Volume Title: Scanner Data and Price Indexes

Volume Author/Editor: Robert C. Feenstra and Matthew D. Shapiro, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-23965-9

Volume URL: http://www.nber.org/books/feen03-1

Conference Date: September 15-16, 2000

Publication Date: January 2003

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

Author: David Fenwick, Adrian Ball, Peter Morgan

URL: http://www.nber.org/chapters/c9729

3

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

David Fenwick, Adrian Ball, Peter Morgan, and Mick Silver

Introduction

The U.K. Retail Prices Index (RPI) 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 program, with the primary aim of ensuring that the best possible statistical methods are used.

A number of studies[1] in the past have pointed to the possibility of scanner data's being used in the compilation of consumer price indexes either as a direct source of price data in its own right or for the estimation of appropriate quality adjustments when collectors are forced to select new items with different characteristics from the original. In addition, it has been suggested that scanner data have the potential to contribute to the effectiveness of probability sampling procedures.

The results of the study highlight the difficulties faced when trying to pro-

The views expressed in this paper are those of the authors and not the Office for National Statistics.

David Fenwick is director of the Consumer Prices and General Inflation Division of the U.K. Office for National Statistics. Adrian Ball is head of research and methodology for the Retail Prices Index at the U.K. Office for National Statistics. Peter Morgan is senior lecturer in quantitative analysis at Cardiff Business School, Cardiff University. Mick Silver is professor of business statistics at Cardiff Business School, Cardiff University.

1. These include Silver (1995), Bradley et al. (1998), and Richardson (chap. 2 in this volume).

duce a representative sample for use in consumer price indexes. The comparison of prices for six products obtained from sampling procedures of the U.K. RPI, when compared to unit values from scanner data, highlight difficulties in using each data set. In particular, existing sampling techniques run the risk of not having a representative sample, whereas scanner data have the opposite problem of including price quotes for items not wanted in the RPI, such as damaged goods or closeout sales.

We try to identify an approach that takes the advantages of each type of data and combines them to produce a more accurate sample from which pricing information can be taken. In this case, we recommend either using the scanner data to inform selection of items for the RPI data collection or, alternatively, using the expenditure weights implicit in scanner data to post-weight the RPI collection data into a representative formula.

3.1 The Retail Prices Index Data Collection System[2]

This section gives details of the RPI data collection system. It highlights the main characteristics of the collection system. Some of these characteristics will be the subject of further comment in the context of the results of the comparative analysis of the RPI and scanner data samples mentioned above and the subject of this paper.

3.1.1 Background

Data are collected for the RPI in two ways: local collection by price collectors who visit shops to determine prices available in each location and a central collection for those stores with a national pricing policy or for items for which a local collection would not be cost effective. Since 1995 the local collection of price data has been contracted out to a private-sector company. The tendering process leading up to the award of the contract acted as a catalyst for a number of initiatives. These included the move from a purposive sample to a random sample of outlets and the introduction of handheld computers for the collection of price data in the field.

3.1.2 Sampling Procedures for Local Price Collection

Current methodology for the selection of locations from which we collect local prices, introduced in 2000, aims to give each shopping center in the United Kingdom 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 is required for local price collection, and the number to be selected within each of the regions is determined by taking a pro-

2. Details of the RPI methodology are given in Baxter (1998).

portion equal to the proportion of total U.K. 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, it is not cost effective to collect from areas too small to provide a reasonable proportion of the full list of items, therefore, locations that had fewer than 250 outlets were excluded. Second, and for similar reasons, 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 used as a single location in the probability sampling, producing a final sample of locations.

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, in contrast to outlet sampling, 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 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 in the area of that particular item. 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 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. In addition, it spreads the sample to include a wider range of products and varieties than would be covered if a very tight description were employed.

3.1.3 Sampling for Centrally Collected Prices and Prices Obtained over the Telephone

In some instances prices are collected centrally without resorting to the expensive activity of sending price collectors into the field. Central price collection is done for two distinct groups:

1. 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.
2. Central items, for which there are a limited number of suppliers and purchases of which do not normally take place at local outlets. Examples of these include gas, electricity, and water, whose prices are extracted from tariffs supplied direct by the head offices of the companies involved. These data will be used to create subindexes that are combined with other subindexes to produce the all-items RPI.

