Price Collection and Quality Assurance of Item Sampling in the Retail Prices Index: How can Scanner Data Help?

David Fenwick & Adrian Ball.
UK Office for National Statistics,
1 Drummond Gate, London SW1V 2QQ,
United Kingdom
Email: david.fenwick@ons.gov.uk & adrian.ball@ons.gov.uk,

Peter Morgan and Professor Mick Silver
Cardiff University

1.0   Introduction
The UK Retail Prices Index is an important and widely used macro-economic indicator both in the formation and monitoring of economic policy and for the indexation of welfare and other state benefits. Its accuracy is of paramount importance. A 0.1 percentage point overstatement or understatement of the inflation rate would affect Government expenditure and receipts by about £100 million a year and could mislead managers of the economy. The Office for National Statistics continually seeks improvements in the methodology used to compile the RPI, and so has a continuous research programme.

A number of studies in the past have pointed to scanner data being used in the compilation of consumer price indices, either as a direct source of price data or for estimating the value any change in quality when there has been an unavoidable replacement of an item in a fixed basket.. It has also been suggested that scanner data has the potential to contribute to the effectiveness of probability sampling procedures. To test this a joint research project was set up with Cardiff University to explore the potential for scanner data to be used as a diagnostic tool for the identification of potential deficiencies in the sample of items used for the RPI and to provide solutions.

2.0   The RPI data collection system

2.1 Background
Data are collected for the RPI in two ways: local collection by price collectors who visit shops to determine prices charged in each location; a central collection for those stores where a local collection would not be cost effective and for those items, such as utilities where the purchase does not usually take place at a local outlet. Since 1995 the local collection of price data has been contracted out to a private sector company. It is the local collection which is the focus of this work.

2.2 Sampling procedures for local price collection
Current methodology for the selection of locations from which we collect local prices aims to give each shopping centre in the UK a probability of being selected for the price collection equal to its proportion of total consumer expenditure. This is achieved using a two stage hierarchical sampling frame based on geographical regions. A total of 141 locations are required for local price collection and the number to be selected within each of the regions is determined by taking a proportion equal to the proportion of total UK expenditure that each region attracts. Within regions the appropriate number of locations are chosen at random using a probability proportional to size technique, using number of employees in the retail sector as a proxy for retail expenditure in each area. Each selected location is then enumerated by price collectors to produce a sampling frame from which outlets are randomly selected. Multiple and independent retailers are separately identified.

The selection of representative items to be used to calculate the RPI is purposive (i.e. judgmental not random). All categories of expenditure on which, according to the household expenditure survey, significant amounts of money are spent are arranged into about eighty sections and items are chosen to be representative of each section

The selection by the price collector of the products and varieties to represent the selected items is also purposive and carried out in the field. Price collectors are instructed to choose the product or variety in the selected shop that most represents sales in the area of that particular item

2.3 Coverage of the RPI
The RPI is an average measure of the change in the prices of goods and services bought for the purpose of consumption by the vast majority of households in the UK. The reference population is all private households with the exception of a) pensioner households which derive at least three-quarters of their total income from state pensions and benefits and b) "high income households" whose total household income lies within the top four per cent of all households.

3.0 Characteristics of scanner data
Scanner data are based on electronic point of sale (EPOS) data recorded by bar-code readers at the time and point of purchase. As more shops move over to bar-code readers the potential benefits to compilers of consumer price indices increase. Scanner data provides the potential to deliver up-to-date and accurate information on: sales volume, sales value, characteristics of the product and characteristics of the outlet.

In reality the market coverage of scanner data varies between different shop types and products and the amount and detail of data actually available can vary depending on the commercial source and which product is being examined. In addition definitions may not be compatible with index compilation. For example, the average transaction "price" recorded by scanner data relates to unit values (i.e. average revenue generation). It does not take into account the specific needs of index compilers to measure according to a strict set of pre-determined rules that disallow certain discounts such as those relating to damaged stock or management discretion. The latter should be excluded from the RPI but will be included implicitly in scanner data (see next section).

