The use of supermarket scanner data in the Dutch CPI

Heymerik van der Grient and Jan de Haan

THE USE OF SUPERMARKET SCANNER DATA IN THE DUTCH CPI

Abstract: In January 2010 Statistics Netherlands has introduced a new method for computing price indexes based on scanner data in the CPI. At the same time the number of data providers was expanded from one supermarket chain to six chains. This paper describes the new method, including the index number formulae, and the scanner data used. Some results are also presented and compared with index numbers computed using a recently proposed 'benchmark' method.

Keywords: consumer price index, index number theory, scanner data.

1. Introduction

The CPI-Manual (ILO, 2004; 54, 92, 478) notes that "scanner data constitute a rapidly expanding source of data with considerable potential for CPI purposes. …. Scanner data obtained from electronic points of sale include quantities sold and the corresponding value aggregates on a very detailed level. …. Scanner data are up to date and comprehensive."

Some countries in Europe are already using scanner data in the compilation of their CPIs, albeit in different ways. Statistics Norway has been exploiting scanner data since August 2005 to compute the sub-index for food and non-alcoholic beverages (Rodriguez and Haraldsen, 2006). Statistics Netherlands introduced supermarket scanner data into the CPI in June 2002 (Schut, 2002), initially for two chains but after some time only one was left. In Norway and the Netherlands both prices and expenditure weights (for a large sample of items from each product group) are derived from the scanner data. The Swiss Federal Statistical Office follows a more pragmatic approach. Scanner data of some major retail chains are used as an additional source for price collection; prices taken from scanner data simply replace the prices formerly collected in the outlets without changing the underlying principles of computing the price indexes (Becker-Vermeulen, 2006).

In January 2010 Statistics Netherlands has expanded the use of scanner data for the compilation of the Dutch CPI. Seven more supermarket chains have been found willing to co-operate and supply scanner data on a regular basis; the data of five of them were incorporated in January 2010. The six chains for which scanner data are currently utilized have a market share of around 50% and account for slightly more than 5% of the CPI-weight. It is anticipated that

scanner data from the other two supermarket chains will be implemented during 2010 or in 2011 at the latest.

The potential advantage of using scanner data for Statistics Netherlands is twofold. First, the quality of the CPI and HICP can be improved since both prices and quantities are available for all transactions on a very detailed level. Second, cost efficiency will increase. Visiting supermarkets is a major cost component of producing a CPI in the traditional way. A reduction of around 15,000 price quotes each month has been attained for all six supermarket chains together. The use of scanner data is beneficial for the data providers as well. They will no longer be bothered by price collectors who walk around in the stores or ask staff for help. Lowering the response burden for enterprises is a key issue for Statistics Netherlands.

The wider application of scanner data is part of an overall re-design of the Dutch CPI, a project that started at the beginning of this century (De Haan, 2006). Key words in this project are: quality, efficiency, and flexibility. The implementation of new practices began in 2007 when the usual five-yearly updating of the base year and the expenditure weights was replaced by annual updating and chaining. The weights (at the upper level of aggregation) are now taken from the national accounts instead of the household expenditure survey. In 2009 a two-dimensional weighting scheme was implemented. In addition to the well-known COICOP-classification for the goods and services (the first dimension) a classification of branches was introduced (the second dimension). In the traditional sense a branch encompasses all companies or stores that belong to the same type of trade, such as supermarkets, department stores, butchers, hairdressers, etc. A branch can alternatively be defined as a single retail chain for which statistics are calculated. This is the choice made for scanner data in the Dutch CPI.

This paper is structured as follows. The former procedure for treating scanner data from one supermarket chain will be outlined in section 2. This procedure turned out to be very labour intensive and too costly to extend it to six or more retail chains. Statistics Netherlands has therefore chosen to implement a new computation procedure which requires much less manual work. The new method is explained broadly in section 3. Section 4 describes the available scanner data and a number of activities, like data cleaning, that are carried out before computing the index numbers. Section 5 presents the formulae for the new method, followed by some first results in section 6. Section 7 compares the results with those found using a promising method recently developed by a group of academic researchers.

2. Why a new method?

2.1 The former method

In the old method a large sample of items was selected at the beginning of the year which was representative for the previous year. Each item is identified by the European Article Number (EAN) and was given a weight representing its relative importance, i.e. its turnover share within the supermarket chain.[1] The monthly price index for an item was calculated as ratio of the unit value in the current month[2] and the unit value in the base year. Next, elementary price indexes were computed for each 4 digit COICOP group as a weighted average of the constituting items' indexes. That is, during a calendar year the product category price indexes were calculated according to the Laspeyres formula. This was also true for price indexes at higher aggregation levels. At each level of aggregation these short-term indexes were subsequently chained in December to create long-term index series.