In addition, the prices of some locally collected 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, for which there is no outlet as such, and entrance fees to leisure centers, for which there are unlikely to be any ambiguities over pricing and in cases in which a trip to the center may be relatively time-consuming for the collection of just one price. These prices are combined with data obtained by price collectors as necessary.

3.1.4 Price Reference Day

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

3.1.5 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 United Kingdom. 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 4 percent 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, for which a user cost approach is adopted.

3.2 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 indexes increase. Scanner data provide the potential to deliver up-to-date and accurate information on

1. The number of sales over a chosen period of individual products uniquely identified by the bar code number,
2. The total value of those sales and by implication the average transaction "price" or unit cost, and
3. An analysis by the characteristics, outlet type, and geographical location of the individual products concerned.

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" (unit cost) recorded by scanner data does not take into account the specific needs of index compilers to measure according to a strict set of predetermined rules that disallow certain discounts, such as those relating to damaged stock. 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 reused bar codes, in order to make them usable for statistical purposes.

3.2.1 Main Definitional Differences between Scanner Data and Data Collected Locally for the Retail Prices Index

The main differences between the two data sets are the following:

1. The RPI is a sample that covers all transactions conducted in retail outlets by private households for private domestic consumption. Scanner data cover EPOS sales (coverage of prices for outlets not using bar code

scanning is dependent on a survey and is not of equal quality) and may exclude "own" brands. It does not distinguish between commercial customers and others in the sales figures it provides.

2. RPI data measure individual transaction prices according to RPI conventions, mainly by taking display price. They therefore exclude conditional discounts (for example, where a "club" card is required), two-for-one offers, personal discounts offered on a one-time basis by shop managers, and discounts on discontinued or damaged stock. Scanner data measure average revenue generated after discounts given by whatever method; they will include discontinued or shop-soiled stock and may 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).

3. RPI data relate to prices charged in a set sample of outlets and therefore do not include the effects of outlet substitution. Scanner data, on the other hand, relate to current transactions in all outlets and therefore include outlet substitution.

The numerical impact of these differences is not known. However, it is clear that the impact will not necessarily be constant over time but, rather, will vary with market circumstances, and that differences are likely to be greater for some goods than others.

In addition to the main differences, other characteristics of the two data sources need to be borne in mind when one compares display prices in shops and corresponding scanner data. In particular,

1. In the case of prices collected from shops there is the potential for a relatively large sampling error due to the small number of prices that may be gathered for a particular product variety (the RPI sample is not designed to provide reliable information at this level of detail, particularly for goods and services for which there has traditionally been a wide variation in price). Scanner data can provide almost total coverage.

2. The RPI records prices for a particular day in the month, whereas the scanner data used for this exercise cover a whole month.

3. The sample for local price collection is designed to be self-weighting, and therefore the data set of prices does not distinguish between different types of retailers such as multiple and independent. This is unlike the unit values available from scanner data, which can identify different outlet types separately (although the detail of the categorization varies between market sectors). This means that there is a potential problem with differences in the mix of outlet types between the two data sources, both in a single month and varying over time. This can lead to inconsistencies in the comparisons that cannot be easily corrected for. (This is countered by the fact that scanner data provide full weighting information so that actual, rather than implicit, weights can be applied.)

3.3 Research Design

The research consisted of two separate but related exercises: (a) the benchmarking of RPI product and variety selection against corresponding scanner data, and (b) 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.

The benchmarking exercise involved a comparison of the relative distributions by product and variety for each of five preselected items: televisions, washing machines, vacuum cleaners, dishwashers, and cameras.

3.4 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.

3.4.1 Summary of Results

Table 3.1 shows a comparison between the proportionate coverage of scanner data and data collected for the RPI. The figures are ordered to show the top ten sellers for each product group in September 1999 according to sales volume from scanner data, alongside which is the proportion of quotes that are taken for the RPI collection for that item.

The results show some very interesting patterns. In general collectors tended to choose items that were good sellers, although frequently they overcollected 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 percent of quotes, and the seventh most popular, which only accounted for 4 percent of sales, was represented by over 20 percent of quotes. This pattern was repeated in other items.

