Transactions Data: From Theory to Practice

Khanh Moore

Australian Bureau of Statistics

Abstract: The opportunity that big data presents to statistical agencies is the potential to produce more relevant and timely statistics than traditional data sources such as sample surveys. Hence, big data has received interest from the Australian Bureau of Statistics (ABS) as an input into the compilation of official statistics. An example of the ways that the ABS can make use of big data is incorporating transactions data into the compilation of key economic statistics such as the Consumer Price Index (CPI). This paper outlines how the ABS has approached the various practical and technical issues encountered while attempting to use transactions data to construct price indexes. Other challenges faced when utilising very large volumes of detailed, high frequency data are also examined.

Keywords: big data**,** consumer price index, CPI, transactions data

1. Introduction

Big data refers to the large volume of structured or unstructured data that organisations generate and store. It is characterised as data that generally contain high volume, high velocity and/or high variety information and demands cost-effective, innovative ways of processing for enhanced insight and decision making[1].

The opportunity that big data presents to statistical agencies is the potential to produce more relevant and timely statistics than traditional data sources such as sample surveys. As an input into official statistics, either for use on its own, or combined with more traditional data sources, Big data could help position National Statistical Offices (NSOs) to improve the accuracy of their measures or the quality of the statistics produced. It can also help improve the comprehensiveness of official statistics by addressing existing data gaps.

An example of Big data is transactions data from major retailers obtained from the electronic capture of product information at the point of sale. Transactions data contain detailed information about the business name and location of the transaction, date and time, quantities, product descriptions, values of products sold as well as their prices.

These data are collected and aggregated by retail businesses and may be used to estimate unit values, changes in unit values, and household expenditure on the items offered for sale by the business. The ABS sees transactions data as a rich source of information. Access to electronic data files will help reduce provider burden and the associated costs of physically collecting data.

This is particularly relevant for improving the coverage of products and outlets sampled in the CPI. Transactions data allow conceptual as well as functional changes in price measurement [ILO Manual, 2004]: "*In theory, to construct a perfectly accurate CPI the prices statistician would need to record the price of every variety of every good and service purchased by all the relevant consumers according to the expenditure in scope*". By reusing administrative data collected by the retail businesses rather than having price collectors make special visits to stores enables prices for many more products to be collected from a much greater number of outlets.

Transactions data contains information on both quantities and revenues for each item, which present opportunities for compiling superlative price indexes[2]. However when using high frequency and high volume data, current methods based on superlative indexes are known to have fundamental weaknesses.

[1] http://www.gartner.com/it-glossary/big-data/

[2] ILO Manual (2004, p2) describes superlative index as one that 'treats the prices and quantities in both periods being compared symmetrically. Different superlative indices tend to have similar properties, yield similar results and behave in very similar ways'. Superlative indices are generally the favoured formulas for calculating price indices.

The superlative indexes from period to period must be chained together to form a continuous series which results in a weakness known as chain drift[3] which can upwardly bias the index (de Haan and van der Grient, 2009). Price oscillation or bouncing and the associated quantity shifts are thought to contribute to chain drift (Hill, 1993). Research on transactions data obtained from supermarkets (Ivancic et al. (2009), Krsinich (2011)) has found evidence of substantial chain drift for monthly-chained superlative price indexes. This is because prices for supermarket items tend to frequently go on sale for short periods and then return to their pre-sales price.

Attempts to find a solution to chain drift have led to the development of a family of approaches internationally, adapted from the multilateral method of Gini (1931), Eltetö and Köves (1964) and Szulc (1964) (GEKS). The GEKS method takes the geometric mean of the ratios of all bilateral indexes (calculated using the same index number formula) between a number of entities.

The GEKS is known from spatial price comparisons and for spatial indexes these entities are generally countries, while for price comparisons across time, the entities are time periods (usually months). A problem with this method is that the results for all time periods will change when the observation period is extended and new data are added. By consequence, historical index numbers will be subject to revision, a situation that is unacceptable for a CPI.

