Thirteenth meeting
Copenhagen, 1-3 May 2013
Topic 4: Sampling and Data Collection

ISSUES ON THE USE OF SCANNER DATA IN THE CPI

Prepared by Muhanad Sammar, Anders Norberg and Can Tongur, Statistics Sweden

1. INTRODUCTION AND BACKGROUND

In the coming years, more and more national statistical offices, NSIs, will be able to secure and study scanner data, also called cash register data. Three motives for using scanner data are to reduce the cost for price data collection, to increase sample sizes and thus reduce or even eliminate sampling variance, and to eliminate measurement errors existing with other data collection modes.

In 2011 a project (SCB, 2011) was initiated with the objective of studying the possibilities of improving the price collection by replacing manually collected prices with scanner data. Thorough comparisons between manually collected prices and scanner data were made to ensure that the quality was good enough for price index calculations. Some of the main questions in the analysis were the extent of agreement between the two data sources and the need for data cleaning. Following the analysis phase, the project implemented a system for using scanner data in the regular daily necessities price collection 2012.

2. SCANNER DATA

2.1 What is scanner data?

Scanner data are big files of transactions, identified uniquely by some product code such as EAN-code (also called GTIN = Global Trade Item Number).

For companies to be able to communicate well with each other and act efficiently in the market, some tools are required. There is a worldwide system that simplifies the global and local trading using unique identification numbers for products and locations, meaning that all parties in the value chain can exchange information. The standard EAN bar code was from the beginning introduced so that the retailers would not have to price tag all goods in stores and to streamline the manual handling at the cash desk. Today you can get information about the contents of products, how and when they are made with the GS1 System, GS1 (2013).

2.2 Who are the owners of EAN-codes?

The source of the information in this section is www.gs1.org

The EAN-barcode[1] (International Article Number, formerly called European Article Number) is mostly a 13 digit combination (12 data and 1 check). The first three digits usually identifies the country were the manufacturer is registered. The country code is followed by 9 digits, of which the first part is a company prefix and the rest is the company numbering of their articles. An EAN code for a product that is no longer available on the market can be reused for another product after a few years.

GS1 is an international not-for-profit association with member organizations in over 100 countries. They are dedicated to the design and implementation of global standards and solutions to improve the efficiency and visibility of supply and demand chains globally and across sectors. The GS1 system of standards is the most widely used supply chain standards system in the world, see GS1 (2013). GS1 provides a classification system that is structured logically with hierarchical levels, Global Product Classification (GPC). An increasing number of trade items identified with GTIN are classified according to GPC by the manufacturers. The purpose of GPC is to give manufacturers and retailers a common

[1] http://en.wikipedia.org/wiki/International_Article_Number_(EAN)

language for grouping products in the same way. The "brick" level is the lowest classification level and consists of a group of narrowly defined products (e.g. perishable milk and milk products).

Master data about trade items is in many cases shared between trading partners by using an interconnected network of databases that are certified by GS1. This is called the GDSN (Global Data Synchronization Network). A lot of products are covered in the GS1 system, and the best coverage is currently for groceries and general merchandise. Some producers do not publish their product attributes to any database. National organizations in each country decide the structure of the database, but the way data is synchronized is standardized in detail for which GS1 can provide recommendations. In Sweden GS1 provides a package of services (Validoo) that handles article information (such as ingredients, package size and packaging) and facilitates exchange of information between manufacturers and retailers. The Validoo service is connected to other GDSN data pools in other countries.

In general, the code structure for attributes is the same within a country, but companies can enter different characteristics. It should be noticed that the GS1 standards for item identification and classification have been developed primarily for supply chain applications. This means that in some cases it may not be perfectly optimized for statistics. However, in these cases it is possible to submit work requests to GS1 to further develop the standards to meet the needs of statistics agencies.

2.3. Co-operation

To get access to scanner data, a possibility would be to purchase data from a market research company, but this has appeared as expensive even compared to manual collected prices. In 2008, Mr. Muhanad Sammar initiated contacts with the head offices of the three largest daily necessity retailer chains in Sweden to start negotiations for obtaining scanner data.

