# Scanner Data, Time Aggregation and the Construction of Price Indexes

by

Lorraine Ivancic[1], Kevin J.Fox[2]* and W. Erwin Diewert[3]

**Abstract**

The impact of weekly, monthly and quarterly time aggregation on estimates of price change is examined for nineteen different supermarket item categories over a fifteen month period using scanner data. We find that time aggregation choices have a considerable impact on estimates of price change. When chained indexes are used the difference in price change estimates can be huge, ranging from minus 1.42% to minus 25.78% for a superlative (Fisher) index and an incredible 17.22% to 9,548% for a non-superlative (Laspeyres) index. The results suggest that normal index number theory breaks down when weekly data with severe price bouncing are used, even for superlative indexes. However, indexes constructed using quarterly time aggregation appear to be largely free from chain drift.

*JEL Classifications: C43, E31*
*Key words: Price indexes, aggregation, scanner data.*

---

[1] Centre for Applied Economic Research, University of New South Wales
[2] School of Economics and Centre for Applied Economic Research, University of New South Wales
[3] Department of Economics, University of British Columbia

*Corresponding author:
Kevin J. Fox
School of Economics and CAER
University of New South Wales
Sydney 2052 Australia
Tel: +61-2-9385-3320
Fax:+61-2-9313-6337
K.Fox@unsw.edu.au

Acknowledgements: The authors gratefully acknowledge the financial support from the Australian Research Council (LP0347654 and LP0667655), and the provision of the data set by the Australian Bureau of Statistics.

## 1. Introduction

Aggregation is fundamental to the construction of any price index. Prior to index number calculation decisions must be made as to how price and quantity data are to be aggregated in order to estimate price change. Aggregation decisions are generally limited by the use of regular but infrequent surveys to collect data used in the compilation of the Consumer Price Index (CPI). However, the advent of high-frequency electronic-point-of-sale "scanner data" has made increasingly detailed and comprehensive data on consumer purchases available to price statisticians. The use of more detailed data means that aggregation issues become even more complex when attempting to estimate price change. There are a number of dimensions (e.g. items, stores, time) over which data can potentially be aggregated before an index is calculated. In this paper we are primarily concerned with the issue of *time* aggregation on estimates of price change.

Only a handful of authors have used scanner data to examine this issue, including Reinsdorf (1999), Hawkes (1997), Bradley et al. (1997), de Haan and Opperdoes (1997) and Dalen (1997). Reinsdorf (1999) found that the use of different aggregation methods over time resulted in estimates of price change which differed by as much as 7.9% while de Haan and Opperdoes (1997, p.10) found that 'taking unit values [average prices] over one week every month instead of unit values over the entire month as the price concept leads to differences in the formula that exceed by far the differences due to alternative elementary aggregate index formula'. These results indicate that time aggregation decisions are likely to be extremely important, particularly when high frequency data are used.

A limitation of existing studies is that they typically use data on a small number of product categories. For instance, Reinsdorf (1999), Hawkes (1997), de Haan and Opperdoes (1997) all had information on only one product category (coffee), while Dalen (1997) had information on four product categories (fats, detergent, breakfast cereal and frozen fish). This makes it difficult to draw broad conclusions or make generalisations. A major benefit of the current study is that we have information on 19 supermarket item categories. This allows us to examine whether results found in other studies hold for a larger set of products and whether regularities, resulting from different aggregation methods and the use of different index number formulae, can be identified across different item categories.

The paper is set out as follows. Section 2 provides a brief overview of the properties of aggregation units, known as unit values. The aggregation and index number methods used to estimate price change are outlined in Section 3. A brief description of the data set is given in Section 4 as is a discussion of the index number results. In Section 5 we attempt to quantify the degree of chain drift in the estimated indexes; i.e. the degree to which the process of chaining causes the indexes to be biased. Section 7 concludes.

## 2. Aggregation and the construction of unit values

Aggregation in this context refers to the calculation of average prices and total quantities which are to be used as inputs in the compilation of a price index. Aggregation over quantities is relatively straightforward. Once the unit to aggregate over has been chosen the quantities relevant to that unit are simply added up.[6] Aggregation over prices involves the construction of what is known as a unit value. A unit value is, in effect, the average price over some unit such as time, product group or stores. Comparing the unit values between two periods results in a unit value index of price change.

From a theoretical perspective, the use of unit value indexes is a somewhat contentious. The source of this controversy largely stems from the failure of unit values to satisfy two axiomatic properties which are used to evaluate index number formulae (Balk, 1998).[7] These are the “identity” and “dimensional invariance” axioms.

The *identity axiom* states that ‘if the price of every good is identical during the two periods, then the price index should equal unity, no matter what the quantity vectors are’ (ILO, 2004). This can be regarded as somewhat controversial as this test does not take into consideration shifts in the quantities purchased between the two periods. *Dimensional invariance* refers to the idea that the

[6] See Hawkes and Piotrowksi (2003) for a range of potential aggregation units.
[7] For a more detailed explanation of the axiomatic approach to index number theory see Chapter 16 of the CPI Manual (ILO, 2004).

price index should not change if the units of measurement for each commodity are changed. A unit value index fails to be invariant to the units of measurement used. These failures do not necessarily rule out the use of unit values, as many of the well-known indexes fail to satisfy all 'reasonable' axiomatic tests and there is no general agreement on the 'ideal' list of axioms. As such, unless the failure of the identity and dimensional invariance axioms is thought to be crucial, unit values indexes may still provide a useful representation of average price change.

Balk (1998) showed that a unit value index is equal to a partial Cost-of-Living index (COLI)[8] if 'base and comparison period expenditures on the commodity group are optimal with respect to the prevailing prices…and only if the underlying preference ordering can be represented by the simple sum utility function' (p. 8). Bradley (2005) argues that Balk's finding only holds when items in the commodity group are either perfect substitutes or sub-utility is Leontief, and that these two cases are 'extreme and most often do not hold'. He goes on to say that the use of unit value indexes, where the goods are not pure complements or perfect substitutes, will lead to inconsistent estimates. The problem with this view from a practical perspective is that, as Diewert (1995, p. 20) notes, 'at some level of disaggregation, bilateral index number theory breaks down and it becomes necessary to define the average price and total quantity…using what might be called a "unilateral" index number formula'. This break down may occur if, for example, the time period which we aggregate over is so short that it leads to zero quantities purchased in the second period. In other words, in order to calculate a bilateral price index, time aggregation of some sort is needed.

It may be argued that rather than using unit values, a handful of what are thought to be 'representative' price quotes could be used. However, this course of action would involve a loss of much of the information on consumer purchases that scanner data has to offer. Furthermore, Diewert (1995) argues that 'it should be evident that a unit value for the commodity provides a more accurate summary of an average transaction price than an isolated price quotation'. Balk

[8] A 'partial' COLI refers to a COLI for a particular commodity sub-group.

(1998) shows that a unit value index may actually be more accurate than a single price quotation.[9]

The use of unit values at some level is necessary. However, the use of a simple aggregate unit value index over all goods should be avoided in favour of index number formulae which use unit values as representative prices for a particular commodity class, where the goods within each class are seen as being similar enough for their quantity units to be aggregated. This paper examines the time unit over which the unit values should be defined for each commodity class, the additional impact of aggregating over outlets in defining the unit values, and the differences between price change estimates from using these unit values in different index number formulae.

## 3. Estimating Price Change using Scanner Data

To examine these issues, a number of different index number formulae were used to calculate price change. The commonly used base period weighted Laspeyres index and its' current period weighted counterpart, the Paasche index, were calculated. The theoretically more attractive "superlative" indexes (Fisher, Törnqvist and Walsh indexes) were also calculated (Diewert, 1976). As the estimates of price change were not noticeably affected by the use of different superlative indexes, results presented in this paper are for the Fisher index.[10]

The (fixed base) Laspeyres price index can be written as follows:

$$Laspeyres_t = \sum_i w_{i0}\left(\frac{p_{it}}{p_{i0}}\right), \qquad (1)$$

where $p_{i0}$ is the based period price of item $i$, $p_{it}$ is the price of item $i$ in period t, for t = 1,…, T, and $w_{i0}$ is good $i$'s share of total expenditure in period 0. In practice, the prices are unit values for commodity class $i$ for each period t of some pre-specified length (e.g. weeks, months or quarters). Note that equation (1) aggregates unit value indexes by using appropriately defined share weights.

