Electronic Data Collection Methods; the Use of Transaction Data for PPI

Heiðrún Guðmundsdóttir, heidrun.gudmundsdottir@statice.is,
Rósmundur Guðnason, rosmundur.gudnason@statice.is,
Guðrún R. Jónsdóttir, gudrun.jonsdottir@statice.is.[1]

Statistics Iceland
Borgartun 21a
150 Reykjavik
IcelandWebsite: www.statice.is

Abstract

This paper describes electronic data collection methods and the Icelandic PPI, where electronic transactions data sets are used to calculate a superlative index. PPIs measure the rate of change in prices of goods and services bought and sold by producers. Electronic methods are used to collect data from the information systems of firms or scanner data from the point of sales. Electronic data sets from the accounts of firms include information about sales of producers' goods. They originate from sales data of the firms but the buyer is always known and that information can be utilized in the index calculation. A byproduct of the superlative PPI price index calculation is a quantity index for industrial production, a measurement of the changes in the quantity vectors used.

Key words: Electronic data collection, producer price index, superlative indices, Fisher index, transaction data, scanner data, industrial quantity index.
JEL: C43, C81, D11, E31.

[1] The authors like to thank Lára G. Jónasdóttir, Stefán L. Sigurþórsson, Hallgrímur Snorrason, Dr. Tsutomu Watanabe and the participants at the 2008 World Congress on National Accounts and Economic Performance Measures for Nations, Washington DC 12-17 May 2008, the Joint UNECE/ILO Meeting on Consumer Price Indices, Geneva 8-9 May 2008 as well as the participants at the 24th meeting of Nordic statisticians, Reykjavík 25-28 June 2007 for helpful comments.

Electronic data collection methods

Two types of electronic data sets can be collected either directly from the books of firms or as scanner data from the point of sale of retailers.

PPIs measure the rate of change in prices of goods and services bought and sold by producers. In the case of the Icelandic PPI, the electronic data sets, from the accounts of firms include information about sales of producers' goods. They originate from sales data of the firms, the buyer is always known and that information can be utilized in the calculation of the index.

Electronic information is collected from the wage records of firms and subsequently wage statistics are compiled by Statistics Iceland. An ongoing development program at Statistics Iceland aims at designing a new data collection system for foreign trade in services using electronic methods for collecting information directly from firms. However in many cases, the detailed information in the accounting systems of these firms turns out to be insufficient for statistical needs. Consequently, to make such data usable the firms in question have to improve and expand their data records.

The CPI measures price changes of goods that consumers buy. The scanner data used in this case originate at the point of sale. It is very detailed but the buyer of the good is not known. Electronic data sets are used for example in the Netherlands and Norway for CPI compilation.

Scanner data can also be collected through receipts in household expenditure surveys and in such cases information on expenditures is directly connected to the household and the buyer is always known (Guðnason 2004).

Electronic data collection for the PPI

The Icelandic economy is very small. It has two important goods production sectors, the seafood sector and the power-intensive industry sector. As there are few large firms covering a considerable part of the production, firms are chosen judgmentally into the sample.

In the beginning of the Icelandic PPI development, the aim was to use as detailed data from the firms' accounts as possible, without losing sight of the necessary effectiveness of the collection method. This objective was facilitated by collecting electronic data from the firms' accounting systems. The method has resulted in low collection cost and light response burden

for the participating firms, but more importantly it has made it feasible to collect transaction prices. This level of data precision is essential to the quality of the index compilation.

The conduct of Statistics Iceland is in accordance with Icelandic laws obligating firms to deliver requested data. This obligation is without any payment or retribution to the firms for the use of their data. However, all statistics published based on these data are available for anyone to use gratuitously. Statistics Iceland must adhere to a strict code of conduct in handling and protecting the data with confidentiality. Communication at peer level with the firms is considered to result in higher quality data and better response. This is why one of the most important issues to the incorporation of new firms is one of building trust and conveying total confidentiality to every new member in the sample.

When firms were approached in the beginning it proved to be an easier task than expected to convince them to participate. Some reluctance was anticipated given the delicacy of the data as they contain business transactions in full detail, on monthly bases. Besides prices and quantities, the data shows product numbers and customer identifiers as well as business terms for the customer at the time of the trade, allowing unit prices to be calculated with discounts taken into account. It has worked in the favor of Statistics Iceland, as a statistical authority, to be highly respected nationwide.

Since the history of the index in Iceland is short it turned out to be an obstacle that the firms did not realize its usefulness at first; however, as the general knowledge and use of the index has spread, firms have been participating more willingly. Some firms get reimbursed for their contribution with a feedback in the form of index results specifically compiled for their own production. The firms have considered this to be quite valuable especially in comparison to aggregate results; therefore this action has increased general contentment and strengthened relations between the firms and the Statistical Office. Services of this kind are available to all participants.

By collecting the data electronically, the time spent by the firms on preparing and handing them over is minimized as well as the risk of wrong information being reported. In order to ensure this, a data report from the firms' accounting systems needs to be structured at the outset, but once it has been established the data delivery is effortless. Technical barriers to the collection method have proven to be surmountable as most firms have the ability to prepare

the necessary reports. Exceptions to this are few and concern mainly small firms. Occasionally Statistics Iceland has assisted by buying technical services in such cases.