There is however no need to publish the revised numbers. Since the time series is free of chain drift, we may use the change in the GEKS index between times T +1 and T, (which are both computed on the data of periods 0,...,T +1), as the chain link to update the time series. Ivancic, Diewert and Fox (2009) suggest a further refinement, the Rolling year GEKS (RGEKS) approach, to address this problem. The RGEKS method builds on the multilateral method of (GEKS) and uses a moving window to update the observation period as data for new periods become available and calculates the price change between the two most recent periods as described. This window limits the effect of the earlier data in the sample on the current index, while removing most of the chain drift (Ivancic, Diewert and Fox, 2009).

The RGEKS method is not without its limitations. The RGEKS makes optimal use of the matches in the data as it uses all matched items' prices and quantities to construct price indexes. The procedure combines bilateral superlative indexes[4]. However, because the method uses only items existing in both the periods relating to each bilateral index, it does not account for any price change associated with new or disappearing items (i.e. unmatched items).

International efforts to resolve this issue led to an extension of the RGEKS method being introduced by de Haan and Krsinich (2012) called the imputation Törnqvist RGEKS (ITRGEKS). The method

[3] Chain drift is defined by the ILO (2004, p.445) as when an index 'does not return to unity when prices in the current period return to their levels in the base period'.

[4] Typically, for accurate representation of consumer behaviour and substitution effect, bilateral indexes within a RYGEKS are computed using the Fisher or Törnqvist index formula.

utilises bilateral imputation Törnqvist indexes as inputs into the RGEKS process. The ITRGEKS method implicitly imputes price movements for the unmatched items by using prices predicted from regression models.

The ABS has been collecting transactions data from several providers since December 2011 and has undertaken a significant amount of work assessing the viability of applying the RGEKS theory in practice. Work is still evolving to develop an internationally agreed method that makes use of all available data on prices and expenditure and results in price indexes that are relatively free from chain drift. The ABS has not yet decided if any of these internationally developed methods will be implemented in the Australian CPI because there remain a number of methodological challenges that have yet to be sufficiently resolved.

A small number of NSOs have utilised transactions data to compile their CPI using methods that vary from replacement of field collected prices[5] to implementation of new index construction methods[6]. The benefits of using transactions data to compile the Australian CPI cannot be disregarded; therefore, the ABS has decided to introduce a method of directly sampling from transactions data to replace field collected prices. For the quarter beginning 1 January 2014, the ABS has replaced around 20,000 field collected prices with prices derived from transactions data for a subset of items within the total CPI sample where analysis has shown transactions data provides a reliable measure of product prices.

This paper discusses the ABS experience with using transactions data in the CPI. Section 2 outlines the various practical and technical issues encountered by the ABS. Section 3 presents the current ABS approach to incorporating transactions data into the compilation of the Australian CPI. Transactions data present other opportunities to improve official statistics; section 4 discusses other areas of official statistics that may benefit from transactions data. Some concluding remarks are provided in Section 5.

[5] Sammar, M, Norberg, A. and Tongur, C. (2013),' Issues on the use of scanner data in the CPI', Paper presented to the Thirteenth Meeting of the International Working Group on Price Indices, Copenhagen, Denmark

[6] van der Grient, H.A. and J. de Haan (2011), Scanner Data Price Indexes: The 'Dutch Method' versus RYGEKS, Paper presented at the Twelfth Meeting of the International Working Group on Price Indices, Wellington, New Zealand.

2. The Challenges of Moving from Theory to Practice

Using transactions data to compile the CPI has received considerable attention from prices statisticians internationally and there has been much collaboration on this subject. The focus has been on the development of internationally agreed methods. Typical of very large datasets, there are challenges associated with distilling information from these data for use in a statistical production process. The various practical and technical challenges faced when utilising very large volumes of detailed, high frequency data to construct price indexes are examined in this section.

2.1. Securing the Data

One of the key challenges is securing the ongoing and timely supply of transactions data. Some NSOs have gained access to transactions datasets through the purchase of those data from market research companies. In some countries, there is legislation in place that may provide NSOs the right to collect transactional information from businesses for statistical purposes. Transactions data often has high commercial value and there are usually significant costs to businesses in providing the data. This may lead to some businesses being resistant to providing these data and/or seeking funds to cover extraction costs. Even with the provision to collect data under legislation, NSOs might need to demonstrate the statistical purpose for accessing the data and be prepared to negotiate mutually agreed principles for the provision of data with the data providers.