[9] Balk (1998) argues that 'if the unit value index is appropriate for a certain commodity group then it is equal to each single price ratio, and all those price ratios are equal.' 'In practice, however, there may be small distortions'. A unit value index is able to capture these price distortions whereas a single price quote cannot.
[10] Diewert (1978) noted that as all standard superlative indexes approximate each other to the second order it should not matter which superlative index is used.

A common counterpart to the Laspeyres price index is the Paasche price index, which can be written as follows:

$$Paasche_t = \left[ \sum_i w_{it} \left( \frac{p_{i0}}{p_{it}} \right) \right]^{-1}, \qquad (2)$$

where $w_{it}$ is good $i$'s share of total expenditure in period t, for t = 0,…,T.

The Fisher index formula is the geometric mean of the Laspeyres and Paasche indexes, $Fisher_t = [Paasche_t \text{ x } Laspeyres_t]^{1/2}$.

For each index number formula,

1. average prices and total quantities were aggregated in turn, over weekly, monthly and quarterly intervals; and
2. items were in turn, treated as different items if they were not located in the same store (*no* item aggregation over stores) or treated as the same good no matter which store they were in (item aggregation over stores).

The issue of whether or not to aggregate items over stores was considered in tandem with the time aggregation problem as it is of interest to know if such aggregation mitigates the effects of the choice of time aggregation. Statistical agencies currently aggregate items over stores, "perhaps because search theory was not a prominent part of economics before Stigler (1961)" (Triplett, 2003, p. 153). However, if consumers do not substitute across outlets it is clear that "unit values across stores are not the prices actually faced by households and do not represent the per-period price in the COLI, even if the unit values are grouped by type of retail outlet" (Triplett, 2003, p. 154). Hence, the impact of such aggregation is of practical interest.

Direct and chained indexes were also estimated for all of these combinations. For direct indexes, the basket of goods over which the price index is constructed is held fixed over time, while for chained indexes the base period is incrementally updated. Two types of chained indexes were estimated in this study. First, an index we refer to as a 'fixed basket' index was estimated using a basket of items which was matched with the direct index — no new items which appeared in the sample period were incorporated into this index over time. This type of index provides a 'pure' comparison with the direct index as it is not affected by new items or quality change. Second, a 'flexible basket' index which incorporated new items which became available over time was also estimated. It is of interest to see how this index behaves relative to the 'fixed chain' as new items "may experience price changes that differ substantially from the price changes of existing items" (ILO, 2004).

One of the important features of chained indexes is that the basket of goods is able to be constantly updated as new and disappearing goods are able to be incorporated into estimates of price change over time. However, chained indexes may suffer from what is known as chain drift. Chain drift occurs when an index 'does not return to unity when prices in the current period return to their levels in the base period' (ILO, 2004, p.445). An objective method to test for the existence of chain drift, known as the *mutiperiod identity test*,[11] was proposed by Walsh (1901; 401) (1924; 506) and Szulc (1983; 540). This test is defined as follows:

$$P(p^1,p^2,q^1,q^2)P(p^2,p^3,q^2,q^3)\,P(p^3,p^1,q^3,q^1) = 1. \qquad (2)$$

In equation (2), $P(p^1,p^2,q^1,q^2)$ and $P(p^2,p^3,q^2,q^3)$ are price indexes between periods 1 and 2, and then 2 and 3, respectively. Their product gives the chained price index between periods 1 and 3. Therefore, each index is also referred to as a chain link. Note that there is an additional link in the chain in equation (2), $P(p^3,p^1,q^3,q^1)$, a price index between periods 3 and 4. However, the period 4 price and quantity data are the same as the period 1 data, so $P(p^3,p^1,q^3,q^1)$, takes us from period 3 directly back to period 1. The price index formula P will have *no chain link bias* if the product of all of these factors equals 1.[12]

Chain drift is thought to result from what is known as price oscillation or bouncing and the associated quantity shifts (Hill, 1993). Price bouncing is often observed in supermarket scanner data as prices for supermarket items tend to frequently go on sale for short periods and then return to their pre-sale price. Scanner data not only captures price bouncing due to sales but also captures associated quantity shifts due to sales. Triplett (2003) argues that quantity shifts (due to

[11] Diewert (1993; 40 and 53) gave the test this name. Note that the number of intermediate periods can be extended indefinitely.
[12] Thus in the present context, perhaps a more appropriate name for Walsh's test would be the *chain link bias test*.

sales) may be largely due to two types of shoppers: shoppers who only buy when items are on sale and shoppers who stock up when an item is on sale; see also Feenstra and Shapiro (2003). Empirical work by de Haan (2008) using scanner data has shown that quantity shifts in response to sales are substantial. Therefore, it is of interest to see if our estimates of price change suffer from chain drift.

Direct and chained indexes were estimated over a 15 month period as follows:

1. quarterly estimates of direct price change compared prices in quarter 1 with quarter 5; chained estimates compared prices in all quarters, from quarter 1 to quarter 5;
2. monthly estimates of direct price change compared prices in month 1 with month 15; chained estimates compared prices in all months, from month 1 to month 15; and
3. weekly estimates of direct price change compared prices in week 1 with week 65; while chained estimates compared prices in all weeks, from week 1 to 65.

## 4. Results

We use a scanner data set collected by A.C. Neilson, which contains information on four supermarket chains located in one of the major capital cites in Australia. In total, over 100 stores are included in this data set with these stores accounting for approximately 80% of grocery sales in this city (Jain and Abello, 2001). The data set contains 65 weeks of data, collected between February 1997 and April 1998. Information on 19 different supermarket item categories, such as bread, biscuits and soft drinks are included. A large number of observations on transactions exist for all item categories, with a minimum of 119,565 observations for the item category 'bread' and a maximum of 2,639,642 observations for the item category 'juices'. An observation here refers to the average weekly price (weekly unit value) and total weekly quantity sold of each item transacted in each store in each week. For example, from table 1, there were 2,452,797 sales observations on biscuits over the 65 week period.

Within each item category the data set contains price and quantity information on all of the brands which belong to an item category sold in all stores; for example, from table 1 there were 1,322 brands of biscuits traded in all stores over the period. Additional information includes the product brand name, a unique 13 digit identifier (known as the European Article Number/Australian Product Number (EANAPN)) and, where relevant, the physical weight of the item.

Price change estimates are presented for Fisher, Paasche and Laspeyres indexes, and for direct and chained indexes using the methods described in section 3. In general, the results point to a high degree of variation in index number estimates across different methods of time aggregation and different index number formulae; see tables 2 to 7. The results are presented in index terms with a base of 100, so that e.g. 100.21-100 = 0.21% price change over the period. The results indicate that more time aggregation leads to more stable estimates of price change, for all types of indexes. However, the degree of the instability varies considerably across the different indexes.

The impact of time aggregation is extremely pronounced when chained indexes are used. This is particularly true for the Laspeyres index, where some price change estimates appear to 'explode' as the frequency of chaining increases. For example, from table 5, Laspeyres price change estimates for the item category toilet paper based on quarterly, monthly and weekly time aggregation (with no item aggregation over stores) range from a somewhat reasonable (106.71-100=) 6.71% (quarterly, fixed basket) to a massive (11,955-100=) 11,855% (weekly, fixed basket) over the 15 month period.[13] Considering comparisons between the intermediate periods as well, chained indexes can result in a massive range of estimates of price change between any two periods, from minus 1.42% to minus 25.78% for a superlative (Fisher) index and an incredible 17.22% to 9,548.23% for a non-superlative (Laspeyres) index.

With item aggregation over stores and using flexible-basket chained Laspeyres indexes (table 2), over the 19 product categories the average absolute difference between weekly and quarterly price change estimates is approximately 297%. When we look at indexes where items have been disaggregated over stores (table 5) this becomes 3,176%!

