

## Panel on Improving Cost-of-Living Indexes and Consumer Inflation Statistics in the Digital Age

### **Open Session: Innovative use (or planned use) of alternative data by statistics agencies in price measurement programs**

**October 7, 2020**  
6-9am (US eastern), 11am-2pm (UK), 9-12pm (Canberra)

**Session goals:** The panel is charged with examining the potential to improve the CPI by supplementing (or in some cases replacing) traditional survey-based data collection with an approach that blends multiple (survey and non-survey, government and commercial) data sources. The focus of this session is on elementary indexes (improved timeliness, greater detail, reduced costs, etc.) but the study is also covering other stages of CPI methodology, such as construction and updating of upper level category weights. Presenters will update the study committee about the initiatives of their statistical offices to replace survey sample prices with alternative data, expand samples, and develop methodologies that exploit new data opportunities. Questions may include: What approaches were used for evaluating quality and incorporating new data? What were the motivations (cost reduction, detail of information, timeliness, etc.)? And what are the statistical office's next steps in data infrastructure development?

#### Presentations (20 minutes each, give or take)

- **Rob Cage, Anya Stockburger**, BLS. Overview of BLS data plans and priorities. Status report on the agency's big data projects; outline of their research agenda.  
[Support materials: Konny, Williams, and Friedman, 2019, *Big Data in the U.S. Consumer Price Index: Experience & Plans*]
- **Tanya Flower, Helen Sands**, ONS. Ongoing and planned CPI data transformation at ONS, including integration of scanner and scraped price data. Discussion of approach developed by ONS for dealing with issues associated with high product churn. Role of new data sources for measuring inflation during the pandemic.  
[Support materials: *Introducing alternative data sources into consumer price statistics: May 2020*, ONS; *Research and developments in the transformation of UK consumer price statistics: September 2020*, ONS]
- **Ken Van Loon**, Statistics Belgium. Use of scanner data and web scraping, and implementation of a multilateral method (GEKS).  
[Support materials: "Integrating big data in the Belgian CPI," by Ken Van Loon and Dorien Roels; "Evaluating multilateral index methods on scanner data," by Ken Van Loon]
- **Leigh Merrington, Catherine Smyth**, ABS. Discussion will include:
  - A brief history of the ABS's use of scanner data in the CPI. How has the agency gone about implementing data changes (e.g., communication with external users, research conducted,

- input from international experts). Benefits of using scanner data - increased accuracy, lower collection and processing costs.
- A brief summary of use by ABS of web-scraped data. New insights from use of alternative data sources.

[Support materials: “An Implementation Plan to Maximise the Use of Transactions Data in the CPI,” ABS; *A detailed list of types of data sources used in the Australian CPI* (excel file) “Utilizing big data and multilateral index methods to produce the Australian CPI: Theory, implementation and empirical results,” by van Kint, de Haan, Webster]

- **Heidi Ertl**, Statistics Canada. Statistics Canada is in year 3 of its CPI Modernization plan, which is largely focused on the use of alternative data sources. Discussion will highlight the agency’s experiences and lessons learned in the following areas:
  - Use of alternate data sources: including scanner data, web-scraped and application programming interface (API)
  - Machine learning: processing and mapping to the CPI classification structure
  - CPI basket updates: experimental work to adjust expenditure weights using new data sources, increased frequency
  - Services and the digital economy: development work to enhance representativeness of the CPI basket (digital products, services, digital platforms, shipping costs, etc.)

Discussion. Synthesis and takeaways from the above presentations, focusing on data quality issues and statistical problems with non-survey data. (15 minutes each)

- **Jens Mehrhoff**, Deutsche Bundesbank. Statistical assessment of probabilistic data sets versus scanner and web scraped data sources. Given the state of the literature, and work taking place in other countries, what are some productive paths forward for BLS in using scanner/scraped data.
- [Support materials: “Consumer price indices: from traditional to new data sources and techniques” (presentation by Jens); “The Value Chain of Scanner and Web Scrapped Data,” by Jens Mehrhoff]
- **Kevin Fox**, UNSW. Synthesis of material presented by the statistical agencies; observations about what has been done to date and what is being planned, including reflections on which approaches are theoretically preferred, while noting some practical considerations.
- [Support materials: “Substitution Bias in Multilateral Methods for CPI Construction” by Diewert and Fox]
- **Claude Lamboray**, Eurostat. Elementary indexes, weighting issues. Data quality/statistical issues when alternative data are used as a substitute for, or are blended, with traditional data. Thoughts on the challenges facing various statistical agency initiatives related to alternative data.

Open discussion with panel, BLS, and invited participants (30 minutes)

Additional experts invited to participate in this discussion (some of whom will feature in the panel’s next meeting on work taking place beyond the statistical agencies) include: **David Byrne** (FRB), **Alberto Cavallo** (Harvard) **Abe Dunn** (BEA), **Erica Groshen** (Cornell), **David Johnson** (Michigan), and **John Stevens** (FRB).