Even if we investigate a cumulative distribution, problems are evident. In all cases the proportion of RPI quotes that represent the top ten selling models are significantly lower than their sales figures. The reasons for this are not obvious but may be illustrated by an example. In September there is a particular model of washing machine that attracts almost 10 percent of RPI quotes, whereas scanner data indicate that no sales of this particular model took place. This is clearly an anomaly and represents a real difficulty in maintaining the representativity of the sample. It should be noted that

Table 3.1 **Top-Ten Selling Items According to Scanner Data, and Associated Percentage of RPI Quotes**

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	1.0	16.2	10.5	30.1	18.7
Model 2	13.9	25.0	12.8	4.4	13.2	3.0
Model 3	11.0	1.9	11.7	1.8	8.7	1.2
Model 4	8.5	28.6	10.2	8.8	5.7	1.2
Model 5	8.2	3.8	10.1	31.6	4.4	0.6
Model 6	6.9	4.8	10.1	3.5	4.1	20.5
Model 7	6.6	1.9	6.1	8.8	4.1	0.6
Model 8	4.9	4.8	5.6	0.8	3.8	1.2
Model 9	4.4	1.0	4.1	1.7	3.5	0.6
Model 10	3.9	3.8	1.8	1.7	3.4	6.6
	Cameras		Dishwashers		Washing Machines	
	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	28.4	38.4	17.2	2.2	12.0	6.5
Model 2	13.6	1.2	17.1	16.3	11.2	20.3
Model 3	11.9	12.8	9.4	11.9	11.2	2.3
Model 4	7.6	3.5	7.8	5.9	9.8	5.8
Model 5	6.7	1.2	7.3	6.7	6.9	1.4
Model 6	5.6	2.3	5.8	0.7	5.1	4.3
Model 7	4.4	15.1	5.1	23.0	5.1	2.9
Model 8	4.3	3.5	5.1	0.7	4.4	1.4
Model 9	4.0	1.2	4.8	3.0	4.2	1.4
Model 10	3.4	1.2	4.1	0.7	4.1	4.3

the data relate to September, and it is quite possible that a model chosen by a price collector at the start of the year is still on the shop floor being priced but may have limited, if any, sales.

3.4.2 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, it does seem to indicate two things:

1. Despite the instruction to the price collector to chose a product variety that is representative of the sales of that item in each area, often through asking the shopkeeper which is the best-selling item, the pricing of items can apparently be skewed toward products and varieties that scanner data indicate have relatively small sales. Conversely, there is the nonselection of some big-selling items. This at first sight seems odd, given that the instructions to price collectors would encourage the selection of the big sellers and, therefore, may be more to do with outlet selection. Initial indications suggest that another cause may be brand loyalty on the part of collectors. Collectors identify a popular brand early on in their careers as collectors and tend to stay with it, even when their sales fall.
2. The fixed basket approach, in which products and varieties as well as items are only reviewed on an annual basis (except where a replacement is forced on the price collector because an item becomes obsolete and is no longer found in shops), leads to the sample's becoming increasingly unrepresentative as the "fixed" selection of goods in the basket ages over the year. This is not surprising but does raise the issue of whether, for certain items for which models change very quickly, updating of the basket should occur more frequently than once a year. Certainly it suggests that replacements should be introduced before models disappear and the volume of sales contracts to the point that very few purchases are made.

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.

3.5 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. Resource constraints limited the exercise to the three months from August to October 1999. The process of matching price data from the RPI with scanner data on unit values was not always successful despite a series of reconciliation and validity checks. In part this was due to

the fact that descriptions provided by price collectors in the field were 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, will be required to unambiguously match the product variety with one shown on the scanner list.

3.5.1 Practical Limitations of the Matching Process

The degree of successful matching varied between the five items selected. It was most successful for dishwashers, washing machines, and vacuum cleaners, for which over 70 percent of RPI observations (representing about 50 percent of RPI product varieties) were successfully linked to scanner data. Matching was most problematic for cameras, for which only about one-half of RPI quotes (representing about one-third of RPI product varieties) were matched to scanner data (see table 3.2). 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 in which 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, for which the differences are small. This may suggest that an effect is present within items, although 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 data set and those shown by the matched set. This was explored by calculating Laspeyres, Paasche, and Fisher indexes,[3] both for the full RPI set of price data and for the subsample representing matched observations. The results for a Fisher index are shown in figure 3.1 and indicate that the price changes from the subsample followed similar, but not necessarily identical, patterns to those in the full scanner data.