The Fisher index appears to be relatively less affected by time aggregation than the Laspeyres and Paasche index. Despite this, even the Fisher index shows a degree of variation which seems

[13] For Paasche indexes, the converse occurs, with chained estimates of price change falling rapidly.

to be a cause for concern. For example, from table 7, the Fisher flexible-basket chained estimates of price change for the item category toilet paper (no item aggregation over stores) were calculated at (100.43-100=) 0.43%, (98.61-100=) -1.39% and (79.86-100=) -20.14% for quarterly, monthly and weekly time aggregation respectively. With item aggregation over stores and using the flexible-basket chained Fisher index, on average the absolute difference between weekly and quarterly price change estimates is approximately 8%. When we look at indexes where items have been disaggregated over stores (table 7) the difference increases to approximately 14%. The observed volatility and extreme nature of some of our index number estimates (particularly evident when low levels of aggregation are combined with chaining) are consistent with findings in the existing literature (Feenstra and Shapiro, 2003; Reinsdorf, 1999; Dalen, 1997). It is known that non-superlative (Laspeyres) indexes are prone to drift when price bouncing is evident (Frisch, 1936; Forsyth and Fowler, 1981; Szulc, 1983). Importantly, our results indicate that even superlative indexes, when applied to weekly data, do not seem to be able to deal well with price bouncing behaviour.

It is also interesting to find that index estimates of price change are generally higher for the fixed-basket chained indexes relative to their flexible-basket chained counterparts. This result is quite pronounced when Laspeyres and Paasche indexes are used. When the superlative Fisher index is used, this result is still apparent but considerably less pronounced. This indicates that, for this particular data set, the introduction of new items which are often sold at special low introductory prices, and possible price discounting on exiting items, tends to lower the estimates of price change.[14] If these findings can be generalised to other product categories then this implies that fixing a market basket, particularly for product categories where product turnover is high, could bias price change estimates upwards.

In general, we would expect the impact of time aggregation on direct index estimates of price change to be minimal. Although the impact is substantially smaller than that for chained indexes, the difference in some of the estimates of price change due to time aggregation are still considerable. Over the 19 product categories, the average (absolute value) difference between

[14] Some ‘hidden’ price change may also occur if an item identifier (i.e. EANAPN code) changes for an item which the consumer considers to be essentially the same.

weekly and quarterly Laspeyres estimates of price change is approximately 2.4% (item aggregation over stores) and 3.4% (no item aggregation over stores), with differences going as high as 6.9% for the item category soft drinks (no item aggregation over stores). For the Fisher index, results appear to be marginally more stable, with the average difference between quarterly and weekly indexes estimated at approximately 2.0% with item aggregation over stores and 1.9% with no item aggregation over stores. However, even when using the Fisher index differences go as high as 6.7% for the item category oil (item aggregation over stores). These results show that both non-superlative and superlative direct indexes are noticeably affected by time aggregation.

It is also of concern to find that in a number of cases different aggregation methods used across time lead to a change in the sign of the estimate of price change. Examples of this are seen for both superlative and non-superlative indexes, and direct and chained indexes. For example, when price change is estimated for the item category biscuits with item aggregation over stores using the direct Fisher index — an index which we would anticipate to give us relatively stable estimates of price change — when quarterly time aggregation is used our results indicate that prices decreased by 1.44%, compared with a 0.82% price increase when weekly estimates are used (table 4).

At this point, it seems useful to try to explain in intuitive terms why chained superlative indexes fail us when the unit value time period is short. The problem can be traced back to what happens when outlets have sales or when they discount prices. An example will illustrate the problem. Suppose that we are given the price and quantity data in table 8 for 2 commodities for 4 periods:

**Table 8: Example Data Set**

| Period t | $p_1^t$ | $p_2^t$ | $q_1^t$ | $q_2^t$ |
|---|---|---|---|---|
| 1 | 1.0 | 1.0 | 10 | 100 |
| 2 | 0.5 | 1.0 | 5000 | 100 |
| 3 | 1.0 | 1.0 | 1 | 100 |
| 4 | 1.0 | 1.0 | 10 | 100 |

The first commodity is subject to periodic sales (in period 2), when the price drops to ½ of its normal level of 1. In period 1, we have a "normal" off sale demand for commodity 1 which is equal to 10 units. In period 2, the sale takes place and demand explodes to 5000 units. In period

3, the commodity is off sale and the price is back to 1 but *most shoppers have stocked up in the previous period so demand falls to only 1 unit*. Finally in period 4, the commodity is off sale but we are back to the "normal" demand of 10 units. Commodity 2 is dull: its price is 1 in all periods and the quantity sold is 100 units in each period. Note that the only thing that has happened going from period 3 to 4 is that the demand for commodity one has picked up from 1 unit to 10 units. Also note that conveniently, the period 4 data are exactly equal to the period 1 data so if Walsh's chain drift test were satisfied, then the product of the period to period chain links should equal one.

In table 9, we calculated the fixed base Fisher, Laspeyres and Paasche price indexes, PFFB, PLFB and PPFB and as expected, they behave satisfactorily in period 4, returning to the period 1 level of 1. The chained Fisher, Törnqvist, Laspeyres and Paasche price indexes, PFCH, PTCH, PLCH and PPCH are also listed in table 9. Obviously, the chained Laspeyres and Paasche indexes have chain link bias that is very substantial but what is interesting is that the chained Fisher has a 2% downward bias and the chained Törnqvist has a close to 3% downward bias.

**Table 9: Fixed Base and Chained Fisher, Törnqvist, Laspeyres and Paasche Indexes**

| Period | PFFB | PLFB | PPFB | PFCH | PTCH | PLCH | PPCH |
|---|---|---|---|---|---|---|---|
| 1 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| 2 | 0.69759 | 0.95455 | 0.50980 | 0.69759 | 0.69437 | 0.95455 | 0.50980 |
| 3 | 1.00000 | 1.00000 | 1.00000 | 0.97944 | 0.97232 | 1.87238 | 0.51234 |
| 4 | 1.00000 | 1.00000 | 1.00000 | 0.97944 | 0.97232 | 1.87238 | 0.51234 |

If the data corresponded to months, and they repeated themselves 3 times over the year, the overall chain link bias would build up to the 6 to 8% range, which is significant.

The problem in the tables 8 and 9 is this: when commodity one comes off sale and goes back to its regular price in period 3, the corresponding quantity does *not* return to the level it had in period 1; the period 3 demand is only 1 unit whereas the period 1 demand for commodity 1 was 10 units. It is only in period 4 that demand for commodity one recovers to the period 1 level. However, since prices are the same in periods 3 and 4, all of the chain links show no change (even though quantities are changing) and this is what causes the difficulties. If demand for commodity one in period 3 had immediately recovered to its "normal" period 1 level of 10, then there would be no problem; the use of a superlative index would simply reverse the downward movement in the price index due to the sale and we would be back to the presale situation.[15]

Our empirical results presented above indicate that the degree of time aggregation chosen impacts considerably on estimates of price change. This leads to the question of what type of aggregation is appropriate when scanner data are used. Weekly estimates of price change are, in many cases, found to be volatile to the extent of being implausible and chained indexes appear to suffer extensively from chain drift. The results indicate that the use of weekly time aggregation is inappropriate when combined with scanner data. The results are less clear cut for quarterly and monthly time aggregation. The range of price change presented in tables 2 to 7 makes it very difficult to know how much prices actually changed over the 15 month period. If we are able to determine how much of the chained index values reflect actual price change and how much is due to drift we may come closer to the true level of price change. This would also enable us to understand whether the use of quarterly or monthly chained indexes are appropriate when scanner data are used. Obtaining a measure of the extent of chain drift is the focus of the next section of this paper.

## 5. Estimating chain drift

Multilateral index numbers are often used for price and output comparisons across economic entities, such as countries (see e.g. Kravis et al. 1978, Caves, et al. 1982, Balk 1996 and Diewert 1999). These indexes satisfy a circularity requirement so that the same result is achieved if entities are compared with each other directly, or with each other through their relationships with other entities. Standard bilateral index-number formulae do not satisfy this circularity, or 'transitivity', requirement. The transitive GEKS multilateral index (Gini, 1931, Eltetö and Köves, 1964; Szulc, 1964) is the geometric mean of the ratios of all bilateral Fisher indexes,

[15] Thus it is sales and purchaser storage capabilities that cause the period immediately after the sale to be "abnormal" as compared to other periods. For additional discussion on the problems associated with sales and storage, see Feenstra and Shapiro 2003 and Triplett 2003.

where each entity is taken in turn as the base.[16] Consider the case where there are M entities that we wish to make transitive comparisons across. Let $P_{jl}$ denote a (Fisher) price index between entities *j* and *l*, *l* = 1,…,M, and $P_{kl}$ denote a (Fisher) price index between *k* and *l*. Then the GEKS index between *j* and *k*, can be written as follows:

$$GEKS_{jk} = \prod_{l=1}^{M} \left[ P_{jl} / P_{kl} \right]^{1/M} . \qquad (3)$$

It can be easily shown that this index satisfies the transitivity property, so that $GEKS_{jk}$ = $GEKS_{jl}/GEKS_{kl}$. If we treat each time period as an 'entity' we can make transitive comparisons across time periods using equation (3).[17] It can easily be verified that this index satisfies the multiperiod identity test in equation (2), and is hence free of chain drift.[18]

The advantage of this approach over direct (fixed base) indexes is that we can use the flexible basket approach for each of the bilateral comparisons in the GEKS index. This is the same as with chained indexes, which allow us to update the base each period and use a flexible basket approach. Thus, a comparison between chained and GEKS indexes allows us to isolate any chain drift.

