Sampling in Consumer Price Indices: what role for scanner data?

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 Mick Silver *Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff CF10 3EU.*

Paper presented at the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia, 2-6 April 2001.

ABSTRACT

The quality of a consumer price index depends critically on the quality of the data, and in particular on the representativity of both the sample of retail outlets used to monitor prices and the choice of items priced. This paper looks at the scope for enhancing the quality of a price index by using scanner data as a benchmark to check the representativity of the achieved sample, to control initial sample selection and to adjust after the event for inadequacies in achieved samples. The paper begins by reviewing the underlying principles behind sample selection and the practical choices available to the compiler of an index and the subsequent issues that arise. It looks at the current sampling procedures for the UK Retail Prices Index(RPI), comparing the principals on which it is based with those underlying the compilation of scanner data. It then considers practical issues surrounding the possible use of scanner data to improve current sampling methods. The paper looks separately at two aspects of sampling methodology within the RPI, item selection and outlet selection. For item selection the paper will focus on consumer durables where traditionally the maintenance of representativity has been most challenging. Scanner data is used to benchmark the current sample for consumer durables with high replacement ratios, such as televisions, highlighting differences and presenting solutions, either by controlling the sample through quotas or by post weighting results. The paper also investigates the potential use of scanner data for choosing replacement varieties when the original has disappeared from the shelves and when there is an associated quality change. Also addressed is the timing of the selection and item rotation. For outlet selection the paper highlights the differences in prices that can occur between different outlet types, and points to the need for the RPI to select on a more finely defined stratification to ensure representativity. The paper concludes by developing improved guidelines and quality control procedures for price collection.

Keywords: scanner data, outlets, items, new & old goods, stratification, random & purposive sampling, representativity, modelling, re-weighting, re-sampling, quotas, benchmarking, quality control, guidelines.

1.0 Introduction

A number of studies in the past have pointed to the possibility of scanner data being used in the compilation of consumer price indices either as a direct source of price data in its own right or for the estimation of appropriate quality adjustments when item substitution takes place and the characteristics of the items being priced change. In addition it has been suggested that scanner data has the potential to contribute to the effectiveness of traditional probability sampling procedures.

The potential gains from utilising scanner data are not insignificant, particularly if hedonic regressions and scanner data are used to supplement current practice to better serve, via an integrated approach, the needs of both representativity and quality adjustment.

It is in this context that a joint research project was set up between the ONS and the Cardiff Business School, Cardiff University, to both explore the potential for using scanner data as a diagnostic tool for the identification of potential deficiencies in RPI data collection and to provide solutions. This papers presents the results of the work completed to date.

2.0 Background: RPI target population and sampling procedures

2.1 Target population

The RPI is an average measure of the change in prices of goods and services bought for the purposes 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 that 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. The reference expenditure items are the goods and services bought by the reference population for consumption. Prices used in the calculation of the index should reflect the cash prices typically paid by the reference population for these goods and services. The index is compiled mainly on an acquisition basis, in other words on the total value of goods and services acquired during a given period regardless of whether they are wholly paid for in that period. The main exception is owner-occupied housing where a user cost approach is adopted.

2.2 Price Reference Day

The price reference day is the second or third Tuesday in the month.

2.3 General approach to sampling and price collection

The Office for National Statistics currently follows a traditional approach to sampling, whereby prices are collected locally, from individual shops, and centrally, using nationwide tariffs for utilities or returns submitted by the Head Offices' of retail chains with central pricing policies. The major difficulty with this approach is the lack of availability of a suitable sampling frame to represent the target universe in terms of geography, outlet, product line and individual item. This means that National Statistical Institutes are often obliged to either construct their own sampling frames and random selection procedures or to resort to purposive sampling. These procedures do, of course, need to satisfy representativity in the time dimension. The latter is generally considered less problematical than geography, outlet and product line and item representativity certainly in the context of the price reference period. It should be noted in this context that the choice of price reference day for the RPI was informed by a study of shopping patterns. This concluded that a Tuesday in the middle of the month was likely to be most representative. However, there is another element of the time dimension, namely the deterioration in sample representativity as the "fixed" basket ages as a result of the introduction of new products and outlet, item and variety substitution by consumers. Thus the time dimension is present in all aspects of sampling for a consumer price index.

2.4 Sampling procedures for local price collection

Current methodology for the selection of locations from which we collect local prices was introduced in 2000. This aims to give each shopping location in the UK a probability of being selected for price collection equal to its share 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. This is the first stage of the sample and is based on information obtained from household expenditure surveys. Within each region locations are selected on a probability proportional to size basis, using the number of employees in the retail sector as a proxy for expenditure. Practical considerations mean that this basic principle is modified in two ways. Firstly, because it is not cost effective to collect from areas too small to provide a reasonable proportion of the full list of items, we exclude locations which had fewer than 250 outlets. Secondly, and for similar reasons related to cost effectiveness, out of town shopping areas, in which a high level of expenditure takes place, but from which it is not possible to obtain all items, are paired with smaller locations nearby from which the rest of the items can be obtained. This joint location is then treated as a single location in the probability sampling.

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. This processes is performed on a rotation basis, so that the whole sample is refreshed every five years.

In contrast to outlet sampling, 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 number of representative items for each section depends on both the weight given to that section and the variability of the prices of the items covered by that section. Around 650 representative items are chosen centrally by commodity specialists and reviewed each January to ensure that they continue to be representative of the section. New items are chosen to represent new or increasing areas of expenditure, or to reduce the volatility of higher level aggregates. Other items are removed if expenditure on them falls to insignificant levels. Decisions are informed by market research reports, newspapers, trade journals and price collectors in the field. This enables the basket to be kept up to date but it does not, on its own, guarantee sample representativity. The descriptions are generic rather than prescriptive leaving the price collector with the task of choosing the precise product or variety to be priced.

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 of that particular item in that particular shop. In practice the price collector will normally get the assistance of the shopkeeper to help in this process by asking which is the best selling product or variety. This is, in most cases, the one that is chosen as the representative item for price monitoring. This shop based sampling procedure has the advantage of increasing the achieved sample size by overcoming the problem of particular shops not stocking a particular product or variety. Also, in theory, it spreads the sample to include a wider range of products and varieties than would be covered if a very tight description were employed.

2.5 Sampling for centrally collected prices and prices obtained over the telephone

In some instances prices are collected centrally, without resort to the expensive activity of sending price collectors into the field. Central price collection covers two distinct sets of circumstances:

- **Central shops** where for cost-effectiveness prices are collected direct from the headquarters of multiples with national pricing policies. These prices are then combined with prices collected locally from other outlets in proportion to the number of outlets originally chosen in the selected locations;

- **Central items** where there are a limited number of suppliers and where purchases of the item do not normally take place at local outlets. Examples of these include gas, electricity and water where prices are extracted from tariffs supplied direct by the Head Offices of the companies involved. These data are used to create sub-indices that are combined with other sub-indices to produce the all items RPI.

