Utilising big data and multilateral index methods to produce the Australian CPI

Marcel van Kints: Prices and Economic Research Branch, Australian Bureau of Statistics

Jan de Haan: Division of Corporate Services, IT and Methodology, Statistics Netherlands, and OTB, Faculty of Architecture and the Built Environment, Delft University of Technology

Michael Webster: Methodology Division, Australian Bureau of Statistics

The views expressed are those of the authors and do not necessarily reflect the views of the Australian Bureau of Statistics and Statistics Netherlands.

Outline

1. Introduction to scanner data
2. Practical considerations
3. Multilateral index methods
4. Implementation
5. Empirical results

Consumer Price Index: Data Sources

- Measures change in prices of goods and services that households consume (ILO)
- Prices are collected via several modes, including
  - Personal visits
  - Telephone
  - Online
  - Scanner data

Scanner data: Description

- Sourced from retailers (originally)
- High or full item coverage
- Item level information
  - Expenditure, quantity, description, provider's classification
- Generally pre-aggregated by location (outlet) and time period (week, month)

Practical considerations: Acquisition

- Several statistical offices have negotiated supply directly from retailers and/or from intermediaries and market research companies
- Different benefits and challenges in negotiating with retailers vs intermediaries
- Negotiations may take some time

Practical considerations: Preparing for use

- Challenges include
  - Development of IT systems
  - Item definition and classification
  - Quality assurance
  - Skill development

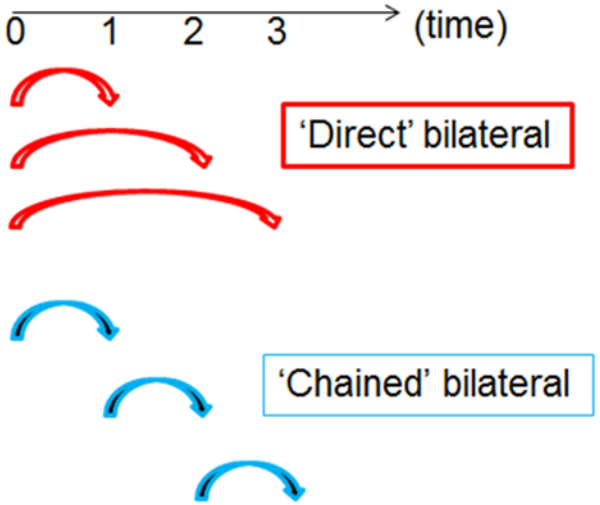
Scanner data: Potential uses

- Improvements to sampling
  - Confront with, replace, or expand samples
- Improvements to weighting
  - Update weights for categories or items
- Challenges relate to dynamic nature of data
  - Rapid item turnover, fluctuations in expenditure

Bilateral approaches

- Desirable to produce weighted index using all items, such as Törnqvist

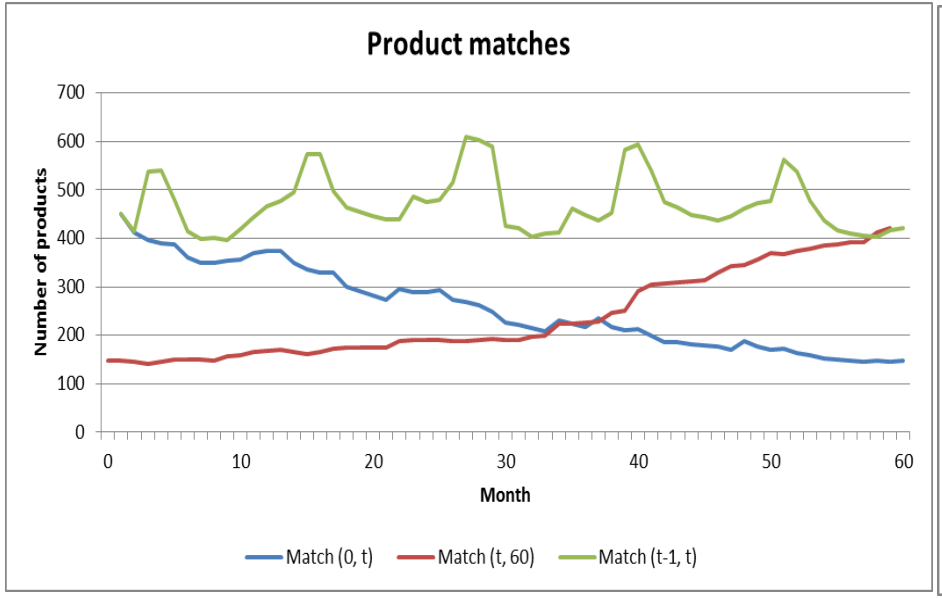
$$P_T^{0,t} = \prod_i \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}}$$



