“Evaluating unit-value price indices in a dynamic item universe”

16th Meeting of the Ottawa Group, 8-10 May 2019
Rio de Janeiro, Brazil

Li-Chun Zhang
Statistics Norway/University of Southampton/University of Oslo

Ingvild Johansen
Statistics Norway

Ragnhild Nygaard
Statistics Norway
Background

• Strategy: Increase the use of scanner data in Norwegian CPI/HICP

• Current official index methods do not fully make use of scanner data potential
  ◦ Until now mostly focus on stable long-lived items
  ◦ Possible to use scanner data of **dynamic items** in regular production?

• Requires **new methods** to be implemented

• Aims of Eurostat grant 2018-2020:
  1. Generic calculation method to be applied across different commodity groups
  2. Incorporate expenditure shares at most detailed level

• No international consensus on scanner data calculation method – NSIs choose differently
TEF – Total Effect Framework

• A systematic approach in order to empirically evaluate different choices

• Measure effects by defining a set of generic diagnostics

• Not just a question of formula - three necessary choices: 1) Homogenous products 2) Reference universe 3) Index base

• The choices dependent on the dynamism of item universe

• Dynamic items: 1) Replacement items 2) Regeneration items 3) Strongly seasonal items
Homogenous products (HPs)

• How to define the product?
  ◦ Defining the product at item code (GTIN) level may be too detailed

• Creation of HPs mostly motivated by replacement items

• For markets with high item churn the effects of **missing replacements** can be large and systematic

• In general, HP-based indices are not more volatile than the item-based, hence no strong indication of classification bias
Homogenous products (HPs) II

- How item codes are classified key choice in HP formation
- HP formation depends on available metadata – often limited
- Approach implemented by Statistics Norway for price indices related to sport clothing and equipment
  - Scanner data from one major sport clothing and equipment retailer
  - HPs defined by **brand blocking** – *Store concept|Raincoat men|Helly Hansen (brand)*
  - In cases with too high heterogeneity only using metadata – “normal price” as additional classification criteria
  - The majority of the HPs entirely based on metadata
  - Fixed HPs within the short-term link
Reference universe

- Multilateral index formula does not dictate the choice of reference universe

- Multilateral reference universe **responsive to regeneration items**
  - Given bilateral fixed index base a new item/HP will not be captured until next time the base month is updated

- Diagnostic: Testing the sensitivity to the choice of reference universe
  - Compare bilateral index to multilateral index (Fixed base monthly expanding window) with **fixed item universe 0 to T**
Reference universe II

- Diagnostic: Testing the sensitivity of different window lengths

- Sensitivity of the GEKS index varies across the commodity groups

- The GEKS increases with window length
Index base

- Given that HPs capture replacement items, the choice of index base is most affected by strongly seasonal and regeneration items.
  - **Fixed base month index** – to capture strongly seasonal items will normally require imputation of both price and quantity.
  - Generally, **moving base month t-1**, cannot eliminate chain drift completely.
    - May lead to heavy drift in chained *bilateral* indices, and may create bias in case of strongly seasonal items if first-appearance prices are not appropriately accounted for.
  - **12 month base** - may be the best option for covering strongly seasonal items, but not necessarily for the inclusion of regeneration items.
Index base II

- Diagnostic: Check the actual dynamics in given markets (expenditure shares)

- For commodity groups with seasonal patterns there can be large difference (bicycles, pork, fresh berries etc)

- Greater sensitivity of fixed base month than 12 month base

- No ideal choice of index base to all commodity groups
Splicing options

- Tested different splicing options;
  - “Movement splice”
    - Rolling window of 13 months
    - Month-to-month movement spliced on to existing time series
  - “Fixed base moving window (FBMW) - December splice”
    - Combine rolling window with December linking
Splicing options II

• “Movement splice” captures the **long-term effects** of strongly seasonal items
  ◦ Ex. strawberries only available in June and July
  ◦ More drift in “movement splice” compared to “FBMW with December splice”?

• “**FBMW December splice**” does not capture the long-term effects of strongly seasonal items not present in base month

• Trade off between drift and capturing permanent effects of strongly seasonal items?
Splicing options III

- Marginal total effects on COICOP 01

- “Movement splice” seems to pull the price growth marginally up, while the effects of seasonal items (and regeneration items) pull in opposite direction

- Empirically difficult to conclude – in general small differences between the options

- “Mean splice” - geometric mean of the indices by using every possible splicing alternative - suggested solution, but results in complex production routines
Index formula

- An additional separate choice - set of different index formulas tested and compared
  - GEKS, GUV (Generalized Unit value) indices, official Jevons index (dynamic method) and direct bilateral superlative index

- All necessary choices matters, but the effects of using HPs seem to outweigh the other choices (more systematic)
Conclusions

• Choices are many and answers are not obvious

• In favor of using HP in order to capture replacement items in dynamic universes

• Likely to implementing a multilateral price index formula in order to capture regenerations items in a timely manner

• A fixed length 13-months window seems to be a good choice
  ◦ Splicing – a trade off?

• Own research as well as experiences and advises from NSIs and the statistical community very important for making final conclusion
Thank you!