EUROPEAN FREE TRADE ASSOCIATION

EFTA Statistical Office Publication:

Transaction data in the consumer price index

In recent years, a massive increase in the availability of data has created new opportunities for National Statistical Institutes (NSIs). Utilizing the new data sources has the potential of both increasing quality and lowering the cost of producing official statistics. One of the areas of official statistics where the incorporation of new data sources presents most opportunities is the consumer price index (CPI), where especially data on individual purchases by consumers are increasingly becoming available and shows high promise. Such transaction data has been used by some NSIs for some years now, to varying degrees, and its usage is spreading rapidly. In this article, we will examine the state of play in the CPI in the EFTA Member States Iceland, Norway and Switzerland[1] for the most common form of transaction data – namely scanner data.

Consumer price index

The objective of the CPI is to measure price changes for goods and services purchased by individuals. It is generally published monthly, and as a measure of inflation, is an important economic statistic. The CPI is used for a number of purposes, such as rent adjustment, general contract adjustment, and as input for monetary policy. In many countries it is one of the most important determinants of the central bank interest rate.

To accurately reflect the development of prices each month, a large number of prices must be collected. Traditionally, this has mostly been achieved by either sending a price collector in person to the shops, or by sending a questionnaire to the businesses, or a combination of the two.

With the digitalization of society, this job has become more complex. Not only do new goods and services enter the market at high frequency, the way we purchase them has also changed. It is not that long ago that we did not have smartphones, for example. Now an increasing amount of people use their phones to stream entertainment, to purchase clothes and other goods, to book transport

---

[1] The fourth EFTA Member State, Liechtenstein, does not produce a CPI.

and to book stays in homes rented out by other individuals. In parallel to this, the importance of data for customer insight has increased.

The vastness and speed of changes entails both challenges and opportunities for the consumer price index.

Advantages and challenges

The digitalization of society means that the amount and scope of data has increased and is continually increasing. Making use of these new opportunities has become not only a possibility for the compilation of a good CPI, but a necessity. One new type of data has become especially important in recent years – transaction data.

The term transaction data technically refers to all data detailing transactions, but this article focuses on scanner data, a certain kind of business-to-consumer transactions, which is defined by ILO et. al. (2004) as "detailed data on sales of consumer goods obtained by 'scanning' the bar codes for individual products at electronic points of sale in retail outlets."

Scanner data has a number of advantages for the production of the consumer price index. Firstly, the sample size increases – we have more prices for more products sampled from more outlets over a longer period of time. Secondly, after the initial set-up, the economic cost is generally lower than with the traditional methods: Compared to employing a price collector, the cost for the NSI goes down; if the alternative was sending questionnaires, the cost for both the NSI and the businesses goes down, since sending scanner data can be fully automated.

While there are many advantages of using scanner data in the CPI, there are also some challenges that need to be mentioned. Firstly, there is some additional work burden for the NSIs, as well as initial cost of setting up the system. The reason for this is that there is likely to be an increased need for data cleaning and -processing, and that obtaining the data and creating a data pipeline for the transfer of data between businesses and NSIs require resources in the initial phase. Secondly, it increases the NSI's dependence on individual retailers, meaning that non-delivery of data will have bigger consequences than before. This problem can be abated by formulating good contracts and having a good relationship with the data providers. Thirdly, introducing new data sources in the CPI increases the complexity of the statistical production. This means that it might be more difficult to keep an overview of the data, and it requires more advanced IT competence from the CPI staff than what may have been required previously.

The use of scanner data in EFTA Member States

In this section we will briefly discuss the state of play for scanner data for the EFTA Member States that produce a consumer price index. This includes a history of their usage of scanner data, and how scanner data is used today in the countries, without going into technical detail.

Iceland[2]

Statistics Iceland has used scanner data in the production of CPI for several years. The work began in 2015, when Statistics Iceland started receiving and testing scanner data from three major grocery store chains. After about a year of analysing the data and preparing for implementation, Statistics Iceland implemented scanner data for groceries in the production of CPI in April 2016 (Ottawa paper).

Apart from the data source, the index calculation method itself did not change with the implementation of scanner data. This allowed Statistics Iceland to benefit from the expansion of data, while minimizing any operational risk associated with changing both the data source and the fundamental methodology at the same time. This also gives Statistics Iceland the opportunity to gain experience with using scanner data, and to analyse the data further, thereby being well-prepared to increase utilization of data in the future.