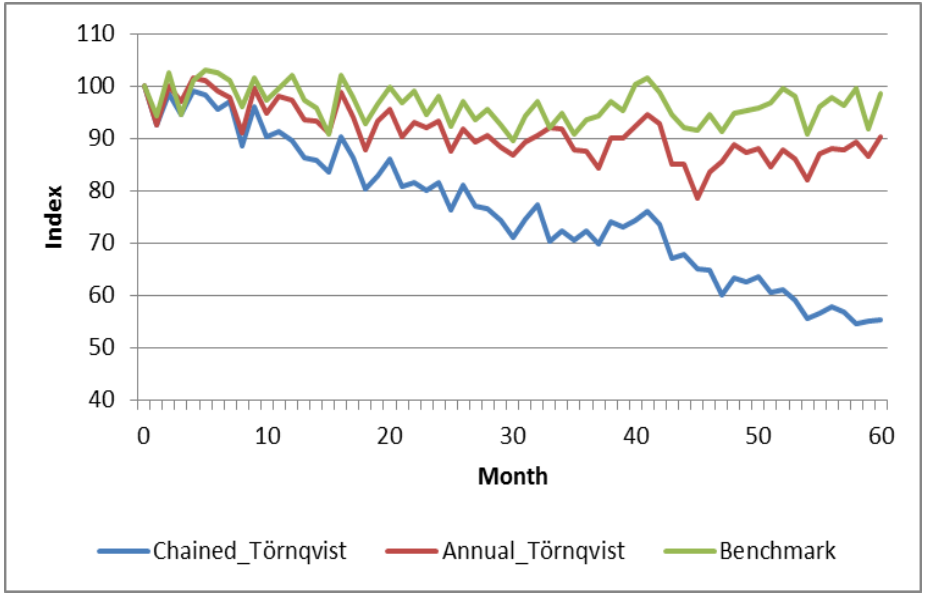
- Bilateral approaches not always suitable, however…

Challenges with bilateral approaches

- Reliability of direct indexes diminishes over time

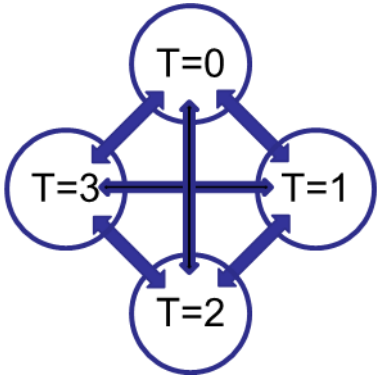


- Chained indexes can drift over time



Multilateral indexes

- Ivancic, Diewert and Fox (2011) proposed using multilateral indexes with scanner data
- Simultaneous price comparisons between multiple time periods
- These methods have traditionally been used for spatial comparisons



