[1]Use of cash register data

Rósmundur Guðnason
Statistics Iceland
Skuggasund 3
Is - Reykjavík 150, Iceland
rosmundur.gudnason@statice.is

Hallgrímur Snorrason
Statistics Iceland
Skuggasund 3
Is - Reykjavík 150, Iceland
hallgrimur.snorrason@statice.is

1. Summary

Taking advantage of the huge amount of available electronic data is becoming an increasingly important task in the field of official statistics, especially in CPI work.

The use of data coming from the bar code cash receipts collected in the 1995 Icelandic Household Budgets Surveys (HBS) has rendered more accurate estimates of private household consumption than previous traditional surveys and given better information about types and brands of goods purchased and on outlets. This utilisation will probably lead to new developments in weighting procedures as well as facilitating the use of scanner data collected directly from outlets.

Statistics Iceland will in future yearly HBS´s put a great deal of effort to collect more and better information from these receipts for obtaining more reliable and up to date household budget statistics, not least for rebasing the CPI.

2. Cash register data from HBS´s

Through Household Budgets Surveys (HBS´s), individual consumption can be recorded at the point of sale by using bar code cash receipts. This opens up a whole range of new possibilities for collecting more exact information from such surveys.

The use of bar code cash receipts was tried in the 1995 Icelandic HBS (Guðnason, 1997). The main aim of using this method was to lighten the response burden of the survey respondents. Instead of itemising the goods bought, the respondents only needed to write into their survey diaries their purchases and to put the itemised receipt received into a special pocket in the diary. This method could be used as it had been observed that most of the major stores give receipts that itemise the purchase and show the name, brand and the volume (weight, number of pieces etc.) of the goods bought.

The information on the receipt is very detailed and opens possibilities to view many sides of consumer behaviour. The following information is usually available:

- **Name of the outlet**: Opens possibility to evaluate market shares and consumer buying patterns.
- **Date and time of purchase**: Facilitates mapping the behaviour of different consumer groups.
- **Description of items, quantity, price and value**: The cash receipts in most larger stores show the type of good purchased, also often the brand, size of package, the quantity

---

bought, the price of the item and the value. On top of expenditure data obtained in this way, this information can for example be used for selecting representative goods, brands and packages for CPI-price collection. In the case of fresh vegetables, fruit and other goods weighed at the store, the receipts will often show the unit price of each item as well as the total item weight and value of purchase. Alternatively, if the unit price is not shown, this allows unit values to be calculated. Finally, the data on the total number of item and purchase value can be used for checking purposes and as information about different market shares.

- **Form of Payment**: The receipts show by what medium the purchases are paid for; by credit cards, debit cards, checks, cash.
- **VAT amount for different VAT rates**: In Iceland the rate for food is lower than on other goods and can be sorted out on the receipts.

Thus, each receipt contains very detailed information which in a HBS is connected to specific households. The results show that this method opens a whole range of new possibilities for collecting data and viewing consumer behaviour from different angles. It also enables precise information to be gathered about consumer activities at much lower level of effort than the traditional survey methods do, and it establishes a much better link than before between the goods purchased and the buyer.

The experience of using this method in the 1995 Icelandic HBS proved very positive. Nearly 41% of all entries in the survey were drawn from the bar code cash receipts. Of all transactions recorded in this way, 64% belonged to the category food and non-alcoholic beverages and 36% to all other goods and services. As was to be expected, the ratio of number of items on cash receipts of the total number of items recorded was much higher in food and non-alcoholic beverages than in other expenditure groups, 53 % in the former as compared with less than 20% in the latter.