SCB now obtains scanner data free of charge by agreements written as a form of a partnership. The companies assume no formal responsibility for the quality of data. In addition, SCB agrees that services from the companies are to be as limited as possible in order to reduce the burden. The agreements also state that any disputes arising out of contract shall be finally settled by arbitration in accordance with the Stockholm Chamber of Commerce Arbitration Rules for Expedited Arbitration.

The agreements are voluntary for the retail companies and are valid until further by a 6 or 12 month notice period. Scanner data are very sensitive and owned originally by each retailer. As it represents a great commercial value SCB has explicitly promised that use of data will be limited to official statistics. Retail companies have been informed of SCB's business and that the data are protected by the rules of the Public Access to Information and Secrecy Act (2009:400) (OSL), notably Chapter 24, 8§ and Chapter 31, 16§.

2.4 "Big data"

More and more data are being produced by an increasing number of electronic devices surrounding us and on the internet. The amount of data and the frequency at which they are produced are so vast that they are usually referred to as 'Big Data'. With scanner data, we are facing three challenges that are characteristic for the analysis of big data.

It is tempting to try to use all the information delivered to NSI. This means that the price index must be computed without us knowing much about vast number of products. The alternative, however expensive to perform, is to apply sampling and estimation. This is evolved in section 3.1.

The data delivered to NSI contain variables such as; EAN-code, price, quantity, name of the product and a product group code from each retailer chain. These product group classifications serve quite another need than the classifications used by NSI do. For the CPI the COICOP[2] classification is relevant. For the survey Production of commodities and industrial services (IVP) and foreign trade with goods (Intrastat and Extrastat) the 8-digit CN[3] is relevant. See section 3.3 on coding scanner data.

Big data can be structured, semi-structured or unstructured. Scanner data are structured data, organized in a way that both humans and computers can read. However received data files don't have to be similar between data providers, see more in section 3.8.

3. DATA AND ELEMENTARY AGGREGATES

3.1 Four ways to use scanner data

There are several methods to use scanner data in the CPI, the following four might be considered as the most realistic for daily necessities:

I. **Replace the manually collected price data with scanner data for the sample of outlets and products.**

The computing of indexes will be equivalent to the computing methods applied within the previous production system of the national CPI, i.e. the Jevons index for elementary aggregates. An aggregation of weekly price data per retail outlet is needed. The full potential of the scanner data is not utilized, although the sample of outlets and products can be made much larger than it has been. The standard error of index can be decreased at low cost if scanner data is obtained for free. This reflects SCB's approach.

II. **Compute index from a census based on all products for which scanner data are available.**

The availability of both price and quantity data for all products in specific outlets would enable the use of a frequently chained index with a superlative index formula such as a Fisher index at the lowest level of aggregation. The first problem here is that the NSI, due to the large amounts of products, has no control of actions taken by market actors. There could be in the interest of a producer to implicitly increase price by decreasing the package size but retaining the price and set a new EAN-code. Secondly a minimum of processing is needed to continuously classify all new products fairly well into product groups. A third obstacle is that bottle deposits for water, soft drinks and beer are not withdrawn from the price, i.e. a change of deposit cost imply change on the index, which is inconsistent with the regulations.

III. **Use scanner data as auxiliary information.**

This is a three-step procedure:

(1) Compute a superlative Fisher price index by using the complete scanner data as in II. Accept errors due to an imperfect price variable adhered with unwanted discounts, incapacity to withdraw deposits and incapacity to deal with market replacements as described in II.

(2) Collect prices for a small sample of retail outlets and a small sample of products through a high quality measurement and replacement method, either from scanner data or by manual price collection (S1). Extract also regular scanner data for the same sample of product offers (S2).

