Discussion on the Treatment of Discounts in the CPI and the Swedish Experience on the use of Scanner Data

Paper presented at the Workshop on Scanner Data
Stockholm, June 7-8 2012

M. Sammar, A. Norberg and C. Tongur*
Statistics Sweden

Abstract: Statistics Sweden has been using scanner data since the start of 2012 in the price index calculations for daily necessities. In this paper, we present our use of scanner data. We also discuss some issues related to that kind of data, e.g. quality of manually collected data compared to quality of scanner data, how to aggregate weekly observations to monthly and how to treat products that undergo changes. The issue of how to treat discounts in the scanner data is especially discussed.

Based on several studies we conclude that scanner data is a promising source of data. Scanner data offers high quality information of actual transactions, which is the target of price collections.

Key Words: cash register/scanner data, data aggregation, mean value computations, discounts, CPI

* Correspondence: Muhanad.Sammar@scb.se; Anders.Norberg@scb.se; Can.Tongur@scb.se
1. Introduction

In the coming years, more and more statistical offices will be able to study and secure cash register data, also called scanner data. In the Swedish consumer price index, named Konsumentprisindex (KPI), scanner data is used in the index for daily necessities.

In this paper, we will highlight a number of practical issues that we believe are of importance for further work. By addressing the issues at an early stage, we hope to come to an agreement on the direction that the National Statistical Institutes (NSI) should take. One issue is the treatment of different discounts. According to what is written in the current Eurostat regulation concerning price collection, prices with member discounts should not be included in the computation of index unless they are available to a significant majority of consumers.

Another question that we raise is the aggregation of weekly data to monthly data. The treatment of products that undergo changes is also discussed in this paper. An area that we do not cover in this paper but still remains open for further discussions, is the question of editing scanner data.

We start in section 2 to clarify what is meant by the term price. Following that we discuss discounts in scanner data. In section 4, we introduce the sampling of products in the Swedish CPI, how price collection is made and finally the index method for elementary aggregates. In section 5, we present some empirical results that motivated our implementation of scanner data in the Swedish CPI. We further discuss compilation and quality aspects. Section 6 concludes the paper.

2. Purchaser prices

In section 2, we will make an attempt to bring some clarity to the definition of price that should be included in the CPI.

2.1 HICP regulations

The Council of the European Union adopted a regulation on the 23th of October 1995 where the legal basis for the harmonized methodology for compiling consumer price indices (CPIs) in EU Member States was set up. The aim of the Regulation was to establish the statistical bases necessary for arriving at the calculation of comparable indices of consumer prices at Community level for all the EU Member States.

Since the first regulation was established concerning the harmonized indices of consumer prices (HICP), several new regulations have been adopted by both the European Council and the European Commission. Frameworks and rules as well as guidelines or non-obligatory statements of good practice have been drawn up in collaboration with Member States for the construction of HICPs.

Regarding prices in the HICP, it is clear that rebates and other discounts should be deducted, if the opinion of the European Central Bank should be accounted for. In the Commission regulation (EC) No. 2602/2000 one can initially read the following:

---

3 Compendium of HICP - reference documents (2/2001/B/5) - GUIDELINES FOR THE TREATMENT OF REDUCED PRICES IN THE HARMONIZED INDEX OF CONSUMER PRICES (HICP) - p 16-19
There is considerable scope for procedural differences in the treatment of price reductions and the relationship between price reductions and purchase price. A harmonized methodology for the treatment of price reductions in the HICP is necessary to ensure that the resulting HICPs meet the comparability requirement of Article 4 of Regulation (EC) No 2494/95. It will also improve their reliability and relevance.

Then further on one can read:

Prices used in the HICP should be purchaser prices actually paid by households to purchase individual goods and services in monetary transactions, including any taxes less subsidies on the products, after deductions for discounts for bulk or off-peak-purchases from standard prices or charges, and excluding interest or services charges added under credit arrangements and any extra charges incurred as a result of failing to pay within the period stated at the time the purchases were made.\(^4\)

In Article 1 in the Council Regulation (EC) No. 1687/98 written on the topic of price deduction the following can be read:

Prices used in the HICP are the purchase prices paid by households to purchase individual goods and services in monetary transactions. Where goods and services have been available to consumers free of charge, and subsequently an actual price is charged, then the change from a zero price to the actual price, and vice versa, should be taken into account in the HICP.\(^5\)

In Article 2 in the Commission regulation (EC) No. 2602/2000 on purchaser prices this can be found:

Unless otherwise stated purchaser prices used in the HICP shall in general take account of reductions in prices of individual goods and services if such reductions:
(a) can be attributed to the purchase of an individual good or service; and
(b) are available to all potential consumers with no special conditions attached (non-discriminatory);
(c) are known to the purchaser at the time when they enter into the agreement with the seller to purchase the product concerned; and
(d) can be claimed at the time of purchase or within such a time period following the actual purchase that they might be expected to have a significant influence on the quantities purchasers are willing to purchase. In particular, reductions in the prices of individual goods and services which are likely or expected to be available again at standard prices or are available elsewhere at standard prices shall be taken into account in the HICP. Standard price means the price without any conditions or qualifications and not described as a special price.\(^6\)

2.2 Consumer price index manual: Theory and practice

The Governing Body of The International Labour Office (ILO) convened The International Conference of Labour Statisticians (ICLS) for the first time in 1923 with 52 participants from 33 countries and the League of Nations. The purpose of the Conference was to create a natural forum for experts to discuss CPI methodology and develop guidelines for CPI producers. In 1998 ILO in cooperation with International Monetary Fund (IMF), the Organization for Economic Co-operation and Development (OECD), the Statistical Office of the European

---


\(^5\) COUNCIL REGULATION (EC) No 1687/98 of 20 July 1998 amending Commission Regulation (EC) No 1749/96 concerning the coverage of goods and services of the harmonised index of consumer prices (*) - Article 1

Communities (Eurostat), the United Nations Economic Commission for Europe (UNECE) and the World Bank, together with experts from a number of national statistical offices and universities came together and developed a manual for CPI compilation. The purpose of the manual is to give an overview of the conceptual and theoretical issues that statistical offices might need when they tackle various problems in the compilation of a CPI. The International CPI Manual was issued in complete form in 2004 and has continuously been revised.  

