Submission to the 16th Series Review of the Consumer Price Index

February 2010

Dr. Lorraine Ivancic[1]
Center for Applied Economic Research
University of New South Wales

[1] The views expressed in this submission are those of the author and do not necessarily reflect the views of the Centre for Applied Economic Research or the University of New South Wales.

Submission to the 16th Series Review of the Consumer Price Index

One of the key issues identified by the ABS for consideration in the Information Paper (Issues to be considered during the 16th Series Australian Consumer Price Index Review) is the frequency of updating the weights in the CPI. The document asks whether the ABS should consider the use of alternative sources of data, such as scanner data, to enable more frequent updating of expenditure share weights. However, the updating of expenditure share weights is only one area in which scanner data could potentially be used to improve the CPI. A number of statistical agencies including Statistics Netherlands, Switzerland, Norway and New Zealand are already using scanner data as an input into their CPI, showing that the use of scanner data is feasible in practice. The primary purpose of this submission is to outline some key areas in which scanner data could potentially be used to improve the quality of the Consumer Price Index. This submission will also consider some of the practical issues associated with the incorporation of scanner data into the CPI.

Using scanner data in the Consumer Price Index

1. Introduction

High quality data on consumer purchases is fundamental to the calculation of an accurate Consumer Price Index. A relatively new source of highly detailed data on consumer purchases, known a scanner data (or high frequency data), has become increasingly available in recent years. Scanner data is the product of what is known as bar-code technology. Perhaps the most familiar form and use of bar code technology to most people is the 'scanning of bar-codes at checkout lines of retail stores' (Feentra and Shapiro, 2003). Currently three statistical agencies, Statistics Netherlands (CBS), Statistics Norway (Statistisk Sentralbyra) and Statistics Switzerland (SCB), use supermarket scanner data in calculating their official Consumer Price Index (CPI) figures. Statistics New Zealand also uses scanner data, in a more limited capacity, to inform expenditure share weights in the CPI.

The most important benefit of scanner data, relative to traditional data sources, is that it provides a detailed and potentially comprehensive source of information on both the prices paid *and* the quantities purchased by consumers. The detailed nature of this type of data has the potential to generate significant benefits to statistical agencies in terms of producing a more accurate CPI. Accuracy here refers to a number of aspects of the CPI, including capturing of the actual prices paid by consumers and

potentially more accurate expenditure share weights, improving sample representativity and the potential to reduce new goods bias. Rodriguez and Haralsan (2009) from Statistics Switzerland state that the ability to produce better indexes 'increases dramatically' with the use scanner data. However, utilising this type of data is also likely to generate some costs, at least in the short term, for statistical agencies. In the longer term there is the clear potential for statistical agencies to derive cost savings from the use of scanner data. Some of the key benefits and costs of incorporating scanner data into the CPI are outlined below.

2. Potential Benefits Associated with the Use of Scanner Data

The most important benefit of scanner data sets are that they contain information on both prices paid by consumers and quantities purchased in each period. Importantly, the prices collected by bar-code technology are different to those currently collected for use in the calculation of the CPI by most statistical agencies. Scanner data sets contain the actual price paid by consumers at each point in time. This means that both 'regular' and sale prices (and the associated quantities purchased) are recorded. This is in contrast to current price collection methods by most statistical agencies, where price collectors record the shelf price of a sampled good at a particular point in time. Traditional point-in-time price collection means that the price collected may not be representative of the average price paid by consumers for the good in that particular period, i.e. consumers may not have actually bought the item at that particular price. Therefore, scanner data is able to provide accurate information on the prices that consumers actually paid for an item. As a result, scanner data brings us closer to the objective of the CPI which is to measure the change in the price paid by consumers in different periods.

In Australia the Household Expenditure Survey which is used to update expenditure weights in the CPI is undertaken every six years.[2] Hawkes and Piotrowski (2003) note that the use of 'out-of-date' expenditure weights is 'likely to result in the overstatement of inflation' (p.21). In contrast scanner data provides information on the quantities sold of each item in the current period. This information can be used to continuously update expenditure share weights used in the CPI, right down to the lowest level of aggregation in the CPI, i.e. the elementary aggregate level (Silver and Webb, 2002; Reinsdorf, 1999). At the elementary aggregate level this is an important improvement over existing methods which currently (implicitly) weight all items equally.

[2] Prior to 2003/04 the Household Expenditure Survey had been conducted in 1974-75, 1975-76, 1984, 1988-89, 1993-94, 1998-99 and 2003-04 (ABS, 2009).

