Estimating Price Movements for Consumer Durables Using Electronic Retail Transactions Data: An Empirical Study on Televisions

First draft for NBER conference on scanner data and price indexes

Arlington, Va., Sept 15-16 2000

By

Robin Lowe and Candace Ruscher
Prices Division
Statistics Canada

Views expressed are those only of the authors and not necessarily those of Statistics Canada.

1.0 Introduction

The emergence of electronic data records[1] that are kept by retailers, recording every transaction, provide new opportunities to price index makers. This paper describes the work that has been conducted in recent years at Statistics Canada, with the aim of improving the Consumer Price Index. Research[2] has documented some of the costs and benefits in the use of scanner data for food items. However, the benefits may be even greater for products where there is more rapid product development such as consumer durables, especially those based on electronics: audio and visual equipment, cameras, and home computers (Silver et al., 1995, 1997). The discussion in this paper focuses on televisions, but the same approaches should probably apply to other similar goods.

The principal advantage of scanner data is that they record sales actually made, for an extensive array of products. Current practice for price index construction at Statistics Canada consists of selecting a small number of representative products, and monitoring their prices. How many, if any, sales are made, however, is unknown, nor is anything known about the diversity of price changes in the group of products that the sampled products represent. Scanner data should be able to improve information in that area. In current practice it usually happens that the exact item being observed becomes unavailable. For consumer durables this is almost inevitable as manufacturers routinely modify their products. When this happens a replacement must be made in the sample, and the comparison between the qualities of the replaced item and its replacement often triggers an adjustment to the index measurement. This is where one of the weaknesses of scanner data lies, as the larger number of records may preclude giving the same level of care to the quality adjustment process. A critical issue then is one of weighing the advantages of additional observations provided by the scanner data against the disadvantages of paying less attention to changes in the quality of the sampled items.

What follows is an analysis of the price indexes that were calculated for televisions. It compares two different approaches—the current practice at Statistics Canada and variants using scanner data. The first part of the paper describes the sample and the quality change evaluation process for televisions over the last nine years, and identifies the main issues for improvement. The second part of the paper examines the results using scanner data for the period December 1997 to November 1999. There are at several ways in which these sources might be used. Some – using the database to periodically evaluate quality differences, or using it to periodically to assist sampling and weighting representative products - require only occasional use of these data to complement the existing methodology. Most attention, though, is given to the use of these data to replace the existing methodology. At this time the work is still in the investigative stage, but if these data were to be used in production comparative costs would be important and they are touched on briefly. None of the ways of handling the data seems entirely satisfactory without detailed examination of the micro-data, so it appears that a practical use of this source will involve using only a part of it.

[1] For simplicity, electronic data records will be called scanner data in the rest of the paper.

[2] See Reinsdorf, Dalén, Hawkes, de Haan and Opperdoes, and Scobie.

2.0 CPI index construction for televisions

Televisions exhibit relatively uncommon price behaviour, as their prices have fallen steadily, both in nominal terms and as measured by price indexes. The available data set starts in June 1990 and figures are shown through November 1997. For most of this period there are two representative products: a 20" and a 27" colour TV, but a television in the 32” to 36” ranges was added in June 1999. Deviations from the preferred specification are allowed, for example, 19" screens are permitted in the first representative product, and 25" to 28" screens are accepted in the second. Since the most common screen sizes are 20" and 27", they will be featured in the rest of the paper.

Over the study period, the total Canadian CPI sample averaged about 220 observations until early 1995, when it dropped to about 140 - two price observations from each of about 70 outlets. Items were replaced on average, about once a year, causing a quality change evaluation each time. Quality change adjustments were forced either because the existing item became unavailable. There were no changes to the item selection initiated by the statisticians during the study period other than the addition of 32” televisions.

Routinely the specific items being surveyed have to be replaced in the sample. Comparisons between the replaced and replacing item may affect the index. There are two kinds of replacement. In the first, the replacement is directed; the old model has been updated, often with some minor changes, and the manufacturer has assigned a new model number. Often there may be no other changes to the product at all. The second kind of replacement occurs when the product is no longer available and there is no updated version available either at that outlet. This may be because the manufacturer has stopped making that product, or just that the retailer is not selling it any more. In these cases the price collector finds a replacement that may be quite different from the old item.