It should be noted that the use of the Fisher formula in the GEKS approach has a strong connection with economic theory.  Diewert (1999) termed the GEKS method a *superlative method*; i.e., if households (or more generally, purchasers of the product group under consideration) had square root quadratic preferences either in the primal and dual[19], then the

[16] Sometimes the term 'GEKS', or commonly 'EKS', is used to refer to the method of making any bilateral index number formula transitive using the same geometric averaging technique. Here we employ the more common usage of the term so that it refers to the multilateral index based on the bilateral Fisher index formula.
[17] This approach is typically not used for constructing indexes across time due to the loss of characteristicity (Dreschler, 1973). Characteristicity refers to the 'degree to which weights are specific to the comparison at hand' (Caves et al., 1982).  Dreschler (1973, p. 17) noted that 'characteristicity and circularity are always…in conflict with each other." This conflict is usually resolved by imposing chronological ordering as the unique ordering so that the issue of transitivity is not considered.
[18] Another way of viewing GEKS indexes is as follows: pick any one of the time periods in the sample, call it the base period and construct Fisher indexes comparing all other time periods with this fixed base period.  This generates one set of price indexes for all periods.  Now carry out this procedure for all possible choices of the base period and take the geometric average of the resulting price indexes.  If all items were in all outlets for all periods, it is very easy to construct GEKS indexes.  But this is not the case and so the difficulty in constructing GEKS indexes for the flexible basket case is to match the items that are present in the outlet for every pair of periods.
[19] See Diewert (1976) for the details.

GEKS method is exact for these (flexible functional form) preferences under optimizing behavior assumptions.

GEKS indexes were calculated using information from two product categories (soft drinks and butter) and the following four aggregation methods:

1. quarterly time aggregation, with item aggregation over stores;
2. quarterly time aggregation, with no item aggregation over stores;
3. monthly time aggregation, with item aggregation over stores; and
4. monthly time aggregation, with no item aggregation over stores.

These aggregation methods are consistent with those used to estimate the price indexes presented in Section 4.

GEKS and chained indexes were calculated between periods 1-2, 1-3, 1-4 and 1-5, and the results are shown in figures 1 to 4 for data aggregated at quarterly intervals and 5 to 8 for data aggregated at monthly intervals.

The plots indicate that for the product categories toilet paper and butter, there is very little chain drift occurring in quarterly indexes; differences between GEKS and chained indexes are small, ranging from approximately 0.1% to 0.3%. Interestingly, ‘dips’ in the chained price indexes such as those seen for toilet paper in quarter 4 are mirrored by the GEKS index, indicating that movements in the chained index  reflect actual movements in prices over the period rather than drift. For monthly data the chain drift is somewhat more apparent, particularly for the item category butter; we can see the results of chain drift from about month 11 onwards in figures 7 and 8. Differences between the GEKS and chained indexes for price indexes constructed with and without aggregation over stores are approximately 1.11% and 1.56% respectively by the end of the 15 month period. The GEKS index appears to mirror the movements in the chained index for toilet paper quite closely, particularly for the series constructed with aggregation over stores. When we look at results for indexes constructed with no store aggregation there seems to be a fairly constant level difference in estimates of price change, which at the end of the series is approximately 1.61%; it is only for a few periods that there is any significant difference between the indexes. The nature of the difference between the GEKS and chained indexes numbers (i.e.

relatively constant over time) indicates that this difference is not due to drift. Similarly, the monthly price change series for toilet paper does not seem to suffer from chain drift (figure 5).

Overall, our results indicate that indexes constructed using scanner data and quarterly time aggregation appear to be relatively unaffected by chain drift. Quarterly aggregation over time appears to be able to sufficiently smooth out the price and corresponding quantity bouncing behaviour that is captured in scanner data and leads to chain drift. However, many statistical agencies produce monthly indexes and as such, quarterly aggregation may not be practical. The results for monthly chained indexes are mixed. For the item category butter, chain drift seems to be present whereas this is does not seem to be the case for the item category toilet paper. Thus at this stage, our results indicate that for some items, statistical agencies with access to scanner data could safely compute monthly chained superlative indexes that would be free of chain drift. However, our results indicate that this strategy would not work for some items. In the latter case, we recommend the use of a *rolling year GEKS approach*; i.e., the GEKS approach that we implemented in this paper for butter and toilet paper could be computed using data for the last 13 months and the resulting measure of price change going from month 12 to 13 could be used in a real time CPI. As the scanner data for the following month became available to the statistical agency, the data pertaining to the earliest month would be dropped from the rolling year and a new set of rolling year GEKS indexes could be computed and the resulting measure of price change going from month 13 to 14 could be used in the CPI. We believe that these suggestions may be of some use to Statistics Netherlands as they are planning to use scanner data from retail chains in their CPI in the near future.

A potential drawback of the GEKS method is that there are no standard errors on our index series. The use of an alternative approach, the 'Country Product Dummy' method,[20] again borrowed from the international comparisons literature with appropriate adaptation, could provide indexes free of chain drift and standard errors on the indexes. This is left for future research.

[20] See Summers (1973).

## 6. Conclusion

One of the key results of this work has been to show that, when using high frequency data, decisions about how to aggregate and whether or not chaining is used can have a huge impact on estimates of price change. It is known that when price bouncing is present the use of chained indexes in combination with non-superlative indexes tend to exhibit drift. However, the extent of drift seen for many item categories over what is a relatively short time period is, to say the least, surprising. In addition, it is also of concern to see that indexes which we would typically consider to be much more stable, such as direct indexes and superlative indexes, show a troubling degree of volatility. These results indicate that traditional index number theory appears to break down when high frequency data are used.

This suggests that using unit values defined over months or quarters are preferable to unit values defined over weeks. Whether or not items are aggregated over stores in constructing the unit values appears to be a relatively minor consideration compared to the choices of time aggregation, index number formula, and fixed or flexible basket.

An additional contribution of the paper is the proposition of a novel method for estimating the degree of chain bias in the chained indexes. The results for two commodity classes (toilet paper and butter) suggest that chain drift is not a major problem using monthly and quarterly unit values. Whether or not this result generalizes to other commodities is left for further research, along with the examination of alternative methods for estimating chain drift.

**Table 1. Data: Descriptive statistics**

| Item Category | Observations | Number of brands |
|---|---|---|
| Biscuits | 2,452,797 | 1,322 |
| Bread | 119,565 | 427 |
| Butter | 225,789 | 78 |
| Cereal | 1,147,737 | 548 |
| Coffee | 514,945 | 149 |
| Detergent | 458,712 | 177 |
| Frozen peas | 544,050 | 227 |
| Honey | 235,649 | 113 |
| Jams | 615,948 | 389 |
| Juices | 2,639,642 | 1,125 |
| Margarine | 312,558 | 98 |
| Oil | 483,146 | 314 |
| Pasta | 1,065,204 | 706 |
| Pet food | 2,589,135 | 1,062 |
| Soft drinks | 2,140,587 | 964 |
| Spreads | 283,676 | 102 |
| Sugar | 254,453 | 114 |
| Tin tomatoes | 246,187 | 164 |
| Toilet paper | 438,525 | 128 |

