New data sources in the UK consumer price statistics

Jack Philips

UN CPI Group of Experts Meeting, Geneva, 7th - 9th May 2018
Initial research stage into using alternative data sources finished in August 2017 (summarised here)

Focus was on using web scraped data to replace or supplement the traditional collection of prices

Recommendations fed into development of the Consumer Prices Data Collection Strategy
Strategic direction…

Support current and future analysis needs

Keep pace with how items are purchased in practice

Consider alternative data sources first

Provide value for money

Be based on resilient systems and processes
Future collection methods

• Point of sale scanner data for the largest retailers (including online transactions)
• Web scraped data for other products which are dominated by a few smaller retailers or for which attribute data is also required.
• Manual collection, as now, where point of sale and web scraped data are not available/efficient.
Work plan

Four key areas of further work will be required:
1. Obtaining robust sources of alternative data
2. Development of systems to support the inclusion of new data sources
3. Methods research
4. Assessing the impact on our aggregate consumer prices statistics
Robust data sources

• Previous research focused on web scraped data collected via in-house methods
• High resource costs (maintenance and expansion of collection)
• Decision to procure web scraped data from a third party supplier to begin May 2018
• Continue to engage with retailers to access scanner data (new legislation)
Open methodology questions…

- Coverage of basket
- Extending classification structure
- Expenditure weights for web scraped data
- Implications for local collection sampling framework
- Clearance prices
- Product definition/ relaunches
- New index methodology

…
Coverage of the basket

• How can we identify which sources of data to target for which items?

• Web scraping:
  • Items which are dominated by collection from a few specialised retailers (e.g. books)
  • Evidence of national pricing policies
  • Requirements for detailed attribute information (e.g. clothing, technology goods)
Comparison of prices (clothing)
Extending the classification structure

Current classification structure based on UN COICOP hierarchy.

Two possible extensions:
1. Expand the list of items that are captured underneath the COICOP5 level
2. Create a new layer of hierarchy at the sub-item level for certain items
Extending the classification structure

COICOP4: Fruit

COICOP5: Fresh or chilled fruit

ONS Item level: Bananas
- Products that are currently collected
- Products not currently included (e.g. plantains)
- Incorrectly classified items

Other item level categories...

Other COICOP5 categories...

ONS Item level: Peaches
Example: printers

Figure Y: Distribution of web scraped printer prices
Example: printers

Printers: Bilateral Chained Index (Dec 2016 - Jul 2017)
Extending the classification structure

Considerations:
• Expenditure weights for new categories
• System requirements
• Different levels of hierarchy for different items
• Aggregation with traditional price collection
• Item definitions
Can we identify any quick-wins?

- Hedonic models currently used in 4 areas:
  - Laptops, smart phones, tablets, PCs
- Burden on collectors to gather enough attribute information to update these models
- Efficiency saving – can we use the new web scraped data for these models instead?
Next steps

- Continue working to secure access to scanner data (new legislation)

- New web scraping data arriving May 2018 – start exploring this data and using it to test the methodology development (for example, expenditure weights for web scraped data)

- Consider a staged approach to implementing these new data in our consumer prices indices
Thank you for your attention..