The following is said in the manual regarding rebates and discounts.

In paragraph 3.138 in the Consumer price index manual one can read:

“CPIs should take into account the effects of rebates, loyalty schemes, and money-off vouchers. Given that a CPI is meant to cover all the reference households, whether in the country as a whole or in a particular region, discounts should be included even if they are available only to certain households or to consumers satisfying certain payment criteria.”

In paragraph 6.80 one can find:

“One of the principles relating to consumer price indices, which is applied with few exceptions (such as owner-occupier housing costs), is that only transaction prices, that is prices actually paid by individuals or households, should be included in the index. This may differ from the advertised price if, for example, a discount is offered.”

Under the same paragraph the authors continue to write:

“In practice, however, discriminatory discounts, which are available only to a restricted group of households (as opposed to non-discriminatory discounts that are available to all), are generally excluded on principle. For example, money-off coupons and loyalty rewards for previous expenditure are normally ignored and the non-discounted price is recorded.”

How should “are generally excluded on principle” be interpreted in this context?

Under paragraph 6.82 this text can be read:

“Discounts available only to a restricted group of households should be disregarded because they are discriminatory, unless they are significant and are available either to the vast majority of the population or to identifiable subgroups who qualify for such discounts on the basis of demographic or other characteristics not requiring action by the individuals concerned at the time of purchase. In the latter case, they should be treated as stratification or coverage issues in item sampling.”

The lead term here, “unless they are significant” can be a window of opportunity for discussions on how to treat discriminatory discounts which are available to a significantly large, but yet restricted, group of households.

We will in the following section highlight the most common discounts and rebates that can be found in the retail industry. We will also clarify how the discounts are registered in the cash register system in one of the retail chains.

---


3. Discounts

Statistics Sweden would like to emphasize that it is important to set guidelines on how cash register data can and should be treated. In the long run, this might lead to cost savings and quality improvements for many countries. One question however, that ought to be treated very soon, is the handling of discounts in scanner data.

Scanner data may contain various sorts of discounts. For instance, there may be e.g. membership discount and bonus offers. Since it is a new data collection method, it may require newer regulations on how discounts should be considered. We believe that regulations were written with respect to practical conditions for manual collection of prices which implies that only discounts available to all consumers should be measured without services in return. Price collectors can not know the discounts available at the collection moment and for whom they are available.

On the contrary, the prices in scanner data reflect all the transactions made including rebates and other discounts i.e. the actual price paid by the population of households. To avoid a conflict in the understanding of what price should be collected, a review of the policy on how to treat discounts in scanner data is needed.

In the following, we will discuss the different kinds of discounts and at the same time display how they are treated in scanner data for two retail chains.

3.1 Member card discounts

“Membership” offers are for the moment a steadily growing marketing strategy used by retailers to attract consumers. This trend has grown and decentralized to new industries mostly because of a competitive market. By offering households “membership” within their retail chain, the household will gain different sorts of benefits. Some examples of member benefits are premium checks, good membership prices and membership discounts in shops and supermarkets as well as offers on travel, hotels and various entertainments.

Further, two of Sweden’s retail chains within daily necessities do not just offer membership to individuals but also a wide range of financial services, which ties their customers more tightly to them. The banking services are conducted under the supervision of the Financial Supervisory Authority.

Discussions between Statistics Sweden and two retail chains have rendered the following information. Both the chains have millions of households tied to them through membership cards which are used actively, however there is of course a large overlap of households having membership in both chains. Households might also have additional cards through the same chain, e.g. wife holds membership and provides additional card to the husband. According to one of the retail chains, 80% of all their sales, registered in 2011, were made by members. That number has increased in recent years largely because the retail chain has made it more profitable to their members.

If a purchaser is offered a discount through her member card on specific products, the rebate will be subtracted from the regular price of the specific product (similar to a sale). Refunds or other membership benefits relating to the purchase are subtracted from the total costs and are not allocated to specific products listed on the receipt.

---

9 Membership is often attached with an entrance fee.
According to the International CPI Manual membership discounts on specific product should not be included “unless they are significant and are available either to the vast majority of the population or to identifiable subgroups who qualify for such discounts on the basis of demographic or other characteristics not requiring action by the individuals concerned at the time of purchase.” A similar text can be found in the HICP guidelines. Relating this information to the preceding section of CPI compilation in the presence of discounts, we consider 3.2 million members to be a vast majority of the population.

In scanner data
For one of the studied chains, the price variable includes discounts from member cards and for the other chain it does not (due to technical reasons there may be failures of this standard).

3.2 Bonus offers (extras) like "buy 3 pay for 2"
"Buy 3, pay for 2" is a common marketing strategy. The customers are offered to buy three products for the price of two.

From what is written in the International CPI Manual Bonus offers, extras and free gifts should be excluded from the price. According to the HICP guidelines for the treatment of reduced prices, the product concerned should be disregarded if they are not significant. This type of discount is included in the price variable for one of the studied chains and thus cannot be identified in the data. The question is if an exception can be made for the use of scanner data?

Another discount that is treated the same way in the cash register data as "buy 3 pay for 2" is multiple discounts (e.g. a store offers customers discounts for purchasing multiple items of same brand or product).

In scanner data
Prior to a campaign launch, the store personnel types in the EAN-codes of all three products that are included in the campaign into the cash register system. At the same time they register the total amount of the discount for the three products. The total discount for the three products will then be split proportionally over the goods. That is, when a customer buys three products for the price of two, the discount will automatically be registered and distributed over all three products. The same rule is applied when the component products have different regular prices. Note further that the offer does not necessary have to include the same sort of product.

3.3 Coupon discounts, issued by the seller or the supplier
A coupon issued by the seller or the supplier is treated like any other means of payment.

In scanner data
The discount is deducted from the total sale. The price variable in scanner data does not include this sort of discounts on specific product items.