When the change is minor the price evaluators can value them quite easily. In most cases they have brochures from the manufacturers to help them assess changes in the specifications of models. Many times, a small change in a specification accompanies a model change — in the type of remote control supplied, in the warranty coverage, or in the number or placement of jacks, for example. In the second situation, where the replacement is not a modification of the old item, it is more difficult. The price collector is asked to find a similar item if possible, but one that is a volume seller. The volume seller requirement sometimes results in a model that is quite different from the previous one. Price evaluators have some guidelines for taking into account the price difference due to slightly different screen sizes, from mono sound to stereo sound, and for other common improvements, so they can make reasonable comparisons in many cases. However, the comparison is complicated when it involves a change of manufacturer. While it is recognised that manufacturers do vary in quality, it is difficult to compare and estimate by how much. A frequent shift in sourcing is a phenomenon particular to certain retail outlets. Most retailers carry certain manufacturers and change rarely. Some, however, switch manufacturers frequently, making the best buys they can each time.

In evaluating replacements the prices of the old and new item and their descriptions are compared. If the price of the replaced item had been specifically discounted immediately before its disappearance that price is restored to its normal level before making the comparison. Analysis of the results of quality change evaluations during the nineties shows certain patterns. In a little more than half the cases the result is to link the replacement into the index to show no price change.[3] There is a difference depending on whether the replacement was minor or major. Arbitrarily, we use a 10% change in the reported price to distinguish between minor changes, mostly modifications, and major changes. Doing that, we find that minor changes are much more likely to induce a judgement – over 60% of all instances of a minor change. In a significant portion of these the editor judged that quality had not changed at all, all the nominal price change was pure price change. On the other hand, when the difference between the prices of the old and new item is more than 10%, implying a major change, a judgement is only invoked about 45% of the time, and even less, less than a third of the time, if the manufacturer has changed. The single best improvement in quality change evaluation would be a better way to evaluate such large changes.

Given this programme of quality assessment, the questions arise: what is its impact on the index, and what would be the result if different treatments were used? We have recalculated the index numbers for the study period using a number of scenarios reflecting different treatments of quality change.

Applying different assessments of quality change

The first scenario replicates current practice.[4] The second scenario stretch-links[5] all quality changes—an option that is easy to adopt. The third scenario splices large price changes only, but keeps the existing results for small changes. The fourth scenario is the reverse of the third—adjustments for large price changes are kept, but small changes are spliced. The purpose of scenarios 3 and 4 is to see the relative impacts of large and small changes on the index. The fifth scenario, a simplistic one, calculates the index with all quality changes ignored and all price changes accepted as pure price change. This is what would be obtained if all quality changes were ignored. Finally, there is a variant, scenario 1a, which does not reflect an alternate method of assessing quality changes, but employs a different computational handling of links. In this scenario, most stretch-linked quality changes are excluded from the matched sample for the month in which the quality change occurs – that is, only identical items have their price matched from month to month, a procedure that underlies some analysis of scanner data.

---

3 *This method is similar to working with matched records from month-to-month except for timing. As usually there is no overlap period where both replaced and replacing items are available simultaneously, the method assumes additionally that there is no price movement between the period the old item is last available and the replacing item is first available. Both methods assume implicitly that the ratio of prices equals the ratio of qualities.*

4 *The first scenario does not recreate the historical indexes exactly for a number of reasons, including that the regional strata were simplified for these calculations.*

5 *The term stretch-link is used to describe the case where the replacement in time t+1 shows no pure price change from the replaced item in time t. In earlier papers this has been described sometimes as a splice, but splicing implies both prices are available in the same period*