In addition the prices of some items are collected over the telephone, with the retailer being visited in person only occasionally to ensure that the quality of response is being maintained. Such prices include electrician's charges, where there is no outlet as such, and entrance fees to leisure centres, where there are unlikely to be any ambiguities over pricing and where a trip to the centre may be relatively time consuming for the collection of just one price. These prices are combined, as appropriate, with locally collected data.

2.6 Critical factors

The procedures for sampling locations and shops are, on the whole, statistically rigorous leaving limited opportunity for problems to arise. The view is therefore taken that the potential for problems of non-representativity to materialise is most likely to be associated with the selection of items - more so given the relatively high item turnover for some products. Therefore it is clear that success in achieving a representative sample in the context of the UK RPI is particularly dependent on:

- The procedures for the initial purposive sampling of items in the field;
- The procedures used for selecting forced replacements when items disappear from shops' shelves;

- The procedures in place to update the sample selection to reflect the general turnover in products and varieties.

It was with these issues in mind that an exercise was undertaken to benchmark the achieved RPI sample, for a selection of electrical and hi-tech goods, with corresponding scanner data and to compare the relative price levels and price movements.

Before presenting this exercise it is worthwhile reminding ourselves of the main characteristics of scanner data, especially as scanner data itself is not specifically designed for the compilation of consumer price indices and therefore has its own problems . The characteristics of scanner data are reviewed in the next section.

3.0 Characteristics of scanner data

Scanner data is compiled from 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, scanner data increasingly provides the potential to deliver up-to-date and accurate information on:

- number of sales over a chosen period of individual product varieties uniquely identified by the barcode number;
- the total value of those sales and by implication the average transaction "price";
- a listing of the individual characteristics of the individual product varieties concerned;
- geographical and other characteristics relating to the outlet.

In reality the current market coverage of scanner data varies between different shop types and commodity groups and the amount and detail of data actually available can vary depending on the commercial source and on the individual product or product group. Also because scanner data is a by-product of a financial accounting and stock system it is not specifically designed with the price statistician in mind, and this has implications for its use in index compilation. Firstly, definitions may not be compatible with the definition of the index. For example, the average transaction "price" recorded by scanner data includes discounts such as those relating to damaged stock, not normally included in consumer price indices. Secondly the coding of data may not be in a readily useable form, and compatible with international standards. This applies, for example, to the categorisation into commodity headings.

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

3.1 Main definitional differences between scanner data and data collected locally for the Retail Prices Index.

The main differences between the two data sets are:

- RPI data covers transactions conducted in retail outlets by private households for private domestic consumption. Scanner data covers only EPOS sales, usually supplemented by surveys to cover shops where bar coding is not used. It often excludes “own” brands but includes sales to commercial customers;

- RPI data excludes conditional discounts (for example, where a “club” card is required), two-for-one offers, personal discounts offered on a one-off basis by shop managers and discounts on discontinued or damaged stock. Scanner data measures average revenue generated after discounts given by whatever method, it will include discontinued or shop-soiled stock and will attribute discounts to the scanner code rather than to the transaction (for example, free video tapes given away with a recorder will be shown as a reduction in average revenue for video tapes);

- RPI data relates to a fixed selection of outlets and therefore excludes the effects of outlet substitution. Scanner data relates to current transactions and therefore includes outlet substitution.

Whilst the numerical impact of these differences is not known, it is clear that the impact will not necessarily be constant over time and will vary with market circumstances and commodity type.

Other characteristics of the two data sources need to be borne in mind when comparing display prices in shops and corresponding scanner data, including:

- The sampling error associated with sample surveys, particularly at the level of product variety which is investigated in this paper (the RPI sample is not designed to provide reliable information at this level of detail). In contrast, scanner data provides total coverage for those retail segments included;

- The RPI records prices for a particular day in the month whilst the scanner data used for this exercise cover a whole month;

- Scanner data distinguishes between different types of retailers such as multiple and independent whilst RPI data doesn’t (there is no need because the sample for local price collection is designed to be self-weighting). This means that there is a potential problem of lack of homogeneity in comparisons between the two data sources if the mix of outlet types varies between the two data sources and changes over time.

4.0 Research design

The research consisted of three stages:

- The benchmarking of RPI product and variety selection against corresponding scanner data. This involved a comparison a relative distributions of sales proportions, and proportions of quotes;

- A comparison of RPI average unit prices and price changes with the corresponding unit values (i.e. average revenue generation) and unit value movements obtained from scanner data;

- An investigation of possible options for enhancing the performance of traditional sampling techniques by utilising scanner data in standard data collection procedures and for adopting an integrated approach to representativity and quality adjustment.

Investigations focussed on five pre-selected items: televisions; washing machines; vacuum cleaners; dishwashers; and cameras. Related work was also carried out on the same database to investigate hedonic regression techniques for explicit quality adjustment and for identifying key item characteristics that need to be taken into account when making forced replacements for items that have disappeared from shops shelves. It has become increasingly clear during the course of the work that sample representativity and quality adjustment are inter-linked. We return to the latter towards the end of this paper.

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 unrepresentative samples of products and varieties being chosen for pricing. It looked at overall distributions obtained from the selection procedures used in the RPI and compared these with the overall distributions given by scanner data. Monthly data were compared for the period from August 1999 to October 1999. This was done at an aggregate level, there was no individual linkage of data.

5.1 Summary of results

In table 1 below the distributions of price quotes by model are ordered to show the top 10 sellers for each product group in September 1999 according to sales volume from scanner data. Alongside are the corresponding proportions of quotes represented in the RPI collection for that item.

Table 1: Top 10 selling items according to scanner data, and associated percentage of RPI quotes September 1999 (cumulative percentage in brackets).