During the year the sample of items shrinks. The magnitude obviously differs between product groups, depending on the particular market circumstances. To enhance representativity and keep the sample size fixed, new items were selected as replacements for the 'old' ones. This raised the question whether or not to adjust explicitly for any quality differences. In practice, implicit methods, typically bridged overlap, have mainly been used; only in a limited number of cases quantity adjustments were carried out.

### 2.2 Considerations underlying the new calculation method

One problem with the former method was that similar items were selected as replacements for most disappearing items. While this minimized the need for quality adjustment, it is not good practice since really new items will not be included in the sample, at least not until the next yearly sample revision. But quality change does not seem to be the most important issue in this respect. What may be more important is the lack of representativity. A quick look at scanner data reveals that market dynamics are substantial. Each month many items disappear from the existing set of 'supermarket goods' and many new items are introduced on the market. A fixed item basket – which is what the former method essentially boiled down to – will therefore rapidly loose its representativity. An example of the high attrition rates of goods observed in scanner data is given in figure 1.

---

[1] A fraction of the expenditures could be business or foreign household expenditures rather than domestic household expenditures. We assume that this fraction is negligible.

[2] Based on the expenditures and quantities sold in the first two full weeks of the month.

**Figure 1. Number of matched items; detergents**
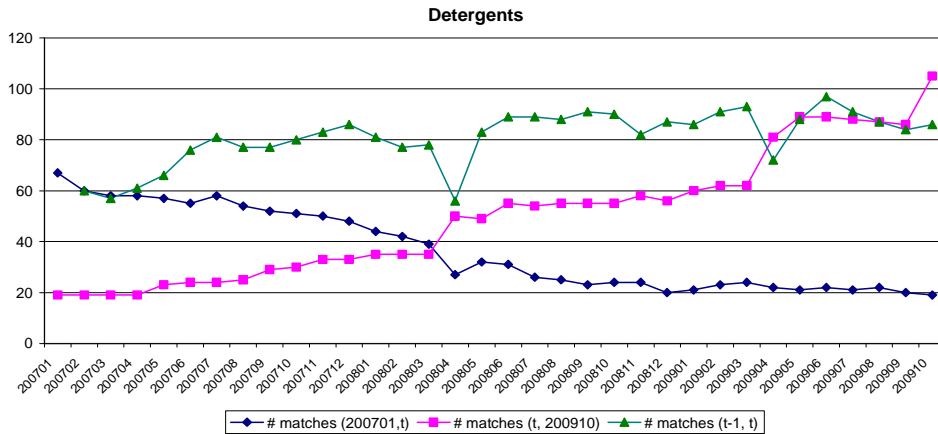
This figure displays the number of matched items for monthly data on detergents in three ways. The downward sloping curve shows how the set of items at the beginning of the period (January 2007) shrinks over time. Only 19 out of the 67 initial items can still be purchased by the end of the period (October 2009). The upward sloping curve should be read in reverse order: it depicts the number of matches between the last month and each earlier month. A comparison with the downward sloping curve indicates that the total number of different types of detergent increases over the years. Apparently there have been more entries than exits. The third curve depicts the number of monthly matched items, i.e. items which are available in adjacent months. In the short run some marked changes occur. For example, it seems as if in April 2008 the supermarket chain refreshed a part of its detergents assortment.

Another, more practical problem is that the annual construction of the item sample and the ongoing selection of replacement items appeared to be a very time-consuming exercise due to the large number of items involved. It laid a heavy burden on the capacity of the CPI department at Statistics Netherlands. Our conclusion was that the application of this method to more than one or two supermarket chains would be impossible given the prevailing time and cost constraints. Dealing in an efficient way with scanner data from several supermarkets requires having a computation method with much less manual interference. This suggests the use of a monthly chained matched-item index at the elementary (product group) level. Such an approach makes the annual construction of a basket of items superfluous and would also actively follow the market dynamics.

As expenditures are available for each item in each month, at first glance it seems quite natural to use a superlative index number formula like the Fisher or the Törnqvist for the computation of price indexes based on scanner data. However, when chaining is applied frequently, such indexes can suffer from

what is known as chain link bias or chain drift (ILO, 2004; 283). This type of drift arises from the fact that prices and quantities 'bounce' as a result of promotional sales; households tend to stock up during sales periods and consume from inventory at times when the goods are not on sale.[3] The quantities sold at times of sales can rise to a hundredfold of those at times when the goods are sold at regular prices. A typical example of price and quantity bouncing is shown in figures 2 and 3 for a certain type of dishwashing tablet sold in a particular supermarket chain. The most striking feature is that purchases made at the regular price (of approximately 6.5 euros) are negligible.