The Icelandic PPI

The Icelandic PPI currently covers production of marine products and all domestic manufacturing production with the exception of NACE-classification 22.1 (publishing) and 35.1 (building and repairing of ships and boats). Services are currently not included in the PPI and neither is energy production. The marine products industry has the biggest production share with nearly 37% of the index, followed by the power-intensive industry with more than 20% and food production which is just around 17%. The sample of firms for the marine products cover about one third of the total production and in manufacturing the sample covers approximately 45% of the total production. Around 100 thousand price quotations are collected every month and used in the calculation of the PPI index.

Weights and compilation at the elementary level

When the present PPI was first compiled in 2004 it was published quarterly but since the beginning of the year 2007 it has been published monthly. From the start, detailed information on prices, quantities and discounts has been collected, and a superlative approach has been applied in the compilation at the elementary level. By using the Fisher ideal index formula, changes in unit prices enter the PPI calculation whenever a product is sold to the same customer in two consecutive months.

The Paasche and the Laspeyres inputs in the Fisher formula are given by:

$$\text{Paasche index} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} \quad , \qquad \text{Laspeyres index} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$$

Where the transaction prices, p, are weighed by the transaction quantities, q. The raised indicator refers to the evaluation month, 1, or the base month, 0, and the lowered indicator refers to products. The Fisher index is then calculated as a geometric mean of the Paasche and Laspeyres indices.

The method of collecting detailed transaction records is particularly useful in firms which have similar production each month, such as in the food industry and the power-intensive

industries. Difficulties arise when firms with specialized production are included, such as printing or machinery production where the production lines are changing all the time. In these cases model pricing is used and in some cases component pricing by pricing units used in production or typical types of work conducted, such as a printing job for a book. In all cases, however, information about total production is gathered even if the strict condition for using all prices is not fulfilled. In such cases production information is used to calculate quantity vectors for the weights of the superlative index.

Weights and compilation at the upper level

When the PPI was introduced in 2004 the index at the aggregate level was calculated as a fixed base Young index. The weight information was gathered from the yearly industrial production survey (PRODCOM) which was available with a lag of two years. In the beginning of the year 2006 the calculation methods were changed and a superlative compilation with the Fisher formula was extended to the aggregated level. At that point the two most recent quantity vectors available were two and three years old. The weights are still based on the annual PRODCOM surveys which are now released approximately six months after the end of the year. The base for the Paasche weight is derived from the latest available release, whereas the base for the Laspeyres weight is obtained from the preceding release.

The production value, w, that company, j, gets allocated from the relative PRODCOM surveys, is extrapolated to the evaluation month, t, and becomes, w_j^t. The extrapolation is based on the monthly value of sold production, V, from the firms in the PPI sample. The monthly weight can then be described as follows:

$$w_{j,Paasche}^{t} = w_{j,Paasche} \frac{\sum_{k=1}^{12} V_j^k}{\sum_{Paasche} V_j^k}$$

$$w_{j,Laspeyres}^{t} = w_{j,Laspeyres} \frac{\sum_{k=13}^{24} V_j^k}{\sum_{Laspeyres} V_j^k}$$

The denominator sums up the total production value in the same year as the respective PRODCOM survey. The numerator sums up the production value over a twelve month period. In the Paasche weight this period contains the evaluation month plus the previous eleven months. The numerator in the Laspeyres weight sums up the same months from the previous year.

Further development and implementation difficulties

Ongoing is further development to increase the coverage of the PPI, e.g. by including the energy production sector (electricity production) and to increase coverage to services.

The main benefit of collecting electronic data directly from the accounts of firms is that it enables the calculation of a superlative PPI index on a monthly basis. The fact that product quantities are included in each transaction has also led to the development of a monthly quantity index for industrial products. This intriguing byproduct of the PPI data is still at an experimental stage but has shown promising results.

The basic directive for the weight derivation is to use the most recent and the most fitting information available. However, various circumstances may cause some implementation difficulties. Information on production value may not always be available, for example where component prices are collected. Moreover, information on the production value, when a company is added to the sample, may not span as far back as is necessary for the extrapolation of the base weights. One option, to avoid this issue, would be to use production value series from a company in similar production. Another option would be to use export statistics or value added tax records. For instance, weights for marine products rely on detailed foreign trade statistics.

Conclusion

Electronic data collection methods offer intriguing new possibilities in processing data and compiling indices. With the firms' accounts as a source, information on products, prices, and quantities as well as buyers is available and it becomes feasible to use a superlative approach. The Icelandic PPI uses the Fisher ideal index formula on both levels of the compilation.

References

Diewert, W. E., (1976), Exact and Superlative Index Numbers, *Journal of Econometrics* 4, 115-145.

Guðmundsdóttir, H., Guðnason, R., Jónsdóttir, G.R., (2007), The use of transaction data records in the Icelandic PPI, *24th meeting of Nordic Statisticians, Measurement problems in statistics,* Reykjavík, Iceland, 25-28 June 2007.

Guðnason, R., (2004), The Receipts Approach to the Collection of Household Expenditure Data, *SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity* Vancouver, Canada; June 30 - July 3, 2004.

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

International Monetary Fund (2004), Producer Price Index Manual: Theory and practice, P. Armknecht (ed.), ILO/IMF/OECD/UNECE/Eurostat/The World Bank, Washington, D.C.

Statistics Iceland, (2007), Producer price index 2003-2007, *Statistical Series, Hagtíðindi, 2007:4,* 22 March 2007.

Statistics Iceland, (2008), Producer price index 2007, *Statistical Series, Hagtíðindi, 2008:2,* 6 March 2008.