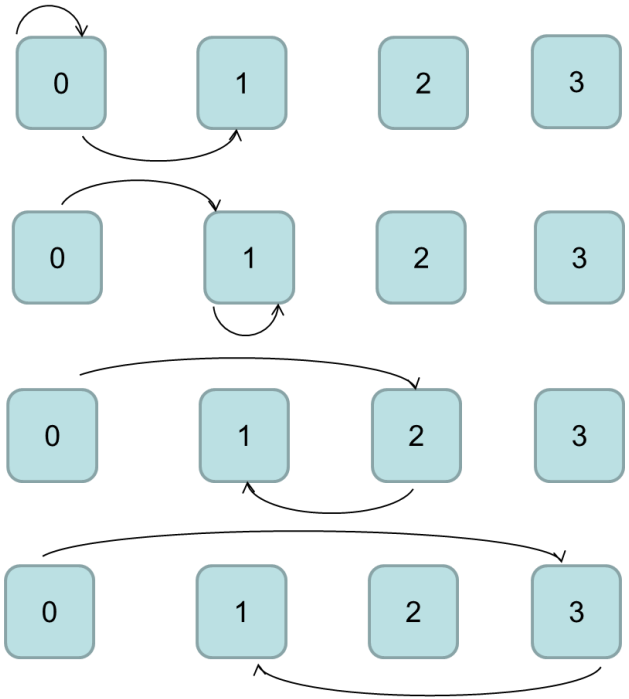
Multilateral comparison

GEKS and CCD methods

- GEKS: calculate bilateral (Fisher) price comparisons between each pair of periods then combine:

$$P_{GEKS}^{0,t} = \prod_{l=0}^{T} \left(P_F^{0,l} \times P_F^{l,t} \right)^{\frac{1}{T+1}}$$

- Caves, Christensen and Diewert (1982) suggested using Törnqvist index in place of Fisher (yields CCD or GEKS-Törnqvist)



Geary Khamis (GK) and Time Product Dummy (TPD)

temporal price level

- Basic model: $p_i^t = P^t \times \pi_i$ or $p_i^t / P^t = \pi_i$

price

item price level

Multiplication by q_i^t and summation yields

$$\sum_i \left(p_i^t \times q_i^t\right) = \sum_i \left(P^t \times \pi_i \times q_i^t\right)$$

and

$$\sum_t \left(\frac{p_i^t}{P^t} \times q_i^t\right) = \sum_t \left(\pi_i \times q_i^t\right)$$

which define the GK index.

Exponentiation by s_i^t and multiplication yields

$$\prod_i \left(p_i^t\right)^{s_i^t} = \prod_i \left(P^t \times \pi_i\right)^{s_i^t}$$

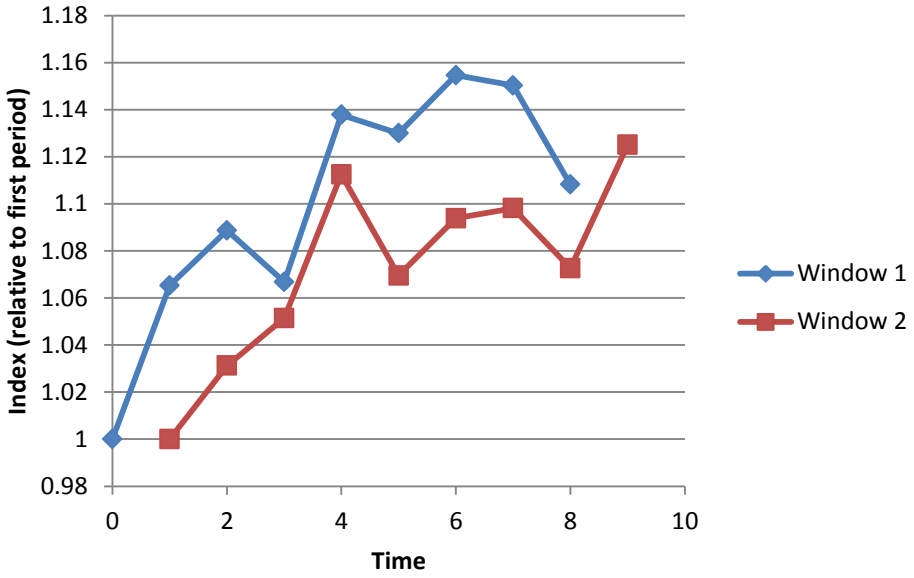
and

$$\prod_t \left(\frac{p_i^t}{P^t}\right)^{s_i^t} = \prod_t \left(\pi_i\right)^{s_i^t}$$