Model	14" Televisions		21" Televisions		Vacuum Cleaners	
	Percentage of scanner data	Percentage of RPI quotes	Percentage of scanner data	Percentage of RPI quotes	Percentage of scanner data	Percentage of RPI quotes
Model 1	17.7 (17.7)	1.0 (1.0)	16.2 (16.2)	10.5 (10.5)	30.1 (30.1)	18.7 (18.7)
Model 2	13.9 (31.6)	25.0 (26.0)	12.8 (29.0)	4.4 (14.9)	13.2 (43.3)	3.0 (21.7)
Model 3	11.0 (42.6)	1.9 (27.9)	11.7 (40.7)	1.8 (16.7)	8.7 (52.0)	1.2 (22.9)
Model 4	8.5 (51.1)	28.6 (56.5)	10.2 (50.9)	8.8 (25.5)	5.7 (57.7)	1.2 (24.1)
Model 5	8.2 (59.3)	3.8 (60.3)	10.1 (61.0)	31.6 (57.1)	4.4 (62.1)	0.6 (24.7)
Model 6	6.9 (66.2)	4.8 (65.1)	10.1 (71.1)	3.5 (60.6)	4.1 (66.2)	20.5 (45.2)
Model 7	6.6 (72.8)	1.9 (67.0)	6.1 (77.2)	8.8 (69.4)	4.1 (70.3)	0.6 (45.8)
Model 8	4.9 (77.7)	4.8 (71.8)	5.6 (82.8)	0.8 (70.2)	3.8 (74.1)	1.2 (47.0)
Model 9	4.4 (82.1)	1.0 (72.8)	4.1 (86.9)	1.7 ((71.9)	3.5 (77.6)	0.6 (47.6)
Model 10	3.9 (86.0)	3.8 (76.6)	1.8 (88.7)	1.7 (73.6)	3.4 (81.0)	6.6 (54.2)

Model	Cameras		Dishwashers		Washing Machines	
	Percentage of scanner prices	Percentage of RPI quotes	Percentage of scanner prices	Percentage of RPI quotes	Percentage of scanner prices	Percentage of RPI quotes
Model 1	28.4 (28.4)	38.4 (38.4)	17.2 (17.2)	2.2 (2.2)	12.0 (12.0)	6.5 (6.5)
Model 2	13.6 (42.0)	1.2 (39.6)	17.1 (34.3)	16.3 (18.5)	11.2 (23.2)	20.3 (26.8)
Model 3	11.9 (53.9)	12.8 (52.4)	9.4 (43.7)	11.9 (30.4)	11.2 (34.4)	2.3 (29.1)
Model 4	7.6 (61.5)	3.5 (55.9)	7.8 (51.5)	5.9 (36.3)	9.8 (44.2)	5.8 (34.9)
Model 5	6.7 (68.2)	1.2 (57.1)	7.3 (58.8)	6.7 (43.0)	6.9 (51.1)	1.4 (36.3)
Model 6	5.6 (73.8)	2.3 (59.4)	5.8 (64.6)	0.7 (43.7)	5.1 (56.2)	4.3 (40.6)
Model 7	4.4 (78.2)	15.1 (74.5)	5.1 (69.7)	23.0 (66.7)	5.1 (61.3)	2.9 (43.5)
Model 8	4.3 (82.5)	3.5 (78.0)	5.1 (74.8)	0.7 (67.4)	4.4 (65.7)	1.4 (44.9)
Model 9	4.0 (86.5)	1.2 (79.2)	4.8 (79.6)	3.0 (70.4)	4.2 (69.9)	1.4 (46.3)
Model 10	3.4 (89.9)	1.2 (80.4)	4.1 83.7)	0.7 (71.1)	4.1 (74.0)	4.3 (50.6)

It should be noted that the RPI sample for September represents the sample produced from the combined effect of the original sample selection (in theory up to five years old), the annual update of the basket (in this instance new price quotes introduced in January 1999 when a quarter of outlets was replenished) and forced replacements since January as old models disappear from the shelves.

The results 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.

Even if we investigate a cumulative distribution, problems remain evident. In all cases the proportion of RPI quotes that represent the top 10 selling models are significantly lower than their sales figures. In the case of dishwashers the top ten models which account for 74.0% of sales according to scanner data are represented by just 50.6% of price quotes in the RPI sample. Over the three months studied these results are fairly stable, though with enough variations to suggest some deterioration in the sample over the period.

The reasons for these apparent anomalies, which are not obvious, are investigated later on in the paper with a more detailed in depth analysis. That said it is not necessarily solely related to deficiencies in the RPI data. For example, in September there is a particular a model of washing machine that attracts almost 10 per cent of RPI quotes, while scanner data indicates that no sales of this particular model took place. As it is difficult to believe that collectors are gathering the price of a machine that doesn't sell at all in a particular month one can speculate whether sales of the machine are taking place in a particular market segment not covered by scanner data. Unfortunately we have been unable to follow this line of thought through due to a lack of information on the actual outlets covered by scanner data..

5.2 Interpretation

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, they do seem to indicate that the pricing of items can apparently be skewed towards products and varieties which scanner data indicate have relatively small sales, despite the instruction to the price collector to chose a product variety that is representative of the sales of that item in that particular shop. Conversely there is the non-selection of some big selling items. Possible causes include:

- The fixed basket approach - where products and varieties as well as items are reviewed at most on an annual basis - leads to the sample becoming increasingly unrepresentative as the "fixed" selection of goods in the basket ages over the samples life. This is not surprising but does raise the issue of whether, for certain items where models change very quickly, updating of the basket should be more frequent than every year. Certainly it suggests that replacements should be introduced before models disappear and the volume of sales contract to the point where very few purchases are made;

- Weaknesses in the approach where a "similar" product or variety is chosen when a replacement is forced on the price collector because an item becomes obsolete and is no longer found in the shop. This approach can contribute to the ageing of the sample but has the advantage of reducing reliance on quality adjustment procedures. It emphasises the need for an integrated approach to representativity and quality adjustment;

- Adequate product and variety selection undermined by unrepresentativeness in outlet selection. This is considered the least likely cause given the sampling regime used, although it is instructive to note that scanner data shows a large variation between outlet types in unit values and monthly changes in unit values. Thus a relatively small bias in outlet sample selection could have a disproportionate impact on the reliability of the measured inflation. (see section 7.0).

The extent to which these findings are a cause for concern depends, at least in part, on whether there is a noticeable impact on the published index and the measured rate of inflation. The second stage of the research designed to test whether this is the case is reported in the next section.

6.0 Average unit prices and price changes

This part of the investigation involved observing, for specific product varieties, the extent to which the price levels and changes differ between those derived from data collected by price collectors in the field and those shown by scanner data. In order to do this, data for specific models of each product in the scanner data had to be carefully matched with data for the same models in the RPI data. This work involved considerable resources as detailed data had to be extracted from the computer files storing archived RPI data and a series of reconciliation and validity checks carried out before the data could be used. It was for this reason that the exercise was limited to the three months from August to October 1999.

6.1 Practical limitations of the matching process and the degree of success achieved

It should be noted that problems remained unresolved despite the checking processes described above. These mainly arose from price collectors' descriptions being inadequate for the process of matching (although generally adequate for the identification of product varieties in shops). For instance, a maker's name and a select number of attributes may be all that is required to identify a product variety in a shop but the model number, which in many cases will not be listed by the price collector, will be required to unambiguously matched the product variety with one shown on the scanner list.