11 Compendium of HICP - reference documents (2/2001/B/5) - GUIDELINES FOR THE TREATMENT OF REDUCED PRICES IN THE HARMONIZED INDEX OF CONSUMER PRICES (HICP) - p 353 Practicability (6)
13 Compendium of HICP - reference documents (2/2001/B/5), GUIDELINES FOR THE TREATMENT OF REDUCED PRICES, IN THE HARMONIZED INDEX OF CONSUMER PRICES (HICP) - p 352 Practicability (5)
3.4 Stamps

A discount similar to the “bonus offers” discounts is stamps. After a purchaser has made \( x \) purchases of a same product in different occasions, she will get the \( x+1 \)th product for free.

**In scanner data**

Uncommon in supermarkets and hypermarkets, but if they existed they would be treated the same way as the “Bonus offers”.

3.5 Sale

Sales prices are most frequent for clothing and footwear but might appear in any product group.

**In scanner data**

Since scanner data are actual transaction data, prices due to sales are included and are thus not in conflict with compilation rules such as paragraph 6.80 in the International CPI Manual.

3.6 “Close to best before date” - rebates

Sometimes products close to best-before-date are sold at reduced prices. Most often the same product with good best-before-date is sold at regular price in parallel.

**In scanner data**

We assume that it is not possible to identify any such transactions but for practical reasons the outlet has the regular price in the cash register system and the reduction is withdrawn manually.

4. Sampling, price collection and elementary aggregates in the Swedish CPI

The following section gives a short introduction to the sampling of outlets, products and product offers in the Swedish CPI. It ends with an introduction to the elementary aggregates of index.

4.1 Sampling of outlets

The Business Register (FDB) at Statistics Sweden is used as a sampling frame for outlets. The data on outlets in the register include industry, number of employees and location (address).

The method for selection of outlets uses stratification with panels and sequential Poisson sampling with selection probabilities proportional to the size, \( \pi_{ps} \), see SCB (2001). Thereby the size measure used is the number of employees plus one. The latter adding of one ensures that the size measure is always non-zero, even when there are no employees, and it may for small shops represent the owner. The sample is drawn by SAMU, a tool for coordinated sampling in business surveys.

The SAMU sampling method is based on Permanent Random Numbers (PRN), which are randomly generated from uniform distribution, in the interval \((0,1)\) and assigned to the outlets. New units, births, are assigned new PRN independently of the already existing numbers. Discontinued units are deleted. The system facilitates the sampling of panels.

For each unit in the sampling frame the ratio between the permanent random number and the size is calculated. The frame is ordered by strata, and within strata by these ratios. The sample is
taken to consist of the first units in each stratum, as many as the planned sample size in the stratum.

The sample is annually renewed for 20% with a method known as RRG, the random rotation group method. Each unit in the sampling frame has not only a PRN but also one of five RRG codes 1-5, randomly set at birth. In year 1 the PRNs for units in the RRG Group 1 are reduced with 0.1, the PRN numbers hereby becoming negative are increased by 1.0 so that they again are in the range (0.1). In year 2, units in the RRG Group 2 are changed in the same way. After five years, all PRN numbers have been reduced by 0.1 or increased to 0.9. The small units that have a probability of selection less than 10% will most probably be sampled at most one five year period, while larger units can be sampled for consecutive years.

4.2 Sampling of products

For about 30 years now, SCB has used probability samples of specific products for daily necessities except for fresh food, such as vegetables, fruits and meat. The advantages of probability sampling in general are well known, as the method has a strong scientific basis. A problematic point, on the other hand, is one that is also present for scanner data. Namely, there is a risk that price changes are hidden in the index if sampled products cannot be replaced in price collection due to strict sampling methodology aspects. Consequently, purposive methods are used to replace products that no longer are available on the market.

The Swedish CPI is using sales data for constructing sampling frames. CPI is provided aggregated retail sales from all outlets of the three major retail groups, annually, based on scanner data. Such data are estimated to be 80% of all goods sold in supermarkets.

The sample is drawn by sequential Pareto πps selection within strata (with no annual rotation), see SCB (2001). Negative coordination of samples between the three groups of outlet chains is accomplished coordinated permanent random numbers that are used. In short, the selection process is as follows:

With the help of the outlet groups’ own commodity classifications and cross-referencing between the three blocks, article records are classified into product strata.

The target inclusion probability is

$$\lambda_{hi} = n_h \cdot \frac{x_{hi}}{\sum_{i \in I} x_{hi}}$$

where \( n_h \) is the desired net sample size in stratum \( h \) and \( s_{hi} \) is the size (turnover) of product \( h_i \), \( i = 1, 2, ..., N_h \), in stratum \( h = 1, 2, ..., L \). If \( \lambda_i \) is greater than one the article is selected with certainty. A ranking variable is computed as

$$Q_{hi} = \frac{R_{hi} \cdot (1 - \lambda_{hi})}{\lambda_{hi} \cdot (1 - R_{hi})}$$

where \( R_{hi} \) for article \( i = 1, 2, ..., N_h \) and stratum \( b = 1, 2, ..., L \), is a permanent random number drawn from uniform distribution on (0,1). The records are sorted by stratum \( h \) and the ranking variable \( Q_{hi} \). For each stratum the first \( n_h \) articles considered to be available for price collection are selected for the sample.
Three different product samples of approximately 500 products each are created. The three samples are negatively coordinated, i.e., they have minimal overlap. The product samples are then matched to the outlet sample. Only product-offers that are available in the sampled outlet in December (base period) and/or January are included in the target sample.

4.3 Price collection

Statistics Sweden decided to use scanner data from one retail chain in price index calculations of the daily necessities in CPI as from January 2012. The decision was based on several studies, see Norberg et al. (2011) and it seems not to have resulted in any breaks in the time series but rather indicates higher statistical quality. As a preventive measure however, Statistics Sweden decided to do parallel manual price collection besides the scanner data collection during December 2011 and January 2012 for the specific retail chain whose data was used. This dual collection allowed one final cross checking of data, similar to the preceding thorough analyses of scanner data for the years 2009 and 2010. Comparisons were made of the January 2012 index for daily necessities in the two collection modes and there were no apparent differences found between the two modes of data, neither on micro nor aggregated levels, i.e. index, for the 88 product groups containing daily necessities. Due to this final cross checking, scanner data was accepted as the data collection mode as from year 2012.