(3) Adjust either

[2] COICOP = Classification Of Individual Consumption by Purpose

[3] Combined nomenclature, also used by the Customs

a) the superlative Fisher price index from (1) by the average ratio of the high quality measured prices (S1) and regular scanner data prices (S2) for the sample in base and current months, or

b) a Jevons price index on high quality sample data (S1) by the average ratio of the superlative Fisher price index from (1) and a superlative Fisher index for sample scanner data (S2).

IV. **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 exist, depending on interviewer training, price collection instructions, measurement device, auditing etc. Unfortunately we have little empirical data on these errors.

For other product groups, other methods could be preferable e.g. for electronic devices, a hedonic approach might be appropriate because then one can account for product characteristics that can be found in scanner data.

3.2 Sampling of outlets

The Business Register (FDB) at SCB is used as a sampling frame of outlets. The information on outlets in the register includes industry, number of employees and address. The method for selection of outlets is stratified sequential Poisson sampling with selection probabilities proportional to their size, see SCB (2012). The sample is drawn with the SAMU technique for coordinated sampling of business surveys, see SCB (2003).

3.3 Product frame population

Scanner data are mainly used in COICOP groups 01 (Food except for perishable fruits, vegetables and meat) and 02.2 (Beer and tobacco). We have coverage of detergents, hygiene articles etc. in COICOP 02.1.3, 05.5, 05.6, 06.1, 09.3 and 12.1. Scanner data cover approximately 14% of CPI.

For many years, the retail organizations of daily necessities have delivered annual total values of sales of each and every EAN-coded product to SCB. These files nowadays make up a total of about 360 000 unique products. Roughly two thirds of the EAN-numbers remain in data from one year to the next, hence approximately 100 000 products are introduced each year.

The files are coded automatically and by SCB staff with COICOP and national CPI code in a multi-step procedure. A statistical approach is used all the way implying that the quality of the final CPI statistics is important but not every data value in the sampling frame. First a code is set, using a key between the product group of the retailer and the CPI-classification. After this stage there are many misclassified products. By automatic recognition of pieces of the product name, an association to a more correct CPI-code can be made for some significant groups. Finally, for products available in more than one retail chain, the preliminary codes are compared. Deviations are flagged and those with high transaction values are prioritized for manual follow up. This coding activity is nowadays performed for the purpose of CPI and for producing the annual statistics "Food sales".

3.4 Sampling of products

For about 30 years now, SCB has used probability samples of specific products for daily necessities. The advantages of probability sampling have a strong scientific basis. There is, however, a risk that price changes are hidden in the index if sampled products cannot be replaced, like consumers do, due to strict enforcement of sampling methodology. Consequently, purposive methods are used to replace products that no longer are available on the market (method I in section 3.1).

Samples are drawn by sequential Pareto πps selection within strata, see SCB (2012). Three product samples of approximately 800 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 products that are available in the sampled outlet in December (base period) and/or January are included as product offers in the sample.

3.5 Monthly average price

Scanner data are received on a weekly basis. The weekly average prices are then aggregated to a monthly average for each product offer (combination of product and outlet).
Below is argued three different ways to aggregate weekly average prices to a monthly average price. Let us note that the transaction is the unit of interest and the transaction price is the obtained variable. The sum of all individual transactions that are made by customers during a period of three weeks constitutes the population for each store and product.

1) With a simple arithmetic average of all transaction prices during a three weeks period, we get a relevant measure of the consumer population ability to adapt to changes in relative prices. A quantity weighted arithmetic average of weekly aggregates of transactions generates the same results.

2) The parameter “unweighted arithmetic average of prices offered in a month” can be interpreted as the average price level for consumers who are completely insensitive to price levels and campaigns. This might not be seen as the ideal definition for the CPI. Further, there are practical problems to receive daily scanner data and to compute this average.

3) The “unweighted geometric average of prices offered in a month”, regardless of how many units consumers buy each day, is a parameter with no relevant interpretation. This aggregation can, however, be motivated by its resemblance to the formula we use for other product groups to estimate average prices based on a sample of shops, products and timing. The unweighted geometric average is generally considered the best way to reflect (estimate) the substitution effect of price changes. It is used for most other product groups in the CPI, regardless of data collection method. If it fits well for sample data it could be good enough for a population.

