

Past, present and future of scanner data with focus on Statistics Sweden

By **Muhanad Sammar**

For presentation at the Workshop on Scanner Data for HICP
Lisbon, Portugal
26-27 September 2013

Statistics Sweden's position in this area is not completely established so some of the opinions and ideas presented in this paper are my own and do not necessarily reflect those of Statistics Sweden.

Introduction

Approximately 14% of all collected data used in the Swedish Consumer Price Index are from Scanner data.¹ Scanner data are mainly used in COICOP groups 01 (Food and non-alcoholic beverages, except for perishable fruits, vegetables and meat) and 02 (Alcoholic beverages and tobacco). We also have coverage of detergents, hygiene articles etc. in COICOP, 05.5 (Tools and equipment for house and garden), 05.6 (Goods and services for routine household maintenance), 06.1 (Medical products, appliances and equipment), 09.3 (Other recreational items and equipment, garden and pets) and 12.1 (Personal care).

Two of many reasons why the Price Unit at Statistics Sweden decided to replace manually collected data with scanner data for the daily necessities were that scanner data were more cost efficient and had better quality than the manually collected data.²

Statistics Sweden is presently working on securing more scanner data. During 2013, meetings have been held with a major furniture retailer, a Swedish railway operator, the Swedish Post and Telecom Authority and with two international leading research companies, among others.

Several studies over the last years have shown that it has become increasingly difficult for the National Statistical Institutes (NSIs) to collect enough accurate data to ensure the reliability of the survey results. In the next section, I will highlight two of many different studies written on non response rate. One of the papers was written by Statistics Sweden.

Empirical studies

Statistics Sweden, similar to other NSIs, has during the last decade been struggling with the problem of non responses in telephone surveys. One of the telephone surveys that is carried out by Statistic Sweden is the Labour Force Survey (LFS). In July 2013, Statistics Sweden had a non response rate of 32,5% while in the year 2000 the non response rate for the same survey

¹ In 2013.

² Sammar, M, Norberg, A and Tongur, C (2013), "Issues on the Use of Scanner Data in the CPI." Available at <http://www.dst.dk/da/Sites/ottawa-group/agenda.aspx>

was at 15%. Another survey that has had a similar development is the Household Budget Survey. In 2000 the non response rate was 48%; in 2012 the non response rate increased to 60%.

In a study conducted by Statistics Sweden, some of the main explanations of the steady decline of response rates over the years were found to be external and internal factors:³

- External factors
 - Difficulties finding relevant telephone numbers
 - Increasing reluctance towards answering among sampled units
 - Increased competition between different survey organisations and telemarketing companies
- Internal factors
 - Increasing workload at the Interview Unit
 - Poor interviewer training
 - Poor contact strategies
 - Inefficient work procedures

A second study, published in year 2010, was conducted by the Swedish Consumer Agency (Konsumentverket) with the scope of reviewing the price information in Swedish supermarkets.⁴ A total of 13 500 product offers were examined in 291 stores. The research was conducted in the late summer of 2009 with the help of consumer advisors in 35 municipalities across the country. Two main findings of the study are:

- 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.

These two studies indicate that there is a high risk that both telephone surveys and manual price collection might become outdated in the near future. Besides the fact that both methods are costly, they may also be subject to measurement errors, particularly observed through studies of scanner data, as price tags sometimes differ from the payable prices.

The most favourable choice of method for the NSIs must still be the method that best reflects the true picture. If a new data collection method is discovered and is as good or of better quality than the current method, then the new method must be seen as the most preferable one. If the new collection method is less expensive to maintain compared to the current one, then there is another significant argument for why an NSI should consider changing method.

³ Hörngren, Jan. Statistics Sweden's Overarching Project – Measures to Reduce Nonresponse in Individual and Household Surveys. 2011-05-30.

⁴http://www.konsumentverket.se/Global/Konsumentverket.se/Best%C3%A4lla%20och%20ladda%20ner/rapporter/2010/2010_02_Prisinformation%20inom%20dagligvaruhandeln.pdf

Scanner data

One area worth exploring further is data stored at companies' own databases. As most companies in Europe operate in a competitive market, one can assume that companies obtain, collect and store large amounts of sophisticated data.