6.2 Price levels

Table 2 gives an overview of the success of the matching process. It should be noted that the degree of successful matching varied between the five items selected. The process was most successful for dishwashers, washing machines and vacuum cleaners where over 70% of RPI observations (representing about 50% of RPI product varieties) were successfully matched to scanner data. It was most problematical for cameras where only about a half of RPI quotes (representing about a third of RPI product varieties) were matched. These differences could, clearly, have an influence on the conclusions of the research. In particular, differences between the price levels and price changes for the matched sample and the full RPI dataset could cause biases if the match sample was selected in such a way as to be unrepresentative.

A number of observations can be made:

- Significant differences can exist between the mean **average price level** for a product variety based on the full set of RPI quotes and the subset successfully matched with scanner data. This was most marked for television sets and washing machines;

- 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 a non-random effect is present within items, though this is difficult to test with a weighted mean, and a serially correlated sample;

- Differences occur between **average price changes** shown by the full scanner dataset and those shown by the matched set. This was explored by calculating Laspeyres[1], Paasche[1] and Fisher[1] indices for the full RPI set of price data and for the sub-sample representing matched observations. The results for a Fisher index indicate that the price changes from the sub-sample followed similar, but not necessarily identical patterns, to those in the full scanner data (see Figure 1).

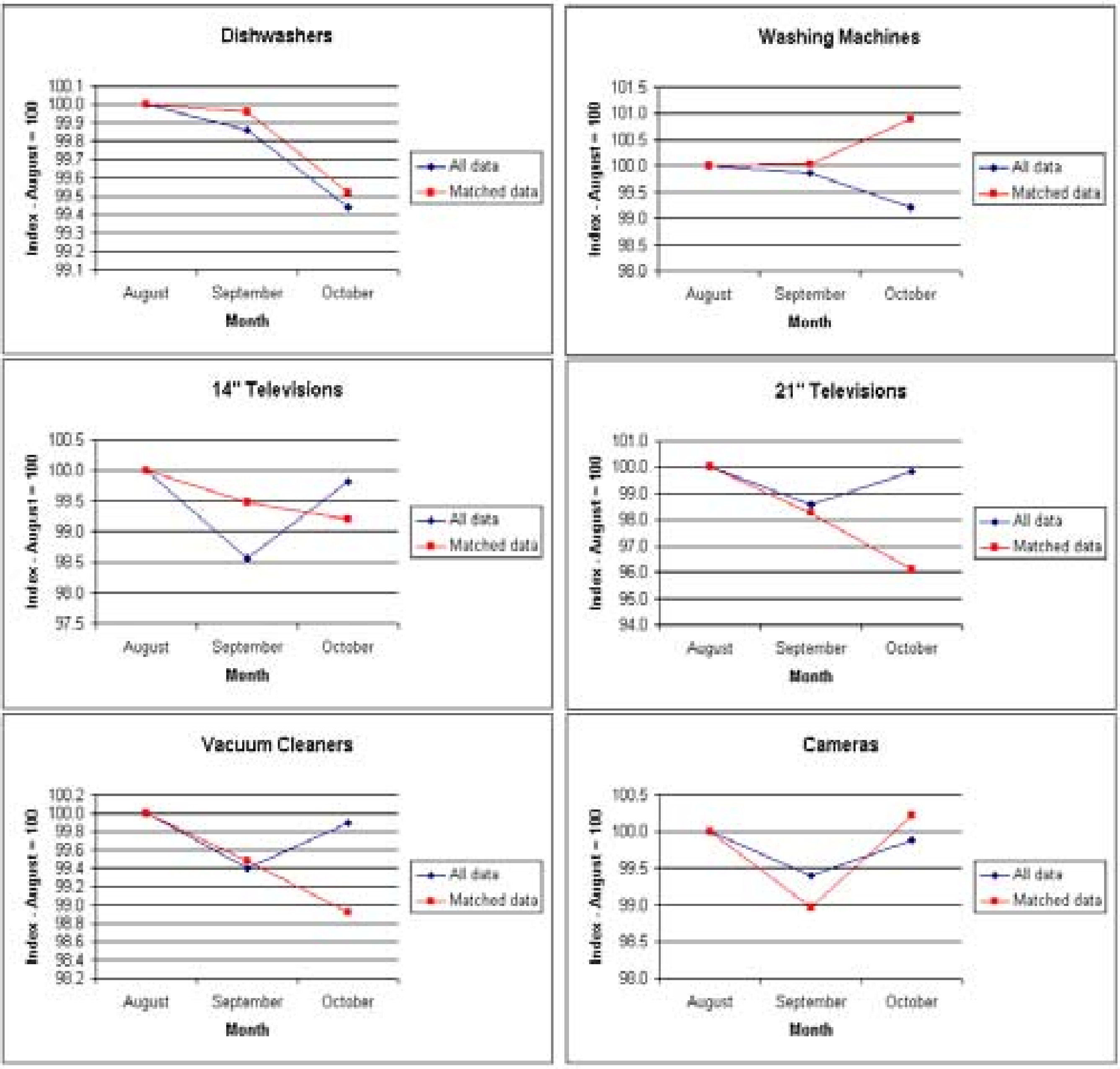
These results clearly show that there are real differences between the full and the matched datasets, most specifically in relation to the price of the item. It is difficult to be certain of the reasons for these differences as testing them from the RPI system is problematical. However, it is possible that data from some store types are better specified and this, combined with the differences in price described in later analyses, causes the effect. However, whatever the cause, there is clearly a real effect and this needs to be borne in mind whenever the results of the comparisons are analysed.

Table 2: Percentage coverage of matched data and comparison between means of prices for the whole RPI and the matched sample. August to October 1999 (means in £s).

	August			September			October		
	% matched	Mean of all RPI quotes	Mean of matched sample	% matched	Mean of all RPI quotes	Mean of matched sample	% matched	Mean of all RPI quotes	Mean of matched sample
14" Televisions	39	135.5	146.7	46	130.8	148.9	46	129.2	150.7
21" Televisions	48	249.7	291.3	56	246.5	283.8	58	240.1	268.4
Vacuum Cleaners	76	129.5	129.1	77	130.0	130.9	78	128.9	130.2
Cameras	55	55.4	56.9	50	56.5	59.9	53	57.3	56.4
Dishwashers	71	339.5	332.3	73	337.9	330.8	69	333.3	328.5
Washing Machines	81	345.3	349.7	75	354.0	323.2	76	348.9	317.8

[1] See Appendix

Figure 1: Price Indices for each item calculated using all scanner data and the matched subset, using a Fisher Index. August to October 1999



6.3 The results

Despite the limitations to the exercise arising from problems of matching, the results are nevertheless instructive. The first observation to be made is that, in all cases, the average price produced by RPI quotes is higher than the corresponding unit value produced by scanner data. That this is the case should not come as a surprise, and arises from the different bases underlying the data collection. The RPI sample collects data for a fixed basket of goods, taking no account of product or outlet substitution. In addition it is restrictive in the types of discount that are allowed to influence prices, in particular end-of-line or clearance sales are specifically excluded. In contrast scanner data directly estimates the prices actually paid be consumers for there goods by measuring the value and volume of goods bought. Because of this it tracks consumers' efforts to get the lowest prices for goods, and consequently includes the effects of substitution in its estimates. This will always produce a lower average price. In addition all discounts are included, however they arise, a factor that also reduces the average price implied by the unit cost.

