Using transactions data to enhance the Australian CPI

Alexander Howard, Kristan Dunford, John Jones and Marcel van Kints[1]

Prices Branch

Australian Bureau of Statistics

Kristy Naylor[1] and Rory Tarnow-Mordi

Methodology Division

Australian Bureau of Statistics

The authors would like to thank Professor Jan de Haan for his input and advice during the preparation of this paper; and Frances Krsinich for providing the programing code to calculate the FEWS and RYGEKS indexes.

Abstract

The availability of transactions data to compile official statistics provides the opportunity to re-examine the traditional approach to compiling the Consumer Price Index (CPI), and in doing so, creates new avenues to enhance aspects of the Australian CPI. There is also the potential to achieve these enhancements at a lower cost.

This paper details the work undertaken by the Australian Bureau of Statistics (ABS) to integrate transactions data into the Australian CPI. There is a focus on the methodological approach as well as the actions undertaken to reduce manual interventions. Current research by the ABS is focused on maximising the use of transactions data to compile the CPI. This involves using transactions data to expand pricing samples, for weighting at the lowest levels of the CPI and developing new methods to compile the CPI. This paper also promotes a phased implementation of transactions data into the CPI.

[1] The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Australian Bureau of Statistics (ABS).

Introduction

1. The Australian Consumer Price Index (CPI) is a quarterly measure of household inflation. The Australian CPI utilises the acquisitions approach[2], is compiled quarterly and the expenditure weights, sourced primarily from the Australian Bureau of Statistics (ABS) Household Expenditure Survey (HES), are updated every 6 years. The methods, data sources and compilation frequency of the CPI have served Australia well over many decades.

2. While the CPI is a fit-for-purpose temporal price index for *consumption* goods and services acquired by Australian resident households, there are always aspects of the CPI that can be enhanced.

3. Consultation with users of the CPI, particularly during the 16th Series CPI review[3], has highlighted areas of the CPI that could be enhanced. These areas can be grouped into: (1) Compilation frequency; (2) Frequency of expenditure weight updates; and (3) Evolutionary enhancements to various components of the CPI.

4. The ABS has, from time to time, suggested approaches to enhance aspects of the CPI. These approaches have typically been traditional in nature, which would require ABS field price collectors obtaining more prices more frequently. These traditional approaches are generally expensive and require new funding.

5. The availability of transactions[4] data to compile official statistics provides the opportunity to re-examine the traditional approach to compiling the CPI, and in doing so, creates new avenues to enhance aspects of the Australian CPI, most likely at lower cost.

6. The ABS Prices Branch has begun the *Enhancing the Australian CPI* project. One major component of the project will examine how transactions data can enhance aspects of the Australian CPI.

7. The potential benefits of transactions data to National Statistics Offices (NSOs) generally, and for price statistics in particular, are well documented. The data are high frequency, there is (usually) a full enumeration (as opposed to a sample) of product data from all locations within a retail business, and product revenues and quantities can be used to calculate unit values for weeks, months or quarters rather than using point in time pricing. The presence of timely revenue data also introduces the

---

[2] The acquisitions approach is defined as 'The basket of goods and services priced can be defined as consisting of all consumer goods and services actually acquired by households during the base period.'
[3] ABS Cat.no. 6469.0 - Outcome of the 16th Series Australian Consumer Price Index Review, Dec 2010
[4] Transactions data are a description of the interactions of institutional units buying and selling products on terms mutually agreed by the buyer and seller.

possibility of calculating superlative indexes.

8. The remainder of this paper focuses on the methodological approach undertaken to date to introduce transactions data into the Australian CPI; as well as the actions undertaken to automate the CPI sample maintenance. Further, there is a description of current ABS research to maximise the use of transactions data to compile the CPI. This involves using transactions data to expand pricing samples, for weighting at the lowest levels of the CPI and developing new methods to compile the CPI. This paper proposes a phased implementation of transactions data into the CPI.

The traditional ABS approach to compiling the Australian CPI (prior to March quarter 2014)

9. Traditionally, most prices used to compile the CPI were collected by personal visits to selected businesses. These personal visits were made by ABS field officers who observe prices as well as discuss discounts, special offers and volume-selling items with the sampled businesses. The field officers record this information during the visit in handheld computers. The regular personal visits to businesses enable the ABS field officers to actively monitor market developments and observe product quality change.

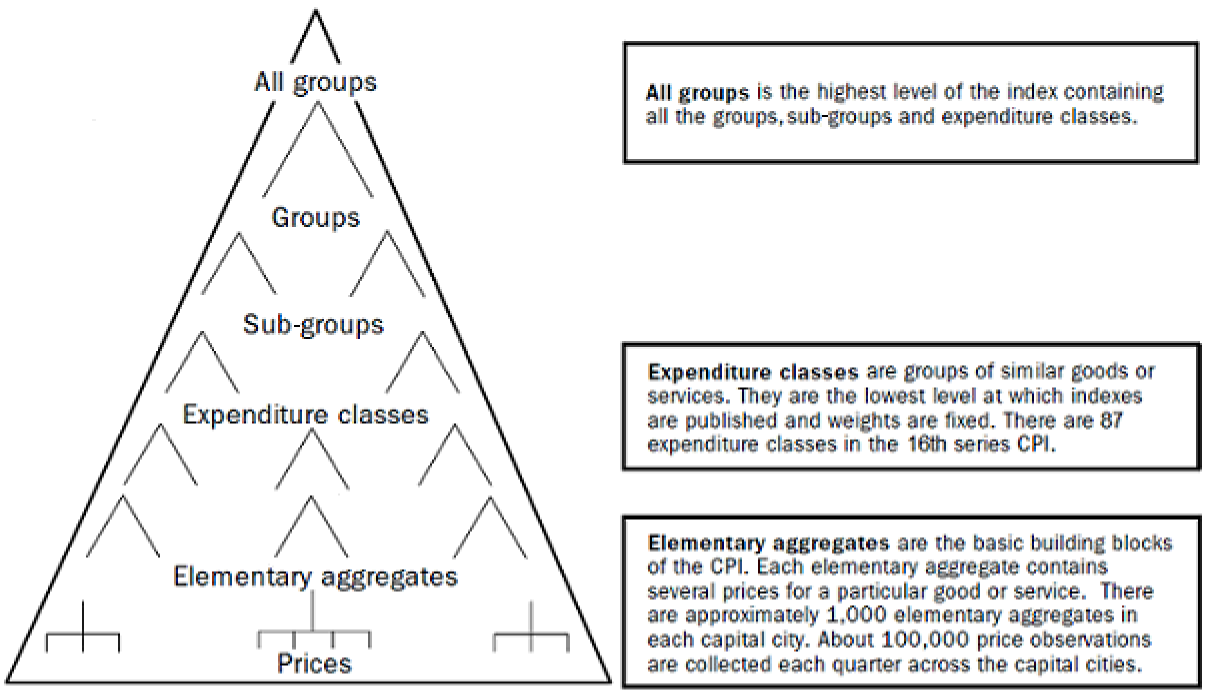
10. ABS field officers aim to measure price changes for identical or equivalent items in successive periods. However, products change; their components or ingredients may change which may result in a change in quality. As the characteristics of products are altered, the ABS field officers collect descriptive information that enables the effects of a quality change to be separated from the price change, so that the CPI measures pure price change.