In addition experience indicates that a great deal of expertise and effort is needed to clean scanner data, adjusting for such things as re-used bar-codes, in order to make it usable for statistical purposes.

4.0 Research design
The research consisted of two separate but related exercises:

Testing representativity through benchmarking the RPI product and variety selection against corresponding scanner data;
A comparison of RPI average unit prices and price changes with the corresponding unit values and unit value movements obtained from scanner data.

The benchmarking exercise involved a comparison of the relative distributions by product and variety for each of five pre-selected items: televisions; washing machines; vacuum cleaners; dishwashers; and cameras, chosen specifically because of their high turnover of models and features, reflecting a high level of technical innovation.

5.0 Representativity of product and variety selection
The purpose of this stage of the research was to determine the extent to which current selection practices may lead to the choice of an unrepresentative sample of products and varieties for pricing. It looked at overall distributions obtained from the selection procedures used in the RPI and compared these with the overall distributions of sales given by scanner data. Monthly data were compared for the period from January 1998 to December 1998. This was done at an aggregate level, RPI and scanner data were not linked in any way to facilitate this exercise..

5.1 Summary of results
Figures for the top 10 sellers for the five product groups in September 1999, according to sales volume from scanner data, show some very interesting patterns. In general collectors tended to choose items that were good sellers, though frequently they over collected from models that were only mildly popular. Some of the most obvious examples of discrepancies were within dishwashers. Here the top selling model, which accounted for around one fifth of sales, was represented by just 2 per cent of quotes, and the seventh most

popular, which only accounted for 4 per cent of sales was represented by over 20 per cent of quotes. This pattern was repeated in other items.

5.2 Interpretation

Any interpretation of the results clearly depends as much on the quality and coverage of the scanner data as on the representativity of the RPI sample. However, assuming the initial selection of outlets is satisfactory (and there is no evidence to suggest it is not), it does seem to indicate two things: despite the instruction to the price collector to chose a product variety that is representative of the sales of that item in each area, the pricing of items can apparently be skewed towards products and varieties which scanner data indicate have relatively small sales; the fixed basket approach - where products and varieties as well as items are only reviewed on an annual basis leads to the sample becoming increasingly unrepresentative as the "fixed" selection of goods in the basket ages over the year.

But do these things matter? Clearly this depends on the extent to which there is a noticeable impact on the published index and the measured rate of inflation. The following section reports on the second stage of the research designed to test whether this is so.

6.0   Average unit prices and price changes

The purpose of this stage of the research was to observe for specific product varieties the extent to which the price levels and changes observed by price collectors in the field differed from the price levels and changes shown by scanner data. This required scanner data and RPI data to be matched by models within each product. The success of this was variable, depending on the item descriptions given by the collectors, being most successful for dishwashers, washing machines and vacuum cleaners where over 70% of RPI observations (representing about 50% of RPI product varieties) were successfully linked to scanner data. and worst for cameras, where only about a half of RPI quotes (representing about a third of RPI product varieties) were matched to scanner data. Further analysis indicated that in some instances there were significant differences between the mean **average price level** for the full set of RPI quotes and the subset where there was a successful match with scanner data for a product variety. This was most marked for television sets and washing machines.

The figures suggest that, in general, there is no pattern across the items as to whether the matched sample had a higher or lower mean price than that for all RPI quotes. However, within an item the direction of the difference remained the same over time, with the sole exception of cameras where the differences are small. This may suggest that an effect is present within items, though this is difficult to test with a weighted mean, and a serially correlated sample. Differences were also detected between average price **changes** shown by the full scanner dataset and those shown by the matched set. These need to be borne in mind when analysing the matched data

Despite the limitations to the exercise arising from problems of matching, the results are nevertheless instructive. an overview of matched comparative prices and unit value suggests that, for a particular product variety, the average price recorded by price collectors was higher than that for the scanner data unit values in more than fifty per cent of cases. This was most notably the case for washing machines where in 72% of cases the collected data produces a higher average. Looking at the average price for the whole product range reinforces this point. In all cases, except in October for 21" televisions, the average recorded price was higher than the corresponding figure from scanner data.