Figure 2. Weekly unit values; dishwashing tablet XYZ
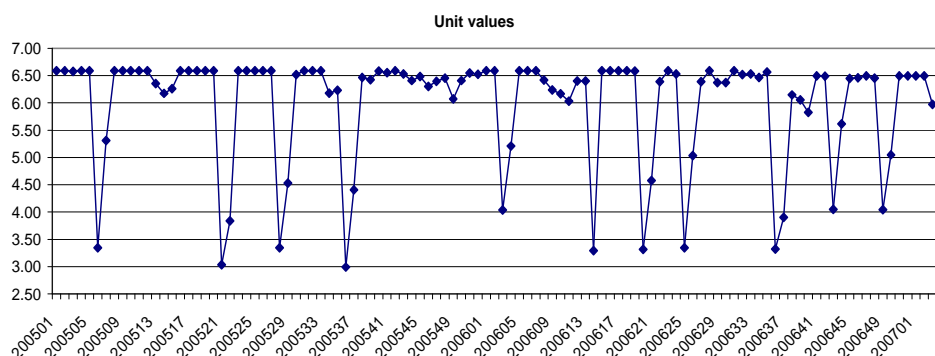
Figure 3. Weekly quantities sold; dishwashing tablet XYZ
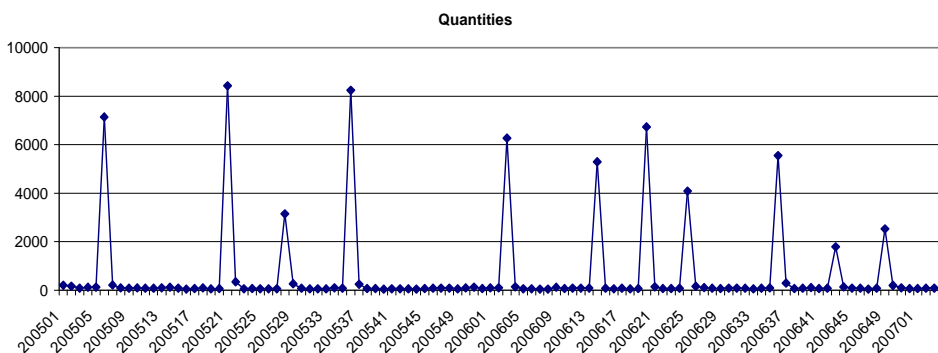
The potential problem of drift in monthly chained weighted indexes has led to the choice for using the unweighted geometric mean or Jevons formula at the elementary level. In sections 3 and 4 we will point out that a number of adjustments had to be made in order to get acceptable results.

---

[3] The underlying problem is an asymmetry in the weights of items that are on sale. Suppose that some items are on sale in month *t*. If their weights in the post-sales month *t*+1 would return to their values in the pre-sales month *t*-1, then a monthly chained superlative index would not be subject to chain drift. For a more detailed description, see De Haan and Van der Grient (2009).

5

3. An overview of the new method

The distribution of expenditures across the items within a product category is usually highly skewed; it quite often happens that 40-50% of the items count for less than 10% of aggregate expenditure. Put differently, a relatively small number of items may be responsible for the majority of expenditures. The use of the (unweighted) Jevons index would ignore this. We decided to introduce a crude type of implicit weighting by applying cut-off sampling: important items of an elementary aggregate are included in the sample with certainty whereas unimportant items are excluded. More specifically, an item will be used in the computation of the index between two consecutive months if its average expenditure share (with respect to the set of matched items) in those months is above a certain threshold value. The threshold was chosen such that roughly 50% of the items in an elementary aggregate will be selected, representing 80-85% of the expenditures.

A drawback of a strict matched-items method is that temporarily unobserved items are excluded from the computation. This means that the price changes of those items occurring between the last month they were in the sample and the month they re-enter the sample would be ignored. Those 'missing prices' are therefore imputed by multiplying the last observed price by the (Jevons) price index of the matched items within the same elementary aggregate, as usual. In a way a panel element is forced onto the dynamic matched-items approach. Price changes that occur after a period with missing prices are now included in the index.

Like any other matched-items method, the new method does not explicitly take quality changes into account. Since implicit quality-adjustment methods have been most prominent in the Dutch CPI in the past, in this respect the new method is similar to the former one. However, the newly-built computer system does allow for making explicit adjustments, just in case. In particular, quantity adjustments for changes in package size or contents could be made when deemed necessary, but we expect this feature to be used infrequently.

The procedure for computing indexes at higher levels of aggregation has not changed. They remain to be calculated as yearly chained Laspeyres indexes, where the previous year serves as the index- and weight reference period for the short-term index series. To make optimal use of the available expenditure information, the elementary aggregates are defined as detailed as possible. To avoid manual recoding of EANs a lower boundary is set by the classification provided by the supermarket. As an elementary aggregate should be robust and contain enough items on a structural basis, regularly these lowest possible product groups have been grouped together. In practice the elementary

level which is still supermarket-specific in most cases, is comparable to the six digit COICOP level.