11. Prior to the March quarter 2014, prices for a small number of products were obtained from transactions data. These prices were used to compile the CPI. Automotive fuel prices, for example, were obtained from electronic funds transfer transactions in each capital city. Prices from businesses across all areas of each capital city were obtained each day, including weekends and public holidays. Prices were recorded for a range of automotive fuel types.

12. The ABS used non-probability sampling to compile the CPI. This sampling approach selects representative sets of products for regular pricing from a selection of businesses. The selected products (and prices) and sampled outlets are chosen to be representative of the CPI population.

13. The traditional index structure of the CPI is presented in Figure 1.

Figure 1: CONSUMER PRICE INDEX STRUCTURE



14. Expenditure estimates, used for weighting the CPI, are obtained from the ABS Household Expenditure Survey (HES) for total annual expenditure of private households in each capital city for each of the 87 expenditure classes in the CPI. Of note, some expenditure data required by the CPI which cannot be obtained from households via the HES (e.g. the New Dwelling Purchase by Owner Occupiers EC based on Building Activity Survey and population data), or adjustments are required to the HES data (e.g. alcohol and tobacco data) to ensure an accurate measure of total expenditure.

15. While these expenditure class aggregates are derived for well-defined categories of household expenditure (e.g. bread), they are still too broad to be of direct use in selecting price samples. For this purpose, expenditure aggregates need to be subdivided into as fine a level of commodity detail as possible.

16. The more detailed level of commodities are referred to as ‘elementary aggregates’. Once price movements are calculated for each elementary aggregate, they can be used to derive the index numbers at any level of the CPI.

17. In the Australian CPI, the change in average price for the Elementary Aggregate is predominately calculated using the Jevons formula (geometric mean). The Jevons formula calculates price change by taking the geometric mean of price relatives. The Jevons index can be written as follows:

$$I^{0:t} = \left(\prod_{i=1}^{n} \frac{p_i^t}{p_i^0}\right)^{\frac{1}{n}} \times 100 \qquad (1)$$

18. A Lowe index is used above the EC level. The Lowe index using the quantities from period b (the most recent HES) can be written as follows:

$$I_{CPI}^{0:t} = \sum_{i=1}^{n} I_i^{0:t} s_i^0$$
$$where\ s_i^0 = \frac{p_i^0 q_i^b}{\sum_{i=1}^{n} p_i^0 q_i^b} \qquad (2)$$

Part 1: Work undertaken to date to incorporate transactions data in the Australian CPI

Background

19. The ABS has, in recent years, been successful in negotiating the supply of transactions data directly from retail businesses. The aim for the ABS is to be able to use these data to support the compilation of official statistics. As with all "big data", electronic point of sale data collected directly from retailers has significant potential for improving official statistics, but realising these benefits requires careful implementation.

20. After securing transactions data from retail businesses, early work by the ABS centred on an assessment of the transactions data against the ABS quality framework to determine their robustness and fitness for purpose. Liaison with retail businesses has resulted in the receipt of transactions data that has well-structured metadata; and data that are of sound quality. This has resulted in transactions data being used by the ABS to price products that represent approximately 25% of the weight of the Australian CPI (as at the December quarter 2014)[5].

21. In addition to ensuring the transactions data are of high quality, the ABS commenced a research program examining new methods to compile the CPI using transactions data. This program commenced because it was clear that current methods and processes used to compile the CPI could not be applied to transactions data. It is apparent that

[5] See Attachment 1 for product groups where some prices are derived from transactions data

new methods and processes will need to be developed to maximise the use of transactions data.

22. The ABS began compiling a number of experimental indexes using various methods, with the aim of understanding the weaknesses and strengths of each approach. These methods ranged from sophisticated superlative indexes with a Rolling Year GEKS (RYGEKS)[6] adjustment to straightforward replacement of point-in-time prices with unit values calculated over a week, month or quarter.

The Current use of transactions data by the ABS

Phase 1: Price replacement

23. The ABS significantly increased the use of transactions data to compile the CPI from the March quarter 2014.

24. Given the lack of international consensus on the best approach to fully integrate transactions data; and the research that had been undertaken internally; the ABS adopted a phased approach to the implementation of transactions data in the Australian CPI.

25. The first phase commenced on 1 January 2014. The ABS replaced some field collected prices with prices derived from transactions data. ABS field officers no longer visited businesses that have provided the ABS with transactions data.

26. The price for an individual product, identified from the transactions dataset by the product's Stock Keeping Unit (SKU), is calculated by dividing a SKU'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 more representative of prices paid by consumers over the reference period than point-in-time pricing.

27. The price samples of items/products which were collected in the field have been matched with the same or similar items in the transactions datasets. There have been no changes to the lower level aggregation formulae, with an equal weighted Jevons index still being calculated. In essence, the ABS changed the method of price collection without changing any of the underlying compilation practices used to construct the CPI.

28. Before discussing the Phase 1 method in more detail, it's useful to clarify what the SKU is, and how it relates to an individual product. SKUs are codes used by Australian

---

[6] An introduction to the RYGEKS method in a price index context can be found in Ivancic, L, Diewert, W.E, Fox, K (2011). Scanner data, time aggregation and the construction of price indexes. Journal of Econometrics 161, 24-35.

retailers. In general, the ABS has found the SKUs to be of sufficient detail to enable individual products to be identified from the transactions datasets. However, the SKUs differ in quality and detail between retailers and occasionally within divisions of the same retailer.

29. For example, one SKU may be 'Store brand chips, thinly cut, original flavour, 150g', while another may simply be 'meat'. Obviously, SKUs with insufficient detail, such as 'meat', could not be used and alternate pricing approaches are needed.

30. The ABS considered alternatives to using SKUs, however these all had drawbacks. Global Trade Identification Numbers (GTINs or barcodes) for example would typically be too low a level of detail, differentiating products by product aspects which would typically be considered irrelevant to consumers, such as supplier identity and packaging. Ultimately, SKUs were judged the most appropriate code to use in the Australian context.