Looking at the data in more detail it was found that not only were the average prices recorded by RPI collectors for each product generally higher than the average unit value from scanner data, but more often than not the average price recorded by price collectors for a particular product variety was also higher than the corresponding unit values from scanner data. However, a comparative analysis of absolute and percentage absolute deviations between RPI quotes and scanner data unit values (Table 3) indicates that a large proportion of this difference is caused by a relatively small number of high or low prices or unit values appearing in the comparison. Thus the deviations of the medians are in all cases significantly lower than the corresponding deviations of the arithmetic means.

Table 3: Absolute and percentage absolute deviations between averages for RPI quotes and scanner data unit values, using both mean and median differences: Average of August to October 1999.

	Absolute Deviation (£s)		Percentage Absolute Deviation	
	Mean	Median	Mean	Median
Dishwashers	29.4	21.1	9.99	6.35
Washing Machines	34.8	21.3	10.45	7.58
Vacuum Cleaners	13.3	7.7	9.71	6.07
14" Televisions	14.9	9.7	13.95	7.84
21" Televisions	30.0	16.6	9.60	6.05
Cameras	9.2	5.9	16.10	10.36

The coefficients of variation given in Table 4 provide a useful overview, as they discount the impact of the different levels of the mean for the different products. Dishwashers have the highest coefficient of variation for the difference between average price and average unit value when expressed as a percentage of the average unit value. Vacuum cleaners and 21" television sets have high coefficients of variation both for the price difference expressed in monetary and the difference expressed in percentage terms. Clearly, there is a case for enlarged samples where, as in the above cases, means are particularly vulnerable to outliers.

Table 4: Coefficients of variation

	Coefficients of variation	
	Monetary Absolute Deviations	Percentage Absolute Deviations
Dishwashers	0.92	1.32
Washing Machines	1.09	0.99
Vacuum Cleaners	1.41	1.19
14" Televisions	1.07	1.12
21" Televisions	1.23	1.23
Cameras	1.04	1.04

6.4 Price changes

A corresponding analysis of monthly price changes (Table 5) indicates that there is no evidence of recorded price **changes** consistently exceeding unit value **changes** or vice versa except for:

- washing machines and vacuum cleaners where price falls recorded by scanner data are consistently higher than those seen in the RPI sample;
- cameras, where, RPI data shows the same pattern of price movements, though the movements are more extreme.

Table 5: Index (August = 100), and month to month price changes for recorded RPI quotes and matched scanner data. August to October 1999.

	August		**September**		**October**	
	Index	Change on Previous month	Index	Change on Previous month	Index	Change on Previous month
Dishwashers						
RPI Quotes	100	-	102.2	+2.2%	104.7	+2.5%
Scanner data	100	-	101.6	+1.6%	106.0	+4.4%
Washing Machines						
RPI Quotes	100	-	98.4	-1.6%	97.0	-1.4%
Scanner data	100	-	96.6	-3.4%	98.3	-1.7%
14" Televisions						
RPI Quotes	100	-	99.9	-0.1%	140.6	+4.4%
Scanner data	100	-	100.5	+0.5%	101.2	+0.6%
21" Televisions						
RPI Quotes	100	-	93.5	-6.5%	91.5	-2.1%
Scanner data	100	-	94.5	-5.5%	99.2	+5.0%
Vacuum Cleaners						
RPI Quotes	100	-	97.1	-2.9%	94.3	-2.9%
Scanner data	100	-	96.6	-3.4%	92.5	-4.3%
Cameras						
RPI Quotes	100	-	109.8	+9.8%	101.5	-7.6%
Scanner data	100	-	105.5	+5.5%	100.8	-4.5%

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 month to month with the introduction of sale prices and special offers. This should not necessarily be a cause for concern as the RPI is not designed to measure price changes of individual product varieties. However, in other instances the difference is difficult to explain. One reason may be differences in the mix of outlets and in particular the fact that scanner data will pick up outlet substitution, i.e. the resulting changes in average prices paid as customers seek the cheapest. This problem of lack of homogeneity was referred to earlier and can potentially have a significant impact because of large observable variations in price levels and price trends between different outlet types. This can be seen from the analysis of unit values from scanner data given in Table 6.

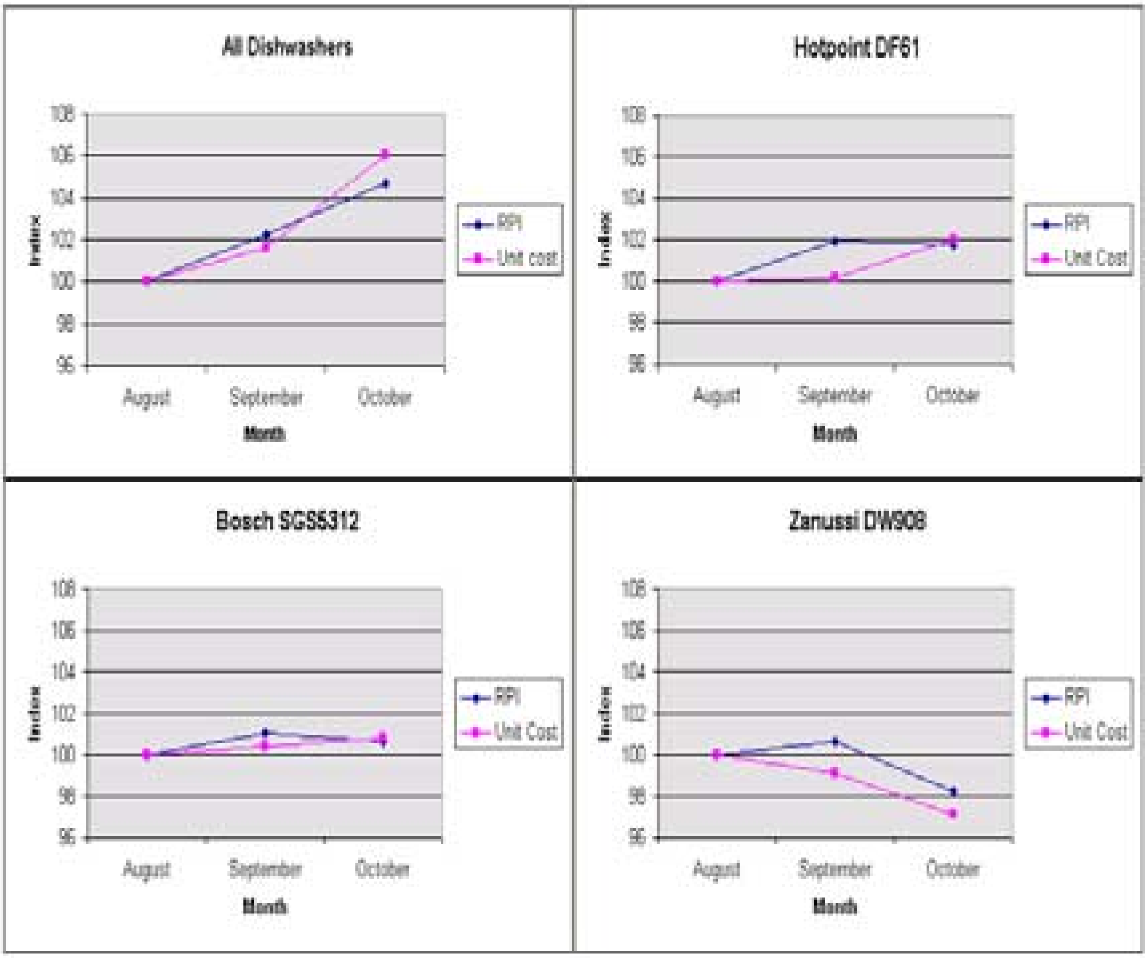
Table 6: Effect of shop type on scanner data unit values of individual brands of dishwasher.