Table 1. Items recorded by sources in the Icelandic 1995 HBS

COICOP	Group	In percentage of total records from diaries	In percentage of total records from bar code cash receipts	In percentage of total records
1	Food and non-alcoholic beverages	47,5	52,5	63,9
11	Food	45,5	54,5	55,5
111	Bread and cereals	50,5	49,5	11,7
112	Meat	38,4	61,6	4,9
113	Fish	49,6	50,4	1,3
114	Milk, cheese and eggs	40,5	59,5	14,8
115	Oils and fats	40,1	59,9	1,7
116	Fruit	43,2	56,8	4,2
117	Vegetables, including potatoes etc.	39,2	60,8	6,2
118	Sugar, chocolate etc.	65,5	34,5	6,8
119	Food products n.e.c.	37,6	62,4	3,9
12	Non-alcoholic beverages	60,3	39,7	8,5
121	Coffee, tea and cocoa	42,1	57,9	1,0
122	Mineral waters, soft drinks, juices	62,9	37,1	7,4
	Other groups	80,4	19,6	36,1
	Total	59,4	40,6	100,0

For a HBS, it is essential that the cash receipts are well itemised. In the 1995 Icelandic survey the bar code cash receipts were fully satisfactory. Initially, there where certain doubts about applying this method and about the quality of the receipts. In particular, there were worries that receipts would get lost, that the itemised receipts and the amounts entered into the diaries would not match properly and that a large ratio of the receipts would not be itemised enough. This was not the case. The cash receipts were seldom missing and the diaries showed these purchases clearly. Furthermore, a very little part of the receipts (1.5%) had information that was not detailed enough which is somewhat lower than in diaries in former HBS's.

3. Using HBS cash receipt data for weighting

With the aim to check the results of the 1995 Icelandic HBS, a comparison was made between the survey data and information based on scanner data from the biggest supermarket chain for that year (Guðnason, 1998). The scanner data from the supermarket was aggregated covering about 400 groups and items for the comparison. The data involved expenditure on food, beverages, and a few other groups, altogether accounting for 19% of the total CPI weight.

The results from the HBS and the information from the supermarket chain for the year 1995 gave similar results. In two groups, tobacco and prepared food, the scanner data showed higher expenditure percentages than the HBS data set. For these particular groups, there may be two explanations. Firstly, there is probably a tendency to underreport these items in the HBS and secondly, tobacco and prepared food may be included in the information not coming from the bar code cash receipts. On the whole this comparison showed that the results from the HBS compared well with similar data from the supermarket.

The information gathered from the HBS can be used as a source of expenditure weights for different outlets. Supermarkets have their biggest share of sales in oil and fats (COICOP group 0115), fruit and vegetables (0116), food products n.e.c. (0119) and coffee (0121). In these groups the big supermarkets have about 30% of their total sales. Their lowest coverage is in fish and tobacco as shopping for these goods takes place in smaller specialised outlets. On the other hand the relative composition of the sales differs considerably between the supermarket chains.

The results indicate that a different method could be applied for weighting the food category in the Icelandic CPI. There are now four regional indices calculated for different parts of the country. The average price change for the whole country is calculated as a weighted average of the change in these four indices. The weights are partly based on regional sales figures. The survey indicates that instead of this, the very detailed expenditure shares by outlets could be used as weights. The average price change would then be measured as a weighted average of the expenditures in the different type of outlets independent of geographic location.

Further, shopping habits of households as mapped in the HBS could be used as a source for weights. This would be done by utilising information on detailed expenditure of typical costumers in each type of outlet. Calculation of the average price change would then be based on the expenditures of different households in the outlets. So for each outlet there would be different indices calculated for all types of households.

4. Possibilities and limitations of using scanner data

Cash register data consists of electronic data records that are either given as printed receipts to buyers or stored as scanner data in data bases of sellers. This information is generated for different purposes as the consumer gets information on his purchases from the receipt but the seller receives the same information at the point of sale. In other words, the same basic information may be viewed from the consumers side and from the point of view of the seller. Both types of data include quantities, values and types of goods but whereas the data from receipts can be connected to a specific type of household, the sales scanner data cannot.