**Table 2. Laspyeres Index: price change estimates – item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly |
| Biscuits | 98.89 | 100.74 | 101.94 | 96.21 | 109.04 | 185.77 | 98.50 | 101.66 | 166.95 |
| Bread | 104.33 | 106.69 | 108.87 | 104.88 | 114.05 | 562.24 | 104.91 | 113.76 | 615.50 |
| Butter | 100.95 | 102.91 | 100.11 | 101.91 | 106.85 | 145.14 | 101.50 | 107.48 | 145.60 |
| Cereal | 100.27 | 102.00 | 104.02 | 100.65 | 107.45 | 215.57 | 100.94 | 107.01 | 210.04 |
| Coffee | 111.14 | 112.38 | 115.70 | 111.49 | 126.21 | 274.76 | 111.57 | 125.72 | 267.83 |
| Detergent | 102.71 | 105.71 | 105.25 | 102.64 | 112.31 | 165.05 | 103.09 | 111.54 | 164.11 |
| Frozen peas | 100.78 | 100.73 | 101.75 | 100.94 | 108.25 | 202.12 | 101.28 | 107.24 | 195.92 |
| Honey | 104.77 | 105.93 | 105.52 | 104.42 | 108.14 | 120.40 | 104.87 | 107.27 | 119.30 |
| Jams | 100.49 | 101.52 | 102.08 | 100.09 | 107.29 | 174.15 | 100.99 | 105.47 | 167.01 |
| Juices | 101.74 | 101.77 | 104.21 | 101.90 | 110.82 | 332.11 | 102.69 | 109.65 | 318.52 |
| Margarine | 104.29 | 102.80 | 104.10 | 106.81 | 124.53 | 1606.77 | 106.86 | 124.86 | 1562.35 |
| Oil | 92.93 | 90.87 | 87.37 | 92.82 | 100.48 | 141.16 | 93.48 | 100.05 | 142.56 |
| Pasta | 100.88 | 101.16 | 104.88 | 100.30 | 110.46 | 347.14 | 101.22 | 109.38 | 342.19 |
| Pet food | 100.46 | 101.64 | 103.52 | 100.82 | 106.17 | 165.54 | 101.11 | 105.64 | 161.59 |
| Soft drinks | 104.13 | 106.41 | 108.65 | 105.83 | 132.27 | 1074.89 | 105.95 | 132.21 | 1024.45 |
| Spreads | 104.86 | 107.88 | 107.14 | 104.70 | 111.163 | 122.84 | 104.98 | 110.64 | 121.94 |
| Sugar | 106.37 | 107.20 | 106.71 | 106.09 | 111.39 | 149.44 | 106.07 | 111.43 | 149.47 |
| Tin tomatoes | 101.33 | 98.93 | 101.68 | 101.142 | 110.51 | 165.82 | 101.95 | 109.42 | 164.62 |
| Toilet paper | 100.61 | 99.62 | 100.46 | 103.67 | 125.71 | 1656.92 | 103.99 | 124.69 | 1571.90 |

**Table 3. Paasche Index: price change estimates – item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Quarterly** | **Monthly** | **Weekly** | **Quarterly** | **Monthly** | **Weekly** | **Quarterly** | **Monthly** | **Weekly** |
| **Biscuits** | 98.44 | 99.68 | 99.71 | 97.24 | 91.93 | 48.12 | 96.38 | 88.75 | 45.28 |
| **Bread** | 102.83 | 102.89 | 101.66 | 102.79 | 97.14 | 19.33 | 102.35 | 94.48 | 16.91 |
| **Butter** | 100.30 | 101.25 | 99.07 | 99.84 | 97.74 | 66.45 | 99.98 | 97.85 | 66.50 |
| **Cereal** | 100.23 | 100.73 | 102.64 | 99.33 | 94.98 | 43.82 | 99.12 | 94.47 | 43.87 |
| **Coffee** | 109.30 | 110.04 | 111.23 | 108.71 | 98.70 | 35.00 | 108.62 | 98.43 | 35.57 |
| **Detergent** | 102.39 | 104.67 | 103.82 | 101.89 | 97.83 | 61.16 | 101.52 | 96.68 | 59.82 |
| **Frozen peas** | 100.33 | 100.32 | 100.21 | 100.11 | 93.65 | 44.32 | 99.86 | 92.77 | 44.94 |
| **Honey** | 104.37 | 105.30 | 104.52 | 104.12 | 102.82 | 89.54 | 103.84 | 102.42 | 89.13 |
| **Jams** | 100.39 | 100.73 | 98.18 | 99.67 | 95.49 | 46.62 | 99.04 | 94.24 | 46.37 |
| **Juices** | 100.69 | 99.43 | 98.65 | 100.15 | 91.77 | 27.29 | 99.12 | 90.23 | 27.37 |
| **Margarine** | 103.14 | 97.96 | 102.39 | 101.37 | 80.57 | 5.52 | 100.72 | 80.31 | 5.59 |
| **Oil** | 91.05 | 87.72 | 83.21 | 90.07 | 75.76 | 42.41 | 88.93 | 74.02 | 39.83 |
| **Pasta** | 100.37 | 100.63 | 100.92 | 99.78 | 92.05 | 25.75 | 99.25 | 90.25 | 24.17 |
| **Pet food** | 100.56 | 99.88 | 101.84 | 99.92 | 95.35 | 59.10 | 99.65 | 94.73 | 59.74 |
| **Soft drinks** | 102.77 | 102.31 | 103.32 | 101.33 | 80.19 | 6.06 | 101.01 | 79.36 | 6.22 |
| **Spreads** | 103.91 | 105.87 | 105.57 | 103.81 | 103.123 | 88.23 | 103.73 | 102.85 | 87.85 |
| **Sugar** | 106.14 | 106.99 | 106.23 | 105.93 | 101.23 | 66.06 | 105.97 | 101.25 | 66.06 |
| **Tin tomatoes** | 101.32 | 98.16 | 98.73 | 100.46 | 89.45 | 53.31 | 99.5892 | 88.64 | 51.96 |
| **Toilet paper** | 99.32 | 96.58 | 87.06 | 96.61 | 76.65 | 3.68 | 96.70 | 76.67 | 3.82 |

**Table 4. Fisher Index: price change estimates – item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly |
| Biscuits | 98.66 | 100.21 | 100.82 | 97.87 | 100.12 | 94.55 | 96.29 | 94.99 | 86.95 |
| Bread | 103.58 | 104.77 | 105.20 | 103.85 | 105.25 | 104.25 | 103.61 | 103.67 | 102.03 |
| Butter | 100.62 | 102.08 | 99.59 | 100.67 | 102.19 | 98.20 | 100.94 | 102.56 | 98.40 |
| Cereal | 100.25 | 101.37 | 103.33 | 100.13 | 101.02 | 97.19 | 99.88 | 100.54 | 95.99 |
| Coffee | 110.22 | 111.20 | 113.44 | 110.13 | 111.61 | 98.07 | 110.05 | 111.24 | 97.61 |
| Detergent | 102.55 | 105.19 | 104.53 | 102.49 | 104.82 | 100.48 | 102.08 | 103.84 | 99.08 |
| Frozen peas | 100.55 | 100.52 | 100.98 | 100.70 | 100.68 | 94.64 | 100.40 | 99.74 | 93.83 |
| Honey | 104.57 | 105.61 | 105.02 | 104.49 | 105.45 | 103.83 | 104.13 | 104.81 | 103.12 |
| Jams | 100.44 | 101.12 | 100.11 | 100.33 | 101.22 | 90.10 | 99.56 | 99.69 | 88.00 |
| Juices | 101.21 | 100.59 | 101.39 | 101.41 | 100.84 | 95.21 | 100.50 | 99.47 | 93.37 |
| Margarine | 103.72 | 100.35 | 103.24 | 104.08 | 100.16 | 94.17 | 103.72 | 100.14 | 93.44 |
| Oil | 91.99 | 89.28 | 85.26 | 91.76 | 87.25 | 77.37 | 90.86 | 86.05 | 75.35 |
| Pasta | 100.62 | 100.90 | 102.88 | 100.50 | 100.84 | 94.55 | 99.77 | 99.36 | 90.95 |
| Pet food | 100.51 | 100.76 | 102.68 | 100.51 | 100.61 | 98.91 | 100.23 | 100.04 | 98.25 |
| Soft drinks | 103.45 | 104.34 | 105.95 | 103.62 | 102.99 | 80.70 | 103.39 | 102.43 | 79.80 |
| Spreads | 104.39 | 106.87 | 106.35 | 104.39 | 107.07 | 104.11 | 104.22 | 106.67 | 103.50 |
| Sugar | 106.26 | 107.10 | 106.47 | 106.00 | 106.19 | 99.36 | 106.03 | 106.22 | 99.36 |
| Tin tomatoes | 101.32 | 98.55 | 100.20 | 101.20 | 99.43 | 94.02 | 100.363 | 98.48 | 92.49 |
| Toilet paper | 99.96 | 98.09 | 93.52 | 100.23 | 98.16 | 78.13 | 100.13 | 97.77 | 77.51 |