3.5.2 The Results

Despite the limitations to the exercise arising from problems of matching, the results are nevertheless instructive. Table 3.3 gives an overview of matched comparative prices and unit values expressed in terms of both monetary amounts and the percentage of product varieties for which the mean collected for the RPI is higher than that produced by scanner data unit values. We found that for a particular product variety the average price

3. See appendix.

Table 3.2 **Percentage Coverage of Matched Data and Comparison Between Means of Prices for the Whole RPI and the Matched Sample (means in £s): August to October 1999**

	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

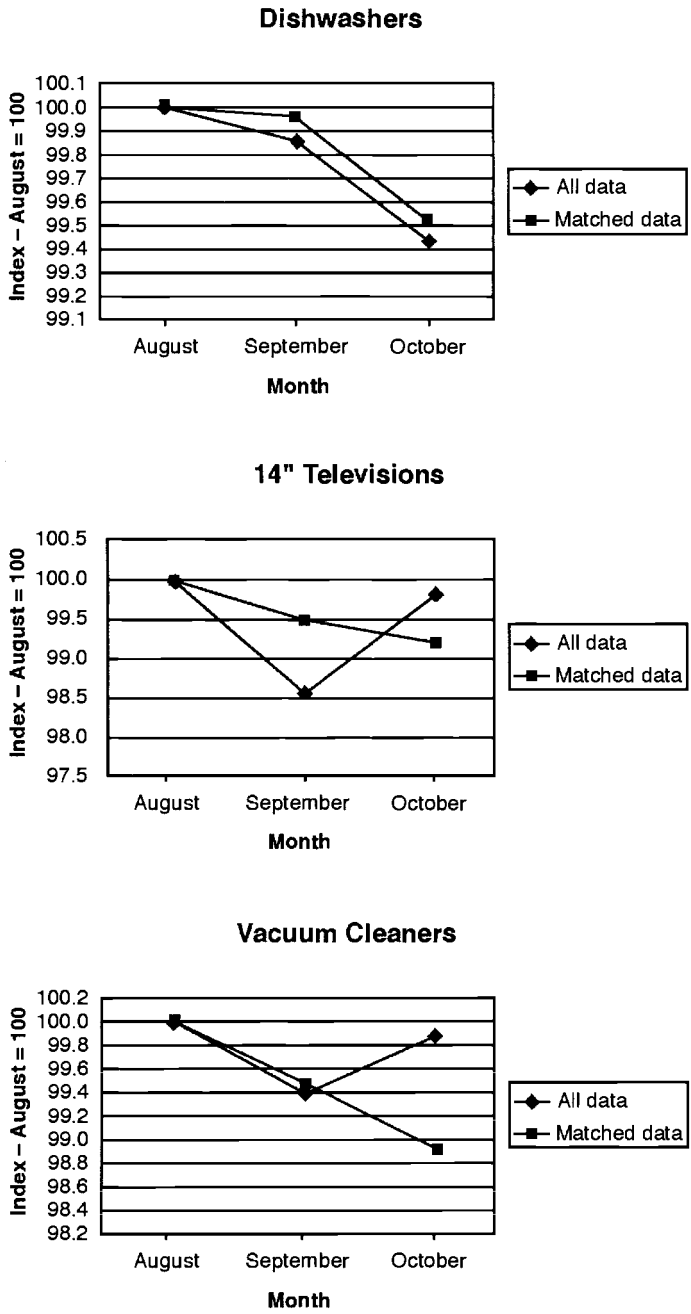

Fig. 3.1 Price indexes for each item calculated using all scanner data and the matched subset, using a Fisher index: August to October 1999