Some NSIs might argue that not all companies have either good quality or sufficient data. The author agrees with this, but what prevents the NSIs from requesting this type of data? This leads to the next section, the challenges that NSIs are facing in the future.

The tasks ahead for the NSIs

For an NSI to be able to take advantage of a company's scanner data, it is necessary to have an open dialogue with retailers and their organisations. It is also very important to make considerable efforts to create good relationships with people at various positions within the retail organisation and to ensure that companies that provide data receive outstanding treatment. For example, the NSIs could offer the data providers guidance and support on how they can build their own indices and indicators for their own analyses. NSIs could also give feedback in return to the data providers.

Another possible task for the NSIs could be to coordinate workshops for the providers of the scanner data and member organisations. The concept of these workshops could be to:

- Influence and inspire
- Share experiences
- Get insight of developments made by the companies (if a change was made by a company then the NSIs would thus be prepared for that change.)

Here follows a sample of topics that could be discussed at the workshop:

- How to ensure better quality data
- Which variables might be useful for further analyses
- How prices are set on different products

NSIs should also organise a more academically oriented workshop at least once a year. The participants at these meetings should be people from the academic community and experts within the subject. The topics of the discussions should be methodological issues (such as sample design, choice of index formula) and other practical problems.

An NSI that engages other organisations to co-operate and at the same time strives to become more effective is consistent with what the public is expecting. Organising these types of workshops, according to the author, would also signal to the public that the NSIs take their mission very seriously.

The tasks ahead for Eurostat

Eurostat should be more actively engaged in the NSIs and European Central Bank simply to capture changes and new ideas. Eurostat should provide solutions, financial and technical support and, to large extent, remove barriers. I also would like to see that Eurostat took a greater economic responsibility for the NSIs. To exemplify: instead of placing the responsibility of each country to develop its own production system, the optimal solution could be that Eurostat itself developed a system and distributed it among the member states within European Union.

When it comes to scanner data, Eurostat should take the lead to develop a manual as soon as possible on how NSIs should ensure the quality of the scanner data they receive from the companies. Which method that is most suited for each survey could be the next topic to undertake. By the time the manual on quality assurance has been finalised, a significant amount of countries might have enough historical data to make empirical studies. Other topics that might be of interest for the NSIs are:

- A creation of a standard system for the use of scanner data
- Other technical aspects
- Customised quality assurance schemes for expanded use of scanner data
- How to manage data from multiple or a large number of companies
- A legal text for data capturing (obligation to provide data to the NSI)

Regarding the legal texts, it is very important for the NSIs, together with Eurostat and the European Central Bank, to agree as soon as possible on a text that obliges companies to provide data to the NSIs. A legal text would facilitate much for the NSIs to obtain scanner data and create good conditions for further research. If that option is not possible, then a second alternative would be for the NSI to get support from its own country's government.

GS1

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

Scanner data are big files of transaction identified uniquely by a product code. The EAN-barcode (International Article Number, formerly called European Article Number) is mostly a 13 digit combination (12 positions for data and 1 position control digit). The first three digits usually identifies the country where 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.

One of the international leading providers of EAN codes is a company named GS1. GS1 is an international non-profit association with member organizations in over 100 countries. Through cooperation, GS1 has developed sector-neutral standards and services which enable efficiency improvements in the flow of information and goods. 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. 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 language for grouping products in the same way, everywhere in the world. The “brick” level is the lowest classification level and consists of a group of narrowly defined products (e.g. perishable milk and milk products). Further note that the GPC scheme is translated to many other languages.

By using the GPC, a classification key between EAN codes and the COICOP classification could be created and that would, in turn, facilitate much of the work done at the NSIs by, for example, reducing the sampling process and facilitating comparisons between similar products, etc.

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 in any database. National organisations in each country decide the structure of the database, but the way data is synchronised is standardised 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 noted 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 optimised 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 the NSIs.

Concluding remarks

Scanner data might today be the most suitable method to collect prices for some surveys at Statistic Sweden, but this does not necessary mean that it will be the optimal solution for all time. As the author sees it, and based on his experience, this is a great time and a exceptional opportunity for NSIs to secure this type of data.

Finally, the author would like to emphasise that rules and practices on the treatment of scanner data must be harmonised across EU Member States as soon as possible. Otherwise, NSIs may make their own interpretations of the existing legislation and guidelines.

Let us inspire each other, let us be inspired by each other!