4.4 Elementary aggregates: Jevons index

At the lowest level of aggregation in index computation, elementary indices are computed for combinations of product group and industry (of outlet). For daily necessities there are about 80 product groups and 2 industries. As in many counties, aggregating prices to indices at the lowest level is made by the Jevons index formula. The latter can be expressed mathematically as follows, disregarding weights that are in some cases available and used:

\[
I_t = \left( \prod_i \frac{p_{it}}{p_{i0}} \right)^{1/n} = \left( \prod_i \frac{p_{it}}{p_{i0}} \right)^{1/n} \frac{\prod_i p_{i0}}{\prod_i p_{i0}}
\]

Index in the current month \( t \) is the ratio of the observed prices \( p_i \) for all product-offers in the current month \( t \) and the observed prices for all varieties in December of the previous year, reference month \( 0 \) for those product-offers that exist in both periods.

4.5 Four ways to use scanner data

There are several ways to use scanner data in the CPI; the following four might be considered the most realistic:

A. **Replace the manually collected price data with scanner data for the ordinary sample of outlets and products.**
   The computing of indexes will be equivalent to the computing methods applied within the current production system of the national CPI, i.e., the Jevons index for elementary aggregation. The scanner data’s full potential is not utilized in such methodical use, although the sample of retail outlets and products can be made much larger than it is at
present time. The standard error of estimate, which is large, can be decreased at low cost if scanner data is delivered for free. This reflects Statistics Sweden’s (SCB) approach.

B. **Use scanner data as auxiliary information.**
   Use scanner data for large samples or total registers as auxiliary information to decrease the standard error of estimated price change from a relatively small sample of manually collected data. This can be divided in a three-step procedure: (1) computing a price index by using scanner data; (2) collect prices manually in a small sample of retail outlets with high quality measurement methods; and (3) adjust the scanner data price index by the average ratio of manually collected prices and scanner data prices.

C. **Compute index from a census based on all products for which scanner data are available.**
   SCB thinks that there might be some problems using this method. SCB is not convinced that NSI officers have enough knowledge to classify over ten thousand products into COICOP-groups with sufficient quality. Another obstacle is that bottle deposits for water, soft drinks and beer are not withdrawn from the price, i.e., a change of deposit cost imputes motion on the index, which is inconsistent with the regulations. Finally substitutes cannot be handled automatically, e.g., when the package size is altered (for example the number of napkins decrease in a package), the EAN-code is altered and the price might remain unchanged then the implicit price increase will not be calculated but linked to show no change if not processed manually.

D. **Use scanner data for auditing and quality control.**
   NSI can use scanner data for review of manually collected prices. Measurement errors in manual price collection exists, the frequencies of which is a function of education, instruction manuals, measurement device, auditing etc. Unfortunately we have little empirical data on these errors.

5. **Empirical studies on the use of scanner data**
   In the following, we briefly mention some, but far from all, empirical findings that we have encountered within this field.

5.1 Sources of error with manual price collections

In year 2010, the Swedish Consumer Agency published a report (Konsumentverket, 2010) with the scope of reviewing the price information in Swedish supermarkets. A total of 13 500 product offers were examined in 291 stores. The research was conducted in late summer 2009 with the help of consumer advisors in 35 municipalities across the country. Here follows some of the results from the study:

- For 9% of the items in the survey, the prices were hard to find or could not be found at all. The lack of price information was larger in smaller shops.
- For 6% of the examined products, the prices on the shelves and packages were different from the purchase prices.

Another source of error is the manual collection as such. Human deficiency should be considered as a substantial source of error and should not be neglected. One of the advantages with using scanner data is that it eliminates both sources of errors that were mentioned above.
5.2 Comparison of information

Manually collected data and scanner data were compared for the years 2009 and 2010 with respect to the CPI sample in one outlet chain. There are sources of differences that are known in advance and adjusted for on time, such as deposits for beverages. We found that for the subpopulations of data where both manually collected data and scanner data were relevant, about 85% of the prices were equal. In about 5% of the cases, no scanner data price could be found. This is because no package was sold in an outlet in a week. For 7-9% of the cases scanner data prices are not equal to manual collected prices, but there is a symmetry between the two collection methods.

Table 5.1 Scanner Data (S.D.) and Manually Collected Prices (M.C.P.) in comparison.

<table>
<thead>
<tr>
<th>Matching categories in percent.</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neither in M.C.P. or S.D.</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>In M.C.P. but not in S.D.</td>
<td>4.5</td>
<td>5.3</td>
</tr>
<tr>
<td>In S.D. but not in M.C.P.</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>M.C.P. = S.D.</td>
<td>83.4</td>
<td>86.2</td>
</tr>
<tr>
<td>M.C.P. &gt; S.D.</td>
<td>4.3</td>
<td>3.7</td>
</tr>
<tr>
<td>M.C.P. &lt; S.D.</td>
<td>4.8</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Number of comparable product-offers is 36 102 and 38 786 respectively.

5.3 Price variation within weeks and months

An analysis of the first twelve weeks in 2012 for the CPI product sample, which amounts to 100 000 price observations, rendered that more than 98% of observations appear as being weekly prices without variation between days, whereas about 1% of observations seemed to be averages of two or more prices during the week. On monthly basis there were some 35 000 observations of which around 9% showed to have different prices in the three weeks in the month.

5.4 Unweighted geometric mean versus weighted arithmetic mean

Scanner data is collected on weekly basis three times a month, which means that there is a need of aggregating the (maximally) three data points for each combination of month, product and outlet into one observation. This is done by a mean value, and two types of mean values appear as more interesting; unweighted geometric mean and quantity-weighted arithmetic mean. These two mean values may be motivated by the following:

- The geometric mean value is used when price quotas are aggregated to the elementary index per month, product group and outlet stratum. Considering week as another dimension (parallel to outlet) that must be aggregated, this is the obvious choice of method.