2.2. Data Management and Classification

A. Data Management

Transactions data represents significantly greater amounts of information coming to NSOs as datasets assembled from transactions data are large. They are also often subject to data management conditions imposed by providers. Effective content management is an important factor when using big data files. Quality measures applied to metadata should go beyond reconciling data fields. It is critical that any enrichment or refinements to the data are applied consistently to maintain the accuracy and usability of the data for different areas of official statistics.

The continued receipt and management of these data will require technical storage facilities. The ABS sees opportunities to develop facilities that will be able to handle many sources of transactions data with a view to a "collect once, use many times" arrangement. However, this represents a significant upfront investment and ongoing requirement to store and maintain the integrity of these big data files.

B. Data Classification and Coding

In Australia, barcode information is not used for a number of reasons. Barcodes uniquely identify products but they are too detailed for statistical purposes. This is because small changes in product

specifications that do not affect consumers' utility may lead to changes in the barcodes. For example, if the source of supply changes from a factory in one country to a factory in another, the barcode will change. Some retailers source products with the same technical specifications from different manufacturers, which may lead to the same product having different barcodes.

The retailers recognise this, and for stock-control purposes, they identify their products not by their barcodes but with their own internal codes. In Australia, these codes are known as Stock Keeping Units or SKUs. SKUs group together like products that have different barcodes.

One key challenge is mapping of the SKUs to the appropriate category in the existing CPI commodity classification. SKUs do not align with statistical classifications very well. It is an exceptionally labour intensive process to map and maintain the relevance of the concordance due to the volume of SKUs and the fact that SKUs are regularly changed. For example, the transactions data provided to the ABS from major retialers contain over 750,000 SKUs, compared to 250 CPI elementary aggregate items (the lowest level of the CPI classification) for the same broad groups of products.

C. Data Editing and Validation

A general problem when dealing with big data is that data errors are difficult to identify. There is also limited scope for querying the data provider about anomalies at the detailed level. The ABS sees the need to have strategies to mitigate this risk through systematic editing approaches and quality assurance procedures. Stringent manual and automated checks should be performed on the data files received before processing of the data commences. For example, checks of file sizes and format changes, automated data validation checks and audits on the data files to detect any unusual activity are performed on each file received.

2.3. System / Capital Investment

The use of transactions data in a CPI is an attractive option for NSO's because of the perceived cost savings by reducing field collection activity. However, there is a need to consider the cost associated with managing and quality assuring large volumes of data. It is clear that undertaking this work requires significant effort and upfront investment in constructing storage facilities as well as modifying existing processes and systems. For the ABS, there are additional costs associated with developing a system to calculate indexes using transactions data and manipulating existing systems to accommodate the outputs.

Feedback from statistical agencies that have commenced using transactions data for their CPI has indicated that potential savings from reduced price collection operations are almost entirely offset by the increase in the cost of managing and quality assuring the large volumes of transactions data. The ABS is hopeful of the savings offsetting the capital investment but this is yet to be determined.

2.4. Methodological Issues

A. Combining Transactions Data with Field Prices

Research performed by the ABS in 2010 estimated that transactions data could potentially be used for components that make up 50% (by weight) of the CPI basket. However, even for those expenditure classes that transactions data are available, complete coverage is not always possible. Transactions data will not eliminate the need for field collection operations.

A significant issue is then how to combine the transactions data outputs with field collected prices. Transactions data contains information on both quantities and revenues for each item. This information allows for the calculation of superlative indexes. Field collected data do not contain information on quantities sold for the sampled items, thereby requiring the use of traditional index formulae to calculate price movements. If these data were combined at the price observation level, this is only feasible by dropping the quantity from transactions data and will enforce the use of traditional index formulae.