**Table 5. Laspeyres Index: price change estimates – no item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Quarterly** | **Monthly** | **Weekly** | **Quarterly** | **Monthly** | **Weekly** | **Quarterly** | **Monthly** | **Weekly** |
| **Biscuits** | 99.77 | 102.11 | 102.99 | 101.60 | 121.16 | 318.33 | 100.65 | 116.05 | 281.30 |
| **Bread** | 104.81 | 108.10 | 112.48 | 106.18 | 125.77 | 3146.25 | 106.16 | 126.05 | 2815.28 |
| **Butter** | 101.26 | 103.22 | 100.78 | 102.59 | 113.99 | 193.00 | 102.80 | 114.15 | 193.21 |
| **Cereal** | 100.77 | 103.56 | 104.53 | 102.54 | 123.24 | 361.49 | 102.36 | 122.85 | 354.71 |
| **Coffee** | 111.97 | 114.25 | 116.98 | 113.70 | 155.80 | 543.34 | 113.72 | 154.65 | 511.04 |
| **Detergent** | 103.27 | 106.61 | 105.69 | 104.15 | 125.14 | 227.96 | 103.50 | 123.70 | 228.01 |
| **Frozen peas** | 101.27 | 101.51 | 102.88 | 102.35 | 119.17 | 300.51 | 101.92 | 117.13 | 273.91 |
| **Honey** | 104.87 | 105.97 | 105.85 | 105.32 | 111.22 | 128.45 | 105.05 | 110.65 | 126.76 |
| **Jams** | 101.50 | 103.28 | 105.61 | 102.23 | 118.08 | 294.13 | 101.40 | 114.53 | 257.39 |
| **Juices** | 102.33 | 102.86 | 106.13 | 104.12 | 124.84 | 821.30 | 103.51 | 123.64 | 764.47 |
| **Margarine** | 105.54 | 106.09 | 107.85 | 111.53 | 182.67 | 13897.59 | 111.94 | 187.85 | 14578.97 |
| **Oil** | 93.00 | 91.10 | 88.33 | 94.18 | 103.21 | 132.41 | 94.10 | 104.66 | 155.57 |
| **Pasta** | 101.28 | 102.61 | 108.07 | 102.44 | 122.15 | 790.75 | 101.97 | 123.78 | 788.53 |
| **Pet food** | 101.32 | 102.01 | 104.82 | 102.93 | 114.15 | 263.49 | 102.53 | 113.264 | 241.45 |
| **Soft drinks** | 106.37 | 108.51 | 113.28 | 111.39 | 175.13 | 46575.10 | 111.82 | 175.88 | 28420.37 |
| **Spreads** | 104.77 | 107.67 | 107.49 | 105.72 | 115.39 | 140.14 | 105.51 | 115.43 | 140.69 |
| **Sugar** | 106.97 | 108.44 | 108.51 | 107.43 | 119.64 | 176.18 | 107.20 | 119.17 | 173.62 |
| **Tin tomatoes** | 102.48 | 101.12 | 103.57 | 103.44 | 119.06 | 212.26 | 103.15 | 117.36 | 208.30 |
| **Toilet paper** | 101.49 | 101.24 | 102.66 | 106.71 | 158.29 | 11955.97 | 107.31 | 162.65 | 11815.05 |

**Table 6. Paasche Index: price change estimates – no item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly |
| Biscuits | 98.25 | 98.99 | 99.07 | 96.37 | 84.02 | 23.68 | 95.25 | 80.41 | 22.67 |
| Bread | 102.63 | 101.11 | 98.53 | 102.113 | 88.35 | 3.20 | 101.87 | 86.54 | 3.50 |
| Butter | 100.00 | 100.47 | 98.52 | 98.95 | 91.88 | 48.46 | 98.91 | 91.93 | 48.23 |
| Cereal | 100.04 | 99.96 | 101.92 | 98.39 | 83.71 | 19.75 | 98.04 | 82.69 | 20.11 |
| Coffee | 108.87 | 108.79 | 110.46 | 107.07 | 79.44 | 13.65 | 106.97 | 79.83 | 15.08 |
| Detergent | 102.09 | 104.06 | 102.61 | 101.43 | 87.81 | 37.90 | 100.64 | 86.46 | 37.11 |
| Frozen peas | 100.37 | 99.97 | 99.97 | 99.65 | 86.20 | 26.71 | 99.20 | 85.79 | 29.23 |
| Honey | 104.18 | 104.89 | 104.27 | 103.66 | 99.90 | 81.14 | 103.38 | 99.54 | 80.94 |
| Jams | 100.86 | 101.19 | 97.60 | 100.21 | 89.29 | 23.92 | 98.49 | 86.80 | 25.79 |
| Juices | 100.57 | 98.89 | 97.17 | 99.21 | 82.54 | 10.51 | 98.09 | 80.96 | 10.82 |
| Margarine | 102.17 | 97.28 | 100.06 | 96.92 | 55.60 | 0.45 | 96.73 | 54.99 | 0.43 |
| Oil | 90.92 | 87.89 | 84.03 | 89.68 | 77.50 | 54.02 | 88.65 | 73.65 | 42.06 |
| Pasta | 100.48 | 99.98 | 97.74 | 99.03 | 83.65 | 8.33 | 98.28 | 79.39 | 7.65 |
| Pet food | 100.44 | 99.25 | 100.90 | 98.85 | 88.78 | 35.64 | 98.48 | 88.18 | 37.41 |
| Soft drinks | 101.76 | 100.50 | 101.23 | 97.46 | 59.76 | 0.12 | 96.74 | 59.49 | 0.19 |
| Spreads | 103.82 | 105.47 | 105.11 | 103.49 | 98.77 | 73.13 | 103.27 | 97.86 | 70.60 |
| Sugar | 106.15 | 106.34 | 105.46 | 105.31 | 95.36 | 46.09 | 105.08 | 94.49 | 46.55 |
| Tin tomatoes | 100.93 | 97.31 | 97.46 | 100.18 | 83.08 | 35.65 | 99.53 | 83.09 | 37.28 |
| Toilet paper | 98.26 | 92.66 | 86.90 | 93.89 | 59.74 | 0.48 | 93.98 | 59.78 | 0.54 |

**Table 7. Fisher Index: price change estimates – no item aggregation over stores**

| | Direct | | | Chained (Fixed basket) | | | Chained (Flexible basket) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly | Quarterly | Monthly | Weekly |
| Biscuits | 99.01 | 100.54 | 101.01 | 98.95 | 100.90 | 86.82 | 97.91 | 96.60 | 79.86 |
| Bread | 103.72 | 104.54 | 105.27 | 104.13 | 105.41 | 100.26 | 104.00 | 104.44 | 99.32 |
| Butter | 100.63 | 101.84 | 99.64 | 100.75 | 102.34 | 96.71 | 100.83 | 102.44 | 96.53 |
| Cereal | 100.41 | 101.74 | 103.22 | 100.45 | 101.57 | 84.50 | 100.18 | 100.79 | 84.47 |
| Coffee | 110.41 | 111.49 | 113.67 | 110.34 | 111.25 | 86.13 | 110.30 | 111.11 | 87.79 |
| Detergent | 102.68 | 105.33 | 104.14 | 102.78 | 104.83 | 92.95 | 102.06 | 103.42 | 91.99 |
| Frozen peas | 100.82 | 100.73 | 101.42 | 100.99 | 101.35 | 89.60 | 100.55 | 100.24 | 89.48 |
| Honey | 104.52 | 105.43 | 105.06 | 104.49 | 105.41 | 102.09 | 104.21 | 104.95 | 101.29 |
| Jams | 101.18 | 102.23 | 101.53 | 101.22 | 102.68 | 83.88 | 99.93 | 99.71 | 81.48 |
| Juices | 101.45 | 100.86 | 101.55 | 101.63 | 101.51 | 92.90 | 100.76 | 100.05 | 90.94 |
| Margarine | 103.85 | 101.59 | 103.88 | 103.97 | 100.77 | 79.26 | 104.06 | 101.63 | 79.35 |
| Oil | 91.95 | 89.48 | 86.16 | 91.90 | 89.43 | 84.58 | 91.33 | 87.80 | 80.89 |
| Pasta | 100.88 | 101.28 | 102.78 | 100.72 | 101.08 | 81.18 | 100.11 | 99.13 | 77.68 |
| Pet food | 100.88 | 100.62 | 102.84 | 100.87 | 100.67 | 96.90 | 100.49 | 99.94 | 95.04 |
| Soft drinks | 104.04 | 104.43 | 107.09 | 104.19 | 102.30 | 75.53 | 104.01 | 102.29 | 74.28 |
| Spreads | 104.29 | 106.56 | 106.29 | 104.60 | 106.76 | 101.23 | 104.39 | 106.28 | 99.66 |
| Sugar | 106.56 | 107.38 | 106.97 | 106.36 | 106.81 | 90.11 | 106.14 | 106.12 | 89.90 |
| Tin tomatoes | 101.70 | 99.20 | 100.47 | 101.80 | 99.46 | 86.99 | 101.32 | 98.75 | 88.12 |
| Toilet paper | 99.86 | 96.86 | 94.45 | 100.10 | 97.24 | 75.79 | 100.43 | 98.61 | 79.86 |