The expenditures at each level of the COICOP classification are available for each supermarket chain and are used to aggregate across the different chains. This detailed weighting information, which was unavailable prior to the use of scanner data, increases the quality of the product category price indexes: the relative importance of product categories appears to differ significantly between the various chains.[4]

4. Scanner data and data cleaning

4.1 Available scanner data

Every week each supermarket chain sends a data file to Statistics Netherlands containing bar code scanning data on expenditures and quantities sold for all individual items, which are identified by European Article Number (EAN).[5] The chains provide these data either for all individual stores belonging to the chain, for a representative selection of stores or in aggregate form across all individual stores. We believe that the second and third option do not pose a big problem. The assumption that stores belonging to the same supermarket chain provide similar services seems reasonable, so that aggregating across stores of a chain (for each item) can be defended. Most supermarket chains in the Netherlands have a nationwide pricing policy; the prices of most items are the same across all outlets. So even in case of a sample of outlets (the second option) it is unlikely that the unit values would differ much from the 'true' values[6].

Each record of a data file pertains to a particular EAN and contains weekly expenditures, quantities sold, and a (mostly short) product description, often including the weight, contents or package size of the item. Especially for the

---

[4] Although the availability of expenditure data that can be used as weights is a big advantage, at the same time it can lead to inconsistencies with data from other sources used to construct CPI weights (national accounts, household expenditure survey, retail trade statistics, etc.). Adjustments will then be necessary to create a consistent overall weighting scheme.

[5] Around 570 Mb of data from the six supermarkets that are currently included in the CPI are received on a weekly basis. The total number of records amounts to 300 mln per year. Note that Statistics Netherlands does not pay for the scanner data.

[6] In the short run however, sometimes small differences in monthly index changes have been observed. In the long run there is no different index development when using a sample of outlets.

compilation of the constant tax HICP it is necessary to know the amount or percentage of alcohol contained in bottles of beer, wine, etc. This information is not always available in scanner data, and estimates then have to be made.

Every supermarket chain adds its own classification code which indicates to what category an EAN belongs. Having a classification code attached to the data is indispensable for an efficient CPI process given the huge attrition rate of EANs. Once the relation between the chain-specific classification scheme and COICOP has been established, EANs can automatically be assigned to the 4 digit COICOP-category they belong to. To prevent regular recoding, a prerequisite is that the chain-specific classifications should be stable through time. Of course they should also be more detailed than the 4 digit COICOP-scheme.

Supermarket chains may group together all kinds of products that are related to occasions like Christmas, Easter or children's birthdays. Products may also be grouped together which serve a certain aim, for example meat, charcoal and sauces used when organising a barbecue. In these cases the EANs cannot automatically be assigned to one of the COICOP-categories, and we therefore decided to exclude them.

Records identified by the same EAN are considered to refer to the exact same (physical) item. This is true for the big majority of products where the EAN has been assigned by the manufacturer. For some products on the other hand, such as fresh fruit, supermarkets may assign an 'EAN' themselves. A specific group of EANs is available for this purpose; shorthand codes are sometimes used at the checkout. This is not problematic as long as the store would use the same code over time for a specific item. However, identical store-specific EANs are irregularly assigned to different products in different months. The obvious problem is that we would not be comparing like with like over time. Fortunately, this phenomenon does not happen often. The number of cases where the prices of different items would be compared appears to be so small that the impact on the results can be neglected. Moreover, if the resulting (wrongly measured) price changes were substantial, they would be eliminated during the data cleaning procedure or during the monthly routine of checking and analysing the results of the index computation.[7]

Different EANs are treated as different products. In some instances the EAN level could be too detailed for CPI purposes. Items with different EANs, but which are identical from the consumers' point of view, should in principle be treated as the same product. If an item (identified by its EAN) disappears and

---

[7] For those supermarkets where it is known which EANs are used for store- or chain-specific coding, we could decide to exclude those EANs in the future from the index calculation.

a completely comparable item (with a different EAN) appears, then the prices should be directly compared. An example could be a package of coffee which is normally wrapped in red paper but, for promotional reasons, is suddenly wrapped in blue paper. The new method however does not match such items. As mentioned before, the possibility to make explicit adjustments is built into the computer system. So if such situations are known[8] to the CPI statisticians, they can choose to make direct comparisons.

In recent decades supermarkets in the Netherlands have expanded their range of products considerably. Nowadays, their assortment includes clothing, glassware, tableware and household utensils for instance. Statistics Netherlands decided to restrict the calculation of indexes based on scanner data to the more traditional product categories. Table 1 lists the COICOP categories for which scanner data indexes are calculated.