One possible approach is to apply the RGEKS method to only the transactions data and the resulting index is then combined at some specified aggregate level with the field data index, which is calculated using a traditional index formula. A key issue with this approach is how to determine weights for the respective indexes, as data on expenditure is only available for one component.

Another hurdle with combing these data sources is the mapping of transactions data items from different retailers to the appropriate category in the existing CPI commodity classification. Currently, coding of store codes directly to the CPI classification below the level in which price indexes are published to sufficient quality is not considered a viable option. The inability to compile lower level indexes leads to at least three issues. Firstly, stakeholders (notably National Accounts) rely on indexes at this detailed level, which will result in fewer detailed indexes available for deflation purposes. Secondly, the lack of lower level indexes will make it difficult to determine what is driving changes to the indexes. Thirdly, the inability to observe change at the lower levels will limit data validation and quality assurance.

B. Treatment of Field Collected Sample

Collecting ‘shelf’ prices by personal visit to businesses is a significant cost and reducing this cost is a key driver for moving to transactions data. In cases where transactions data can be shown to provide consistently reliable price information for the universe of some product groups, this will reduce or even eliminate the need for manual methods of price collection in these areas. A method to assess whether, given the price behaviours and expenditure shares of the two data sources, it would be appropriate to keep or drop the field collected data from the CPI sample will be essential in minimising field collection activities.

C. Accounting for Quality Change

An essential part of price measurement is accounting for quality change and the introduction of new items (ILO Manual, 2004). Transactions data has been demonstrated to exhibit a high level of 'churn' in the specific items available and sold from month to month. There are new models (and versions of models) of products becoming available in the market and old models dropping out of the market as they become obsolete.

The RGEKS makes optimal use of the matches in the data as it uses all matched items' prices and quantities to construct price indexes. However, because the method uses only items existing in both the periods to compile each bilateral index, it cannot account for any price change associated with new or disappearing items (i.e. unmatched items).

The lack of any established mechanism to account for quality change is generally seen as a weakness of the RGEKS method and the ABS has concerns that a failure to capture and account for quality change will be detrimental to the CPI. Only a few studies have examined how to account for quality changes within the RGEKS. De Haan and Krsinich (2012) provided examples of how quality-adjusted RGEKS indexes can be constructed using imputation Törnqvist indexes[7]. The authors concluded that the Imputation Törnqvist RGEKS is a good benchmark index due to its ability to reflect price movements for unmatched items, however it can only be applied to data where a large range of characteristics are available.

Transactions data obtained from Australian retailers generally only have a few or no characteristics of products at the item level available. At most, size, volume or weight and free-text descriptors of the item may be recorded in the data. We have found the methods developed by de Haan and Krsinich (2012) could not be practically applied to the Australian dataset.

D. Analysis of RGEKS Movements

A main disadvantage of RGEKS price indexes is in its complexity. The RGEKS calculation is not restricted to comparing prices from two consecutive time periods. It draws upon prices and expenditures of all the most recent thirteen months. This makes monthly or quarterly index changes difficult to analyse and explain.

Van der Grient (2010) proposed a method to analyse changes in RGEKS index movements. The author suggests that monthly change in the RGEKS index can be analysed by looking at the individual item price changes in the current month and the previous month. The monthly movement changes of the RGEKS indexes and the chained Törnqvist indexes are similar due to the shares of the last two months dominating the result of the movement change. As this method is based on the premise of a

[7] Imputed price indexes adjust for quality changes by imputing the unobservable or 'missing price' of new and disappearing items. De Haan and Krsinich (2012) tested three methods. The first method makes explicit imputations using a hedonic regression model, which is estimated for each time period. The other two methods make implicit imputations; they are based on time dummy hedonic and time-product dummy regression models and are estimated.

fixed set of items, when the set of items between two consecutive months changes in an extreme way the results are less valid. If item prices change significantly (by 50% or more), this can have an impact on the monthly aggregated movement price change.