However, further analysis, using median values, indicates that in most cases the difference between price recorded by the price collector and the average unit value shown by the scanner data was caused by a relatively small number of abnormal high or low prices or unit values. This work, of course, has practical applications in the sampling of items. It is clear that, if means can be influenced significantly by outliers, we need to look closely at the number of quotes sampled, and whether they need to be increased for certain items to reduce this effect.

A corresponding analysis of monthly price changes indicates that there is no evidence of recorded price **changes** consistently exceeding unit value **changes** or vice versa. However significant differences do occur, indicating a possible data collection problem, though not a bias.

In some instances, the divergences that occur in price and unit value trends may be due to the small number of price observations in the RPI for the particular model under investigation - in such circumstances price can fluctuate wildly from one month to another with the introduction of sale prices and special offers. In other instances the difference is difficult to explain but may be due to differences in the mix of outlets, and in

particular, the changes that occur over time in market shares, for example when some outlets have sales and therefore appear with changing weights in the scanner data Despite these possible explanations there yet remains a degree of mystery why some of the differences in price trends occur.

7.0   The effect of different store types

As we have said, one of the main differences between the scanner data and the RPI sample is that the former reflects changes in the mix of store types. An example of this effect is observed in October where the average price of a particular brand of washing machine varied between £301 from multiple chains of stores to £384 for sales from mail order catalogues. Additionally the change in prices between August and September shows a large variation, with the smallest price rise being 0.5% for mail order catalogues and the highest being 1.9% for small independent stores.

Clearly differences in the mix of shops between scanner data and the RPI sample could produce significant differences in both the average price of items, and the average price change. Unfortunately this difference is difficult to test as RPI quote data are not classified as finely as scanner data, though this is clearly an area that we need to investigate further.

8.0   Conclusions and implications for sampling and the collection of price data

The research described in this paper raises questions about current sampling practices in the RPI, not that they are causing a bias, but that they may be increasing the variability of the index. In order to address this it is possible to use the scanner data (assuming we are happy with its accuracy) to improve our sampling techniques. In particular we could:

Introduce some form of quota sampling based on scanner data is likely to help in ensuring a representativity sample;

As an alternative to a quota sample, scanner data could be used to post-weight the quotes obtained by price collectors to produce a more representative final sample;

Use scanner data to check on the deterioration of basket representativity between annual updates;

Where unplanned forced replacements continue to be necessary due to product varieties disappearing from shops scanner data may be helpful in choosing replacements by, for example, identifying replacements that are the closest in terms of characteristics to the disappearing model or with the use of hedonic regression identifying the most important characteristics that feature in determining price;

Scanner data by store type indicates that special care needs to be taken to ensure a proper spread of outlets in the RPI sample and that scanner data may be used for post-stratification where there is reason to believe that the sample achieved under current RPI sampling practices is not totally self-weighting;

Where coefficients of variation suggest that outliers can have an undue influence, we need to re-examine the numbers of quotes taken to see if the situation can be improved by increasing the sample size.

The Office for National Statistics are looking at these issues in more detail as part of its longer-term methodological research programme. Specifically, we are investigating the coverage, and quality, of the scanner data for the six items used in this report, to see whether our use of them as a benchmark is justified. We are also starting practical work to see whether the construction, and use, of a quota sample is feasible in a live price collection. In particular, we are asking a sample of price collectors to try to select a complete basket of new goods for the six items in this study, using a quota sampling technique.

9.0 Summary in French

Cet article examine differentes facons d' utiliser les donnees code-barres pour ameliorer la representativite de l' indice des prix a la consommation (IPC) du Royaume Uni. D' abord les characteristiques de chaque type de donnees sont detaillees, ainsi que leurs avantages et desavantages. Ensuite, pour certains produits, les proportions de donnees collectees pour chaque modele sont comparees avec les chiffres de vente reels. L' article conclut en etablissant des recommendations quant a l' utilisation des code-barres comme marque de reference, et comme aide dans la collecte des prix par la selection d' echantillons representatifs.