As is common in many countries, a large part of the turnover in grocery stores comes from a small number of products. This allows Statistics Iceland to capture about 40 percent of the turnover from grocery stores using only about 4 percent of the items.

Currently, about 16 percent of the total CPI in terms of weight is covered by scanner data.

Norway[3]

Statistics Norway started exploring the possibilities of using scanner data from grocery stores in the late 1990s, and by 2001 received data from all major grocery store chains as well as the state alcohol monopoly. Since then Statistics Norway have gradually expanded its usage of scanner data, first by getting data from petrol stations and pharmacies, before later expanding further by including kiosks and a discount store. The sub-index of food and non-alcoholic beverages has been based exclusively on scanner data since 2005.

More recently, in 2019, Statistics Norway started using scanner data from a large sports retailer, which is used alongside questionnaire data for sports equipment and some relevant clothing items. From 2020, data from a clothing chain is now also used for a number of clothing items, alongside questionnaire data.

Some challenges to the expansion of items beyond groceries have been methodological issues for items with high "churn", and limitations in resources for getting hold of data from new chains.

By items with high churn we mean items that are frequently replaced – a prime example is clothing items. Clothing items generally have a short shelf-life before they are replaced by items for the next season. The problem arises if the items enter the market at a normal price, then leave the market

---

[2] This section is based on Guðmundsdóttir & Jónasdóttir (2016), Guðmundsdóttir (2017), and private correspondence with Heiðrún Erika Guðmundsdóttir.

[3] This section is based on own knowledge from working at Statistics Norway, as well as private correspondence with Randi Johannessen and Ragnhild Nygaard.

after a sale, and do not return. With clothing items, this can for example be a pair of jeans that is introduced before spring, is sold at a discount after spring is over, and replaced by a newer style the next season. When using large-scale transaction data[4] with traditional index calculation methods, this would lead to the index "drifting" down. This means that the price index for jeans would go down from one season to the next, even if jeans cost the same, or even more, than the previous season. This problem is avoided with traditional data collection methods, because product replacement is done when prices are collected.

With new advances in index calculation methods that try to overcome these challenges, it is now possible to utilize scanner data on an increased variety of consumer goods.

Currently, about 22 percent of the weight of the total CPI basket in Norway is covered by scanner data.

Switzerland[5]

The work on incorporating scanner data in the Swiss CPI began in 2006, when the 10 biggest relevant retailers in Switzerland were asked about their readiness and willingness to provide scanner data for statistical purposes. As all 10 retailers indicated that they were both ready and willing to send data, this served as a clear indication that the Swiss Federal Statistical Office (FSO) could start planning to receive and make use of scanner data.

A pilot project in 2008 with the largest retail chain offered the first opportunity to harvest the fruit of previous labour by introducing transaction data in the production of the CPI. A challenge was that FSO did not have a tool for collecting scanner data at this time, so an ad hoc solution was used for this pilot exercise. The next step was to create a generic tool for collecting scanner data, and this work was completed in 2009.

With the new price collection tool, it became possible to introduce more retail chains, and this work started in 2010, when a second retail chain was introduced, before two more followed in 2011/2012, and then another one in 2018.

The currently included retailers cover about 80 percent of the grocery market. Most of the most common and popular items in grocery stores are currently covered by scanner data – food, beverages (alcoholic and non-alcoholic), tobacco, household items, pet food, personal care products, et cetera. Scanner data now cover about 12 percent of the total CPI basket in Switzerland. In total about 7 500 price series are utilized each month.

---

[4] Using small-scale transaction data, it is feasible to manually pick replacement items when items leave the market, thereby imitating traditional data.

[5] This section is based on Eugster & Kleisl (2018) and private correspondence with Jean-Daniel Kleisl.

Developments at European level

As we have seen, the use of scanner data in the CPI varies between EFTA Member States. This is also the case for the Member States of the European Union. Many countries have yet to implement scanner data in the CPI, while in other countries it is already among the most important data sources.

The increased usage of scanner data has been a new challenge in Eurostat's efforts to harmonize the CPI's across the EEA. The reason for this is that it poses new methodological challenges, meaning that for "complex" items (especially items with high churn), completely new index calculation methods need to be developed to avoid problems specific to this type of data. This work is ongoing, and a final consensus as to which method is the best has yet to be reached.