Scanner data sets also contain information on a much broader range of items than currently collected by statistical agencies. They can provide statistical agencies with a census of items sold which would remove sampling error associated with traditional sampling methods. Even if scanner data sets are not used to provide a census of all items, they have the potential to increase the number of items sampled in many item categories. For example, Statistics Netherlands (Bruijn, Mulligen and Haan, 2006) illustrated the potentially large increase in sample size from the use of scanner data for 3 item categories:

Table 1. Comparison of items found in field survey and scanner data*

COICOP	Description	Number of items in scanner data	Number of items in field survey
011140	Bakery products	596	21
011430	Cheese	315	14
012120	Tea	105	1

*Table reproduced from Bruijn, Mulligen and Haan (2006).

As scanner data has the potential to provide a census of information on items it can be used to choose a more representative basket of items or as a check on the representativeness of existing samples (Gudnason and Snorrason, 1999).

A further benefit of the use of scanner data are that they may reduce what is known as new goods bias. New goods bias refers to the 'failure to bring new products into the CPI sample with sufficient speed' (ILO, 2004, p. 213). This failure 'can lead to upward bias if those new products later experience large price reductions that are not reflected in the index' (ILO, 2004, p.213). With scanner data information on new items is collected as soon as a new good is sold in an outlet which uses scanner data technology. As a result, new items can be included in the CPI's basket of goods very quickly (Bradley et al, 1997).

Scanner data can also be used to 'monitor expenditure shifts across outlets' (Bradley, 1996, p.2). This has the potential to inform price statisticians about which 'types' of outlets are important in terms of consumer expenditure and about the changing importance of certain types of stores over time. For example, Hausman and Leibtag (2004) found that in the last decade there has been a significant shift in the type of stores that consumers now shop in, with consumers in the US moving from traditional stores to what Little (cited in Hausman and Liebtag, 2004) described as 'low-price one stop shopping

destinations' and 'convenience stores'. Scanner data can be used to identify and account for these types of shifts more quickly than traditional methods.

Lastly, if scanner data sets are found to be able to provide consistently reliable price information for the universe of some item category/ies then this may reduce or even eliminate the need for manual methods of price collection in these areas, leading to potential cost savings for statistical agencies. Müller (2009) notes that the use of scanner data also reduces the time burden on retail chain staff. However, manual methods of price collection and checks on sample representativeness may still be required where sample coverage of items is incomplete (Abraham, 2003; Bradley, 1996).

3. Potential Costs Associated with the Use of Scanner Data

There are potentially a large number of benefits generated through the use of scanner data. However, the use of scanner data is likely to lead to additional costs for statistical agencies. Perhaps the most important concerns for statistical agencies that have been identified in the literature are the potential costs associated with the purchase of the data, costs associated with the development and implementation of new systems and the timely delivery of data to statistical agencies.

Scanner data sets have typically been made available to statistical agencies and other users by market research companies. These companies usually collect, clean the data sets and put them in an easy-to-use form. An issue for statistical agencies is that the purchase of scanner data from such companies may be quite costly. Statistics Netherlands has been able to mitigate the costs of purchasing scanner data by forming agreements directly with the supermarket chain to provide the scanner data at no cost. Although Statistics Netherlands incurs no cost to purchase the data it has had to take on the role of cleaning the data.

Perhaps the biggest challenge for statistical agencies is the development and implementation of new software and/or systems to incorporate scanner data into the CPI. Two important issues are the development of systems to:

1) match item codes from scanner data sets to the appropriate category in the existing CPI classification system so that items can be allocated automatically to the correct CPI sub-category; and
2) deal with missing items and determine how replacement items are to be chosen and incorporated into the CPI.

In terms of matching item codes, Hawkes and Piotrowski (2003) provide an illustration of some of the differences in item categorization by the US Bureau of Labour Statistics and the scanner data provider ACNielsen for the "food at home" item category; see table 1.

Table 1. How ACNielsen and BLS view the World of Food-at-Home**

	ACNielsen	**BLS**
Division of food-at-home items	6 departments (dry grocery, frozen foods, dairy, deli, packaged meats, perishables); 64 categories (excludes fresh meat and produce); 603 modules (excludes fresh meat and produce)	6 commodity groups (cereals & bakery products, meat, poultry, eggs & fish, dairy, fruits & vegetables, nonalcoholic beverages other); 18 expenditure classes; 53 item strata (of which 34 exclude fresh meat and produce); 65 entry-level items (of which 44 exclude fresh meat and produce)
Division of "ice-cream and related products"	4 modules in 2 categories under "frozen foods"	1 entry-level item in 1 item stratum under "dairy"
Classification of ice cream	frozen food	dairy product

**Table reproduced from Hawkes and Piotrowski (2003, p.23)

It is clear that incorporating scanner data into the CPI will require additional time and resource costs for the ABS but the majority of these costs are likely to be incurred in the initial set-up stages.