- The quantity-weighted arithmetic mean value is the method used by the provider of scanner data to aggregate individual transactions to weekly averages. The reported price per week is the revenue divided by number of packages sold. To use this method to aggregate week to month implies achieving a consistent method to aggregate prices for individual transactions to monthly averages and catches consumption patterns.

- The quantity-weighted arithmetic mean value is also the natural choice from the point of view that CPI shall reflect the development of the cost of households.
There is significant price elasticity in this market, now possible to analyze. The effect of variation in prices appears more frequently in lower weighted arithmetic mean prices than unweighted geometric mean prices.

5.5 The ratio of two mean values

We start by studying basic ratios constructed according to the two methods of interest. Denote the price ratio between the unweighted geometric mean and the weighted arithmetic mean as 

$$R_A^G = \frac{P_t^G}{P_t^A},$$

where $P_t^G$ is the geometric mean value for a product observed in an outlet in the sample in either $t=$January, February or March and $P_t^A$ is the arithmetic analogue. We omit subscripting product and outlet of $R_A^G$. The same analogy goes for the base price ratio $R_b^G = \frac{P_{Base}^G}{P_{Base}^A}$.

We have approximately 8 500 price observations each week, and in the following analysis we used three months of data and the base month, which make four months in all: January – March 2012 with Base in December 2011. In terms of mean values, this makes up about 8 500 price observations each month. Due to potential disclosure issues, we restrict our reporting of the results to the following three clear cases.

- In 2% of cases, the geometric mean price in the base month was significantly larger than the arithmetic mean, i.e. $R_b^G > 1$, while the mean prices the other months were identical, $R_A^G = 1$; $R_b^G > 1$ & $R_A^G = 1$.
- In 3% of cases, the base prices were equal for the two methods ($R_b^G = 1$) while the geometric mean significantly exceeded the arithmetic, $R_A^G > 1$, in any of the other months; $R_b^G = 1$ & $R_A^G > 1$.
- More than 90% of observations were based on unchanged monthly prices, i.e.; $R_b^G = 1$ & $R_A^G = 1$.

Some of the largest discrepancies, in terms of $R_A^G$ and $R_b^G$ were found primarily in product groups Mixed meat products (1233) and Cheese (1418-1419).

We now illustrate the two methods as monthly observations of price ratios of current monthly price in relation to the analogue base price, i.e. $\log(P_t^G / P_{Base}^G)$ and $\log(P_t^A / P_{Base}^A)$. In figure 5.1, the logarithmic price ratio is graphed as a scatter plot between the two methods, i.e. geometric versus arithmetic.
Figure 5.1: Logarithmic ratios of mean prices in current month relative to base period. Unweighted geometric mean on vertical axis and quantity-weighted arithmetic mean on horizontal axis. Eight sectors are numbered for analysis purposes.

Considering figure 5.1, there is a total concordance between arithmetic and geometric mean values for more than 90% of observations and a concentration in the origin exists, not being visible here. The main conspicuous features seem to be the dense areas in sector 1 below the 45 degree “line” i.e. \( \log(\frac{P_A^t}{P_B^t}) = \log(\frac{P_G^t}{P_B^t}) \), on which ratios are equal, and in sector 5 above the 45 degree “line”. These two sectors represent the two features of data as mentioned earlier but with another perspective. In sector 1, we find the 2% of observations where the geometric mean price in the base month was larger than was the arithmetic mean, generally explained by a one week of low prices there. To clarify, the situation in sector 1 occurs when the base price calculated arithmetically is less than its geometric analogue, \( P_B^A < P_B^G \) so the ratio between the two methods becomes smaller than one, \( ((\frac{P_G^t}{P_B^t})/(\frac{P_A^t}{P_B^t})) < 1 \) and thus below the 45 degree “line”. Since the arithmetic mean value base price is lower than the geometric analogue, the price ratio becomes larger for arithmetic mean value.

In sector 5, we find the 3% of observations where the geometric mean price in the January-March month was larger than was the arithmetic mean, \( P_G^t > P_A^t \), which meant that the price increase relative to the base price was larger for the geometric mean than for the arithmetic mean, hence above the 45 degree “line”.

Figure 5.2. Monthly price indices for product groups in supermarkets and hypermarkets, based on geometric and arithmetic mean prices per month.

When index values are aggregated to daily necessities in total, i.e. all price ratios on most detailed level are summarized for each sector, the difference between the two computation methods tend fade out (figure 5.2).

For now, Statistics Sweden is using unweighted geometric mean values for this aggregation purpose but will probably change to quantity-weighted arithmetic mean value from year 2013 on behalf of recommendations from the Swedish CPI board.

5.6 Differences in index due to choice of mean value method

Price ratio comparisons were made, as outlined above, of unweighted geometric mean and quantity-weighted arithmetic mean values. We computed the monthly price index for daily necessities based on this scanner data sample from January – April 2012 and as base December 2011 to calculate geometric and arithmetic mean values, with and without sales quantities as weights. Our calculations yielded that all index values were around 100, with the largest variation due to minor weighting impacts during campaigns, especially in January, see table 5.2.
Table 5.2. Index differences due to method of aggregation (Unweighted Geometric mean=100)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>100</td>
<td>99.815</td>
<td>99.785</td>
<td>100.038</td>
</tr>
<tr>
<td>February</td>
<td>100</td>
<td>99.998</td>
<td>99.996</td>
<td>100.000</td>
</tr>
<tr>
<td>March</td>
<td>100</td>
<td>100.000</td>
<td>100.000</td>
<td>100.003</td>
</tr>
<tr>
<td>April</td>
<td>100</td>
<td>99.969</td>
<td>99.963</td>
<td>100.008</td>
</tr>
</tbody>
</table>

Note on abbreviations: Unw. stands for unweighted, W. stands for weighted.