which define the TPD index

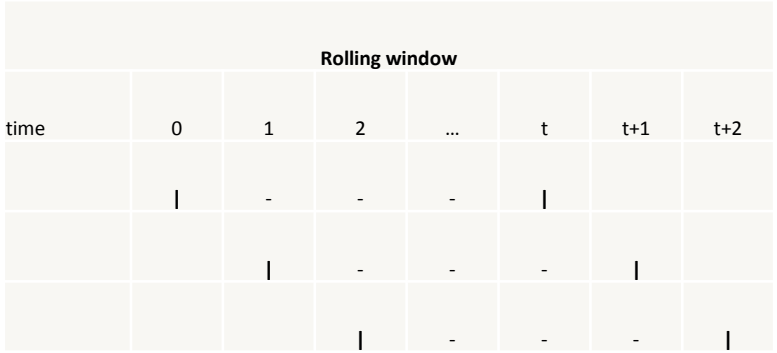
Methods for extending multilateral indexes

- Need to extend index each period
- Data from current period can alter comparisons between earlier periods
- Can't revise index in normal circumstances
- Two questions to address:
  1. How do we form a multilateral "window" incorporating the current period?
  2. How do we splice the results onto previous index levels?

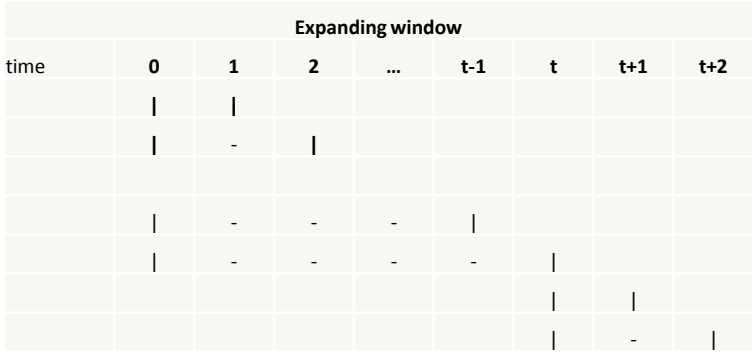


Methods for extending multilateral indexes

- *Rolling* or *expanding window* approaches



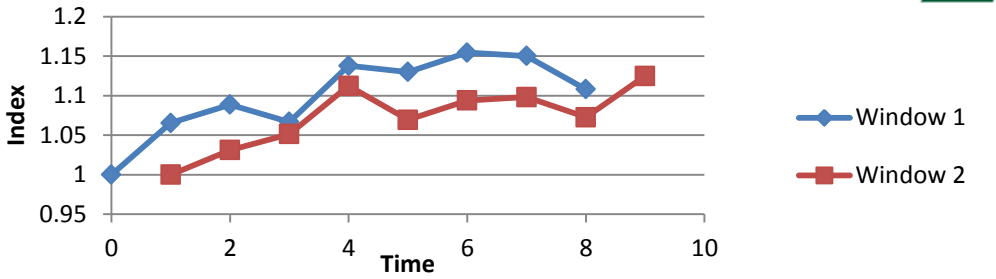
- Fixed length
- Variable start point



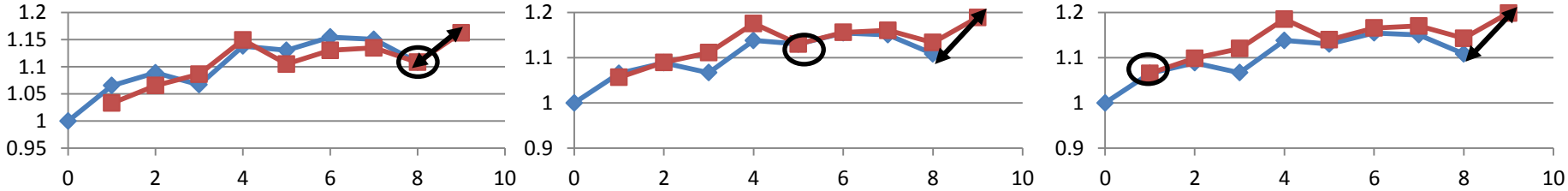
- Variable length
- Fixed start point (can be updated from time to time)

Rolling window extension methods

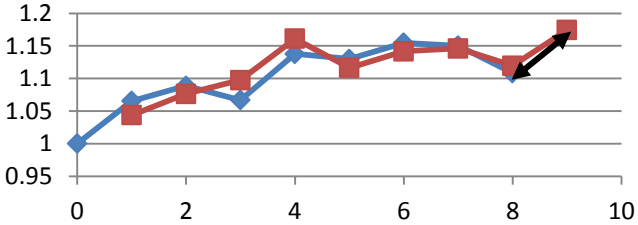
How should we link together successive windows?