Table 1. COICOP groups for which scanner data indexes are calculated

Code	Description
010000	Food and non-alcoholic beverages
021200	Wine
021300	Beer
055000	Tools and equipment for house and garden
056000	Goods and services for routine household maintenance
061000	Not reimbursable medical and pharmaceutical products
093400	Pets, pet foods and products for pets
131300	Appliances, articles and products for personal care

4.2 Data cleaning and preparation

The Dutch CPI figures for month *t* are published in the first week of month *t*+1. The consequence of this timeliness is that data referring to the last week of month *t* cannot be used. Price collection in the field is therefore restricted to the first three full weeks of each month. A similar procedure is applied for scanner data. That is, the unit value of an item is based on the data pertaining to the first three full weeks of a calendar month from all stores within a chain for which scanner data is received.

The prices (monthly unit values) are subjected to two automatic data cleaning procedures. Firstly, month-to-month price changes of a factor greater than 4 are considered implausible and declared invalid. Thus, items for which the

---

[8] The newly built computer system provides the user with indicators pointing to major changes in the assortment of the supermarket.

current price is 300% higher of 75% lower than the price in the previous month will be deleted.

Secondly, an algorithm, referred to as a dumping filter, has been developed to exclude items from the computation which exhibit a strong price decrease in combination with a strong decrease in expenditures. 'Dumping' occasionally occurs in case of stock clearances when an item is sold at an extraordinary low price. As the item will not be available any longer, it does not return to a regular price. The price decreases – without offsetting price increases – can have an unacceptable downward effect on the index of the product category in question, as an analysis showed. In practice the dumping filter may delete some more items than those related to stock clearances only. This should not be viewed as a serious problem as missing prices are imputed.

5. Formulae for index computation

After the data has been cleaned, price index numbers at various aggregation levels are computed. This section presents the index number formulae used. We start with the elementary level, where unweighted geometric mean price indexes are computed.

The following notation will be used. The price and expenditure share of item i in month m of year y are denoted by $p_i^{y,m}$ and $s_i^{y,m}$, respectively. Let a be a certain elementary aggregate and $N_a^{(y,m-1),(y,m)}$ the corresponding number of matched items between months m and m-1 of year y. To introduce a crude type of weighting, each item i is given a probability $w_i^{y,m}$ to be included in the sample for computing the price change going from month m-1 to month m. These inclusion probabilities or implicit weights are given by:

$$w_i^{y,m} = 1 \qquad \text{if } \frac{s_i^{y,m-1} + s_i^{y,m}}{2} > \frac{1}{N_a^{(y,m-1),(y,m)} \chi};$$

$$w_i^{y,m} = 0 \qquad \text{otherwise.}$$

Thus, if the item's average expenditure share in months m-1 and m exceeds the threshold $1/N_a^{(y,m-1),(y,m)} \chi$ then it will be included in the sample. Notice that the sum of all implicit weights determines the effective sample size, that is, $\sum_{i=1}^{N(y,m-1),(y,m)} w_i^{y,m} = n_a^{(y,m-1),(y,m)}$. Parameter χ can be given any positive value, although in practice there is a lower boundary; if its value is too low, the sample will be empty. Based on simulations we have chosen $\chi = 1.25$ for all product categories; there was no need to differentiate between categories. For example, if $N_a^{(y,m-1),(y,m)} = 80$, then items with an average expenditure share greater than 1% will be selected.

The price change between $y,m-1$[9] and y,m for elementary aggregate a is now computed as

$$\pi_a^{y,m/y,m-1} = \prod_{i=1}^{n_a^{(y,m-1),(y,m)}} \left(\frac{p_i^{y,m}}{p_i^{y,m-1}} \right)^{1/n_a^{(y,m-1),(y,m)}}. \tag{1}$$

Equation (1) is a (sample-based) month-to-month Jevons price index. These month-to-month changes are subsequently multiplied or 'chained' to obtain a long-term time series with some reference or starting month y_0, m_0:

$$P_a^{y,m/y_0,m_0} = P_a^{y,m-1/y_0,m_0} * \pi_a^{y,m/y,m-1}, \tag{2}$$

where $P_a^{y,m-1/y_0 m_0}$ denotes the chained matched-items price index going from the starting month to month $m-1$ of year y.