31. Having obtained transactions data and resolved the data quality issues, the ABS – in Phase 1 – implemented processes to: identify and account for quality change; and maintain a representative sample of products to price (including coping with new and disappearing products).

Quality change

32. An essential part of price measurement is accounting for quality change and the introduction of new items (ILO Manual, 2004). This is particularly challenging when using transactions data as these datasets tend to exhibit a high level of churn in the products available 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.

33. Where price change is measured using small samples of field collected product prices, 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 are used to compile the CPI. There are broadly three scenarios where there is a need to quality adjust prices obtained from transactions data. They are:

    a. where new items are brought into the price samples (as replacements);

    b. where there has been a quantity change (eg. change in packet size) and the SKU has changed; and

    c. where there has been a quantity change and the SKU has not changed.

34. The first scenario is the simplest case and requires calculating a previous period price for the new item[7]. 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, as far as possible, that the new product is likely to be an appropriate replacement for the disappearing product (but with a different SKU). Quality adjustment is then actioned manually by price analysts based on the SKU description.

Sample maintenance

35. Traditionally, sampled products are selected for inclusion in the CPI basket by price collectors who examine the shelf space of the products and make judgements about their relative importance.

36. This traditional approach to sampling can be replaced by more scientific methods due to the availability of transactions data. Revenue shares for each product are used to determine the significance of each product within a product group. Products are then selected for inclusion in the CPI sample based on revenue share. These products are then mapped to the appropriate sample within the CPI classification.

37. Over time, however, products in the sample can lose relevance or even cease to exist. In these situations a replacement product is needed, thus maintaining the relevance of the sample. The ABS has developed *relevance tests* which can highlight items in the samples that have become unsuitable and also highlight and rank suitable items as replacements.

38. The main principle behind these relevance tests is that the products should have a stable revenue share (i.e. consistent revenue share compared to other products) within the CPI product group. These product groups are referred to as the Elementary Aggregate or 'EA'). The stable revenue share is particularly important, as items can have large sales when introduced into the market due to novelty or introductory sales prices, have insignificant revenue thereafter, and hence not be representative of the broader market.

39. To mitigate these problems, possible replacement products' revenue must have been stable and significant for the previous six months before they can be considered for inclusion into the price samples. Currently CPI analysts manually check all items which are flagged for replacement and manually select items from a list ranked according to average monthly revenue share over the previous six months. It's worth noting that the ranked list must be used with caution; one particular problem is that the ranking

[7] This approach is described in detail in the later section 'Introducing new products and calculating previous period prices'.

algorithm cannot distinguish between similar products with the same brand.

40. Many food and household items will have varieties of the same base item which have similar if not identical price evolution. A specific brand of canned tuna, for example, is available in many flavours and CPI compilers will be aware that prices for the different flavours from the same brand will behave similarly to one another, going on sale at the same time and changing price at the same time. Having a single flavour in our sample will hence represent the price movement for a much more significant portion of the market than that single flavour's revenues would suggest.

41. The current process used for determining product replacements is manually driven, requiring the analysts to manually select a replacement from this ranked list of potential products that pass eligibility criteria. This was a deliberate strategy and deemed manageable for the existing sample sizes. However, any significant expansion of price samples using this same process would lead to unmanageable workloads.

Introducing new products and calculating previous period prices

42. The ABS price index processing system relies on the calculation of a period to period (short-term) price relative. Hence, when introducing an item into the CPI, the ABS requires prices for the current period and for the previous period. The previous period price should be "representative of the average price at which the item has been sold" (ILO, 9.81). Emphasis is placed on removing the impact of atypical events, such as sales prices.

43. As an example, suppose a product has been identified for inclusion within the index. The product has a recent price history of $10 in the current period (time t), $5 at t-1 and $10 at t-2. Introducing the product at time t will show a 100% price increase and the index will rise accordingly (say, from 100 to 200). Suppose however that the product is introduced at time t-2. Now, it has fallen from $10 to $5 at time t-1. This is a 50% fall in price and the index would fall from 100 to 50. Now when the product returns to $10, the index returns to 100. The former case suggests a doubling in price, whereas the latter case shows the true price evolution.

44. The principle that the ABS has adopted is that previous period prices should reflect the price at a 'normal level of discounting'. This ensures that the index is not distorted by unusual price activity in the previous period when an item is introduced.

45. To calculate this price, the ABS has developed a method based on two values: the maximum price for the product and an average discount.

46. The maximum price is simply the highest weekly unit value for the product in the latest month Using this monthly maximum rather than a longer term average maximum ensures that true price evolution is captured, and price rises and falls during previous months are not incorporated into the calculated price.

47. An average discount is calculated by taking the ratio of the monthly unit price to the monthly maximum price and averaging this across the previous year.

48. A previous period price is then derived as the most recent monthly maximum discounted by the average monthly discount over the previous year. Tolerances are automatically applied to flag prices which appear too low or too high compared with the price history, and these are manually checked by CPI analysts. The flagging criteria are necessarily quite flexible, as there are several ways in which this method is known to fail, resulting in a significant time spent by CPI analysts quality assuring the calculated values.

Proposed future use of transactions data by the ABS

Phase 2 – expanding samples and using quantity data to weight at the lower levels of the CPI

49. The second phase of using transactions data to compile the CPI involves moving beyond the replacement method, or phase 1, by increasing the number of price observations used in aggregation and utilising quantity information from the scanner datasets for weighting purposes.

50. This phase needs to be carefully considered and managed due to resource implications. For example, mapping products from transactions datasets to CPI classifications can be very resource intensive and may result in not all products being able to be mapped to the Elementary Aggregate (EA) level of the index (this will depend on the detail of the product description or SKU).

51. Another barrier to implementing the larger samples for phase 2 is the workload implications, with current manual[8] processes taking some time for analysts to complete each quarter. To address this, processes for selecting new items as well as selecting an appropriate base price have been automated.

52. The automation of the sample maintenance process preserves the basic outline of the process, with the relevancy tests used to identify unsuitable products remaining the same, but automates the approach to two questions: whether a product should be removed from the sample when it is identified as unsuitable and, if so, which product is the best replacement?

53. The automated maintenance process first groups products into 'classes' based on the product's brand and size, and then uses a two-stage approach to find item

---

[8] Items mapped are checked each period and flagged for replacement when they are no longer representative, based on revenue flows. Analysts decide on replacements for the flagged items.

replacements. The first stage uses revenue shares to determine which class the new item should come from, while the second stage determines the exact items which should be selected, again using revenue shares. The two-stage approach addresses the brand problem outlined above.

