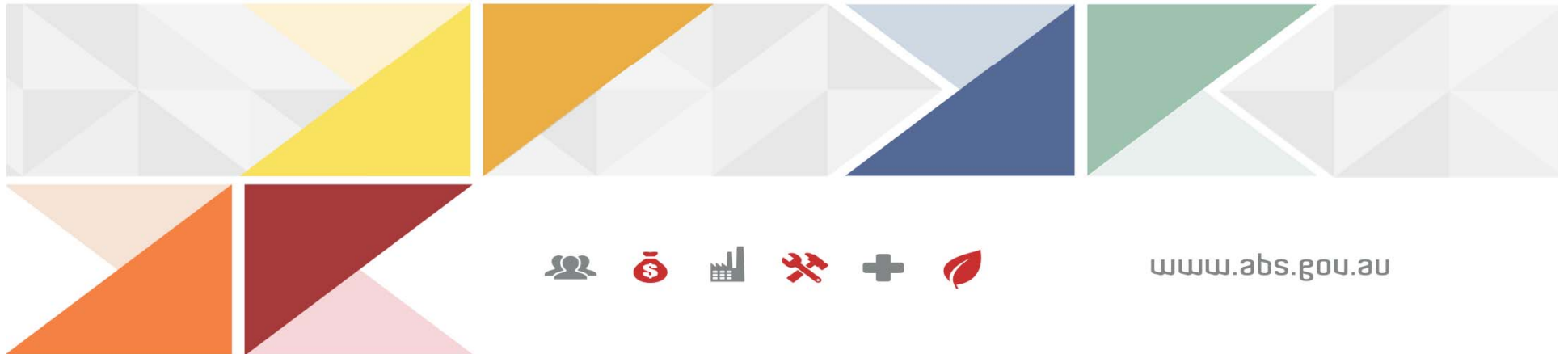




Scanner data and on-line data and official statistics

4th UN Conference on Big Data

Presented by: Michael Holt
(michael.holt@abs.gov.au)





Use of scanner data



- ABS in a transformation environment – seeking ways to utilise ‘big data’ for compilation of economic statistics
- Transactions data contains detailed information about transactions, dates, quantities, product descriptions, and values of products sold
- *Enhancing the Australian CPI: a roadmap* ([ABS 2015](#)) sets out four research priorities
 - Frequency of weight updates

Product ID	Store Location	Product Description	Time Period	Revenue (\$)	Units sold	Unit value (\$/unit)
U0001	Sydney CBD	CARROTS PREPACKED 1KG	Jan-16	5000	2500	2.00
U0001	Sydney CBD	CARROTS PREPACKED 1KG	Feb-16	7000	4000	1.75
U0001	Sydney CBD	CARROTS PREPACKED 1KG	Mar-16	4100	2000	2.05
U0001	Sydney CBD	CARROTS PREPACKED 1KG	Q1-16	16100	8500	1.89
U0002...						

Use of scanner data



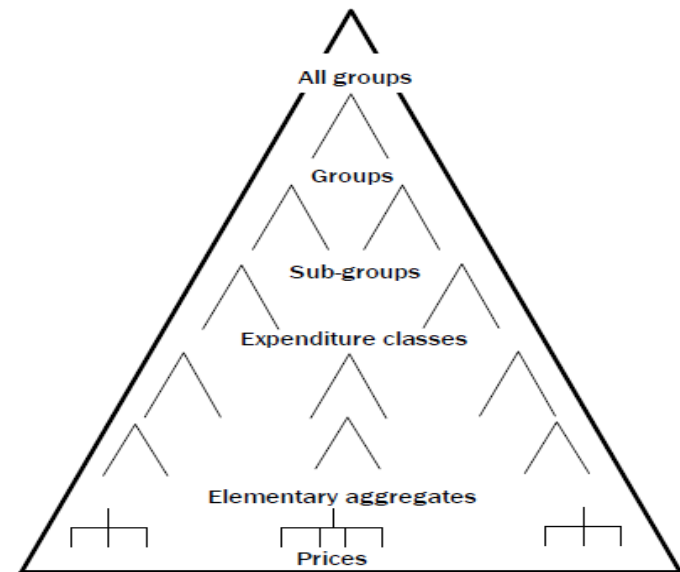
- How scanner data enters the ABS?
- Data negotiated directly with the retailer – enters the ABS under the Census and Statistics Act (1905)
- Deed of Understanding (DOU) is established between the ABS and providers
- DOU is a written formal agreement between two parties that specify the arrangements of the transfer of data between two parties
- Not legally enforceable – the purpose is to build relationships between the two parties