For items which are not sold in month y,m but which were sold in previous periods, a price is imputed[10]:

$$\hat{p}_i^{y,m} = p_i^{y,m-1} * \pi_a^{y,m/y,m-1}. \tag{3}$$

For higher aggregates A short-term price indexes are calculated according to the Laspeyres formula with index reference period $y-1$:

$$P_A^{y,m/y-1} = \frac{\sum_{a \in A} w_a^{y-1} * P_a^{y,m/y-1}}{\sum_{a \in A} w_a^{y-1}}. \tag{4}$$

The weights w_a^{y-1} in (4) are based on the annual expenditures of all items belonging to elementary aggregate a, regardless whether items were included in the sample or not.[11] Next, the short-term series are chained in December (the link month) to construct long-term series with index reference period 0.[12]

$$P_{ch,A}^{y,m/0} = \left(\frac{P_A^{y,m/y-1}}{P_A^{y-1,12/y-1}} \right) * \left[\prod_{\tau=1}^{y-1} \frac{P_A^{\tau,12/\tau-1}}{P_A^{\tau-1,12/\tau-1}} \right] * P_A^{0,12/0}. \tag{5}$$

Short-term indexes $P_A^{y,m/y-1}$ and chained indexes $P_{ch,A}^{y,m/0}$ are computed at all COICOP-categories. This is done for each retail chain separately and across all chains delivering scanner data, each time using formulae (4) and (5).

---

[9] $y,m-1$ is equal to $y-1,12$ in case $m=1$.

[10] The previous price (in month $y,m-1$) can also be an imputed price.

[11] In the calculation of indexes for seasonal items, such as fresh fruit and fresh vegetables, monthly varying weights are applied.

[12] Currently the Dutch CPI has 2006 as the index reference period.

The final step is to combine, using the same procedure again, the aggregate supermarket indexes at each COICOP-level with the price indexes that have been computed for all other relevant branches, in particular specialised shops such as bakeries and butchers. The latter indexes are still computed from data collected in the field.

6. Some results

The methods used to compile price index numbers differ greatly between the traditional approach based on field surveys and the new approach based on scanner data. The prices traditionally collected in the stores are usually shelf prices, whereas scanner data yields average transaction prices (unit values). The sample size, i.e. the number of items observed, for field surveys is typically very small compared to scanner data. Traditional samples are more or less fixed-size panels, while for scanner data a dynamic matched-items approach is followed. Finally, the index number formulae differ. The ratio of unweighted arithmetic average prices (the Dutot formula) has traditionally been used at the elementary level, whereas the unweighted geometric Jevons index is applied for scanner data.

Because of these differences, it is not surprising that the resulting price index numbers differ. Figure 4 compares the index numbers from both methods for food and non-alcoholic beverages (COICOP-category 010000) and for two specific product categories. These indexes relate to the five supermarkets for which scanner data has been introduced into the Dutch CPI in January 2010. The scanner data index for food and beverages is clearly lower than the index based on prices collected in the stores, although the gap narrows during the second half of 2009. For soups and broths the two indexes show a systematic difference in 2009. For pasta products, on the other hand, the trends are quite similar, but there are some marked differences in the short run.

Figure 4. Price indexes; field survey versus scanner data

Coicop 11930: Soups, broths

Coicop 11150: Pasta products
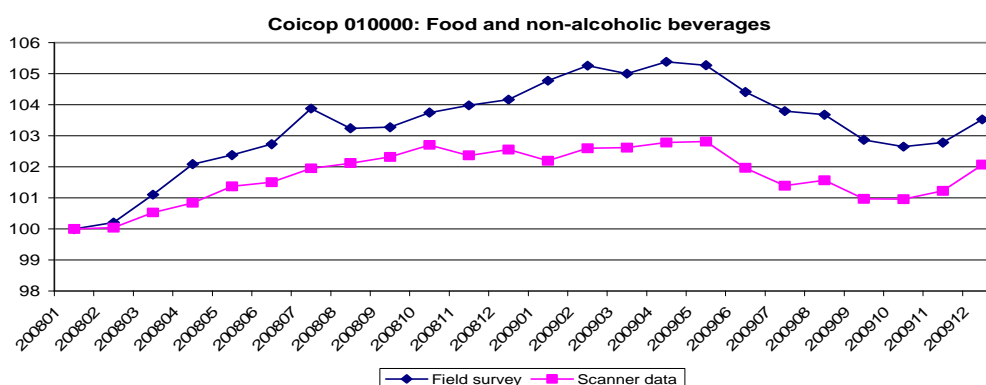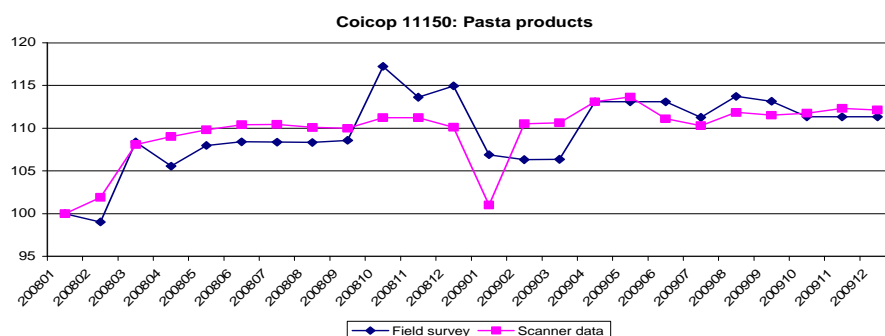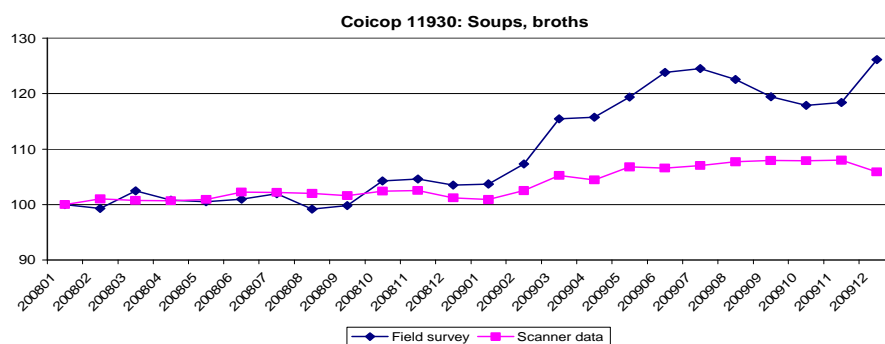
7. Comparison with a benchmark index

