Challenging the CES assumption with scanner data - pitfalls of the fixed basket

Can Tongur, Statistics Sweden

The Ottawa Group Meeting in Rio de Janeiro, Brazil
8-10 May 2019
What is this about?

- Established methodology – recommended and used
- A rather new data source – rendering insights (novel?)
- Some supportive theory – backing up the methodology
- An empirical study – on scanner data
- Some ideas to take home – and to think through?
The Constant Elasticity of Substitution assumption (1)

Elasticity of Substitution

- The simplified situation of consumers discriminating between obviously substitutable items as a response to price changes (substitution in "narrow sense", de Haan (2001))
  Reminisces the ex post Laspeyres v.s. Paasche discussion
- Elasticity of substitution is a concept of what-for-what: how many green apples for red apples, given a change in relative prices
- In practically all cases, it is a parameter of non-negative magnitude (≥0)
The Constant Elasticity of Substitution assumption (2)

Assuming a constant elasticity means that...

- substitution is thought to be equal for all pairs of items in some aggregate under consideration and hence, in all possible baskets
- there is a time invariance concept

And it implies that...

- the universe of items is "closed under sampling" (Laspeyres ⇔ Paasche)
- sampling is a valid approach for including items (randomness is amical)
- homothetic preferences – income levels do not affect choices (timing not an issue)
Estimating the elasticity of substitution

• Balk (1999) derives an expression from which estimation boils down to the application of some numerical procedure, for a basket with n items (c.f. §17.61 in the manual: the Lloyd-Moulton index):

\[ \left[ \sum_{i=1}^{n} s_i^0 \left( \frac{p_i^t}{p_i^0} \right)^{(1-\sigma)} \right]^{1/(1-\sigma)} = \left[ \sum_{i=1}^{n} s_i^t \left( \frac{p_i^t}{p_i^0} \right)^{-(1-\sigma)} \right]^{-1/(1-\sigma)} \]

• There is perhaps an asynchrony in general, as pointed out by Shapiro & Wilcox (1997):

"The mismatch in frequency between the price and expenditure data creates an ambiguity as to how one might best approximate the index formulas prescribed by theory"
A sample of items deemed suitable for analysis

The following set of multi-brand products were analyzed:
1) Sugar free soda beverage, 1.5 Liter (2 varieties, pps-sampled)
2) Dairy product, 1 Liter (2 varieties, pps-sampled)
3) Coffee, 450-500 grams, grounded (all varieties = census)
4) Cheese, packaged, several similar varieties (n most sold varieties, cut-off-sample)

Coverage well*, representativeness well*, by-the-book approach
Two ways of looking at coverage

Coverage for coffee during one year, as used for analysis
How the data was used to render necessary input to the estimation

- *The scanner data is weekly turnover and amount of units sold per item (identified through EAN/GTIN) and per store*
- *Data is aggregated over weeks to a monthly turnover per store and included if it has a match with the base period for the same store (= balance)*
- *Estimations are through "item aggregation over stores”, rendering one aggregate monthly price and expenditure share (summing to unity) per item*
## Summary statistics on estimates of $\sigma$

<table>
<thead>
<tr>
<th>Product</th>
<th>#estimates</th>
<th>Mean</th>
<th>Median</th>
<th>Std. dev.</th>
<th>Share $\sigma &lt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soda</td>
<td>144</td>
<td>3.6</td>
<td>2.05</td>
<td>10.35</td>
<td>22%</td>
</tr>
<tr>
<td>Dairy</td>
<td>72</td>
<td>9.68</td>
<td>1.34</td>
<td>63.1</td>
<td>44%</td>
</tr>
<tr>
<td>Coffee</td>
<td>36</td>
<td>2.56</td>
<td>2.92</td>
<td>2.03</td>
<td>11%</td>
</tr>
<tr>
<td>Cheese</td>
<td>42</td>
<td>4.21</td>
<td>4.05</td>
<td>1.41</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note:** column with #estimates refers to number of estimated $\sigma$ over all time points and included retail chains (one estimate per retail chain and period)
A comparison of four price indices
Laspeyres, Paasche and Lloyd \((\sigma = \text{median})\), as per cent deviation from unweighted (standard) Jevons

<table>
<thead>
<tr>
<th>Period</th>
<th>Soda (\sigma = 2.05)</th>
<th>Dairy (\sigma = 1.34)</th>
<th>Coffee (\sigma = 2.92)</th>
<th>Cheese (\sigma = 4.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5</td>
<td>2.7</td>
<td>2.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>2.6</td>
<td>2.3</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>5.6</td>
<td>6.2</td>
<td>4.8</td>
<td>-1.7</td>
</tr>
<tr>
<td>4</td>
<td>3.8</td>
<td>4.3</td>
<td>3.1</td>
<td>-5.4</td>
</tr>
<tr>
<td>5</td>
<td>6.4</td>
<td>8.0</td>
<td>4.2</td>
<td>-5.8</td>
</tr>
<tr>
<td>6</td>
<td>11.4</td>
<td>16.0</td>
<td>7.8</td>
<td>-5.4</td>
</tr>
<tr>
<td>7</td>
<td>5.8</td>
<td>8.1</td>
<td>2.9</td>
<td>-5.9</td>
</tr>
<tr>
<td>8</td>
<td>5.9</td>
<td>8.1</td>
<td>3.3</td>
<td>-3.2</td>
</tr>
<tr>
<td>9</td>
<td>0.1</td>
<td>0.3</td>
<td>-0.2</td>
<td>-5.7</td>
</tr>
<tr>
<td>10</td>
<td>6.2</td>
<td>8.7</td>
<td>3.6</td>
<td>-8.2</td>
</tr>
<tr>
<td>11</td>
<td>1.7</td>
<td>2.3</td>
<td>1.2</td>
<td>-10.6</td>
</tr>
<tr>
<td>12</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.3</td>
<td>-9.6</td>
</tr>
</tbody>
</table>
Remarks on estimating $\sigma$

- After some consideration, one understands the following conclusion by Henningsen & Henningsen (2012) regarding CES estimation:
  
  “is generally considered problematic due to convergence problems and unstable and/or meaningless results”

- Remember that the limited sample based estimates were questionable to a large extent ($\sigma < 0$)

- Inference should be made carefully – results indicative rather than conclusive!
A fixed basket in a changing universe – realistic?

• This is actually two questions:
  1) a fixed and limited sample based basket, and
  2) a fixed census-like/take-all sample based basket (with the caveat of time *)

• Regardless of the results here, the validity of a limited sample can be discussed when measuring effective prices rather than list prices (offer/over-the-counter)

• (*) The universe of available items is changing

• The problem in estimations also stems from temporary consumption changes due to price campaigns (or perhaps random effects)
Thank you for your attention!

Can Tongur
can.tongur@scb.se

Statistics Sweden
www.scb.se

The Swedish CPI
kpi@scb.se