54. The second part of the automation is around the selection of an appropriate previous period price. In keeping with the principle outlined above, previous period prices should reflect the product's price at a 'normal level of discounting'. This is a difficult problem to solve automatically, as products can have diverse price histories, making the calculation of a normal level of discounting difficult.

55. The ABS used an iterative approach to this problem, testing a number of possible algorithms and comparing them with selections made manually by an experienced analyst. These algorithms were continuously improved until they generally delivered similar results to those of the experienced analyst. A key difference to the current process is that there is no attempt to calculate the actual 'normal level of discount' as this has proven to be quite inflexible and requires frequent analyst intervention. Alternative methods based on longer term average unit values are applied.

56. Also in the second phase, quantity information from the scanner datasets will be utilised for weighting purposes. Specifically, this method utilises an "expenditure matrix" which results in a different CPI aggregation structure to that presented in Figure 1. See Attachment 2 for an example of the expenditure matrix.

57. Essentially, in this approach EAs are now based on respondent specific classification structures; these classification structures are provided in the transactions data. This will increase the number of items which can be included in the CPI (compared to phase 1) as the mapping can be done at the higher Expenditure Class (EC) level. Note that items within the new EAs will still be treated as a sample, and price movements will be determined using the geometric mean of price relatives, as with current samples.

58. This approach also allows for weighting to be done more frequently and based on more timely data. This is due to the new EAs, which are based on expenditure data already being received. This approach also enables transactions and non-transactions data to be explicitly weighted at the outlet or respondent level. An example of the revised aggregation structure based on the Vegetables Expenditure Class can be seen in Attachment 2.

Phase 3 – implementing new methods and approaches

59. New methods will need to be utilised to truly realise the potential of transactions data in the CPI. Contemporary research suggests that the application of multilateral methods represent the most promising way in which this can be achieved. The development and application of multilateral methods represents the final, third phase.

60. The ABS has commenced research which applies multilateral methods to transactions data obtained from retail businesses. Particular attention has been focused on the Rolling Year GEKS (RYGEKS) and the Fixed Effects with a Window Splice (FEWS)[9] methods. These methods use a series of bilateral comparisons spanning across a 13 month (or 5 quarter) period to counter the impact of seasonality. The intention is to monitor the behaviour of these methods over time and refine aspects of the methodology (e.g. length of window splice) prior to making a decision about implementation in the CPI.

Phase 3 Preliminary/Experimental results

61. Research to date has largely focused on establishing whether the RYGEKS or FEWS represent the best method for use in Phase 3. The Imputation Tornqvist RYGEKS (ITRYGEKS) method[10] was also examined, and determined as unsuitable as the ABS transactions datasets lack detailed product characteristic data required for this method.

62. To facilitate comparisons, a "baseline" was required. GEKS was chosen as the benchmark due to it being transitive. However, the authors acknowledge its lack of applicability in a statistical production environment due to the requirement for revisions. For the purpose of this study a Tornqvist base RYGEKS was used with a 13 month or 5 quarter window length. Likewise the window length for the FEWS was set at 13 months or 5 quarters.

63. Results are presented in three sections. The first describes comparisons using real transactions data, for selected ECs, further dissected by respondents. The second section uses simulated data to highlight how the different indexes might respond to plausible (albeit exaggerated) events. The third section brings a mixture of analysis conducted on both real transactions data and simulated data to bear on some of the differences observed in Section 2. It gives particular attention to the FEWS response to product rotation, and manipulating the window splice to mitigate this.

Results from real transaction data

[9] Krsinich, F. (2014). Fixed effects with a window splice: quality-adjusted price indexes with no characteristic information. Paper presented at the meeting of the group of experts on consumer price indexes, Geneva, Switzerland. May 2014.Krsinich F (2014).

[10] de Haan, J., & Krsinich, F. (2012). Scanner data and the treatment of quality change in rolling year GEKS price indexes. Paper presented at the Economic Measurement Group Workshop, Sydney, November 2012.

64. The first analysis undertaken was to produce monthly indexes at the respondent level. Detailed results cannot be released in this paper due to confidentiality restrictions. Discussion is limited to describing the general behaviour of the various indexes.

65. Comparisons revealed that the FEWS and RYGEKS tracked very closely to one another in most circumstances. Differences between the two were most apparent when a product rapidly lost market share while simultaneously experiencing a large change in price. In these circumstances, the FEWS index correctly adjusted. In comparison, the RYGEKS did not quickly adjust, and this led to spurious movements, which were often quite dramatic. This property of RYGEKS suggests that careful adjustment, such as through a clearance filter, must be made to prevent shocks caused by items going on clearance, or items changing price and quantity significantly.

66. The results also demonstrated the drift of the chained Tornqvist leading us to conclude it is not an appropriate index for our purposes. Surprisingly, there was not a noticeable difference between the RYGEKS and the FEWS when there were a number of new and disappearing items which left us unable to draw conclusions regarding the implicit quality adjustment which is theoretically made by the FEWS index.

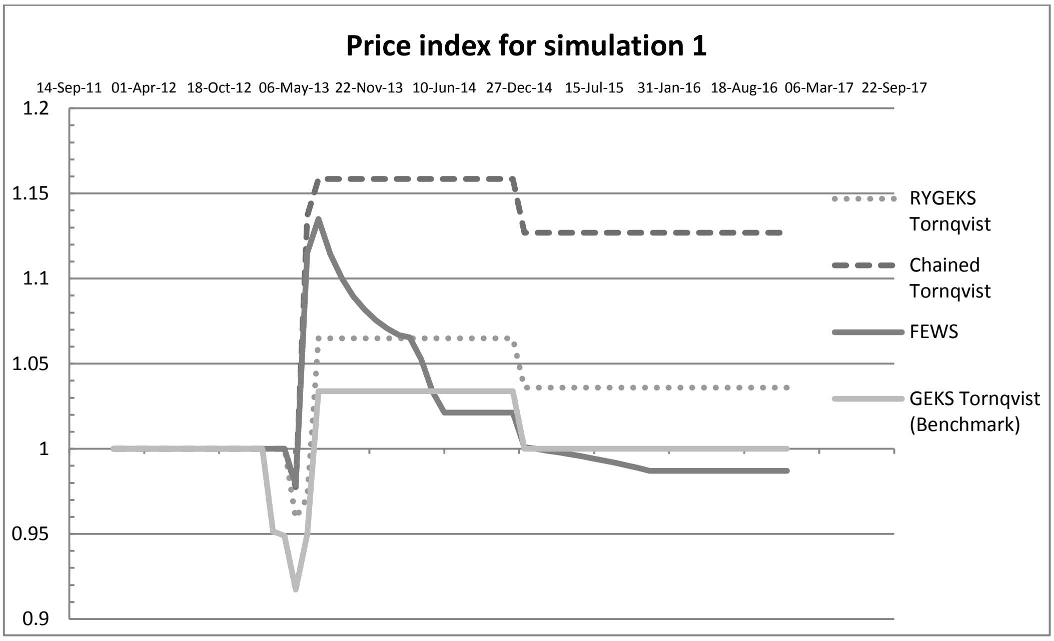
Results from Simulated data

67. The second analysis was undertaken on simulated data, which were used to extract differences between the indexes on entirely controlled datasets. That is, the price and quantity data was modified to elicit different reactions in the indexes, based on observations made in the first analysis. From the following diagrams it is clear that certain characteristics of the price data may significantly separate the RYGEKS, FEWS and Chained Tornqvist, although it could be argued that the simulated data is an exaggeration of real world observations.