3.6 Elementary aggregates: Jevons index

Elementary indices are computed for combinations of product group and industry (of outlet). For daily necessities there are 88 product groups and 2 industries. Aggregating prices to indices at the lowest level is made by the Jevons index formula:

$$I_J^{0:t} = \prod \left(\frac{p_i^t}{p_i^t}\right)^{1/n} = \frac{\prod \left(p_i^t\right)^{1/n}}{\prod \left(p_i^0\right)^{1/n}}$$

Index in the current month *t* is the ratio of the observed prices for product-offers in the current month *t* and the observed prices for product-offers in December of the previous year, the reference month *0,* for the product-offers that exist in both periods.

3.7 Rebates and deposits

When products have a “best-before” date close to the “expiry date” and are on sale, they are labelled with the sales price, and the cashier manually key-enters the sales price. Companies cannot connect two different prices to a specific EAN-code.

Deposits are excluded from the price in our calculations. We identify the articles that have a deposit and then deduct the deposit fee from the price.

3.8 Production system

SCB has through an FTP account established a secured data transmission channel with three retail chains. The routines give SCB enough time to reconnect in cases of a failed data transmission. All chains deliver their data in text files.
A difficulty that might occur is if one of the chains decides to suddenly change from one format to another e.g., ANSI, ASCII, UTF-8, -16, -32. But such changes would be manageable for us to handle because we expect a specific data format.

We have constructed a production system for processing scanner data. This includes data collection, data entry, data editing and aggregation. This is an intermediary system connected to the ordinary CPI production system, and its output is the product sample, delivered in SQL table form, to the ordinary production system.

Our scanner data system has six main stages of production:

1st stage: Initiating a production month
2nd stage: Checking of the scanner data set
3rd stage: Select the data for the product-offer sample and review it
4th stage: Aggregate prices over three weeks
5th stage: Send data to the CPI production system
6th stage: Product life analysis

In the implementation the technology used involves applications based on the SAS System, a dot.Net solution interfacing to the production system (Pi09) for Swedish price statistics, and a robot-based file delivery system. Input files are transmitted in encrypted format to ensure security, particularly essential in view of the sensitive nature of the data.

4. EMPIRICAL STUDIES OF SCANNER DATA

In the following, we briefly mention some empirical findings that we have encountered within this field.

4.1 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. This 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.

4.2 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 i an outlet in a week. For 7-9% of the cases scanner data prices are not equal to manual collected prices, but there is symmetry between the two collection methods, see table 1.

Table 1 Scanner Data (S.D.) and Manually Collected Prices (M.C.P.) in comparison, percent. Product-offers, outlets and weeks. January – December, 2009 and 2010.

Matching categories	2009	2010
Neither in M.C.P. or S.D.	1.5	0.6
In M.C.P. but not in S.D.	4.5	5.3
In S.D. but not in M.C.P.	1.5	0.9
M.C.P. = S.D.	83.4	86.2
M.C.P. > S.D.	4.3	3.7
M.C.P < S.D.	4.8	3.3
All	100.0	100.0

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

4.3 Changes of EAN-codes

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.

By limiting the sample 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) is to be considered small enough to replace the old EAN-code with the new one in sample – 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 within tobacco products due to a package size change.

4.4 Price index with and without discounts

An issue in focus for scanner data is the treatment of discounted sales that may exist in data. By discount it is meant general price reductions that are not of temporary nature but instead targeted to specific consumers.

We have done a study based on the product sample for the year 2012 of daily necessities in the Swedish CPI, applied to data from 2011. The time window between the product frame and the data has not implied a significant loss of transactions. The selected data comprises around 8 000 transactions each week, covering 39 weeks, i.e. three weeks each month starting from the base in December 2010. Index has been computed by using both unweighted geometric means and quantity weighted arithmetic means over the three weeks per month. Both December 2010 and February 2011 are used as base months.