In figure 5.3, it can be seen that most price changes relative to the base seem to be quite small and the distribution appears as log-normal. The graph depicts merely price changes, the majority of monthly prices (> 50%) did not change during this period. Ratios are most often in the span -20% to +20% which implies that the difference between arithmetic and geometric monthly mean values will be very small in index.

Figure 5.3. Distribution of price changes during January – April 2012 with base in December 2011. Unweighted Geometric mean.

5.7 Changes of EAN-codes

We also made analyses on the life time of EAN-codes, which are the key in data to the product. During year 2010, around 70% of the EAN-codes that existed in the base (December 2009) also existed in the last month, December 2010, while 30% expired during the year. This is based on the entire scanner data and not merely the daily necessity sample. The corresponding figures for 2009 were that 80% of the EAN-codes existed the entire year (for this purpose we had merely data to November) while 20% expired during the year (as before November). The same relationship, 80% surviving and 20% expiring EAN-codes also occurred for 2011.

The Swedish CPI uses scanner data from one retail chain for a pps-sample of 538 products, identified by EAN-code. By limiting the sample to this size it is possible for the Price Unit staff to continuously monitor the products in the sample. Occasionally minor changes in package, weight/volume, taste etc. are introduced and a new EAN-code is assigned to the “new product”. A product can be assigned a new EAN-code by the producer even when no change at all is made. The idea of the Swedish construction of the scanner data price index is to monitor all such changes and decide when
the difference between the new product (new EAN-code) and the old product (old EAN-code), i.e. a change, is to be considered small enough to replace the old EAN-code with the new one in sample data – possibly with a recalculation due to quantity change, or

- a product has expired/passed from the market with no similar replacement from the producer/wholesaler.

During the period January - April 2012, the Price Unit actively changed the EAN-code for 35 product codes out of 538 in total for the scanner data sample of daily necessities, i.e. 6.5%. These changes were sometimes multiple, i.e. some products changed EAN-code more than once. Many of the changes were found in tobacco products.

5.8 Quality of data by indicators from output (macro) editing

In the monthly output (macro) editing of the 88 product groups containing daily necessities, on average six product groups failed per month in 2011, i.e. they were flagged to be suspected. In January to April 2012 an average of three products were flagged. Two of the products that were flagged during 2012 were cigarettes and other tobacco products and this was due to changes in taxes on tobacco. This was thus not related to the collection method or temporary effects. A conclusion is that the cash register data result in more stable macro-data than the corresponding manually collected data did last year. One explanation for this is that we usually have three measured prices per month per store and product, rather than only one. Another explanation may simply be a higher quality of data.

6. Concluding remarks

This paper has presented some issues regarding compilation of index based on scanner data. Given that scanner data has been secured, it requires aggregation to monthly levels. Statistics Sweden is for now using unweighted geometric mean values but will change to quantity weighted arithmetic mean values. Some data may contain discounts, which requires some policy on how to deal with when the discounts are available in the data. We have studied the life-span of EAN-codes and found that 20-30% of all codes seem to expire during the price index year (December to December). Our concluding reflection is, based on several studies, that scanner data is a promising source of data and we have started using it in the compilation of consumer price index for daily necessities within the Swedish CPI. Our conclusion is that scanner data offers high quality information of actual transactions, which is the target of price collections.
Discounts, rebates, loyalty schemes and “free” products

3.138 CPIs should take into account the effects of rebates, loyalty schemes, and money-off vouchers. Given that a CPI is meant to cover all the reference households, whether in the country as a whole or in a particular region, discounts should be included even if they are available only to certain households or to consumers satisfying certain payment criteria.

3.139 It may be difficult to record discriminatory or conditional discounts for practical reasons. When only one selected group of households can enjoy a certain discount on a specific product, the original stratum for that product is split into two new strata, each experiencing different price changes and each requiring a weight. So, unless base period expenditures for all possible strata are known, it is not possible to record discriminatory discounts correctly. Similarly, with conditional discounts, e.g. discounts on utility bills for prompt payment, it can be difficult to record the effect of the introduction of such offers unless data are available on the proportion of customers taking advantage of the offer. These kinds of practical problems also arise when there is price discrimination and the sellers change the criteria that define the groups to whom different prices are charged, thereby obliging some households to pay more or less than before without changing the prices themselves. These cases are discussed further in Chapter 7.

3.140 Although it is desirable to record all price changes, it is also important to ensure that the qualities of the goods or services for which prices are collected do not change in the process. While discounted prices may be collected during general sales seasons, care should be taken to ensure that the quality of the products being priced has not deteriorated.

3.141 The borderline between discounts and rebates can be hazy and is perhaps best drawn according to timing. In other words, a discount takes effect at the time of purchase, whereas a rebate takes effect some time later. Under this classification, money-off vouchers are discounts, and as with the conditional discounts mentioned above, can only be taken into account in a CPI if they relate to a single product and if the take-up rate is known at the time of CPI compilation. Since this is highly unlikely, the effect of money-off vouchers is usually excluded from a CPI. It should be noted that the discount is recorded only when the voucher is used, not when the voucher is first made available to the consumer.

3.142 Rebates may be made in respect of a single product, e.g. air miles, or may be more general, e.g. supermarket loyalty schemes where a $10 voucher is awarded for every $200 spent. As with discounts discussed above, such rebates can only be recorded as price falls if they relate to single products and can be weighted according to take-up. Bonus products provided “free” to the consumer, either by larger pack sizes or offers such as “two packs for the price of one”, should be treated as price reductions, although they may be ignored in practice when the offers are only temporary and quickly reversed. When permanent changes to pack sizes occur, quality adjustments should be made (see Chapter 7).

3.143 Given the practical difficulties in correctly recording all these types of price falls, it is usual to reflect discounts and rebates only if unconditional, whereas loyalty schemes, money-off coupons, and other incentives are ignored. Discounts during seasonal sales may be recorded provided that the quality of the goods does not change.