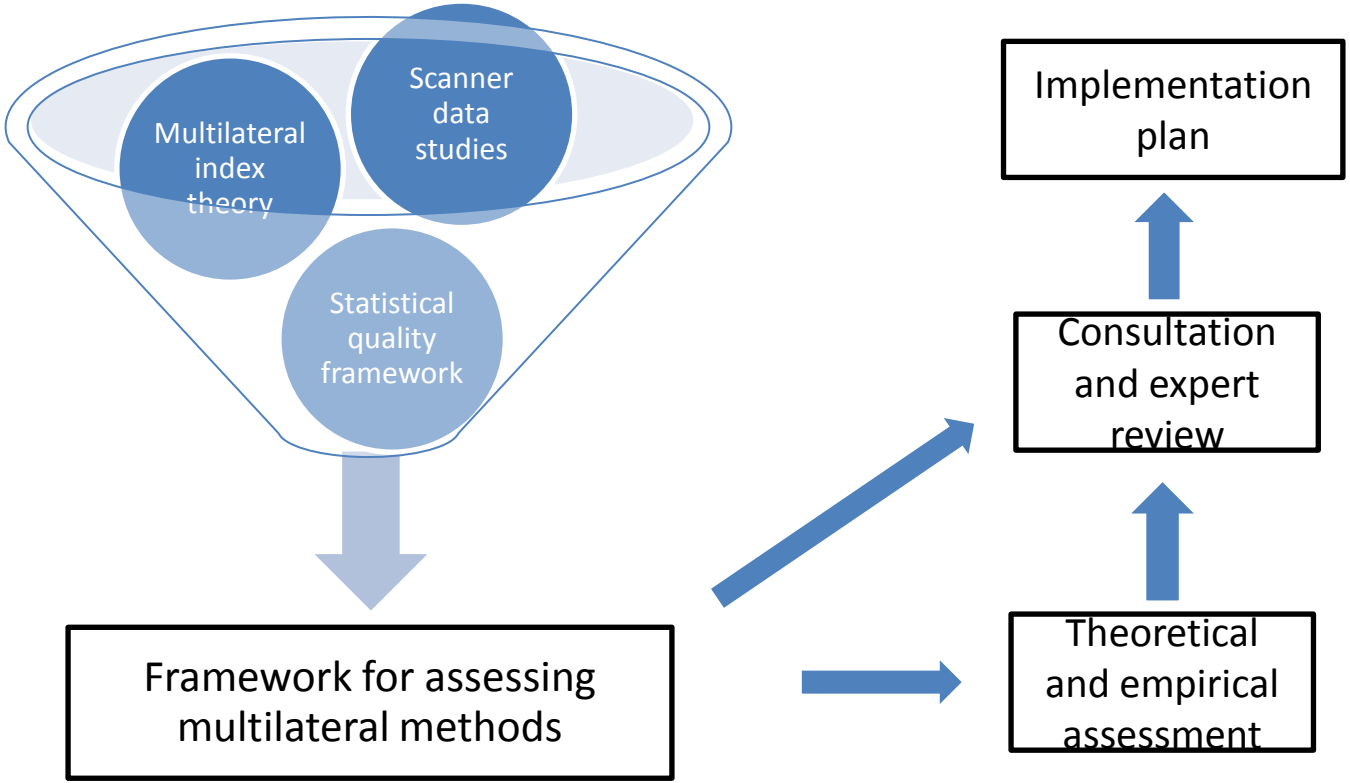
Link in one period (which?)