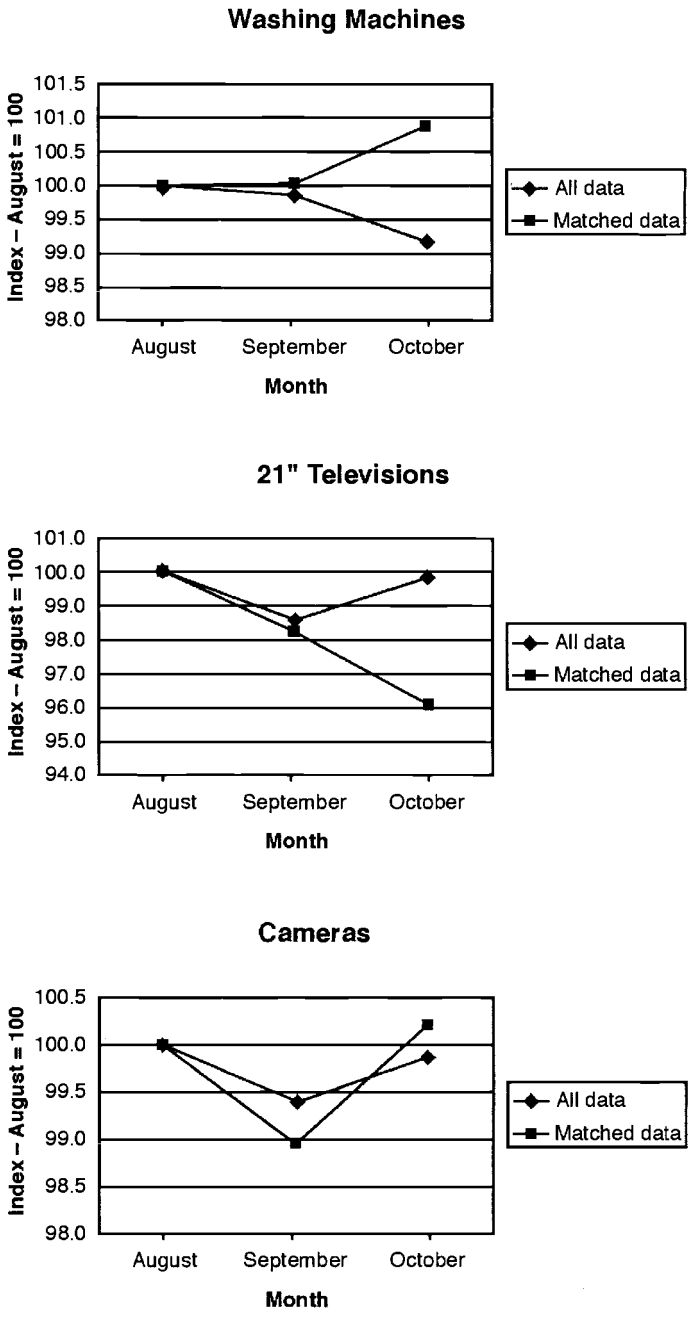
Washing Machines
Index – August = 100
101.5
101.0
100.5
100.0
99.5
99.0
98.5
98.0
August
September
October
Month
All data
Matched data
21" Televisions
Index – August = 100
101.0
100.0
99.0
98.0
97.0
96.0
95.0
94.0
August
September
October
Month
All data
Matched data
Cameras
Index – August = 100
100.5
100.0
99.5
99.0
98.5
98.0
August
September
October
Month
All data
Matched data

Fig. 3.1 (cont.)

Table 3.3 **Average Prices for Recorded RPI Quotes and Scanner Unit Values (proportion of comparisons when the average for RPI quotes exceeded scanner unit values for individual product varieties (means in £s): August to October 1999**

	August	September	October
Dishwashers			
Mean of RPI quotes	339.8	347.4	355.9
Unit value	318.0	323.1	337.2
Proportion of RPI means > unit values	67	67	54
Washing machines			
Mean of RPI quotes	374.0	367.9	362.8
Unit value	348.8	337.0	342.8
Proportion of RPI means > unit values	74	71	67
14" televisions			
Mean of RPI quotes	134.9	134.7	140.6
Unit value	128.1	128.8	129.6
Proportion of RPI means > unit values	52	56	68
21" televisions			
Mean of RPI quotes	307.2	287.2	281.1
Unit value	287.6	271.7	285.4
Proportion of RPI means > unit values	57	69	54
Vacuum cleaners			
Mean of RPI quotes	135.7	131.8	128.0
Unit value	128.9	124.5	119.2
Proportion of RPI means > unit values	42	50	53
Cameras			
Mean of RPI quotes	65.0	71.4	66.0
Unit value	63.2	66.7	63.7
Proportion of RPI means > unit values	62	65	59

recorded by price collectors was higher than that for the scanner data unit values in more than 50 percent of cases for all items other than vacuum cleaners. This was most notably the case for washing machines, for which in 72 percent of cases the collected data produces a higher average. Looking at the average price for the whole product 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 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 can be seen from the analysis given in table 3.4, which shows the deviation of the medians to be much lower than the deviation of the arithmetic mean values.

An indication of the dispersion in the absolute and percentage deviations is given by the coefficient of variation in table 3.5. As a measure of dispersion it is unaffected by the different means for the different products. The results are quite variable, showing substantial variations in price and unit

Table 3.4 **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

Table 3.5 **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

value differences for vacuum cleaners, dishwashers, and 21" televisions, although with less variation for other products.