Simulation 1: Loss of identity in the FEWS index

68. This dataset begins with 5 items with exactly equal prices and quantities. There are then four stages in this data: For the first 13 months prices and quantities remain unchanged, so that there are no initial window effects for FEWS or RYGEKS. Then over the next 5 months an expensive product drops its price, causing an increase in its quantity sold while another product subsequently loses market share (around Jan 2013). The product then raises its price again (around August 2013) – all other products are left unchanged. Prices are then kept constant for a year, and finally (around Dec 2014) the prices and quantities return back to their original values.

Graph 1: Simulation 1 Loss of identity in the FEWS index



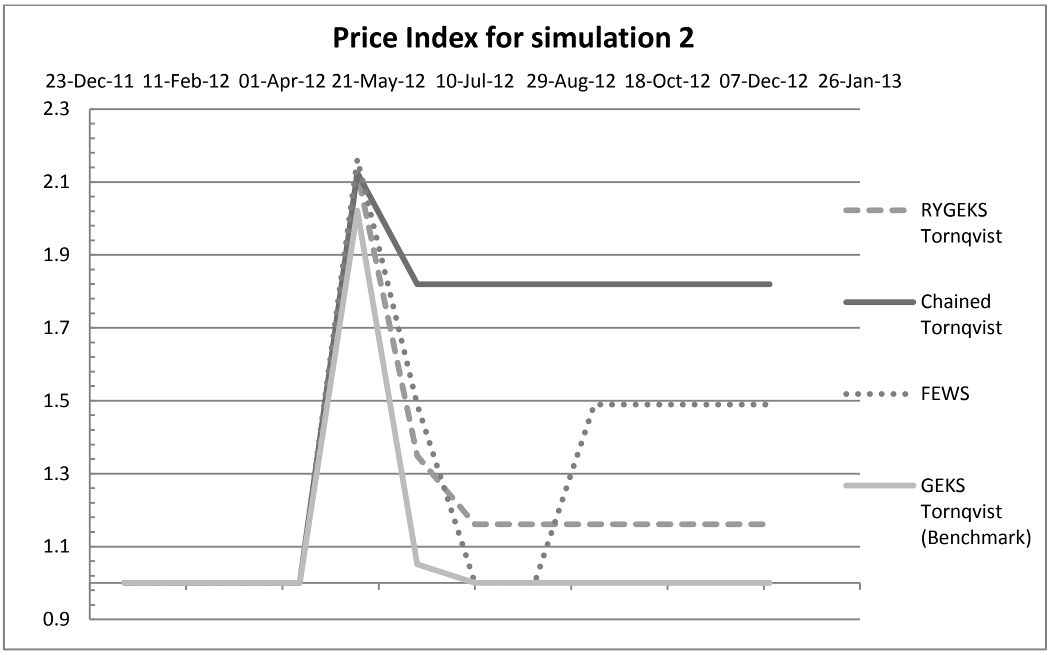
69. This example shows quite clearly that the FEWS adjusts over time - as it collects more information it re-contextualises past movements and corrects for this. This means that the FEWS shows movements in the absence of underlying real world events. This could be seen as undesirable since these movements would be hard to explain to users (and perhaps critically received). As can also be seen in this example, the FEWS somewhat overestimated the first movement compared to the RYGEKS and the GEKS, and only redeems itself by its correction over time. This tendency to exaggerate the first movement and then correct over time was observed in real data. It is worth noting that such retrospective correction is a mandatory quality of any index that adjusts itself to accommodate new products.

70. Of the three indexes compared, the FEWS displays the least drift in this case settling close to the original level, although RYGEKS comes close. As expected, the Chained Tornqvist shows the largest degree of drift.

Simulation 2: Undesirable properties of FEWS due to product rotation

71. In this simulation, the FEWS and RYGEKS have been changed to only have a 4-month window length, a change that was made so that the entire effect could be created within a year's worth of price data. After an initial phase of four months with no price, quantity or product changes, product rotation is performed as follows. Firstly in April 2012, one product experienced a price rise and two new products are introduced. Then the product which increased its price is replaced with another product at the original, lower price (Jun 2012). Finally all products are reset to original (Jul 2012).