...
COUNCIL REGULATION (EC) No 1687/98 of 20 July 1998 amending Commission Regulation (EC) No 1749/96 concerning the coverage of goods and services of the harmonised index of consumer prices (*) - Article 1

THE COUNCIL OF THE EUROPEAN UNION,
Having regard to the Treaty establishing the European Community,
Having regard to the proposal from the Commission,
Having regard to the opinion of the European Central Bank (2),

Whereas Commission Regulation (EC) No 1749/96 of 9 September 1996 on initial implementing measures for Council Regulation (EC) No 2494/95 concerning harmonised indices of consumer prices (3) sets down an initial coverage for HICPs that was restricted to those goods and services covered by all or most national consumer price indices (CPIs); whereas the prices to be taken for the HICP, in particular the treatment of subsidies, rebates and reimbursements, require harmonised definitions; whereas the geographic and population coverage of the HICPs need still to be specified;

Whereas Article 3 of Regulation (EC) No 2494/95 requires that the HICP be based on the prices of goods and services available for purchase in the economic territory of the Member State for the purposes of directly satisfying consumer needs; whereas prices not actually paid by consumers in such purchases or opportunity costs or interest payments are not appropriate for international comparisons of consumer price inflation;

Article 1
Regulation (EC) No 1749/96 shall be amended as follows:

3. subparagraph (a) of Article 2 shall be replaced by the following:

(a)(3) Prices used in the HICP are the purchase prices paid by households to purchase individual goods and services in monetary transactions. Where goods and services have been available to consumers free of charge, and subsequently an actual price is charged, then the change from a zero price to the actual price, and vice versa, should be taken into account in the HICP.


THE COMMISSION OF THE EUROPEAN COMMUNITIES,
Having regard to the Treaty establishing the European Community,

Having regard to Council Regulation (EC) No 2494/95 of 23 October 1995 concerning harmonised indices of consumer prices (1), and in particular Article 4 in conjunction with Article 5(3) thereof,

After consulting the European Central Bank (2),

Whereas:
(1) By virtue of Article 5(1)(b) of Regulation (EC) No 2494/95, each Member State is required to produce a Harmonised Index of Consumer Prices (HICP) starting with the index for January 1997.
(2) There is considerable scope for procedural differences in the treatment of price reductions and the relationship between price reductions and purchase price. A harmonised methodology for the treatment of price reductions in the HICP is necessary to ensure that the resulting HICPs meet the comparability requirement of Article 4 of Regulation (EC) No 2494/95. It will also improve their reliability and relevance.
(3) Prices used in the HICP should be purchaser prices actually paid by households to purchase individual goods and services in monetary transactions, including any taxes less subsidies on the products, after deductions for discounts for bulk or offpeak-purchases from standard prices or charges, and excluding interest or services charges added under credit arrangements and any extra charges incurred as a result of failing to pay within the period stated at the time the purchases were made.

...  

Article 2  
Purchaser prices  
Unless otherwise stated purchaser prices used in the HICP shall in general take account of reductions in prices of individual goods and services if such reductions:
(a) can be attributed to the purchase of an individual good or service; and
(b) are available to all potential consumers with no special conditions attached (non-discriminatory);
(c) are known to the purchaser at the time when they enter into the agreement with the seller to purchase the product concerned; and
(d) can be claimed at the time of purchase or within such a time period following the actual purchase that they might be expected to have a significant influence on the quantities purchasers are willing to purchase.
In particular, reductions in the prices of individual goods and services which are likely or expected to be available again at standard prices or are available elsewhere at standard prices shall be taken into account in the HICP. Standard price means the price without any conditions or qualifications and not described as a special price.

...  

Article 4  
Specification change  
Where specification changes, prices shall be treated in accordance with the rules on specification changes, and in particular those regarding quality adjustment in Article 5 of Regulation (EC) No 1749/96.

Price reductions

6.80 One of the principles relating to consumer price indices, which is applied with few exceptions (such as owner-occupier housing costs), is that only transaction prices, that is prices actually paid by individuals or households, should be included in the index. This may differ from the advertised price if, for example, a discount is offered. In practice, however, discriminatory discounts, which are available only to a restricted group of households (as opposed to non-discriminatory discounts that are available to all), are generally excluded on principle. For example, money-off coupons and loyalty rewards for previous expenditure are normally ignored and the non-discounted price is recorded. Also, it may be difficult to obtain the price paid if this is subject to individual bargaining. It may therefore not come as a surprise that, while the general rule above may appear simple, there are a number of instances requiring special treatment either because of conceptual issues or because of practical difficulties. The following guidelines reflect practices followed by a number of countries. They do not represent a set of rules because the appropriate practice to be followed will be determined by individual circumstances, which might vary between different countries.

6.81 Discounted prices should only be taken if generally available to anyone with no conditions attached; otherwise the non-discounted or unsubsidized price is recorded. In particular, the general practice is to ignore money-off coupons and loyalty rewards. A judgement needs to be made, however, relating to the interpretation of “generally available”. For instance, reduced prices for payment by direct debit may be taken into account depending on the extent to which consumers as a whole have access to and use such a service. A judgement is required in the latter case on the threshold to be set for access, above which action is taken for inclusion in the index. Alternatively, different payment methods may all be priced individually (for example, separate data collection for electricity payments by cash, direct debit and pre-payment) and weighted together to form a single price index for that item.
6.82 Price discrimination. Discounts available only to a restricted group of households should be disregarded because they are discriminatory, unless they are significant and are available either to the vast majority of the population or to identifiable subgroups who qualify for such discounts on the basis of demographic or other characteristics not requiring action by the individuals concerned at the time of purchase. In the latter case, they should be treated as stratification or coverage issues in item sampling. Some judgement is required. Examples of allowable price discrimination may include lower prices offered to pensioners (for example, discounted travel or haircuts) and discounts for people who receive state benefits. Another example of a case where prices are not universally available to all, and where judgement is required, is where a nominal or token membership fee is required by the retail outlet. In these cases, the take-up of such membership – which is widely available to all – needs to be considered in terms of thresholds and general spending patterns of the consumers and the conditions placed on membership which may make the latter restrictive (for example, minimum levels of purchase). Ease of access to the outlets in question may be a relevant factor as well, say, if in practice the customers need to have the use of a private car.