TABLE 1
Percentage change in TV index, Canada, June 1990 to November 1997

Scenario	All TVs	20" TV	27" TV	27" TV less 20" TV
1 Current practice	-25.9	-22.1	-30.4	-8.3
2 All quality changes stretch- linked	-21.4	-19.0	-24.3	-5.3
3 Large price changes stretch-linked	-23.6	-20.7	-27.7	-7.0
4 Small price changes stretch-linked	-23.7	-20.4	-27.3	-6.9
5 All price changes treated as pure price changes	-24.4	-20.9	-28.8	-7.9
1a *Current practice, most splices excluded*[6]	*-26.9*	*-23.1*	*-31.4*	*-8.3*
Range of results Scenarios 1 to 4	**4.5**	**3.1**	**6.1**	

Table 1 shows the separate results for both common sizes of televisions, and for the two combined, for the whole study period. A large part of the drop in prices occurred between 1990 and 1992. The prices used in this study were exclusive of retail sales taxes, so the replacement of the manufacturers' sales tax by the Goods and Services Tax probably accounted for some of this drop.

Three things are apparent in Table 1. First, however treated, prices for 20" TVs have fallen substantially less than prices for 27" TVs. This is true whatever method of handling quality change is used. The differences range from 5.3% to 8.0%. By contrast, the range of results from different treatments of quality change is only 3.1% for 20" televisions and 6.1% for 27" televisions. This suggests that making sure the sample selection is representative is as important as choosing the best quality adjustment technique.

Second, the range of results from different scenarios (3.1% to 6.1%) is small compared to the overall price movement (more than 25%). This is confirmed indirectly by the behaviour of the 89 streams of observations that were in the sample for the whole study period. The greatest amount of price change, about five-sixths, occurred in months where there was no quality change. To some extent this is due to the low inflation during this period. When prices are rising quickly replacements tend to be higher priced items replacing lower priced ones. During this period, the number of replacements where the replacing item was cheaper offset the number of upward ones, so their impacts tended to cancel out.

[6] Those cases where neither the price nor the quality changed were not regarded as quality changes, so they were not taken out of the sample for the calculation.

When we look at the relative impact of large and small changes we see that large changes had twice as much effect. The difference between stretch-linking all changes and the current method was 4.5%. This is the amount the judgement changes lowered the index. The impact of the judgements on small price changes was about the same as on large price changes (2.3% compared with 2.2% for all televisions), even though stretch-linking was more common for large changes. It is clear that the net impact of assessing small price changes was to lower the index, but it was not obvious that the judgements on the larger price changes would have the same impact. There were approximately equal numbers of upward and downward adjustments, and one must wonder how many of the links on large price changes should be replaced by different judgements.

It is curious that the simplistic approach, scenario 5, produces a result fairly close to the official index (scenario 1). This would not be the case for all time periods. Between 1990 and the end of 1991, the index under scenario 5 fell sharply compared to the official index, then rose since the end of 1995. They were virtually equal at the end of 1997. The periods (1990 to 1992, and 1992 onwards) correspond to periods of weakness and recovery in the Canadian economy, and the result is consistent with consumers trading down, then up, accordingly. This provides some validation of the changes in item selection that have occurred over the period, despite the limitations imposed by the specifications.

Third, the computational practice of keeping the linked observation in the matched sample for the month in which it is linked has a fairly significant impact. The impact is about half the impact of applying judgements. Scenario 1a shows that the drag on indexes by stretch-linking was about 1% over the period, while the impact of quality adjustment was 4.5%. As prices were falling for this commodity, stretch-linking has kept the index higher. Preliminary testing on other commodities suggests that this may be a general result, particularly for durable goods whose prices are tending to decline.