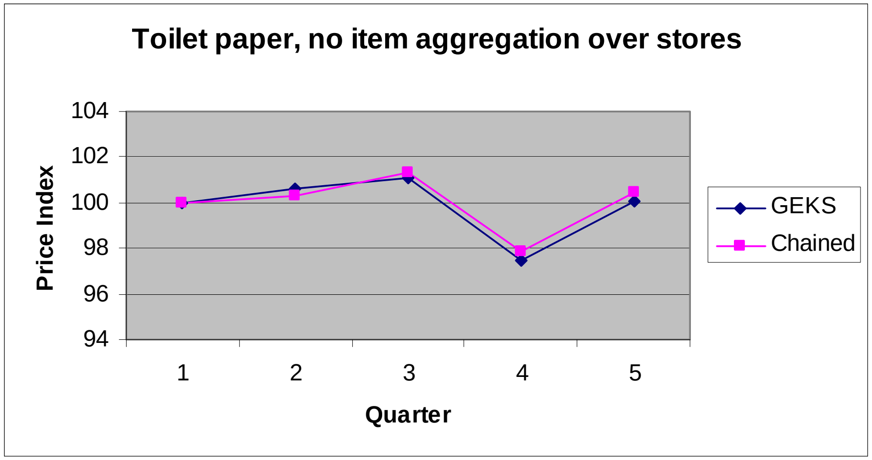


**Figure 1**

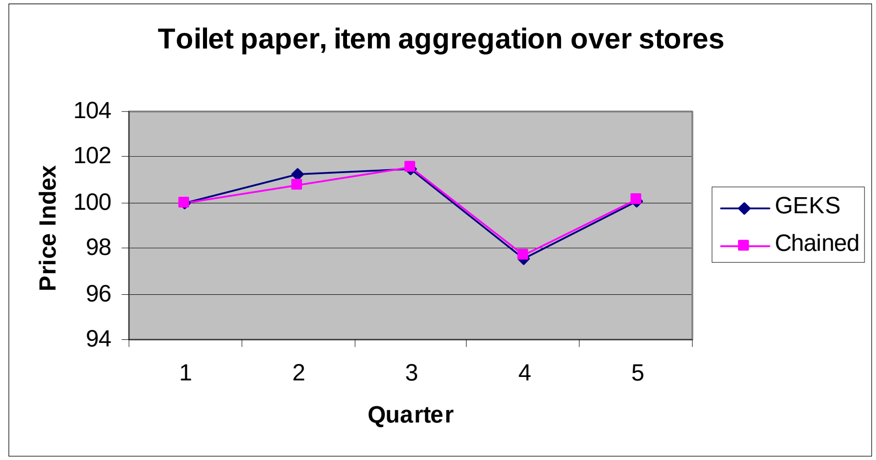


**Figure 2**

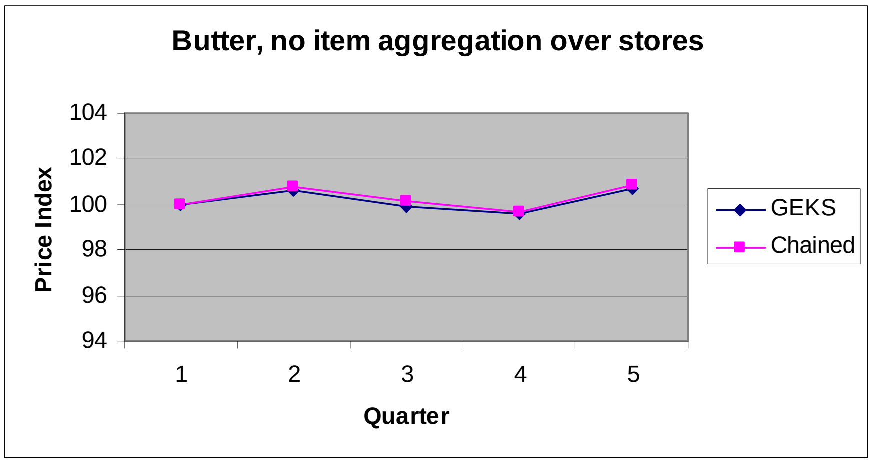


**Figure 3**

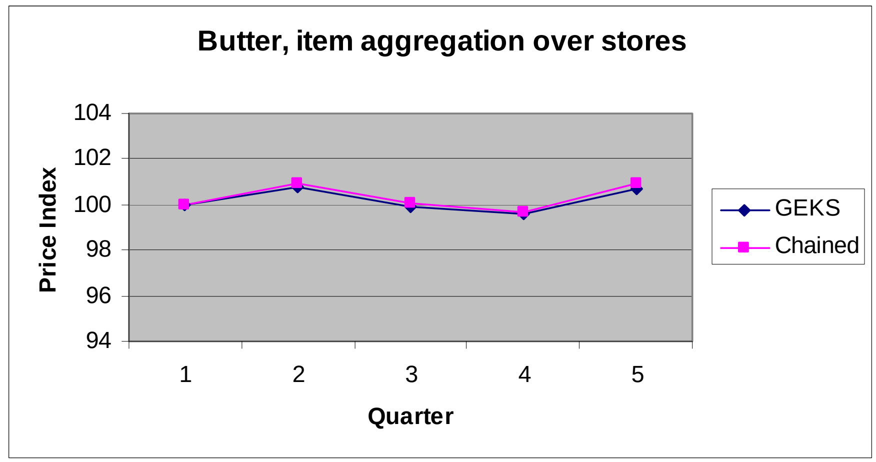


**Figure 4**

Figures (cont.)