6.83 Sale or special offer prices should be recorded if these are either temporary reductions on goods that are likely to be available again at normal prices, or are stock-clearing sales (such as January sales or summer sales). Before designating a price as a “sale” price, however, special care should be taken to ascertain that there is a genuine sale with price reductions on normal stock. On occasion, stock is continually sold below the recommended retail price or advertised as a special offer even though these prices are available all year. In such cases, prices should not be considered as sale prices, but can still be collected. Special purchases of end-of-range, damaged, shop-soiled or defective goods should not normally be priced, as they are likely not to be the same quality as, or comparable with, goods previously priced and are unlikely to be available in future. If the special offer is limited to the first customers, the item should not be priced, as the offer is not available to everyone. Introductory special offers may be included if they are available to all. In reality, however, given the need to price the same “basket” each month, such offers will not be chosen as representative items unless they are introduced at the time of an update of the “basket” or when a replacement item needs to be chosen. Discounts on goods close to expiry dates should be disregarded or treated as specification or quality changes.

6.84 Bonus offers, extras and free gifts. Prices for items temporarily bearing extra quantities (for example, 30 per cent extra free) should not be adjusted to take account of the increased quantity if it is thought that the extra quantities involved may not be wanted by most consumers, will not have influenced the decision to purchase or will not be consumed. Similarly, free items with other purchases (such as buy 2 get 1 free or free gift with every product purchased) should be disregarded. Money-off coupons for future purchases should be disregarded, as these may not be used or wanted. Free gifts such as plastic toys in cereal boxes should be ignored because they are not included in the list for price observations; it is the price to be paid to get the cereal in the box that is relevant. Collectors should be aware that temporary “special offer” weight changes (X per cent extra free) could become a permanent weight change (for example, cans of alcoholic drinks changing size from 440 ml to 500 ml) and should feed the information back to head office as they become aware of it. In this way, head offices can issue new or amended guidance to price collectors about item specifications.

6.85 Stamps. Sometimes purchasers are given special stamps, which can be accumulated and subsequently exchanged for goods and services. If a discount is available as an alternative to such stamps, then the discounted price should be recorded. Otherwise, the stamps should be disregarded.

... 6.90 Irregular rebates or refunds should only be taken into account when they apply to the purchase of an individual product and are granted within a time period such that they are expected to have a significant influence on the quantities purchasers are willing to buy. Loyalty rebates or coupons associated with previous expenditure at the outlet, to be used for similar or other purchases, should generally be disregarded, as they are discriminatory. If they are significant factors, they should be treated as stratification or coverage aspects of sampling (see Chapter 5). One-off rebates (for example, associated with privatization) should be disregarded as they do not relate to the specific time period of the consumption and are unlikely to affect levels of consumption. They can be viewed more as a source of additional income.
(3) 'Inducements in the form of extras’ (extra quantity of the product concerned, inclusion of a different product for 'free’ or other extra favors) should be disregarded if they are not significant. The market value of the inducement may be deducted if known but it should then be added back if the offer is withdrawn.

(a) ‘\( x \) units at a lower unit price than one’ should be disregarded or treated according to the rules applied in the context of specification changes.

(b) ‘Free \( x \) provided with each purchase of \( y \)’ should be disregarded or treated according to the rules applied in the context of specification changes.

(c) ‘Money-off-coupons attached to a certain’ good should be disregarded unless offered to all potential customers at the point of sale and claimable at the time of purchase.

(4) 'Discounts available only to a restricted group of households’ should be disregarded because they are discriminatory, unless they are expected to be significant in which case they should be treated as a coverage or stratification issue in sampling.

(a) 'Lower price for e.g. pensioners’ should be disregarded for the purposes of these guidelines (discriminatory). They are most likely a case of tariff or multiple pricing or potentially a population coverage or stratification issue in sampling.

(b) 'Lower prices negotiated for the members of an organised group, e.g. trade unions members’ should be disregarded for the purpose of these guidelines (discriminatory). They are most likely a case of tariff or multiple pricing or potentially a population coverage or stratification issue in sampling.

(c) 'Card for which people pay, which entitles them to discounts at certain shops’ should be disregarded for the purpose of these guidelines (discriminatory). They are most likely a case of tariff or multiple pricing or potentially a population coverage or stratification issue in sampling.

(5) 'Regular rebates or refunds’ should only be taken into account when attributable to the purchase of an individual product and granted within a time period from the actual purchase such that it is expected to have a significant influence on the quantities buyers wish to buy.

(a) Deposit for ‘money-back-bottles’ should be deducted from the price.

(b) Deposit included in the price of a new car and paid back to the car-owner when the car eventually is handed over by him for destruction should be disregarded as being reimbursed after too long a time to significantly affect the quantities bought at the time of the purchase.

(6) 'Irregular rebates or refunds’ should only be taken into account when they apply to the purchase of an individual product and are granted within a time period from the actual purchase such that it is expected to have a significant influence on the quantities buyers wish to buy. 'Discounts available only to a restricted group of households’ should be disregarded because they are discriminatory, unless they are expected to be significant in which case they should be treated as a coverage or stratification issue in sampling.

(a) 'Loyalty rebates 1, i.e. accumulated points or coupons rewarded for the purchase of a certain good or service can be used to purchase the same good or service at a reduced price (e.g. frequent flyer accounts)’ should be disregarded because they are discriminatory. If they are expected to be important in some population sub-groups and have a significant influence on the quantities bought they should be treated as coverage or stratification issue in sampling to the extent that they are not obtained as a result of business expenditure.

(b) 'Loyalty rebates 2, i.e. accumulated points or coupons can be used to purchase any good or service at a reduced price (e.g. credit card bonus system)’ should be disregarded because they are not relating to an individual good or service and are discriminatory. If they are expected to be important in some population sub-groups and have a significant influence on the quantities bought they should be treated as coverage or stratification issue in sampling to the extent that they are not obtained as a result of business expenditure.
References

Compendium of HICP - reference documents (2/2001/B/5)