2.5. User Community

The CPI is used for a variety of purposes, such as in the development and analysis of economic policy, the adjustment of wages and pensions and in individual contracts. Because of this, there is considerable interest from users in the changes to the methods used to compile the CPI. Major developments in the way the CPI is compiled need to be part of a broad, well communicated plan. NSOs should allow sufficient lead time to prepare, test and review new methods and to notify the user community of methodological changes.

The ABS engages in a considerable amount of communication with users and stakeholders and is aware that sophisticated methods used to compile the CPI using transactions data will be difficult to explain. It is important that the use of transactions data in the CPI is well understood and supported by the Australian user community. Generally, the results of the ABS's investigations are published as experimental price indexes. This provides a starting point for consultation with users to: (i) obtain a sense of the value these statistics offer to users; and (ii) obtain feedback on the methodology used to construct the indexes.

2.6. NSOs Cautious Culture

The CPI is an important economic indicator, which is not normally revised. For this reason, new index construction methods are extensively reviewed, tested and some NSOs will only apply methods that are widely endorsed by the statistical community before they are fully implemented. For example, the suitability of the RGEKS approach has received interest and some support internationally, but uncertainties around the resolution of methodological challenges, including quality change associated with disappearing products and their replacements have deterred the implementation of this approach. There has not been consensus between statistical agencies as to what is the best way to use transactions data and no single method has received international endorsement. Most NSOs, including the ABS, who has access to transactions data, are integrating these data in official statistics cautiously.

Statistics Norway has been using transactions data since August 2005 to compute its index for food and non-alcoholic beverages. Statistics Netherlands introduced supermarket transactions data into its CPI in June 2002. In both Norway and The Netherlands, both prices and expenditure weights for a large sample of grocery items are derived from transactions data. The Swiss Federal Statistical Office also uses transactions data in its CPI. Recently, Statistics Sweden (Sammar, Norberg and Tongur, 2013) and Statistics Denmark (Gustafsson, 2013) have adapted methods of directly sampling from transactions data. Statistics New Zealand also uses transactions data, but only to inform expenditure share weights in the CPI.

3. The ABS Approach

3.1. An overview of the ABS Approach

Transactions data obtained from major retailers were incorporated in the Australian CPI for the quarter commencing 1 January 2014. The ABS uses methods of directly sampling from transactions data to replace field collected prices. Around 20,000 field collected prices have been replaced with prices derived from transactions data for a subset of products within the total CPI sample where analysis has shown transactions data provides a reliable measure of product prices.

The price for an individual product is calculated from the transactions data by dividing a product's revenue by the quantity sold. This price is referred to as a product unit value and represents the price experienced by consumers over a period of weeks or months. A product's unit value is, conceptually speaking, the purest price and is more representative of prices paid by consumers over the reference period than point-in-time pricing.

Larger price samples for individual products are available from transactions data. A unit value price for individual products is collected from an increased number of business outlets in each capital city. Product unit values obtained from transactions data was used to compile a range of CPI Expenditure Class indexes. These Expenditure Classes are listed in Attachment 1.

3.2. Selecting and Maintaining a Representative Sample from Transactions Data

The ABS found that metadata from the transactional datasets varied in quality between data providers. The poor quality of some metadata files (in particular, the product listings) proved problematic for identifying exact product matches for those products already in the CPI sample. For example, duplication of item descriptions in different SKUs and 'dump' SKUs where the descriptions are not clear (e.g. 'mince', 'T bone', 'BBQ sausage'). Overall, the ABS found exact matches for around 80% of CPI sampled products. Where exact matches could not be found a 'similar' high revenue item was selected.

The composition of the CPI basket reflects the consumption preferences of households. Transactions data provide real-time sales information on all products sold at businesses, which has not previously been available to the ABS. Access to this information allows the ABS to improve the relevance of the products priced in the CPI. Each product selected in the CPI sample is evaluated each quarter, undergoing stringent revenue and consistency checks. For those products that fail these tests, a suitable replacement product is selected from a pool of similar products.

3.3. Calculating Unit Values

One advantage of using transactions data is that it allows more flexible pricing options compared to field observations. Ideally, we would use the maximum amount of transactions data to derive unit value prices, however a practical constraint is allowing for sufficient processing time.