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. Alternatively, the use of the geometric mean as an aggregator may be supported, given that it is less influenced by outliers than its arithmetic counterpart.

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, differences occur in (a) dishwashers, vacuum cleaners, and 14" televisions, for which changes in the prices recorded by price collectors are consistently higher than the changes shown by scanner data; and (b) cameras, for which, conversely, the changes shown by scanner data are higher.

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 ex-

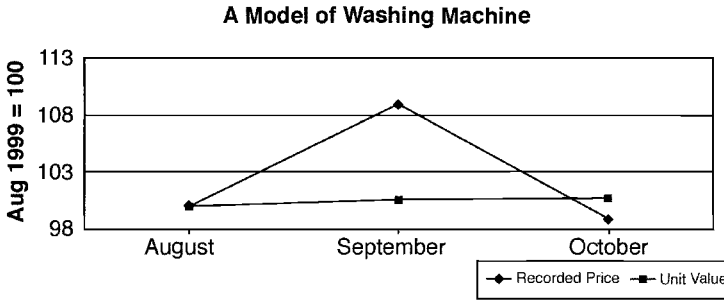

Fig. 3.2 Difference in price movement between scanner data and RPI data for a particular brand of washing machine: August to October 1999

plain but may be due to differences in the mix of outlets and, in particular, the changes that occur over time in market shares and the outlets making sales and therefore appearing with changing weights in the scanner data. This problem of lack of homogeneity was referred to in section 3.2.1. Despite these possible explanations, there yet remains a degree of mystery why some of the differences in price trends occur. This is best illustrated by reference to particular examples. The results for a specific model of washing machine are examined in detail in the paragraph that follows.

Although for washing machines as a whole no appreciable price trend differences were shown between RPI data and scanner data, one model stood out as being very different. The RPI average display price for this model increased by 9.4 percent between August 1999 and September 1999, compared with a much smaller increase of 1.5 percent in average unit value from scanner data before falling back to below its average August price. There was, therefore, a net drop in price over the two months compared with a net increase in price shown by the scanner data (see figure 3.2). Given the fact that there were twenty-six price quotes in the RPI sample, it is unlikely that sampling variability in the matched sample is a major factor. Reference to the outlet type given by the scanner data provides some insight, because this shows not only large differentials between outlet types for prices and price (unit value) movements over time (in scanner data terms unit revenue) but also large variations in the volumes of sales.

3.6 The Effect of Different Store Types

As we have said, one of the main differences between the scanner data and the RPI sample is the difference in the mix of store types. Table 3.6, which has been compiled from scanner data, suggests that, at an item level, there can be significant differences in both unit values and unit value changes for different store types.

For example, in October the average price of a particular brand of wash-

Table 3.6 **Effect of Shop Type on the Unit Value of One Brand of Washing Machine**

Store Type	Unit Value August	Unit Value September	Unit Value October	Percentage Change August to September
Multiple	301.2	304.9	303.3	1.2
Mass merchandiser	306.8	301.1	307.4	1.8
Independent	322.9	329.1	330.0	1.9
Catalog	384.5	388.1	386.7	0.5
All stores	316.8	321.4	320.8	1.5

ing machine varied from £301, in multiple chains of stores, to £384, for sales from mail-order catalogs. Additionally, the change in prices between August and September shows a large variation, with the smallest price rise being 0.5 percent, for mail-order catalogs, and the highest being 1.9 percent, 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, because RPI quote data are not classified as finely as scanner data, although this is clearly an area that we need to investigate further.

For scanner data, clearly, the store mix is changing continually as sales fluctuate between the different sectors. It is important, therefore, that we fully understand the effects of these changes and how they fit into the conceptual basis of the RPI before scanner data can be used to provide weight data for index construction.

3.7 Using Scanner Data to Provide Explicit Weights for Aggregation Formulas

The calculation of indexes for items within the RPI is done using either a ratio of average or average of relatives formula (see the appendix). In neither case is any explicit weighting used in this calculation, and so the implicit assumption for the average of relatives (used in the RPI for these items) is that all quotes are equally important. 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 seen in scanner data. Table 3.7 shows a comparison of indexes of price changes using the ratio of averages and a Laspeyres-based weighted average using scanner data to provide weights.[4]