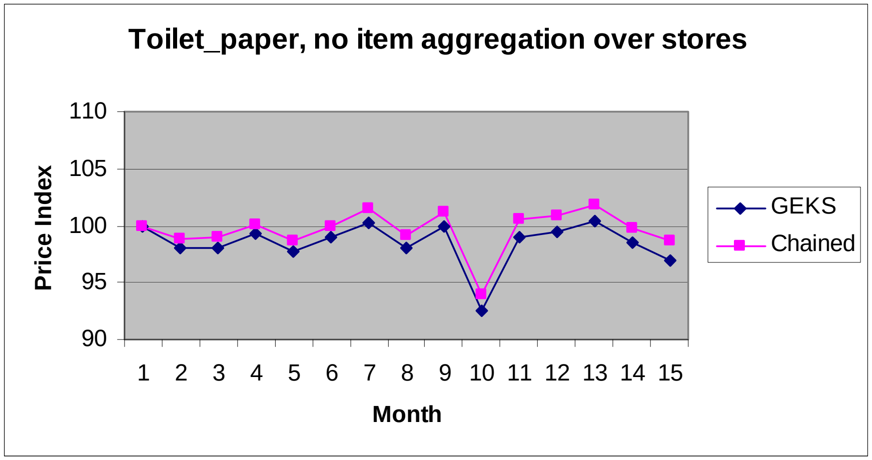


**Figure 5**

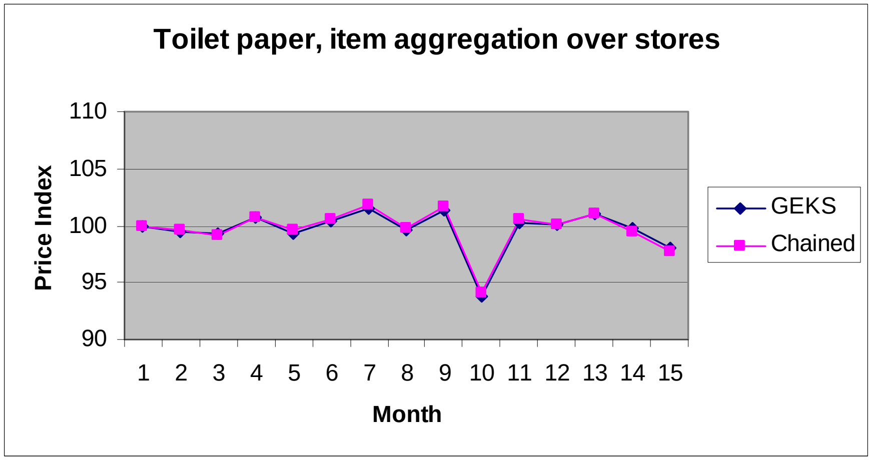


**Figure 6**

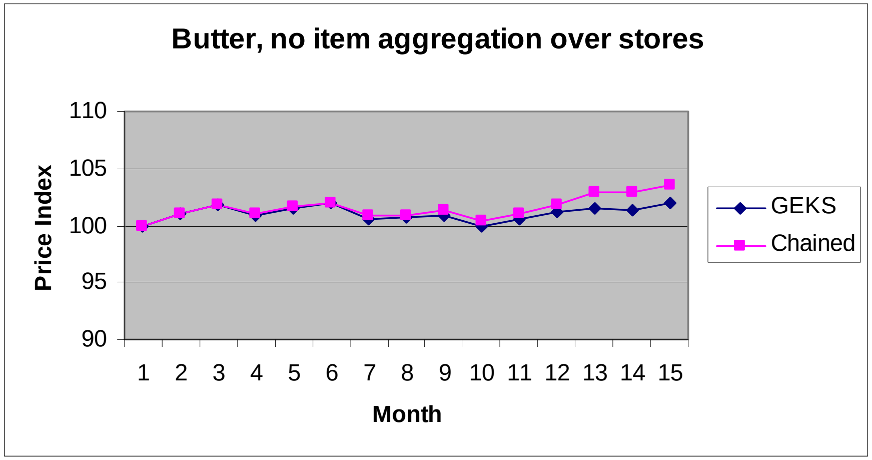


**Figure 7**

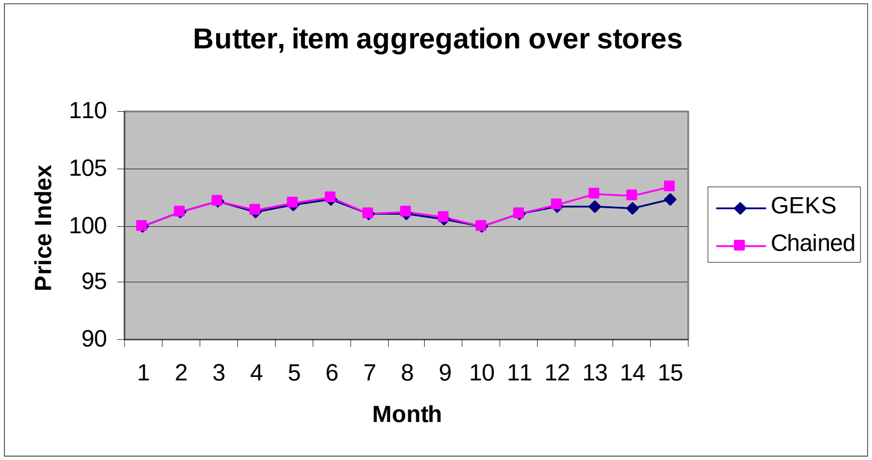


**Figure 8**

## References


Balk, B.M. 1996, 'A Comparison of Ten Methods for Multilateral International Price and Volume Comparisons', *Journal of Official Statistics* 12, pp.199-222.

Balk, B.M. 1998, 'On the use of unit value indices as consumer price subindices', paper presented at the fourth meeting of the Ottawa Group at the International Conference on Price Indices Washington, 22–24 April.

Bradley, R. 2005, 'Pitfalls of using unit value as a price measure or price index', *Journal of Economic and Social Measurement*, vol. 20, pp. 39– 61.

Bradley, R., Cook, B., Leaver, S.G. and Moulton, B.R. 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, 1997, Voorburg, 16–18 April.

Caves D.W., Christensen, L.R. and Diewert, W.E. 1982, 'Multilateral comparisons of output, input, and productivity using superlative index numbers', *Economic Journal* 92, 73-86.

de Haan, J. 2008, 'Reducing Drift in Chained Superlative Price Indexes for Highly Disaggregated Data', paper presented at the Economic Measurement Group Workshop, Sydney, 10-12 December, 2008.

de Haan, J. & Opperdoes, E. 1997, 'Estimation of the coffee price index using scanner data: the choice of the micro index', paper presented at the International Conference on Price Indices, Voorburg, 16-18 April.

Dalen, J. 1997, 'Experiments with Swedish scanner data', paper presented at the International Conference on Price Indices, 1997, Voorburg, 16-18 April.

Diewert, W.E. (1999), 'Axiomatic and Economic Approaches to International Comparisons', pp. 13-87 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Volume 61, Chicago: The University of Chicago Press.

Diewert, W.E, 1995. 'Axiomatic and economic approaches to elementary price indexes', NBER Working Paper No. W5104.

Diewert, W.E, 1993, 'The Early History of Price Index Research', in W.E. Diewert and A.O. Nakamura (eds.): *Essays in Index Number Theory, Vol. 1*, Contributions to Economic Analysis 217, North-Holland, Amsterdam, pp. 33–65.

Diewert, W.E. 1978, 'Superlative index numbers and consistency in aggregation', *Econometrica*, vol. 46, pp. 883 – 900.

Diewert, W.E. (1976), 'Exact and Superlative Index Numbers', *Journal of Econometrics* 4, pp. 114-145.

Eltetö, Ö., and Köves, P. 1964, 'One index computation problem of international comparisons', (in Hungarian), *Statisztikai Szemle* 42, pp. 507-518.

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

Gini, C. 1931, 'On the Circular Test of Index Numbers', *Metron* 9:9, pp. 3-24.

Hawkes, W.J. 1997, 'Reconciliation of consumer price index trends with corresponding trends in average prices for quasi-homogenous goods using scanner data,' paper presented at the International Conference on Price Indices, Voorburg, 16 – 18 April.

Hawkes, W.J. & Piotrowski, F.W. 2003, 'Using scanner data to improve the quality of measurement in the consumer price index', in R Feenstra & M Shapiro (ed.), *Scanner Data and Price Indexes*, University of Chicago Press, Chicago.

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*, International Labour Organization, International Monetary Fund, Organisation for Economic Co-operation and Development, Statistical Office of the European Communities, United Nations, The International Bank for Reconstruction and Development, The World Bank: Geneva.

Jain, M & Abello, R 2001, 'Construction of price indexes and exploration of biases using scanner data', paper presented at the Sixth Meeting of the International Working Group on Price Indices, Canberra, 2–6 April, 2001.

Reinsdorf, M 1999, 'Using scanner data to construct CPI basic component indexes', *Journal of Business and Economic Statistics*, vol.17, no. 2, pp. 152–160.

Summers, R. 1973, 'International Comparisons with Incomplete Data', *Review of Income and Wealth* 29:1, pp. 1-16.

Szulc, B. 1964, 'Indices for multiregional comparisons,' (in Polish), *Przeglad Statystyczny* 3, 239-254.

Szulc, B.J. (Schultz) 1983, 'Linking Price Index Numbers,' in Diewert, W.E. and Montmarquette, C. (Eds.): *Price Level Measurement*, Statistics Canada, Ottawa, pp. 537–566.

Triplett, JE 2003, ‘Using Scanner Data in Consumer Price Indexes: Some Neglected Conceptual Considerations’, in R.C. Feenstra and M.D. Shapiro (eds.): Scanner Data and Price Indexes, NBER Studies in Income and Wealth (Chicago, IL: University of Chicago Press), pp. 151–162.

Walsh, C.M. 1901, *The Measurement of General Exchange Value*, Macmillan and Co., New York.