Use of scanner data

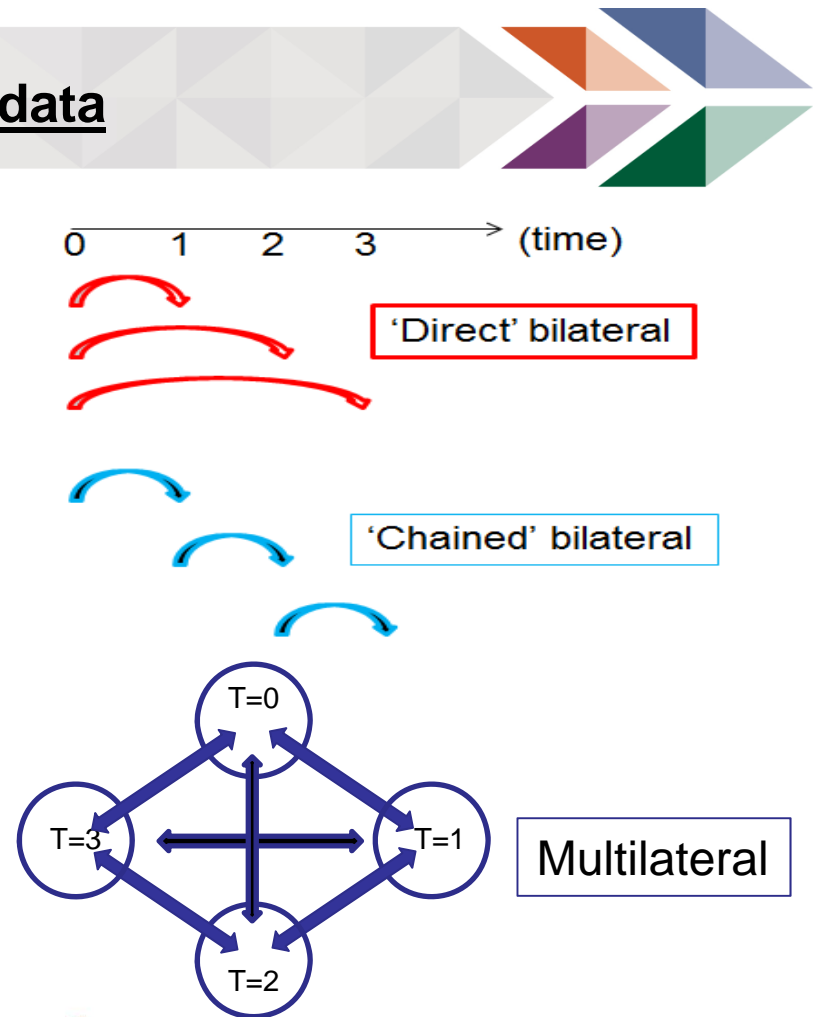


- Transactions data used to compile ~ 25% of CPI
- Stock keeping unit (SKU) defines a product
- Current method directly replaces field collected prices with unit values derived from transactions data within elementary aggregates (Jevons formula)
- Quality benefits: average unit value, increased respondent coverage, informed sampling choices
- Cost benefits: less labour intensive



Use of scanner data

- Desire to use all the information on scanner datasets
- Dynamic nature of scanner data can make theoretically ideal **bilateral** indexes (e.g. Fisher, Törnqvist) perform badly
- A possible solution - use **multilateral** indexes instead (Ivancic, Diewert and Fox, 2011)
 - Simultaneously compare prices over three or more periods
 - Historically used for international comparisons



Use of scanner data



- The ABS has consulted widely with the statistical and user community. Some examples include:
- Peer review of index methods by two international price index experts
- Collaboration with other National Statistical Institutions (e.g. New Zealand, Netherlands)
- Bilateral/multilateral workshops with government (e.g. central bank)
- Presentations at relevant domestic and international forums



Use of web scraped data



- Progressed significantly over the past 18 months
- Web scraping performed within the ABS for product categories such as clothing, audio visual, hardware products from ~ 40 retailers
- Method used replaces field collected prices with a comparative web scraped price using Jevons formula
- How can we maximise the use of web scraped data? Lack of expenditure information makes it difficult to benchmark. Maximising the use of web scraped data is upcoming item on ABS research agenda



Public/private data partnerships



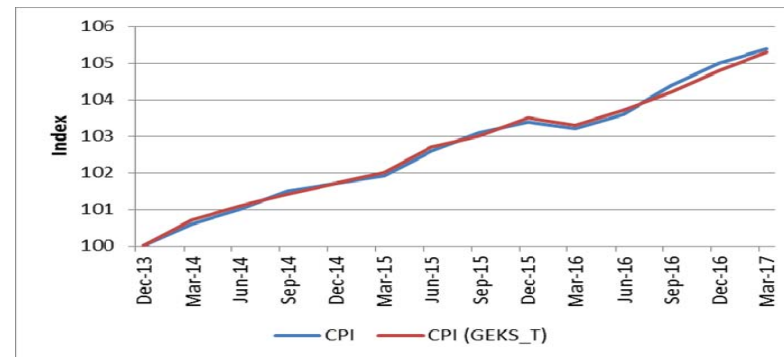
- NSIs typically do not immediately purchase data – procurement at the ABS judged on a case by case (benefit-cost) basis
- Data procured to compile residential property price indexes (Corelogic) and automotive fuel
- Relationships rather than a transaction – private companies certainly add value with cleaning and classification of datasets
- Dialogue with NSIs has reported positive experiences with public/private partnerships



Public/private data partnerships



- Negotiation of scanner datasets – how can we incentivise respondents to report data?
- Scanner data methods have advanced significantly in past ten years - highlights the incredible advancement can be made to official statistics when academics and public institutions work together
- Ambitious in change, cautious in implementation – change in methods requires period of research and consultation



Future of the Australian CPI



- ‘Big data’ such as scanner data and web scraping provide a solution to some long-term challenges for the Australian CPI, including:
 - Producing more outputs with fewer resources
 - More frequent publication (monthly)
 - Non-metropolitan temporal price indexes
 - Spatial price indexes
- What can we offer retailers?
- The required skillset of a (price) statistician – the emergence of the ‘data scientist’