For products that were previously priced quarterly, five options were considered:

1. use unit values based on the middle month of the quarter (this timing resembles the field collection method);
2. use unit values based on the first two months;
3. use unit values based on the first two months plus the first two weeks of the third month;
4. use unit values based on all three months; and
5. use unit values based on a moving three month average (taking an average price from the last month of the previous quarter, and the first two months of the current quarter).

To assess the differences between these methods the ABS looked at the absolute deviation of unit value prices across the five approaches. The deviations for the majority of products were very small and any significant variations could be explained by the timing of product discounting. Option 3 was selected as this enabled the maximum use of the current period data without compromising compilation timetables.

3.4. Quality Change and New Products

Where price change is measured using small samples of products it is possible for field collectors to examine each product in the sample and identify any changes in quality. This direct form of marketplace intelligence is not possible when transactions data is used to compile the CPI. There are broadly three main scenarios, which initiate the need to quality adjust prices obtained from transactions data:

- where new items are brought into the price samples (as replacements);
- where there has been a quantity change (eg. change in packet size) and the SKU has changed and
- where there has been a quantity change and the SKU has not changed.

The first scenario is the simplest case and requires calculating a previous period price for the new item. In the second and third scenario, a quality adjustment factor is calculated to account for the quantity change. The ABS has developed a method to link new and disappearing products. For example, if a product changes in packet size - the SKUs are likely to change. The linking process uses information on the product description, price, revenues, timing (when products appear or disappear on sales listings) and quantity sold. This process identifies that the new product is the same as (or very close to) the disappearing product (but with a different SKU) and adjusts accordingly for the change in quantity.

3.5. Treatment of Seasonal Products

A seasonal product is one with seasonal fluctuations in the quantity purchased throughout the year. These products generally also have seasonal fluctuations in their price during the year. These seasonal variations in quantity and price can be due to changes in demand and/or availability of the product. Such a product may be unavailable for parts of the year. For seasonal products, the ABS decided to select the highest revenue (or best-selling) variety of the product each period. This approach selects the most 'representative' product for that period, which to avoid selecting clearance or unsuitable products, is subject to a minimum monthly revenue threshold.

3.6. Reference Period Price

When field collected prices are replaced by prices derived from transactions data, the shift in the pricing basis has an impact on the price levels of the sample. The price movement in the March quarter 2014 for CPI sampled products for which transactions data are being used was calculated using a reference period price that has also been derived from the transactions data. The resulting price movements are unaffected by the move from point in time pricing (field collected prices) to product unit values.

3.7. Aggregation Process

To maintain the current index structures and published outputs, the procedures for calculating indexes at all levels of aggregation remains unchanged. Lower level price indexes (at the elementary aggregate level) which contain a combination of field collected prices and average unit prices derived from transactions data continue to be calculated using the Jevons formula (i.e. an equal weighted geometric mean index formula).

4. Other Opportunities

The use of transactions data also allows the ABS to examine the frequency and timeliness of the CPI in the future. Key users of the CPI are regularly requesting that the CPI be produced monthly and include regions outside capital cities. The availability of transactions data at high frequency for all outlets in states and territories will allow the ABS to investigate these objectives.

Transactions data may also enable the ABS to compare price movement differences between rural and urban locations, which may provide important information on whether the CPI needs to cover a wider geographic area than the capital cities.

ABS has commenced work to determine wider applications of transactions data in official statistics. The revenue, quantity and price information at item level is potentially useful for economic statistical purposes including: (i) improving the measurement and product detail for retail surveys and elements of household final consumption expenditure in the national accounts; (ii) improving the retail margin element of producer price indexes; and (iii) editing datasets such as the Household Expenditure Survey.

5. Conclusion

The potential of big data such as transactions data as an input into official statistics cannot be ignored by NSOs. The ABS sees these data to have the potential to replace traditional data sources and have a profound effect on the way official statistics are produced. The ABS is continuing to investigate possible ways to increase access to, and use of, Big data sources.