Table 2 Differences in price indices of data with and without discounts for the year 2011. Differences computed as: Difference (I(with discounts) - I(without discounts)).

Period	Base in December 2010		Base in February 2011	
	Geom. diff.	Arith. diff.	Geom. diff.	Arith. diff.
0	-	-	-	-
1	0,64	0,87	-	-
2	0,45	0,63	-	-
3	0,01	0,03	-0,36	-0,51
4	-0,18	-0,04	-0,55	-0,58
5	0,57	0,88	0,11	0,23
6	0,22	0,42	-0,22	-0,19
7	0,69	1,05	0,23	0,40
8	0,63	0,86	0,12	0,17
9	0,72	1,02	0,26	0,35
10	0,62	0,77	0,12	0,09
11	0,22	0,38	-0,24	-0,29
12	0,48	0,73	-0,02	0,05

Geom. = Geometric mean. Arith. = Arithmetic mean.

As is seen in table 2, when the base period is December, which is the ordinary base period, all later periods have higher index value when data contains discounts. This is an indication that the holidays in December imply more or larger discounts and campaigns than in other months. The only period with the opposite situation is April, which may be due to the Easter holiday 21-25 April 2011. Even if three weeks are included each month, this kind of holiday influences the entire period.

By changing base period to for instance February, which is a month that completely lacks within-week holidays, it can be seen that the differences show no specific pattern.

However, it is not possible to make some general conclusions on the monthly variation and implications due to discounts: In December 2011, the impacts of discounts were seemingly similar to a yearly average effect of discounts. The analysis shows that indices based on weighted arithmetic monthly mean prices are affected by discounts to a larger extent than are the geometrically computed indices. This is due to the increase in consumed quantities when prices decrease, which impacts weighted arithmetic means more than unweighted geometric means.

An alternative way to illustrate the differences due to discounts is by computing a ”price index” with running base period, which shows merely monthly effects, given in Table 3. Setting ordinary price (i.e. without discounts) as base, we can see that discounts impact mostly in the base month (period 0) and in April (period 4), but having in mind the calculated uncertainty measure, it cannot be concluded that the index of December 2011 (period 12) differs from the base (period 0), since the uncertainty intervals cover each other. Based on this, we cannot state, with certainty, that the introduction of discounted data will have any systematic effects such as a seasonal impact in December.

Table 3 Index of discounted price relative to ordinary price 2011

Period	Geometric Index	Uncertainty measure	Arithmetical Index	Uncertainty measure
0	97,8	0,4	97,2	0,5
1	98,5	0,3	98,0	0,3
2	98,3	0,3	97,8	0,3
3	97,9	0,3	97,3	0,3
4	97,7	0,3	97,2	0,4
5	98,4	0,3	98,0	0,3
6	98,0	0,3	97,6	0,3
7	98,5	0,2	98,2	0,3
8	98,4	0,2	98,0	0,3
9	98,5	0,3	98,1	0,3
10	98,4	0,2	97,9	0,3
11	98,0	0,3	97,5	0,4
12	98,2	0,3	97,8	0,3

The uncertainty measure has been calculated by repeated half-sampling of stores and products in the sample. These half-samples have then been used for index calculations. Each half sample contains either half the store sample or half the product sample from the ordinary sample. This has been repeated several thousand times and based on these samples, standard errors have been computed in both dimensions. These two standard errors have thereafter been averaged.

The share of discounts in price observations during the study period was 49.2 %, which means that almost half the sample contained discounts. However, we do not have a complete coverage of the time period since we only receive scanner data for the weeks that are included in the collection period. Nevertheless, we can say that one third of the discounts lasts during one week, one third lasts for two weeks and one third at least for three weeks. We also see that the discounts are most frequent during the third week in the collection period, see table 4.

Table 4 The distribution of discount occurrences over price collection weeks 1-3.