Graph 2: Simulation 2: Undesirable properties of FEWS due to product rotation

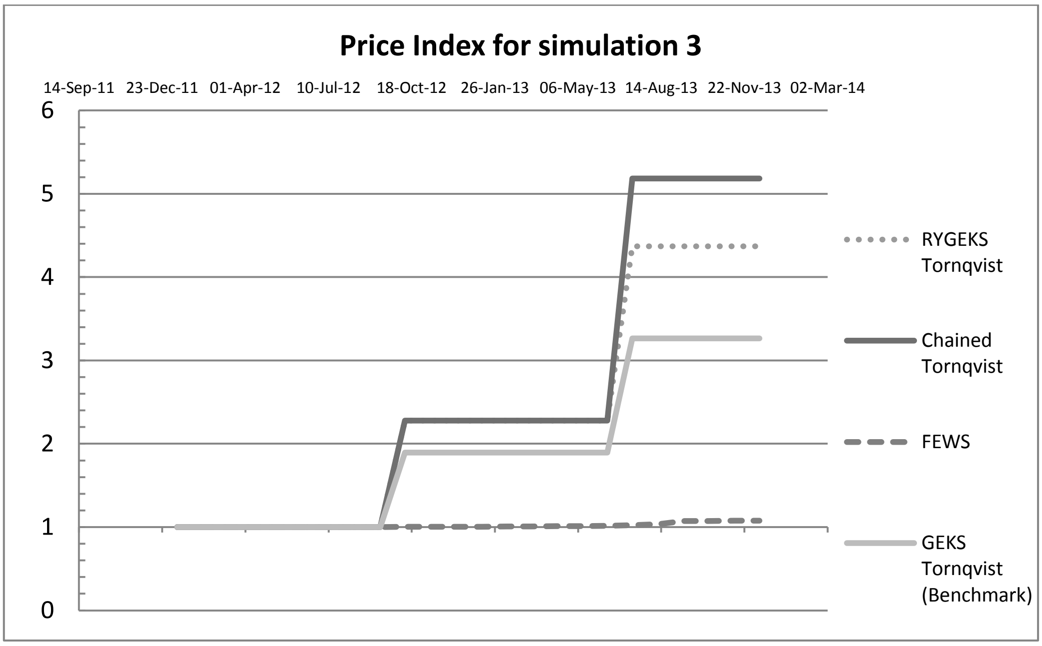


72. In this series we would expect to see a pattern closest to the GEKS: showing an increase in the index for the price rise in one of the products, then returning to normal levels in the next few periods. FEWS does this very well, except for the spurious upward movement a few periods later (around Aug 2012). This movement is caused by the window splice - essentially when the window contains all three periods of the product rotation, the fitted model can compare them to show a drop in prices, however when the window moves on so that it can only detect the two later periods, there are no shared units that give an indication of a drop in price, so FEWS adjusts for this by moving back upwards. This problem was observed in the real data, when products were only available in the first period of the window. A modification of the FEWS window splice was explored as a possible method for resolving this issue. This is described in paragraphs 82-87.

Simulation 3: Undesirable properties of RYGEKS due to loss in market share

73. This simulation dataset was designed to elicit a large spurious response from the Tornqvist-based indexes. It was created by replicating and amplifying the conditions that were observed on the real data: One unit that had a significant proportion of market share suddenly had a large reduction in market share at the same time it experienced a large price change. This is then repeated with a second product (the second jump in June 12). All other products remained constant.

Graph 3: Simulation 3 Undesirable properties of RYGEKS due to loss in market share



74. The Tornqvist-based indexes all react violently to this change, because even though the market share of the unit has become very small, its movement is weighted by the average of its current and previous market shares The inclusion of the previous weight ensures that the impact remains sizable. The FEWS model only weights prices with their contemporary expenditure, so that the sudden irrelevance of a previously important item is fully taken into account, preventing this movement from dominating the series.

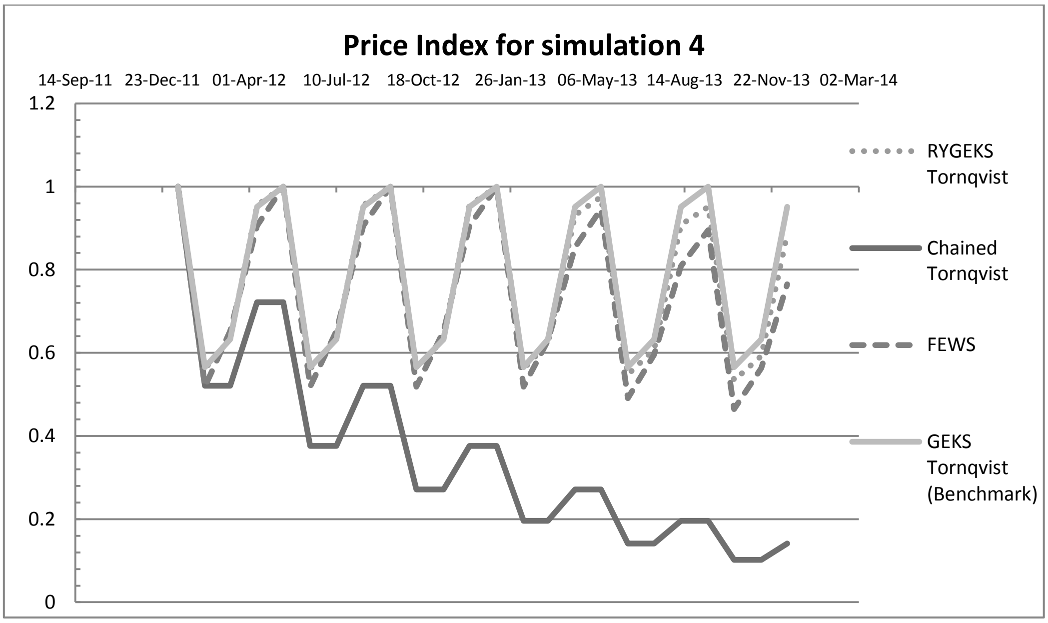
75. Whilst the data used here is simulated, effects like this do occur in the real world. This kind of pattern is common when products are sold in clearances or when market share and price change significantly simultaneously.

Simulation 4: Drift in FEWS

76. This data is taken from an investigation into the effect of seasonal products on RYGEKS indexes (Ribe, 2012 [11]). It features two products, one which holds constant and the other that fluctuates in consumption and price, creating conditions that are associated with drift.

Graph 4: Simulation 4 Drift in FEWS

[11] Ribe, M. (2012), Some Properties of the RGEKS Index for Scanner Data. Paper presented at Statistics Sweden's scanner data workshop, Stockholm, Sweden. 7-8 June 2012.

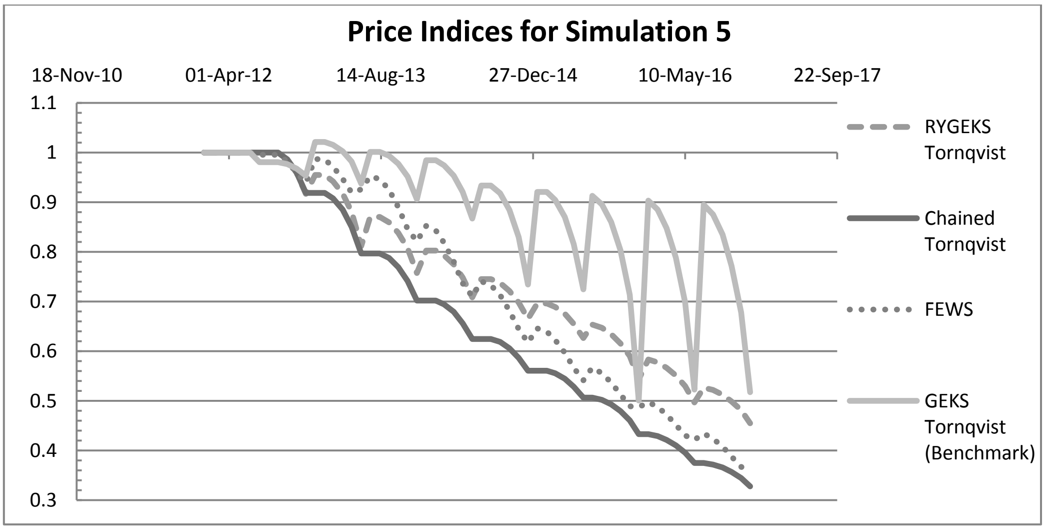


77. As expected, the Chained Tornqvist drifts rapidly. The drift in FEWS and RYGEKS is less prominent, but still evident. Interestingly FEWS appears to drift at a greater rate than RYGEKS.This suggests that seasonal products need further investigation to determine whether a FEWS or RYGEKS are appropriate indexes.