The first steps are the use of transactions data in the compilation of the CPI. To exploit the richness of these data involves much more than developing an internationally agreed methodology. Although the focus on methods is important, that is just the beginning of the challenges to use transactions data in official statistics, as distilling information from transactions data for use in a statistical production process is not easy. This is shown in this paper, which examined various issues encountered by the ABS, with some issues yet to be sufficiently resolved.

The ABS will continue to engage with the international community to contribute to the development of an internationally endorsed methodology. One of our aims is to produce experimental RGEKS indexes to enable an assessment of the method and to gauge the Australian users' community acceptance of this new index construction method.

6. References

Eltetö, O., and Köves, P. 1964, 'On a problem of index number computation relating to international comparison', *Statisztikai Szemle*, 42, pp. 507-18.

Gini C. 1931, 'On the circular test of index numbers', *International Review of Statistics*, Vol. 9.

van der Grient, H.A. 2010, 'How to analyse changes in price indexes when applying the RYGEKS formula', Statistics Netherlands, The Hague.

van der Grient, H.A. and de Haan, J. 2010, 'The Use of Supermarket Scanner Data in the Dutch CPI', paper presented at the Joint ECE/ILO Workshop on Scanner Data,10 May 2010, Geneva.

van der Grient, H.A., and de Haan, J. 2011, 'Scanner Data Price Indexes: The 'Dutch Method' versus RGEKS', paper presented at the Twelfth Meeting of the International Working Group on Price Indices (Ottawa Group), 4-6 May 2011, Wellington, New Zealand.

Gustafsson, N. 2013, 'Drawing a Sample from Scanner Data to Use in the Danish CPI', paper presented at the 13th meeting of the Ottawa Group, Copenhagen, Denmark.

de Haan, J., and van der Grient, H.A. 2009, 'Eliminating Chain Drift in Price Indexes Based on Scanner Data', paper presented at the eleventh meeting of the Ottawa Group, 27–29 May Neuchâtel, Switzerland.

de Haan, J., and Krsinich, F. 2012, 'Scanner data and the treatment of quality change in rolling year GEKS prices indexes', paper presented at the 2012 Economic Measurement Group, Sydney, Australia.

Hill, T.P. 1993, 'Price and Volume Measures', in System of National Accounts 1993 (Brussels/Luxembourg, New York, Paris, New York, and Washington, D.C.: Commission of the European Communities, IMF, OECD, World Bank and United Nations), pp. 379–406.

ILO 2004, *Consumer Price Manual: Theory and Practice*, ILO, IMF, OECD, Eurostat, United Nations, World Bank: Geneva.

Ivancic, L., Diewert W.E., and Fox K.J. 2009, 'Scanner Data, Time Aggregation and the Construction of Price Indexes', Discussion paper 09-09, Department of Economics, University of British Columbia, Vancouver, Canada.

Krsinich, F. 2011, 'Price indexes from scanner data: A comparison of different methods', paper presented at the 12th meeting of the Ottawa Group, Wellington, New Zealand.

Sammar, M., Norberg, A., and Tongur C. 2013, 'Issues on the use of scanner data in the CPI', paper presented to the Thirteenth Meeting of the International Working Group on Price Indices, Copenhagen, Denmark.

Szulc B.J. 1964, 'Indices for multiregional comparisons', (in Polish) *Przeglad Statystyezny 3*, pp. 239-54.

Attachment 1: Expenditure Classes containing Transactions Data

Tobacco
Garments for women
Beef and veal
Bread
Breakfast cereals
Cakes and biscuits
Cheese
Coffee, tea and cocoa
Eggs
Fish and other seafood
Food additives and condiments
Fruit
Ice cream and other dairy products
Jams, honey and spreads
Lamb and goat
Milk
Oils and fats
Other cereal products
Other food products n.e.c.
Other meats
Pork
Poultry
Snacks and confectionery
Take away and fast foods
Vegetables
Waters, soft drinks and juices
Cleaning and maintenance products
Glassware, tableware and household utensils
Other non-durable household products
Personal care products
Tools and equipment for house and garden
Pharmaceutical products
Newspapers, magazines and stationery
Pets and related products
Spare parts and accessories for motor vehicles