Eurostat is active in coordinating methodological efforts and has also created a practical guide for processing supermarket scanner data for use in the Harmonized Index of Consumer Prices (HICP)[6]. The guide includes pointers on what to consider when obtaining data, sampling considerations, as well as considerations about processing the data. Scanner data from supermarkets is "low-hanging fruit" in this context, as there exist good methods for calculating indices for this type of data, and in addition supermarkets generally make up a large part of consumption by individuals.

As of 2017, about one fifth of all EU Member States used scanner data (Eurostat, 2017), but the number is expanding each year. The reasons for not using it varies from difficulties in getting hold of data, to resource constraints and methodological challenges.

Conclusion

To conclude, it is fair to say that the EFTA countries are advanced users of scanner data. There could be many reasons for this. To even get hold of the data, a high level of trust in society is very helpful. All EFTA countries score higher than the EU average on trust (Eurostat, 2019), and this might have helped the EFTA countries in obtaining data and establishing a working relationship with the business community.

Another reason for this may be that the EFTA countries are relatively advanced economies. The GDP per capita can be interpreted as a measure of how advanced a country is, and the EFTA countries all score higher than the EU average on this measure (Eurostat, 2020).

So what is the future for transaction data in the CPI? An increased usage of scanner data from physical stores is sure to take place in the coming years, especially after a consensus is reached on a suitable methodology for processing scanner data items with high churn.

Another trend that is worth mentioning is the continued increase in online shopping, which also brings a number of interesting new challenges for NSIs. Several lessons learned from processing scanner data will be directly applicable to transaction data from online outlets and to web scraped data, but they will also bring their own unique challenges.

---

[6] The HICP is a version of the CPI where comparability across countries is achieved through harmonization of methods and definitions. Read more about the HICP on the Eurostat website: https://ec.europa.eu/eurostat/web/hicp

In conclusion, transaction data enables the production of better quality statistics, at a lower economic cost. Due to the rapid changes in our consumption patterns, the availability of new data sources, and the ever-present challenge for NSIs to produce better statistics at a lower cost, the progress towards more usage of transaction data, and other forms of so-called "big data" is probably inevitable if producers of official statistics are to stay relevant to users in the years to come.

References

Eugster, O., & Kleisl, J.-D. (13.-14.11.2018). Scanner Data in the Swiss CPI: An Alternative to Price Collection in the Field. Presentation from bilateral meeting with Germany.

Eurostat. (2017). Practical Guide for Processing Supermarket Scanner Data. Retrieved from: https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf

Eurostat. (2019). Average rating of trust by domain, income quintile, household type and degree of urbanization. Retrieved from: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw04&lang=en

Eurostat. (2020). Real GDP per capita. Retrieved from: https://ec.europa.eu/eurostat/databrowser/view/sdg_08_10/default/table?lang=en

Guðmundsdóttir, H. E., & Jónasdóttir, L. G. (2016). Scanner Data: Initial Data Testing. Paper for the Meeting of the Group of Experts on Consumer Price Indices, Geneva, Switzerland, 2 - 4 May 2016. Retrieved from: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_1_Iceland_Initial_data_testing.pdf

Guðmundsdóttir, H. E. (2017). How did Statistics Iceland start using scanner data? Paper for the 15th Meeting of the Ottawa Group on Price Indices, Eltville am Rhein, Germany, 10. – 12. May 2017. Retrieved from https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/How%20did%20Statistics%20Iceland%20start%20using%20scanner%20data%20-%20Gudmundsdottir%20paper.pdf

ILO, IMF, OECD, Eurostat, UNECE, & World Bank. (2004). Consumer Price Index Manual: Theory and Practice. Geneva: International Labour Office. Retrieved from: http://www.ilo.org/public/english/bureau/stat/guides/cpi/index.htm

Author:

Kristian Harald Myklatun

Kristian.MYKLATUN@ext.ec.europa.eu

Contact:

EFTA Statistical Office

5, Rue Alphonse Weicker

L 2920 Luxembourg

EFTA-Lux@ec.europa.eu

The **European Free Trade Association (EFTA)** is the intergovernmental organisation of Iceland, Liechtenstein, Norway and Switzerland. The EFTA Secretariat has three duty stations:

- Geneva   in charge of Free Trade Agreements

- Brussels   in charge of the European Economic Area

- Luxembourg   EFTA Statistical Office

The objective of the EFTA Statistical Office in Luxembourg is to sustain and promote the participation of EFTA States as full members of the European Statistical System (ESS) which contributes to a European culture of evidence-based policymaking and monitoring.