What impact does product specification have on a Fisher price index?

Claude Lamboray
Eurostat Unit C4

17th Meeting of the Ottawa Group on Price Indices
7-10 June 2022, Rome, Italy
A three-stage aggregation problem

Stage 1: Individual transactions
- Unit values

Stage 2: Individual products
- Multilateral method

Stage 3: Elementary indices
- Laspeyres-type

Higher-level aggregates
Product specification matters

• Product specification could jeopardize any gains in bias reduction that we would expect from using scanner data (ECB, 2021).
  
  • If too tight: **Matched-model bias**

  • If too broad: **Unit value bias**

• Assignment bias vs. assortment bias (Von Auer, 2017), MARS method (Chessa, 2019).
What is the product to be priced?

1. Aggregation over time
   - Average price over several weeks (for a specific outlet and item code)

2. Aggregation over outlets
   - Average price over several weeks and outlets (for a specific item code)

3. Aggregation over item codes
   - Average price over several weeks, outlets and item codes
Matching, grouping and imputation

Table 1: Matching

<table>
<thead>
<tr>
<th></th>
<th>$p_{t-1}$</th>
<th>$p_t$</th>
<th>Price change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item A</td>
<td>25</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Item B</td>
<td>-</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Item C</td>
<td>40</td>
<td>42</td>
<td>+5.0%</td>
</tr>
<tr>
<td>Item D</td>
<td>30</td>
<td>33</td>
<td>+10.0%</td>
</tr>
<tr>
<td>Item E</td>
<td>20</td>
<td>23</td>
<td>+15.0%</td>
</tr>
<tr>
<td>Geo. avg. price</td>
<td>28.8</td>
<td>31.7</td>
<td>+9.9%</td>
</tr>
</tbody>
</table>

Table 2: Grouping

<table>
<thead>
<tr>
<th></th>
<th>$p_{t-1}$</th>
<th>$p_t$</th>
<th>Price change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hom. prod. (Items A and B)</td>
<td>25</td>
<td>58</td>
<td>+132.0%</td>
</tr>
<tr>
<td>Item C</td>
<td>40</td>
<td>42</td>
<td>+5.0%</td>
</tr>
<tr>
<td>Item D</td>
<td>30</td>
<td>33</td>
<td>+10.0%</td>
</tr>
<tr>
<td>Item E</td>
<td>20</td>
<td>23</td>
<td>+15.0%</td>
</tr>
<tr>
<td>Geo. avg. price</td>
<td>27.7</td>
<td>36.9</td>
<td>+32.5%</td>
</tr>
</tbody>
</table>

Table 3: Imputation

<table>
<thead>
<tr>
<th></th>
<th>$p_{t-1}$</th>
<th>$p_t$</th>
<th>Price change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item A</td>
<td>25</td>
<td>58 (imp.)</td>
<td>+132.0%</td>
</tr>
<tr>
<td>Item C</td>
<td>40</td>
<td>42</td>
<td>+5.0%</td>
</tr>
<tr>
<td>Item D</td>
<td>30</td>
<td>33</td>
<td>+10.0%</td>
</tr>
<tr>
<td>Item E</td>
<td>20</td>
<td>23</td>
<td>+15.0%</td>
</tr>
<tr>
<td>Geo. avg. price</td>
<td>27.7</td>
<td>36.9</td>
<td>+32.5%</td>
</tr>
</tbody>
</table>
Matched, imputation and hybrid indices

• **Matched** Fisher index $P_{MF}$
  
  Index calculated from matched items.

• **Imputation** Fisher index $P_{IF}$
  
  Index calculated from matched, new and disappearing items.

  For new and disappearing items, a price is imputed in the missing period.

• **Hybrid** Fisher index $P_{HF}$
  
  Items are first grouped together.

  Index calculated from the matched groups of items.
Imputation method

| Price of a missing item (in period t) | = | Average price of the group to which the item belongs (in period t) |

- The price of a missing item corresponds to the average price of items of ‘similar quality’ in the same period.
- No additional information is needed apart from the assignment of items into groups.
- Imputation can be modelled as a regression (dummy variable for each group).
  - Imputation Fisher = missing prices are estimated with this model
  - Hybrid Fisher = all prices are estimated with this model
What drives the difference between the matched, imputation and hybrid indices?

- We combine the results in De Haan, 2001 (imputation Fisher) and in Diewert and Von der Lippe, 2010 (hybrid Fisher) in order to disentangle the impact of matched, new, and disappearing items.