Or take the geometric mean over all possible links

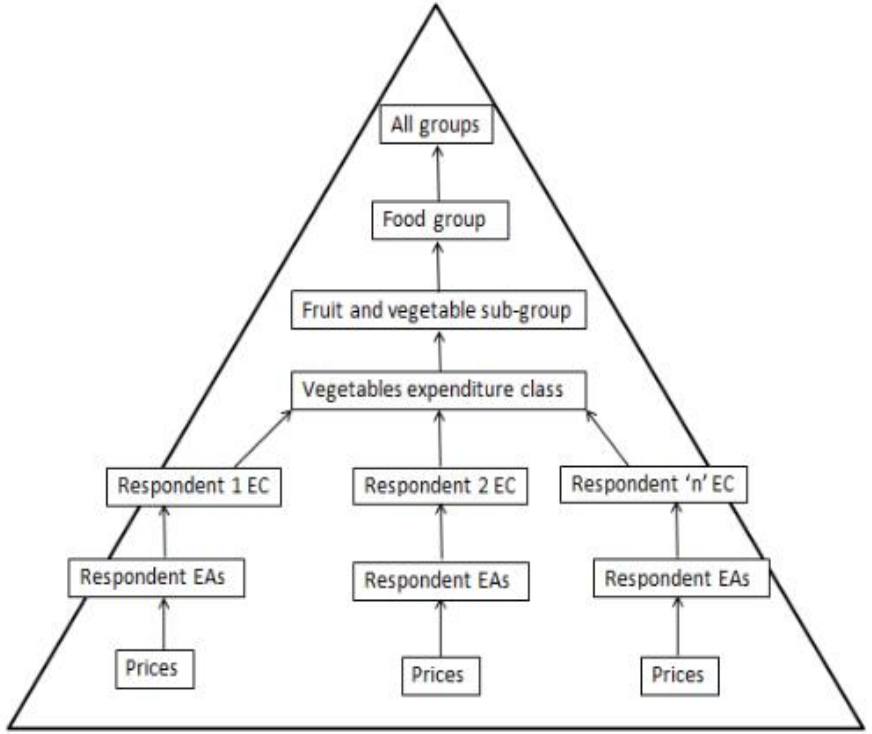


Journey towards multilateral methods: Australian approach

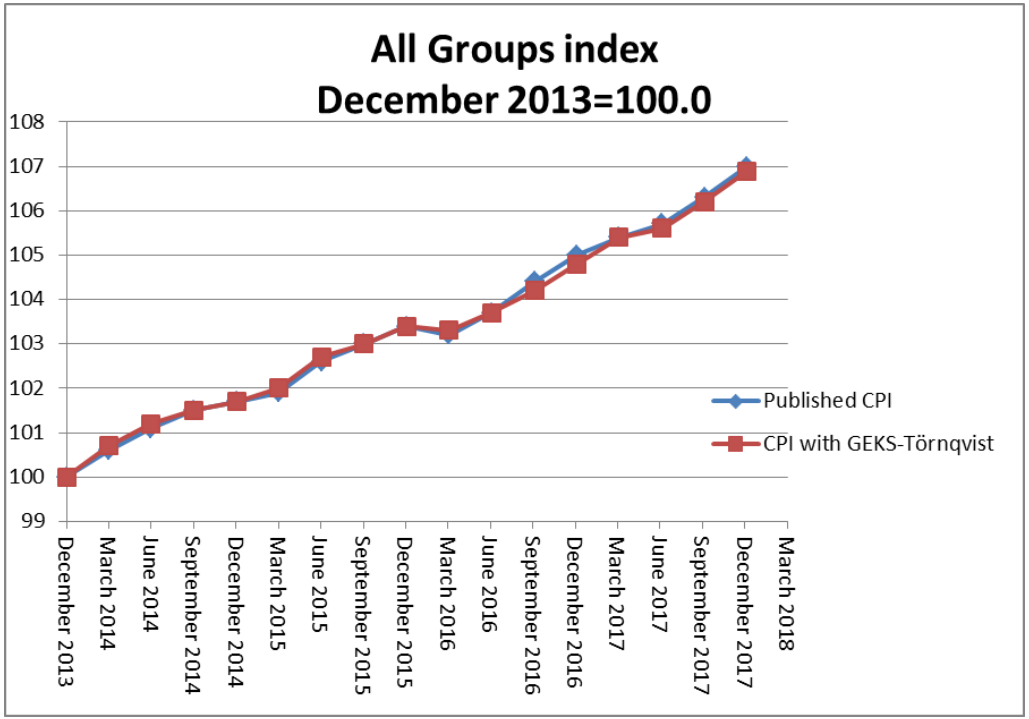


Implementation of new methods: Australian experience

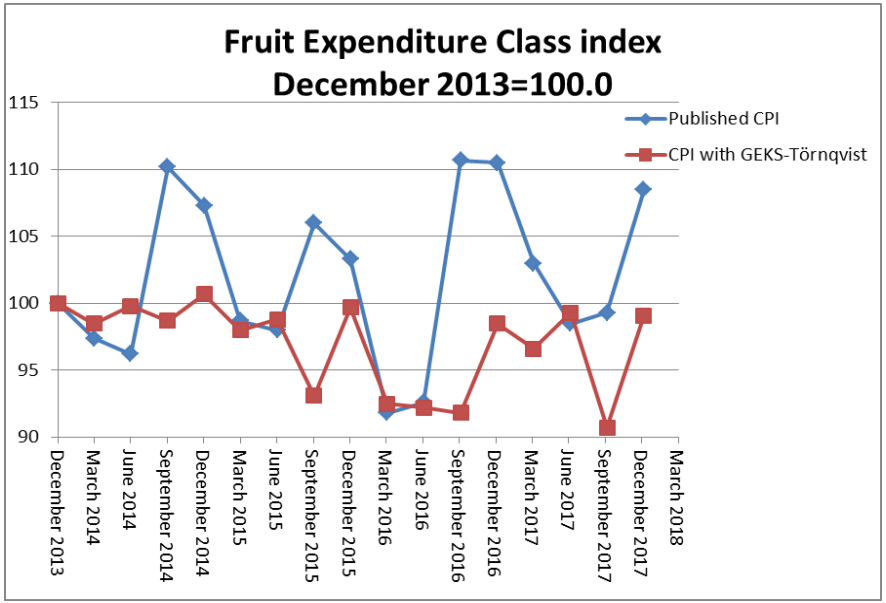
- Started using GEKS-Törnqvist in the CPI in December 2017