A point of criticism that might be raised against Statistics Netherlands' new method for handling supermarket scanner data is the lack of weighting at the elementary level in spite of the fact that expenditure information is available for individual items. As explained earlier, the main reason for this is to avoid chain drift when using monthly chained Fisher or Törnqvist price indexes.[13]

Recently, a group of academic researchers proposed a very promising new approach to construct superlative-type chained price indexes from scanner data which make optimal use of all matches in the data but which are free of chain drift (Ivancic, Diewert and Fox, 2009). Their method is an adaptation of the GEKS-procedure, which is well-known from spatial price comparisons (purchasing power parities), to price comparisons over time. The GEKS-method is 'transitive' by construction, meaning that the chained index equals the direct index, so that it produces a drift free measure of price change. The authors propose a rolling-year version of this method in order to circumvent the problem of revising previously published figures.

---

[13] Schut (2001) mentions that Statistics Netherlands initially intended to introduce a monthly chained Fisher index for scanner data at the elementary level. She also shows some examples of drift in the indexes, which was the reason to implement an annually chained Laspeyres index instead.

In a follow-up paper we applied the rolling-year GEKS approach to a large Dutch scanner data set (De Haan and Van der Grient, 2009), one of the aims being to validate the new method of Statistics Netherlands. Our conclusion was that in general the indexes computed with the new method approximated the rolling-year GEKS price indexes quite well. Different developments were occasionally observed, but these were just temporary. The adjustments made to a strict matched-items Jevons approach – choosing the right value for the parameter (χ) which determines the size of the cut-off sample, imputation of temporarily 'missing' prices, and implementing a dumping filter – appeared to be essential.

Figure 5 compares the price indexes computed according to the old and new method and the rolling-year GEKS method for four product categories. These indexes pertain to the supermarket chain for which scanner data was included in the CPI before January 2010. Using the GEKS indexes as our benchmark measure, the new and old method perform equally well for the category bread and other bakery products. For cereals the old and new indexes deviate from the rolling-year GEKS indexes, albeit in different directions. For syrups and detergents the current method clearly performs better than the former one.

Figure 5. Price indexes; old and new method, and rolling-year GEKS

Coicop 012240: Syrups (Lemonade)