<table>
<thead>
<tr>
<th></th>
<th>Matched</th>
<th>*</th>
<th>New</th>
<th>*</th>
<th>Disappearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{IF}^{01} )</td>
<td>1</td>
<td>*</td>
<td>( \frac{v_{11}^1}{v_{11}^0} )</td>
<td>*</td>
<td>( \frac{\delta_{00}^1}{\delta_{00}^0} )</td>
</tr>
<tr>
<td>( P_{MF}^{01} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{HF}^{01} )</td>
<td>( \mu )</td>
<td>*</td>
<td>( v_{11}^0 \frac{v_{10}^1}{v_{10}^0} )</td>
<td>*</td>
<td>( \frac{1}{\delta_{00}^1 \delta_{01}^0} )</td>
</tr>
<tr>
<td>( P_{IF}^{01} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{MF}^{01} )</td>
<td>( \mu )</td>
<td>*</td>
<td>( v_{11}^1 \frac{v_{10}^1}{v_{10}^0} )</td>
<td>*</td>
<td>( \frac{1}{\delta_{00}^0 \delta_{01}^0} )</td>
</tr>
</tbody>
</table>
Matched-model bias and unit value bias

- **Matched-model bias**
  (Matched compared to imputation index) \( b_{MM}^{01} = \ln \left( \frac{P_{MF}^{01}}{P_{IF}^{01}} \right) \approx \frac{P_{MF}^{01}}{P_{IF}^{01}} - 1 \)

- **Unit value bias**
  (Hybrid compared to imputation index) \( b_{UV}^{01} = \ln \left( \frac{P_{HF}^{01}}{P_{IF}^{01}} \right) \approx \frac{P_{HF}^{01}}{P_{IF}^{01}} - 1 \)

- The difference between the matched and hybrid indices can be explained by these two biases:

  \[
  \ln \left( \frac{P_{HF}^{01}}{P_{MF}^{01}} \right) = b_{UV}^{01} - b_{MM}^{01}
  \]
Matched-model bias and unit value bias

- Implications for product specification:
  - If \( b_{MM}^{01} \) is small and \( b_{UV}^{01} \) is large: use the tight product specification (matched index)
  - If \( b_{MM}^{01} \) is large and \( b_{UV}^{01} \) is small: use the broad product specification (hybrid index)
  - If both \( b_{MM}^{01} \) and \( b_{UV}^{01} \) are large: impute the missing prices (imputation index)
Matched-model bias and unit value bias

• From a bilateral to a multilateral index:
  • GEKS-\textit{matched} (based on the matched Fisher index)
  • GEKS-\textit{imputation} (based on the imputation Fisher index)
  • GEKS-\textit{hybrid} (based on the hybrid Fisher index)

• All these GEKS indices are \textit{transitive} and thereby solve the problem of 'chain drift' caused by the bouncing of prices and quantities.
Matched-model bias and unit value bias

- However, the GEKS indices are not necessarily exempted from the matched-model bias and unit value bias.

- The analysis in the bilateral case extends to the multilateral case:

\[
\ln \left( \frac{P_{GEKS-Hybrid}^{0t}}{P_{GEKS-Matched}^{0t}} \right) = b_{GEKS-UV}^{0t} - b_{GEKS-MM}^{0t}
\]

Average of bilateral unit value biases

Average of bilateral matched-model biases
Matched-model bias and unit value bias

- The matched-model and unit value biases cannot be directly transposed to other multilateral methods such as the Geary Khamis.
  - There is a hybrid and matched variant of Geary Khamis, but not an imputation variant.

- The Geary-Khamis can be seen as a quality-adjusted unit value index.
  - Depending on quality adjustment factors and observed/imputed prices, a quality adjusted unit value index may be closer to the matched Fisher index or to the imputation Fisher index.
Example 1

• Data set (Milk) included in the *PricelIndices* package
• 75 item codes sold in 5 outlets
• Prices in 2 out of the 5 outlets increased by 40%

**Tight** product specification: item*outlet

**Broad** product specification: item (aggregation over outlets)
Example 1 (cont.)

Matched-model bias is low

Unit value bias

Matched index
Example 2

- Data set with 30 items (t-shirts) over 13 periods

**Tight** product specification: Item code

**Broad** product specification: Grouping of items based on their attributes (fabric, sleeves, number of items)
Example 2 (cont.)

Unit value bias

Matched-model bias

Imputation index
Example 3

**Prices for product 1**  
(models 1a, 1b, 1c)

**Prices for product 2**  
(models 2a, 2b, 2c)

**Prices for product 3**  
(models 3a, 3b)

**Tight** product specification: 8 models  
**Broad** product specification: 3 products
Example 3 (cont.)

Unit value bias is low

Matched-model bias

Hybrid index
Conclusions

• Attempt to assess the impact of product specification through formalizing unit value bias and matched-model bias, but:

  • Bias measures depend on imputation method.

  • Unit value bias is only a proxy for measuring the degree of quality differences of products that are grouped together.

  • In practice, there can be many possible product specifications

  • The framework is not adapted to methods that are not responsive to imputed prices (e.g. Geary-Khamis).

  • Instead of calculating a unit value at the level of an individual product, a quality adjusted unit value may be a more accurate target.
Thank you