The comparisons show differences in all cases, with the possible excep-

4. See appendix.

Table 3.7 **Comparison of Indexes Using Unweighted 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

tion of 14" televisions, and in some cases they are quite substantial—for example, 4.5 percentage points for washing machines in September. However, there appears to be no consistent pattern in the magnitude or direction of these differences. That there are differences should not be a surprise. Table 3.1 showed the degree of difference between the proportions of quotes for scanner data and RPI quotes, and it is these differences that are translating into differences in index calculations in the two cases. The reasons behind these differences are unclear, although it is likely that two particular aspects of the data collection contribute significantly. First, collectors are asked to select a variety that is "reasonably representative of the item as a whole." This is a very skilled judgment to make, and we cannot be certain that collectors are making the most appropriate choices in all cases. The second is that we endeavor to follow the same quotes over the period of a whole year. This means that, for a fast-moving technology, even if we start with a representative sample in January there is a real possibility that by August the items are no longer selling as well as at the start of the year and the sample is no longer representative of shopping patterns.

3.8 Conclusions and Implications for Sampling and the Collection of Price Data

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

U.K. RPI. It also points to a number of ways in which scanner data might be utilized to further ensure representativity of item and product varieties in the context of traditional forms of price collection in which prices are observed in shops. The research does not necessarily indicate that current sampling procedures lead to bias, but it does invite the prospect of additional controls and procedures to keep in check the potential for bias.

The starting point of any practical consideration must be whether the items indicated by scanner data truly represent the market they are chosen from. Do they cover a large proportion of the market? Do their changes in price give a true representation of the price changes in the goods they are chosen to represent? If the answer is yes, we can then make the following practical recommendations:

1. The introduction of some form of quota sampling based on scanner data is likely to help in providing a representative sample. In the current practice price collectors are given generic price descriptions and are asked to select for pricing the most representative product variety in the shop being visited. Using a quota sample would, for instance, provide a mechanism for ensuring a better representation of different brands.
2. As an alternative to a quota sample, scanner data could be used to postweight the quotes obtained by price collectors to produce a more representative final sample. Although this has many obvious problems, we will be exploring the principle further.
3. Deterioration in representativity during the life of the "representative" basket, even one that is fixed annually, can be quite marked for some items for which, for instance, the turnover of models is relatively high. In these cases scanner data may provide, at least in those areas where scanner data coverage is good, a useful check on representativity and indicate certain areas for more frequent updating of the representative basket by the introduction of planned "forced replacements." This could be done either by a change in the basket triggered by an algorithm based on scanner data or by perhaps the more practical alternative, prior agreement, in which case manufacturers or retailers warn us that changes are happening.
4. 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 purchasing decisions by consumers.
5. Scanner data by store type indicate that special care must be taken to ensure a proper spread of outlets in the RPI sample and that scanner data may be used for poststratification where there is reason to believe that the sample achieved under current RPI sampling practices is not totally self-weighting.

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

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

Appendix

Formulas of Elementary Aggregates and Index Formulations

$$\text{Laspeyres} = \frac{\sum P_t Q_0}{\sum P_0 Q_0}$$

where P_t = price at time t; Q_0 = quantity sold at time 0; and time 0 = the base month.

$$\text{Paasche} = \frac{\sum P_t Q_t}{\sum P_0 Q_t}$$

where P_t = price at time t; Q_t = quantity sold at time t; and time 0 = the base month.

$$\text{Fisher} = \sqrt{\frac{\sum P_t Q_0 \sum P_t Q_t}{\sum P_0 Q_0 \sum P Q_t}} = \sqrt{\text{Laspeyres} \times \text{Paasche}}$$

where P_t = price at time t; Q_t = quantity sold at time t; and time 0 = the base month.

$$\text{Average of Relatives} = \frac{1}{n}\sum \frac{P_t}{P_0}$$

$$\text{Ratio of Averages} = \frac{\frac{1}{n}\sum P_t}{\frac{1}{n}\sum P_0}$$

References

Baxter, M., ed. 1998. *The Retail Prices Index: Technical manual.* London: The Stationary Office.

Bradley, R., B. Cook, S. G. Leaver, and B. R. Moulton. 1998. An overview of research on potential uses of scanner data in the U.S. CPI. In *Proceedings of the third International Working Group on Price Indices,* ed. Bert Balk, 169–88. Voorburg, the Netherlands: Statistics Netherlands.

Silver, M. 1995. Elementary aggregates, micro-indices, and scanner data: Some issues in the compilation of consumer price indices. *Review of Income and Wealth* 41:427–38.