	Number of weeks with discounts						
Measurement week when the discount starts	1	1	1	2	2	2	3
1	1	0	0	1	0	1	1
2	0	1	0	1	1	0	1
3	0	0	1	0	1	1	1
Share (%)	9.37	7.75	17.11	6.96	12.15	8.54	38.12

4.5 Quality of data by indicators from macro editing

After implementing scanner data, macro editing of the 88 involved product groups in daily necessities showed on practically no impact from the change in collection mode – merely cigarettes and other tobacco products were found to be conspicuous. This was, however, due to tax changes and not because of the mode change. Before scanner data, for instance during 2011, an average of 6 product groups per month failed editing controls, so the conclusion was that overall quality seemed higher, a conclusion reinforced by the fact that now three monthly observations were being used.

5. DEVELOPMENT AND HARMONISATION

Our point of view is that it is of great importance that rules and practices on the treatment of scanner data are harmonised across EU Member States. If this is not discussed and clarified now, there may be a risk that countries make different interpretations of the existing legislation and guidance.

To SCB it seems urgent to be able to take advantage of scanner data to the extent that this is feasible, although the possibilities may currently differ between countries. Manual price collection is expensive and may be subject to measurement errors, particularly as price tags sometimes differ from the payable prices. Scanner data offer promising solutions but call for clarifications in the legislation.

SCB initiated contacts with the company GS1 for discussions on harmonisation of product coding between Eurostat, NSI of member countries and GS1. One idea is that it would be beneficial to establish cooperation at the global and/or European level between the NSIs and GS1.

6. POSSIBILITIES WITH SCANNER DATA – FOR DISCUSSION

a) **Ways to use scanner data**
SCB has since 2012 chosen to replace manually collected price data with scanner data for samples of outlets and products but with larger samples than before (method I in section 3.1). We see problems with computing indices based on all products for which scanner data are available (method II in section 3.1). Have we reasonably considered pros and cons?

b) **Mapping of product classifications**
GS1 has information on products with EAN-codes. A harmonisation between COICOP and the GS1 product groups GPC may not be considered to be a realistic option. As an alternative, a key between EAN codes and COICOP on a 5-digit level, using GPC in the coding process, could be possible. A "one time" mapping effort to create the intermediate key between COICOP and GPC would be required. Note that SCB has produced such keys annually, see section 3.3.

c) **Benefits for data providers**
With scanner data the response burden will decrease for retailers. Can the NSI offer accurate and relevant analysis of price data to data providers in return for their services?

d) **Other benefits**
GS1 standards are being used for other companies and governmental activities, such as customs offices, which in turn gives us an opportunity to cooperate. This means that scanner data may well be used in other statistical area besides the CPI. A possible area may be the Price Indices in Producer and Import Stages. SCB has already used scanner data for parts of the Purchasing Power Parities survey (PPP).

7. REFERENCES

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

GS1 (2013), http://www.gs1.org

ILO/IMF/OECD/UNECE/Eurostat/The World Bank, Consumer price index manual: Theory and practice, Geneva, International Labour Office, 2004

Konsumentverket (2010). Available at http://www.konsumentverket.se/Global/Konsumentverket.se/Best%C3%A4lla%20och%20ladda%20ner/rapporter/2010/2010_02_Prisinformation%20inom%20dagligvaruhandeln.pdf

Norberg, Anders, Sammar, Muhanad and Tongur, Can (2011), “A Study on Scanner Data in the Swedish Consumer Price Index”. Available at http://www.stats.govt.nz/ottawa-group-2011/agenda.aspx

SCB (2003): “SAMU The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden”. SCB, Background Facts on Economic Statistics 2003:3

SCB (2011) “ Livsmedelsförsäljningsstatistik 2011” Process documentation in http://www.scb.se/Statistik/HA/HA0103/_dokument/HA0103_DO_2011.pdf
SCB (2012) “ Konsumentprisindex 2012” Process documentation in http://www.scb.se/Statistik/PR/PR0101/_dokument/PR0101_DO_2012.pdf