	Unit Value (£s)			Percentage Change	Sales (Percent)		
	August	**September**	**October**	**August to September**	**August**	**September**	**October**
Bosch SGS5312							
Multiple	370.1	374.9	379.1	2.5%	707 (31.5%)	853 (34.4%)	681 (35.7%)
Mass Merchandiser	364.0	364.8	363.1	-0.3%	1195 (53.3%)	1288 (51.9%)	944 (49.5%)
Independent	386.5	382.3	386.0	-0.1%	341 (15.2%)	342 (13.8%)	281 (14.7%)
Catalogue	-	-	-	-	0 (0%)	0 (0%)	0 (0%)
All Stores	369.2	370.8	372.2	0.8%	2243	2483	1906
Hotpoint DF61							
Multiple	309.2	310.7	314.6	1.7%	1190 (35.9%)	2361 (54.4%)	1756 (53.5%)
Mass Merchandiser	288.1	296.4	307.8	6.8%	364 (11.0%)	458 (10.6%)	310 (9.4%)
Independent	326.9	328.7	332.9	1.8%	1756 (52.9%)	1513 (34.9%)	1211 (36.9%)
Catalogue	400.0	400.0	346.7	-1.2%	6 (0.2%)	5 (0.1%)	6 (0.2%)
All Stores	315.0	315.6	321.2	1.4%	3316	4337	3283
Zanussi DW908							
Multiple	258.2	261.5	242.2	-6.2%	740 (49.3%)	705 (54.0%)	780 (51.5%)
Mass Merchandiser	264.6	260.6	263.4	-0.5%	236 (15.7%)	287 (22.0%)	210 (13.9%)
Independent	282.1	275.5	286.8	1.7%	463 (30.9%)	265 (20.3%)	476 (31.4%)
Catalogue	313.4	307.9	309.6	-1.2%	61 (4.1%)	48 (3.7%)	49 (3.2%)
All Stores	268.2	265.9	260.6	-2.8%	1500	1305	1515

6.5 A detailed examination of dishwasher product varieties

To understand further why these differences occur requires a detailed examination of each individual product and product variety. Figure 2 shows a comparison between an index for all dishwashers, and those produced for individual models within that group. While, as a whole, dishwashers show no systematic difference in price movements between RPI and scanner data and changes are relatively close, some interesting differences can be seen for individual models.

Figure 2: Price changes between August 1999 and October 1999 for selected brands of dishwasher



- **Bosch SGS5312**

  This dishwasher showed the least difference between price changes from RPI quotes, and changes to unit costs. The reasons for this can be seen from the analysis of shop type prices shown in Table 6. In this case prices, and price changes, for the various store type are similar, with all changes within 1.7% of the mean. These, coupled with there being only minor changes in the distribution between sales by store type, has produced an item index that is similar for the two sources.

- **Hotpoint DF61**

  In this case the index in October is very similar in both the RPI data and scanner data cases. However, the index in September is markedly different. Part of the reason for this can also be found in the store analysis given in Table 6. Between August and September there is a marked move away from purchases from the more expensive independent stores, towards the cheaper multiples, associated with an overall increase in volume. This has, as a consequence, depressed the index for September. However, there is another factor at work as the recovery of the index in October is not accompanied by a shift back of the sales distribution. Part of this will, undoubtedly, be related to the differential increases in prices observed across the groups, though perhaps not all.

- **Zanussi DW908**

  For this dishwasher we see that the index for scanner data, and that for RPI data, diverge between August and September, and though there is a slight narrowing of the gap between September and October, they remain different. Again the initial difference is, at least partly, due to a move away from sales in expensive stores towards sales in less expensive ones. However, in this case, the distributions return almost to their original levels, without a resultant return of the scanner data index back to the level of the RPI index. It is also clear that this is not about differential price changes in the shops, as the shops to which consumers were returning had a higher price rise than the other types. What has caused this difference is unclear, though it is possible that some of the movement may have been due to special offers not captured in the scanner data. We will be investigating these differences as part of the ongoing work.

It is clear from this work, that the selection of outlets is important in ensuring that the RPI produces a representative set of prices. While we are confident that the current system works well it is essential that we are on our guard against changes in sales amongst retailers, particularly over the longer term. Shorter term, outlet substitution, is harder to deal with and is strictly outside the scope of the current RPI. However, we do need to be aware of these changes if in order to better interpret movements in the RPI.

7.0 The issue of implicit weights and aggregation formulae

The calculation of indices for those products which have been the focus of this paper uses the average of relatives formula[1]. Explicit weighting is not used in this calculation but the implicit assumption for the average of relatives is that all quotes are equally important, i.e. they are given equal weight within the elementary aggregate. This is clearly only truly accurate if the mix of quotes taken is representative of sales of brands and models for each item. An alternative approach would be to use the explicit weights available from the volumes of sales of each model as seen in scanner data. Table 7 compares price indices based on current RPI methodology with a Laspeyres[1] based weighted average using a combination of RPI price data plus scanner data relating to August for weights.