An additional issue that has been raised is whether the data can be 'delivered, verified and processed' in time for the release of the CPI (Bradley et al 1997, p.7). Feenstra and Shapiro (2003) note that 'statistical agencies need to be assured of long-term, consistent and timely access to the information need(ed) to construct price indexes' (p.6). Agreement will need to be reached between statistical agencies and the data provider to ensure the timely delivery of data (Schut, 2001). Statistics Netherlands have, in some part, minimized the potential risk of a data hold up by negotiating directly with the supermarket chains rather than going through a market research company. The experiences of statistics Netherland, Norway and Switzerland have shown that, even with the production of a *monthly* price series, scanner data can be incorporated into the CPI in a timely manner. These statistical agencies have also demonstrated that scanner data are a viable long term source of data for statistical agencies, with Statistics Netherlands now successfully using scanner data in their CPI since 2002 and Statistics Norway since 2005.

3. Conclusion

Good data on consumer purchases are essential for an accurate CPI. Scanner data offers statistical agencies the opportunity to obtain highly detailed and accurate data on consumer purchases. For the statistical agencies already using scanner data it seems that the experience has been a positive one, with all agencies committed to expanding their collection of this type of data for use in the CPI. Both Statistics Netherlands and Switzerland have indicated that they will increase their collection of scanner data in 2010 to include data from additional retail chains (Müller, 2009; Statistics Netherlands (Personal communication), 2009). Statistics Norway (Rodriguez and Haraldsen, 2005, p. 233) also notes that 'The work of expanding the use of scanner data from industries that today report prices via paper forms will continue'. This 16th series review provides the ABS with the opportunity to join these agencies in their use of non-traditional data sources to improve price indices.

References

Abraham, K.G (2003), 'Toward a Cost-of-Living Index: Progress and Prospects', *Journal of Economic Perspectives*, 17:1, 45 – 58.

Australian Bureau of Statistics (2009), Household Expenditure Survey, http://www.abs.gov.au/Ausstats/ABS@.nsf/12ce1aabe68b47f3ca256982001cc5da/5f1422f1af472d80ca256bd00026aee6!OpenDocument (accessed January 8, 2009).

Bradley, R. (1996), 'Potential Benefits from the Use of Scanner Data in the Construction of the CPI', American Statistical Association Proceedings, 1997.

Bradley, R., B. Cook, S.G. Leaver, and B.R. Moulton (1997), "An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI", paper presented at the International Conference on Price Indices, Voorburg, 16–18 April.

Bruin de, M., Peter Hein van Mulligen van, P.H., and J. de Haan (2006), "A Matched-Model Geometric Mean Price Index for Supermarket Scanner Data", Presented at the 6th Economic Measurement Group Workshop, 13 – 15 December 2006, UNSW, Sydney.

Feenstra, R.C. and M.D. Shapiro (2003), 'High Frequency Substitution and the Measurement of Price Indexes', pp. 123-146 in R.C. Feenstra and M. Shapiro (eds.), *Scanner Data and Price Indexes*, Chicago: University of Chicago Press.

Hawkes, W.J. and F.W. Piotrowski (2003), 'Using Scanner Data to Improve the Quality of Measurement in the Consumer Price Index', in R.C. Feenstra and M. Shapiro (eds.), *Scanner Data and Price Indexes*, Chicago: University of Chicago Press.

Hausman, J. and E. Leibtag (2004), 'CPI Bias from Supercentres: Does the BLS Know that Wal-Mart Exists?', NBER Working Paper 10712.

Gudnason, R. and H. Snorrason (1999), 'Use of Cash Register Data', paper presented at the fifth meeting of the International Working group on Price Indices, Rejkjavik, 25-27 August.

ILO (2004), *Consumer Price Manual: Theory and Practice*, Geneva: International Labour Organization.

Müller, R. (2009), 'Price Collection with Scanner Data for the Swiss CPI/HICP', Presented at the 11th Ottawa Group Meeting, 27th – 29th May, Neuchâtel, Switzerland.

Reinsdorf, M. (1999), 'Using Scanner Data to Construct CPI Basic Component Indexes', *Journal of Business and Economic Statistics*, 17:2, 152–160.

Richardson, D.H. (2000), 'Scanner Indexes for the CPI', paper presented at the sixth meeting of the International Working Group on Price Indices, Canberra, 2 – 6 April.

Rodriguez, J. and F. Haraldsen (2006), 'The Use of Scanner Data in the Norwegian CPI: The «New» Index for Food and Non-Alcoholic Beverages', *Economic Survey,* 4, 21 – 28.

Schut, C. (2001), 'Using Scanner Data to Compile Price Indices: Practical Problems', paper presented at the Joint Statistical Commission and Economic Commission for Europe/ International Labour Organisation (ECE/ILO) Meeting on Consumer Price Indices, Geneva, 1 – 2 November.

Silver, M. and B. Webb (2002), 'The Measurement of Inflation: Aggregation at the Basic Level', *Journal of Economic and Social Measurement*, 28, 21-35.