Simulation 5: Further example of drift

78. This example was used to test whether FEWS or RYGEKS could cope with with a high degree of product churn and "novelty" (that is, where products have high prices when they're new and low prices before they rotate out). In the real world, such price behaviour is common in products such as fashion or technology goods.

Graph 5: Further example of drift



79. All of the indexes drift downwards. Such drift is clearly unrepresentative of the “real” price behaviour, where new products typically enter with increasingly higher prices. This suggests that these sorts of ECs with high novelty require other methods as suggested in the literature.

Aggregated results on real transaction data

80. A further investigation was undertaken, which aggregated the transaction data across all available respondents to create an All Capital Cities, EC level quarterly index[12]. This index was then compared to the published CPI. Again for confidentiality reasons, the detailed results are not provided here, although a brief outline of the observations will follow. Many of the problem characteristics of the indexes described above were not observed at the aggregated quarterly EC level. This includes the incorrect movement for clearance sales/loss of market share, as well as the correction of FEWS over time. This is promising, however, caution must be exercised in accepting these results due to the short period transactions data has been used. The ABS intends to continue monitoring this over a longer period.

81. In some cases the transaction data indexes tracked well against the published CPI, however there were instances where the set of transaction data indexes moved in opposite directions to the published index. The ABS intends to investigate this further.

Modifications to the FEWS window splice

82. Results from the above suggested that the FEWS was susceptible to instances of product rotation. The FEWS has two primary steps: (1) Estimation of month effects within a certain window, some months prior to the current month. We will call this the "estimation window" (2) Splice of the ratio of month effects across a certain window with the chained estimates up to the beginning of this window. Term this the "splice window".

83. The primary cause of the spurious movement observed was determined to be the fact that one unit rotates out (or at least largely rotates out, in the sense that it's expenditure share drops to almost zero), so that in the final window it appears in, all of its price difference from the mean price is allocated to its product effect and it has no impact on the month effect for that period.

84. In an attempt to remedy this effect, the month effect estimates were calculated on larger windows than the splice window. This means that for products that are in sample and then drop out of sample, there will always be at least two observations to create month effect estimates. This appeared to prevent spurious revisions or “phantom” movements from being detected.

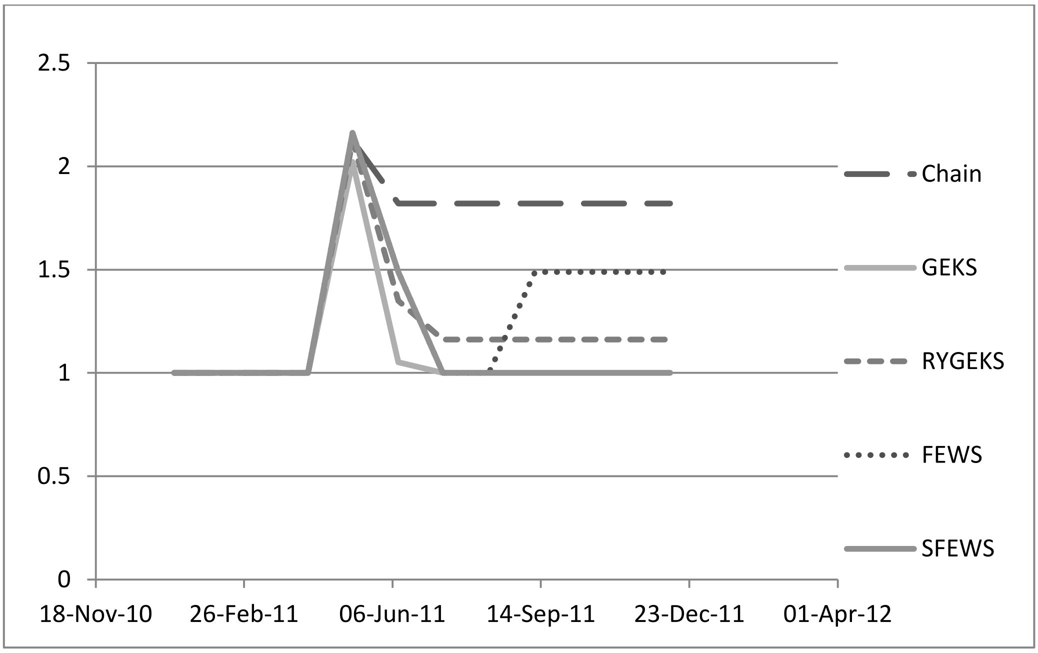
85. In the standard FEWS approach both the estimation window and the splice window are set to periods of 13 months. Our modification reduced the splice window to 12 periods (although likewise we could have increased the estimation window to 14 periods). We believe that if the splice window is made to be shorter than the

[12] Not all respondents provided transaction data, and hence the aggregated index represented the majority of prices in scope of the published CPI but not all.

estimation window then we remove the observed product-death effect and replace it with a slight loss of representativeness, where the splice is further contextualised by time periods before its effect.

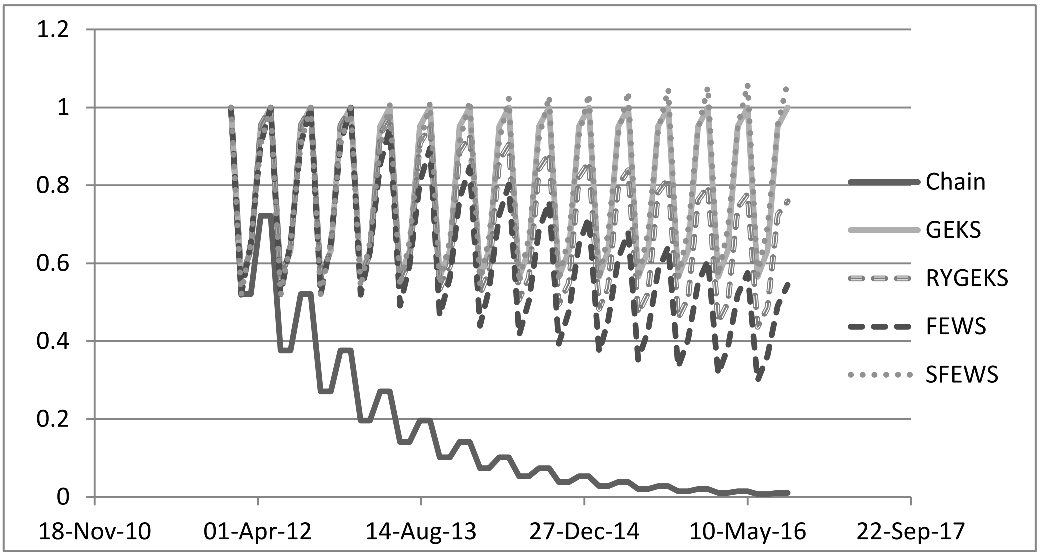
86. The Graph below shows the outcome of this modification (which is labelled as SFEWS). As you can see the original FEWS had a spurious upward jump, whereas the modified FEWS shows no reaction to the product rotation. Apart from this deviation, the modified FEWS follows the original FEWS perfectly. This shows one clear advantage to using the modified FEWS in this circumstance.