- Developed and adopted techniques to minimise manual intervention in resolution of anomalies, quality adjustment etc.
- Modified index structure to weight respondents separately

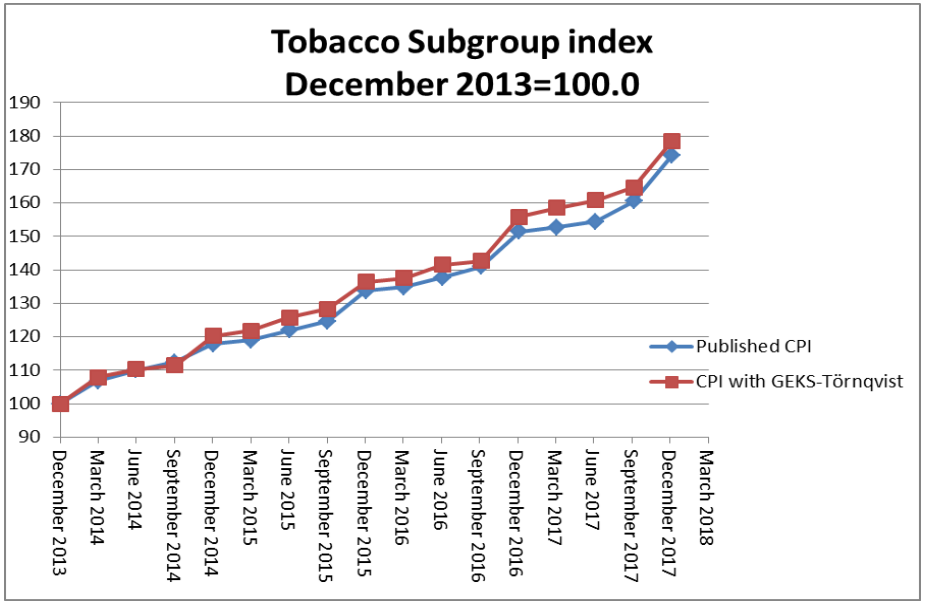


Empirical comparisons



Empirical comparisons





Australian Bureau of Statistics

Questions?