Another difference between HBS data and scanner data from sellers is that the HBS data is based on a sample but the scanner data on the total sale of the outlet. These factors have to be taken into consideration when comparison is made between using bar codes cash receipts in a HBS and the scanner data from shops. One of the advantages of the HBS data use is that it also covers goods bought from shops that are not using scanner data and although the use of electronic data records is increasing every year, parts of the retail trade are not using scanner data. The most common use is in the food and beverages sectors. Scanner data in other field of retailing is increasing such as in drug stores, book stores and in the consumer durable sectors.

Of the total scanner data flow it is easiest to use the data on food and beverages. These products are relatively homogenous and easy to identify. Examples of this kind of products are coffee, soft drinks and detergents. However classifying many other food products can be difficult and special attention has to be given to data on fresh food. Retailers often do not use standardised bar codes for meat, fish, vegetables and other fresh food so for certain items it is easier to use information from the HBS.

The most common use of scanner data is in the field of marketing for evaluation of market shares. This is done by many marketing firms of which AC Nielsen is probably the best known. Information from the HBS bar code cash receipts makes it possible to calculate quantities used by consumers in similar way as the marketing firms do. This means that the information in the HBS for consumption of food can be expressed as quantity vectors instead of value shares and the results for each homogenous group calculated as unit prices and used for measurement of price changes.

Scanner data of this kind have been used intensively in last years for research for example to evaluate the influence of different sampling methods on price measurement (Haan et al., 1997) and other price measurement issues (Silver, 1995, Reinsdorf, 1996, Dalén, 1997). Another issue connected to this is the appropriate price units when detailed quantity information is available (Diewert, 1995).

The data used for some of these research projects is usually based on point of sales information from outlets. As this data comes from sellers and is not connected with specific households it does not necessarily reflect fully the consumption of individuals. This is a limitation as regards HBS, but not as regards research in sampling, such as on the selection of items etc..

The scanner data cover all goods sold in outlets and the amounts of data processed can be gigantic so it is difficult and costly to categorise. Using scanner data for monthly price collection has not been done on a large scale to our knowledge and one of the main reason for that is probably the large amount of data involved. Treating these data sets can be very time consuming and costly. If these are to be utilised sampling methods have to be applied. Hence, the question arises if the one of the keys to such use is the application of HBS scanner data on consumer behaviour and consumption to the enormous scanner data recorded continuously in the databases of sellers.

REFERENCES

Diewert, Ervin W., (1995). Axiomatic and Economic Approaches To Elementary Price Indexes. Department of Economics University of British Columbia, Discussion paper No.: 95-01, Vancouver Canada, January 1995.

Dalén, Jörgen, (1997). Experiments with Swedish Scanner Data. In Proceedings of the Third Meeting of the International Working Group on Price Indices (ed. B.M. Balk), 163–168. Research Paper no. 9806. Statistics Netherlands, February 1998.

Haan, Jan De, Opperdoses Eddy, Schut, Cecile, (1997). Item Sampling in the Consumer Price Index: A Case Study Using Scanner Data. Statistics Netherlands, 1997.

Guðnason, Rósmundur, (1997). Improved methods for the evaluation of the composition and quantity of household goods. In Proceedings of the Third Meeting of the International Working Group on Price Indices (ed . B.M. Balk), 129–131. Research paper no. 9806. Statistics Netherlands, February 1998.

Guðnason, Rósmundur, (1998). Comparison of different sources for weights of selected groups in the Icelandic CPI: Consumers bar-code receipts vs scanner data from supermarkets. In Proceedings of the Fourth Meeting of the International Working Group on Price Indices (ed. W. Lane), 205–209. Bureau of Labour Statistics, Washington, DC, January 1999.

Reinsdorf, Marshall B., (1996). Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee. Bureau of Labour Statistics, Working Paper 277, Washington, DC, April 1996.

Silver, Mick (1995). Elementary Aggregates, Micro Indices and Scanner Data: Some Issues In The Compilation Of Consumer Price Indices. Review of Income and Wealth, series 41, Number 4, 427-438. December 1995.