
<th>Subpopulation (Aspects 1-4)</th>
<th>Measure</th>
<th>Variance (avg.)</th>
<th>$V(I) / V(S)$</th>
<th>+/- 2 std. errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Food $B=400$</td>
<td>Short-term</td>
<td>0.045</td>
<td>0.414</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inflation</td>
<td>0.055</td>
<td>1.360</td>
<td>0.482</td>
</tr>
<tr>
<td></td>
<td>$\Delta$(Inflation)</td>
<td>0.053</td>
<td></td>
<td>0.460</td>
</tr>
<tr>
<td>2. Food: products $B=400$</td>
<td>Short-term</td>
<td>0.028</td>
<td></td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>Inflation</td>
<td>0.037</td>
<td>1.315</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>$\Delta$(Inflation)</td>
<td>0.038</td>
<td></td>
<td>0.388</td>
</tr>
<tr>
<td>3. Food: outlets $B=400$</td>
<td>Short-term</td>
<td>0.014</td>
<td></td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>Inflation</td>
<td>0.017</td>
<td>1.229</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>$\Delta$(Inflation)</td>
<td>0.012</td>
<td></td>
<td>0.216</td>
</tr>
<tr>
<td>(1-2-3) Food: Interaction $B=400$</td>
<td>Short-term</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inflation</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta$(Inflation)</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Clothing and footwear

- Manual price collection, on-site and online
- Outlet sampling; central, register
  - Big multi-store companies form strata of their own
- Product offer selection; locally by collector
- Extreme volatility from campaigns/discounts, season sales
- Hedonic quality adjustments
  - No reduction of variance

Additional information: IQI close to 100% in last 10 years
Clothing: variance in actual & regular price

- **Figure 5.2.1** Clothing and footwear: Short-term index link variance from actual prices, without quality adjustment (dashed) and simple variance (dotted)

- **Figure 5.2.2** Clothing and footwear: Short-term index link variance from actual prices and regular prices variance (dashed)
Table 5.2 Inflation

+/- 2 std. errors = 2.09

Data extraction from homepage:
www.scb.se
Furniture

- On-site, internet, web scraping
- Judgemental quality adjustments halves the variance
- Simple variance estimates appropriate (=no clustering), unlike other cases in the study

Figure 5.2.6 Furniture: Short-term index link variance, with quality adjustment and without (dashed)
Restaurants

- Telephone/internet
- Judgemental quality adjustments (minor issue)
- Establishment characteristics more distinct (cluster effect)

Figure 5.2.11 Restaurants: Short-term index link variance and simple variance (dotted)
Air travel

- Manual online collection → web scraping, schematic according to HICP
- Completely affected by Covid in weights and imputed prices
- Between-destination variance > within-destination variance
- "The carrier issue" as variance driver?
Table 5.3 Aspect 2 Inflation

\( \pm 2 \) std. errors = 10.44

Data extraction from homepage:
www.scb.se
Conclusions

• Bootstrapping viable (when $n>1$) but tedious
  No single setup fits all surveys and years

• Changes in the CPI (basket/method/data source) appear to influence
  Less “traditional” design orientation with new data sources

• Effective sample sizes potentially smaller than presumed
  Simple variance estimators downward biased:
  design effects significant, except in one situation (Furniture)

• Inflation variance/short term variance ratio observed: 1.36 - 1.77
  (yearly averages, dec.-dec. smaller)

• Clothing & footwear disproportionately high variance despite survey costs