Coicop 56120: Detergents
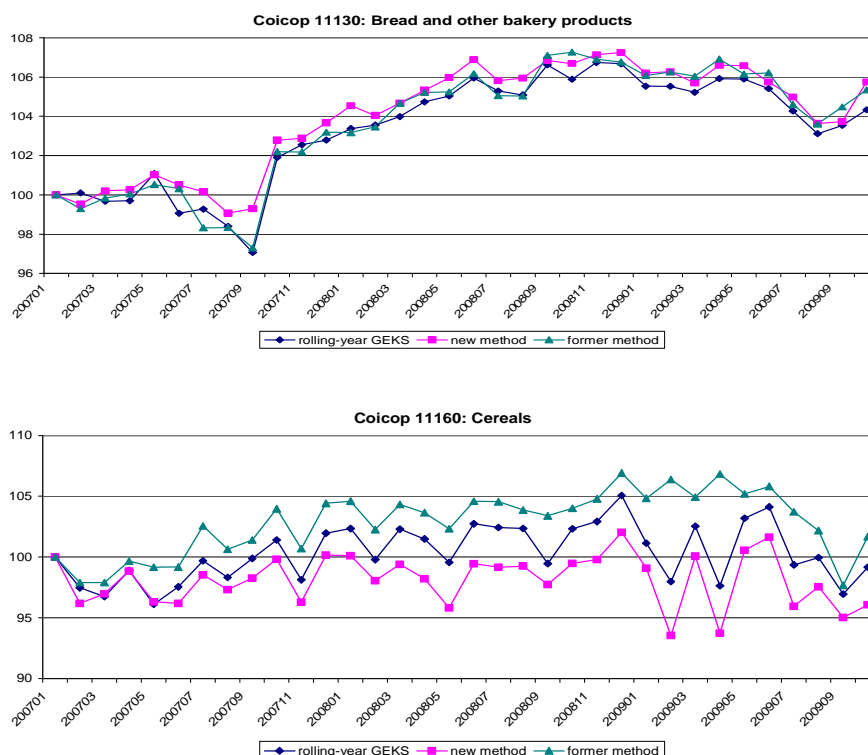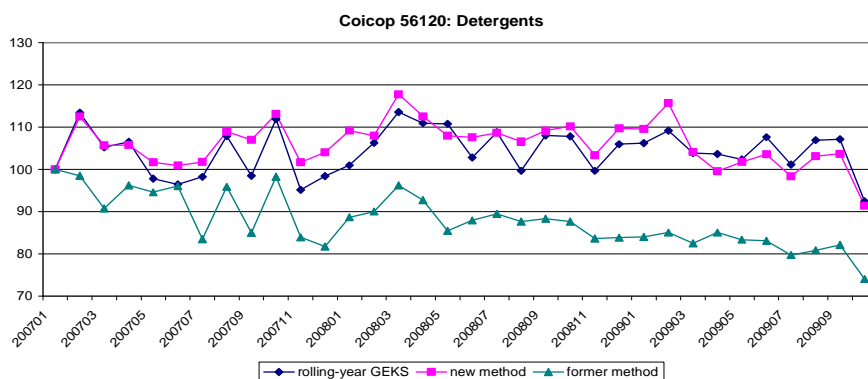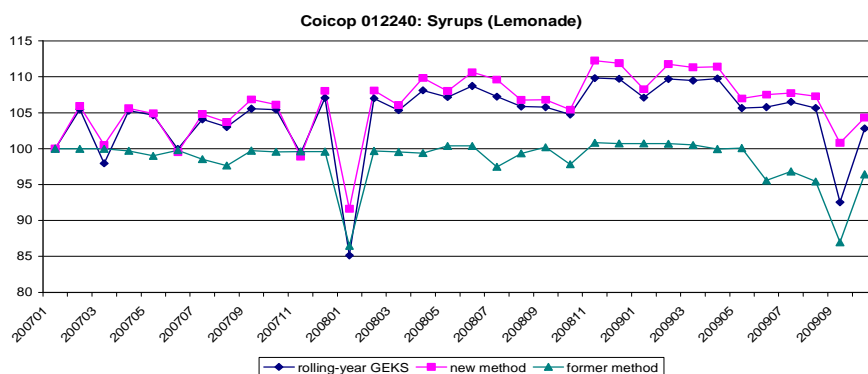
The rolling-year GEKS-approach, while not so easy to explain to users and CPI practitioners, has much to recommend it. Yet, Statistics Netherlands did not implement this approach, amongst other things because of its policy of applying only methods that are widely accepted by the statistical community. When more research on this important topic has been performed by national statistical agencies, Statistics Netherlands might take the decision to change over to the rolling-year GEKS approach.

References

Becker-Vermeulen, C. (2006), Recent developments in the Swiss CPI: scanner data, telecommunications and health price collection, paper presented at the ninth meeting of the Ottawa Group, 14-16 May 2006, London.

Haan, J. de (2006), The re-design of the Dutch CPI, *Statistical Journal of the the United Nations Economic Commission for Europe* 28, 101-118.

Haan, J. de and H.A. van der Grient (2009), Eliminating chain drift in price indexes based on scanner data, paper presented at the eleventh meeting of the Ottawa Group, 27-29 May 2009, Neuchâtel.

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

Ivancic, L., E.W. Diewert and K.J. Fox (2009), Scanner data, time aggregation and the construction of price indexes, paper presented at the eleventh meeting of the Ottawa Group, 27-29 May 2009, Neuchâtel.

Rodriguez, J. and F. Haraldsen (2006), The use of scanner data in the Norwegian CPI: The "new" index for food and non-alcoholic beverages, *Economic Survey* 4, 21-28.

Schut, C. (2001), Using Scanner Data to Compile Price Indices: Experiences and Practical Problems, Paper presented at the Joint ECE/ILO Meeting on Consumer Price Indices, 1-2 November 2001, Geneva.

Schut, C. (ed.) (2002), Gebruik van scannerdata van supermarkten in de consumentenprijsindex, Statistics Netherlands, Voorburg (Dutch only). Available at www.cbs.nl.