These comparisons show some quite substantial difference, (for example 4.5 percentage points for washing machines in September) but no consistent pattern in either magnitude or direction, and reflect in large part the varying proportions of price quotes by model that exists between RPI and scanner data. Clearly these results show the effect on the indices for these items of the distribution differences highlighted in the earlier parts of the paper. Again, we most be careful in applying these results to the index as a whole given the differences seen between the matched data and the full RPI. Despite this, it is clear that we could get noticeably different results for individual product groups if a different approach to selecting items were taken.

Table 7: Comparison of Indices using un-weighted ratio of averages and a weighted Laspeyres calculation: August to October 1999.

	August	September	October
Dishwashers			
Ratio of Averages	100.0	99.2	97.2
Laspeyres	100.0	100.8	100.4
Washing Machines			
Ratio of Averages	100.0	103.3	99.7
Laspeyres	100.0	98.7	99.7
Vacuum Cleaners			
Ratio of Averages	100.0	102.1	101.6
Laspeyres	100.0	101.4	100.2
14” Televisions			
Ratio of Averages	100.0	100.9	100.4
Laspeyres	100.0	101.4	100.0
21” Televisions			
Ratio of Averages	100.0	100.2	94.6
Laspeyres	100.0	96.9	97.2
Cameras			
Ratio of Averages	100.0	100.7	100.0
Laspeyres	100.0	99.2	97.9

8.0 An integrated approach to representativity and quality adjustment

Thus far this paper has focussed on the issue of sample representativity and how this can be tested by benchmarking against scanner data. In practice, it is difficult to detach consideration about sample representativity from issues relating to quality adjustment. In particular, the trade-off both in terms of resources and in terms of the technical quality of the index, between infrequent but large quality adjustments and more frequent but smaller quality adjustments:

- Maintaining sample representativity can impose additional burdens in terms of making explicit quality adjustments. For example, updating the basket more frequently for hi-tech goods by introducing “planned” forced replacements between general updates of the basket will increase the frequency of such adjustment;

- Quality adjustment becomes technically more difficult as the basket gets increasingly unrepresentative. The hedonic variables become less reliable and relevant;

- Some changes in consumer evaluation of quality will not have been captured at the relevant point in time. For instance, where specific characteristics of the old model will have reduced over a period of time to a nominal value;

- The same scanner data source can provide sales information to inform sample selection and characteristics information to perform hedonic regressions for quality adjustment;

- The same hedonic regressions can inform price collectors of the brand and salient characteristics for the selection of a forced replacement as well as provide a basis for explicit quality adjustment.

9.0 Hedonic Regression for sample representativity and quality adjustment.

During the course of the study on sample representativity a parallel exercise was undertaken on the same scanner data, though not matched with RPI data, to estimate hedonic regression coefficients in a live situation. These could then be used both to perform explicit quality adjustment and to inform price collectors of the salient characteristics to take into account when identifying similar items as replacements for goods that have disappeared. The fact that many months of data were available allowed the re-estimation of the hedonic equations to test for stability over time.

A summary of the results of these regressions are given in the paragraphs following. It can be seen that for all five items under investigation it was brand name that exerted most influence on price. This was consistently so for each month under investigation and is not surprising in so far as it confirms the results of similar exercises undertaken in the past by other researchers. Generally for each product there was also a small number of core attributes that remained significant in each time period. In summary, the hedonic regressions produce the following characteristics:

- **Televisions**

  Results for the stability of regressions for televisions were mixed. As for other products brand remained a very significant variable over the period, and within this the different brands with significantly different prices from the benchmark remained virtually unchanged. Doubts, however, emerge when some particular factors are examined.

  The first of these is the coefficient for flat screen technology. In a counter-intuitive result the coefficient for this factor is negative, indicating that it lowers the price of televisions. This is unlikely in real life, but may be due to the fact that most such Televisions are made by Sony, which was used as a baseline. Given the collinearity between the two variables it is possible, and perhaps likely, that much of the difference he being subsumed within the brand variable, making the individual coefficient for this variable unreliable.

The second is that these regressions work best when all televisions are analysed together, and then proceeds to predict that price is highly dependent on screen size. For the RPI we do not collect televisions across screen size, but within specified groups. This means that, perhaps, the most important variable highlighted by these results is of no consequence for the RPI. As a consequence of this the ONS are considering whether prices for televisions should be collected for more screen sizes or, alternatively, whether we should allow collectors to select screen size themselves to ensure that the whole market is covered.

- **Vacuum Cleaners**

  At first sight vacuum cleaners show a good degree of stability over the period studied, the constant term is stable, the same brands remain significant, and there are only minor changes in variables that are consider to have a significant impact on price. However, further analysis reveals problems, most especially within the brand variable.

  In this case the same brands are considered to have significantly different prices from the benchmark. However, in several cases this effect moves from being strongly positive to strongly negative between successive months. For example, Black and Decker moves from a coefficient of –0.22 in September to +0.33 in October, which would have a significant effect on any quality adjustment attempted. The causes of these changes are not clear, but may be due to interdependencies between the variables.

- **Cameras**

  In general cameras show good stability of coefficient over time, with the same variables remaining significant, and good stability of the constant term. However even here there are questions that can be raised about the regressions.

  The first is the very high influence of two of the variables, one brand name Leicka and whether the camera is SLR or not. This leads us to ask questions as to whether we are analysing the data at the right level. In particular, it may be sensible to market segment the data into SLR and non-SLR cameras and repeat the regressions for each separately. This is being investigated.

  The second is that the use of shutter speed to control exposure has a negative effect on price. As this is a feature usually found on more expensive cameras this is a surprise. However this may, again, be being caused by collinearity of this variable with others on the higher specification cameras, and most specifically the SLR variable.

- **Dishwashers**

  The results for dishwashers show a worrying degree of instability over the period studied. Despite this, as for other products, brand remains a strongly significant influence throughout the regression. Indeed even within brand stability is evident, with the same brands being considered as significantly different from the benchmark throughout, the only exception being Zanussi.

For other factors stability is less clear. The most striking example of this being whether the dishwasher is constructed of steel (a more costly option). This goes from being very significantly positive in September, to being no negative in October, though not significant. It is a matter of speculation whether the use of steel is connected with either particular brands or other attributes, thereby producing multicollinearity, and leading to the observed instability.

Another sign of the degree of instability is the amount of erratic movements in the level of the constant term over time. Specifically the constant begins at a level of around eight in August, falls to six in September, before rising to nine in October. This, clearly, does not follow the price changes observed in the data and so is being influenced by other factors. Again, an obvious candidate for the effect is multicollinearity.