Graph 6: Impact of SFEWS modification



87. Another general observation about the modified FEWS is that it seems to be less prone to drift. Using the same periodic data as Ribe (2012), we see that the modified FEWS (denoted as SFEWS) drifts at a much slower rate than either FEWS or RYGEKS. This observation is further supported in actual transactions data.

Graph 7: **SFEWS, FEWS and RYGEK response to Ribe (2012) data**



Summary of results (Conclusions) from empirical study

88. Based on the empirical results of this investigation, it appears the FEWS (or some alteration of the Fixed effects model) is the most suitable method when using prices and quantities from transactions data to compile the CPI. Notwithstanding, the FEWS has a number of undesirable properties which will need to be considered before application of the index method in official statistics.

89. Of most concern, is the FEWS' tendency to show movement even when prices are constant, thus not satisfying the identity test. In these situations the FEWS often exaggerated the initial movement (of a new product entering the market) and then gradually as more information on the product became available, adjusted itself to a more accurate level than the other indexes. Whilst this is a necessary condition for an index that adjusts itself in response to new products, careful consideration will be required before the ABS commits to the implementation of this method.

90. The ABS is planning further work to investigate an alternative method, the Fixed Effects with Movement Splice (FEMS), in comparison to the Ffixed Effects with Window Splice (FEWS) on available transaction data. As well as the most appropriate window length.

91. The ABS is continuing to investigate this pathway, focussing on the interpretability of the method and communication with users. A decomposition of the FEWS to product level contributions is under investigation with work to continue in the future. The ABS is also investigating methods to address the matched-model aspects of the FEWS. In particular a method in which we can match new and disappearing items using the description strings available on the transaction data. A simplistic approach has been developed however the results didn't yield enough high quality matches in the data. The ABS is pursuing machine learning techniques to address the issue. Future work also involves the development of a quality framework to assess and monitor the quality of transaction data received prior to implementation in a production environment.

92. It’s worth noting that whilst the ABS intends to continue research into variants of the FEWS, this method is unlikely to be fit for purpose for all ECs within the CPI.

Conclusions

93. The results presented in this paper indicate that there are significant gains to be made by maximising the use of transactions data in the CPI. This is achieved by increasing pricing samples, using quantity information for weighting purposes and constructing the CPI using multilateral methods. In addition, the results indicate that index volatility is noticeably reduced as a result of these approaches. It is important to note that these findings are experimental and need to be replicated using a larger pool of data.

Further analysis is required before these phases will be considered for implementation in the CPI production environment.

94. This paper also proposes a phased implementation of transactions data in the CPI. This is because the CPI is one of the most important statistical outputs for managing the economy and for indexing government outlays. This should result in a cautious implementation of methodological changes, in particular, that are rigorously tested as well as ensuring changes are well-understood and supported by relevant stakeholders. The phased approach presented here enables a progressive increase in the use of transactions data that demonstrates the benefits of each new phase to users.

Areas for Further Research

95. During this initial investigation the following additional areas for future research were identified in order to maximise the use of scanner data in the CPI.

96. Firstly, in order to utilise quantity or revenue data in phase 2, a practical approach to weighting the respondent level EAs should be developed. The main barrier to utilising the scanner data for explicit weighting purposes at the EA level is acquiring data to estimate the expenditure volumes for the non-transactions data respondents.

97. Secondly, it is anticipated that the increased use of scanner data will result in only a small number of field collected observations being collected in the future. The benefit of these observations under this new context needs to be assessed. It is expected that once expenditure data is secured for the non-transaction data respondents this issue will be largely addressed.

98. Thirdly, in terms of the implementation process, quality gates need to be developed to allow analysts to sufficiently analyse outputs across the various stages in a production environment.

99. Fourthly, methods on how seasonal items will be measured using transaction data need to be established.

100. Finally, the avenues of research identified by research into multilateral methods, such as investigation of the FEWs window splice, text matching, decomposition tools, as well as a general emphasis of understanding the impact on the CPI, will be pursued.

References

de Haan, J., & Krsinich, F. (2012). Scanner data and the treatment of quality change in rolling year GEKS price indexes. *Paper presented at the Economic Measurement Group Workshop, Sydney, November 2012.*

Ivancic, L, Diewert, W.E, Fox, K (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics 161, 24-35.*

International Labour Office (2004). *Consumer Price Index Manual: Theory and practice, Geneva, International Labour Office.*

Krsinich, F. (2014). Fixed effects with a window splice: quality-adjusted price indexes with no characteristic information. *Paper presented at the meeting of the group of experts on consumer price indexes, Geneva, Switzerland. May 2014.*

Ribe, M. (2012), Some Properties of the RGEKS Index for Scanner Data. *Paper presented at Statistics Sweden's scanner data workshop, Stockholm, Sweden. 7-8 June 2012*.

Attachment 1: Product groups where some prices are derived from transactions data

- Bread
- Cakes and biscuits
- Breakfast cereals
- Other cereal products
- Beef and veal
- Pork
- Lamb and goat
- Poultry
- Other meats
- Fish and other seafood
- Milk
- Cheese
- Ice cream and other dairy products
- Fruit
- Vegetables
- Eggs
- Jams, honey and spreads
- Food additives and condiments
- Oils and fats
- Snacks and confectionery
- Other food products n.e.c.
- Coffee, tea and cocoa
- Waters, soft drinks and juices
- Take away and fast foods

- Tobacco
- Garments for men
- Garments for women
- Garments for infants and children
- Glassware, tableware and household utensils
- Tools and equipment for house and garden
- Cleaning and maintenance products
- Personal care products
- Pharmaceutical products
- Medical and hospital services
- Spare parts and accessories for motor vehicles
- Automotive fuel
- Newspapers, magazines and stationery
- Pets and related products
- Property rates and charges

Attachment 2: Current and revised aggregation structure