- **Washing Machines**

  For washing machines the most significant variables remain fairly constant over the period studied. Specifically, those brands considered to be of significantly different price from the benchmark remain the same over the three months. Other significant factors, namely the spin speed, presence of computerised controls and whether the machine is a twin top, top loader or a washer/dryer, all remain in the equation as significant. Other factors, such as width and height of the machine are significant for some months but not for others.

  Despite similar coefficients remaining significant there is evidence of differences in the levels of the coefficients. In particular the coefficient for the Electrolux brand rises notably over the period, from 0.85 in August to 1.25 in October.

  But the main concern is an increase in the constant term, representing the base price, at the same time that general prices of washing machines are falling. Further analysis suggest that this effect may be linked to a fall in the price premiums for features over this period.

It is worth noting that these results are despite the fact that the individual regressions can be considered as well specified. In no case is the adjusted R^2 of a regression lower than 67%, and in many cases is as high as 90%. These represent good diagnostic statistics for individual regressions, and would normally lead to high confidence in the results. It is clear that individual regressions are not enough to validate this type of hedonic technique where questions remain about the stability of the market, or inter-relationships between attributes. Rather, in order to use hedonics effectively, we need to examine a group of results and determine the usefulness of the technique, against general knowledge of the market.

In summary the above outcomes raise a number of issues that need to be resolved before hedonic regressions could be safely applied in practice:

- **Multicollinearity.** This was found to problematical and in part can be associated with the extravagant amount of characteristics information generated by scanner data and the resulting over-specification of models. Thus the availability of scanner data can lead price statisticians into the temptation of submitting all available characteristics data into the regression without first examining it both for common sense (some characteristics can be eliminated as being of either no marketable value to the consumer and/or of no influence on price determination) and for related variables (for example most flat screen televisions are made by Sony) . Whilst multicollinearity will not cause bias in predicted prices if all variables are present, the latter cannot be guaranteed. More importantly, it will result in unstable coefficients which cannot necessarily be taken as representative of the individual price effect of a particular attribute e.g. because of out-of-dateness of the rapidly changing coefficient or because it partly incorporates the value of an associated characteristic. The use of consumer panels, market advice and statistical techniques such as factor analysis may provide workable solutions to sieving out irrelevant data prior to performing regressions;

- **Market segmentation.** As already mentioned this problem became apparent in the poor results in the hedonic regression for cameras where compact cameras and SLR cameras were bundled together despite being associated with two very different markets. In effect, using the same equation is equivalent to forcing a single line through points that in reality represent two separate lines. Market research should be able to assist in determining the appropriate level of market segmentation.

The above issues can clearly be problematical in providing characteristics for price collectors to take into account when choosing "forced" replacements, particularly in rapidly changing markets. Moreover, they can be even more problematical when attempting to use the results of hedonic regression to make explicit quality adjustments:

- For stable markets coefficients need to be stable over time;
- The estimated price effect of a particular attribute must be an accurate representation of current consumer valuation.

Some of the observed instability may due to the constant term increasing as prices increase over time whilst collinearity and inadequate market segmentation can also be problematical as indicated above. Another factor at play may be unstable market conditions leading to price volatility and large fluctuations in volumes of sales (which implies the need to regularly re-run the regressions) . In addition changes in store mix might also come into play if correlated with the characteristics of the products under examination. Accurate representation of consumer evaluation will follow if stability is achieved.

Despite these problems it is clear that hedonic regression has already had an effect on the work of the RPI. Its identification of brand as being a significant variable in determining quality for all products studied as lead us to reinforce instructions to collectors to ensure that this factor is taken into account when forced replacements are made.

10.0 Conclusions and implications for sampling, the collection of price data and quality adjustment

The research described in this paper has raised a number of issues relating to current practices used in the sampling and collection of prices for the UK Retail Prices Index. It also points to a number of ways in which scanner data might be utilised to further ensure representativity of item and product selection in traditional forms of price collection, where prices are observed in shops. The research does not necessarily point to current sampling procedures leading to bias but it does invite the prospect of additional controls and procedures to keep in check the potential for bias.

The starting point in any consideration of the practical implication is the proposition that, in order to reflect the market, representative product varieties should account not only for substantial proportions of the sales for the specified product variety, but also, on aggregate, exhibit similar price changes. We can then make the following practical observations:

- The introduction of some form of quota sampling based on scanner data may act as a useful control for representativity in the context of the current practice, where price collectors are given generic price descriptions and asked to select the most representative product variety in the shop being visited. Such a control would, for instance, provide a mechanism for ensuring better representation of different brands;
- A "representative" basket may deteriorate in its applicability to the market during its life-cycle, even if it is updated annually. This may happen, for instance, in high technology goods where the turnover of models is high. In this case scanner data, in cases where coverage is good, can be used to monitor changes in representativity over time and indicate if, and when, the basket needs to be updated more frequently. The update could be performed using planned "forced" replacements, to avoid the problems of potential bias associated with frequent chain linking. These updates could be trigged either by an algorithm based on scanner data, or more practically at fixed intervals;
- Where forced replacements continue to be necessary, due to product varieties disappearing from shops, scanner data may be helpful in choosing replacements. This would be possible by, for example, identifying replacements that are the closest in terms of characteristics to the disappearing model or, alternatively, by using hedonic regression to identify the most important characteristics featuring in consumers' purchasing decisions;
- The same hedonic regressions can be utilised for explicit quality adjustment, both for traditional replacements, and for the planned "forced" replacements;
- 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;
- Further work will be required on the hedonic regressions themselves before they are robust enough for incorporating in an index. Particular issues to be addressed are multicollinearity and instability over time.

The Office for National Statistics will be looking at these issues in more detail as part of its longer-term methodological research programme.

Appendix 1: Formulae of elementary aggregates and index formulations.

Laspeyres = $\dfrac{P_t Q_0}{P_0 Q_0}$

Where P_t = Price at time t
Q_t = Quantity sold at time t
Time 0 = the base month

Paasche = $\dfrac{P_t Q_t}{P_0 Q_t}$

Where P_t = Price at time t
Q_t = Quantity sold at time t
Time 0 = the base month

Fisher = $\sqrt{\dfrac{\sum P_t Q_0 \sum P_t Q_t}{\sum P_0 Q_0 \sum P_0 Q_t}}$

Where P_t = Price at time t
Q_t = Quantity sold at time t
Time 0 = the base month

Average of Relatives = $\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{P_i^t}{P_i^0}$

Where P_i^t = Price of item I at time t
Time 0 = base month

Ratio of Averages = $\dfrac{\frac{1}{n}\sum_{i=1}^{n} P_i^t}{\frac{1}{n}\sum_{i=1}^{n} P_i^0}$

Where P_i^t = Price of item I at time t
Time 0 = base month

References

Bradley, Cook, Leaver & Moulton [1997] *An overview of Research on Potential Uses of Scanner Data in the US CPI.*

Canberra2201