3.0 Calculations based on scanner data

3.1 Matched records

We have sought scanner data from individual retailers as well as purchasing analyses from market research companies who also collect from retailers. Preliminary analysis of the market research data suggests that there may be differences between its behaviour at the higher level of aggregation provided and reports from individual retailers, but it has been received too recently for extensive analysis to be done. The analysis presented here will be based on one seller’s data. The emphasis is on the change in results from different applications rather than the absolute results. The data contain the number sold and average price for each identified product code by month and by store. The price is the actual transaction price before taxes. Data from stores were aggregated to create one average price and one total quantity for each product code for each month. The product codes distinguish models to approximately the same level of detail as our official CPI survey—for example, a new production run under a different model number will carry a different product code in this database. The number of product codes reporting sales in any month is about 200 for all stores together. The company carries only a few manufacturers, but a full range of products from those manufacturers. The product code description provides enough information to identify the make and model, so by using brochures or consulting manufacturers, the characteristics of each can be obtained.[7] For eventual comparison with market research data, the results shown here are for sales aggregated over two-month periods, starting with December 1997-January 1998, which is used as the reference period in comparisons.

The range of models comprises six groups: 9”, 13”, 20" (19" to 21"), 27" (25" to 29"), 31/32’” and 35/36" Indexes have been calculated for these specifications separately and grouped together.

As there are numbers sold as well as prices for each period, indexes for each pair of periods can be weighted by sales in either period. The results of the chained Laspeyres, Paasche and Fisher indexes over the twelve two-month periods from December 1997-January 1998 to October-November 1999 are given in Table 2.

TABLE 2
Chained bi-monthly indexes, October-November 1999
(December 1997-January1998 =100), all sales

Size	Laspeyres	Paasche	Fisher
13”	74.8	78.5	75.0
20”	76.7	77.2	77.0
27”	78.7	75.4	77.0
31”	75.1	71.0	73.0
35”	71.8	69.1	70.5
All	76.4	73.8	74.5

[7] The market research data provided has similar detail – quantities of individual models, and average price - but aggregated across all outlets reporting to their survey. Further, it has been aggregated over two-month periods, December-January, February-March, etc.

As expected, except for 20" televisions, the larger the screen size the greater is the rate of decline. All of the index declines are substantial. One reason, which biases the measures downwards, is the effect of models at the beginning and the end of their market life. Typically prices fall most at both those portions of their life. At the beginning prices generally start high and drop as the market grows, and at the end, prices are often discounted to clear the stock. For many models, sales are small during these periods, so they have little influence on the overall averages.

However, for some of the most popular models, when one year's model is replaced by the next, sales can be high during the period over which the replacement occurs. This example illustrates.

TABLE 3
Monthly chained index calculations when one TV model replaces an identical one in the market

1997-98	Model 1		Model 2		Combined average price of models 1 and 2 $	Indexes (February 1997=100)			
	Units sold	Average price $	Units sold	Average price $		Laspeyre	Paasche	Fisher	Based on combined average price
Feb.	91	846			846	100.0	100.0	100.0	100.0
Mar.	99	850			850	100.5	100.5	100.5	100.5
Apr.	66	850			850	100.4	100.4	100.4	100.4
May	73	845			845	99.9	99.9	99.9	99.9
June	68	844			844	99.8	99.8	99.8	99.8
July	53	828			828	97.9	97.9	97.9	97.9
Aug.	85	778	15	883	794	92.0	92.0	92.0	93.8
Sept.	63	732	73	874	808	87.3	89.1	88.2	95.5
Oct.	17	697	79	863	833	84.9	87.5	86.2	98.5
Nov.			87	852	852	83.8	86.4	85.1	100.7
Dec.			114	845	845	83.2	85.7	84.4	99.9
Jan.			68	884	884	87.0	89.6	88.3	104.5
Feb.			55	904	904	88.9	91.7	90.3	106.8

The first three index computations are based on the assumption that Models 1 and 2 are not directly comparable. From February to August, and after October 1997, the movements of the three indexes—Laspeyres, Paasche and Fisher—are identical as there are prices of one model only for all of them. Even in August, a true Paasche index cannot be calculated because there is no observed price for Model 2 in July, and in November, a true Laspeyres index cannot be calculated because there is no observed price for Model 1 in that month. The three measures only differ between August and October. The last column in Table 3 is based on recognising that the two product codes describe identical models, so the sales data can be combined. The index is based on the weighted average price each month. The difference is remarkable.

It is clear that this phenomenon must be corrected for. One possibility is to exclude sales at the beginning and end of market runs. Another is to identify families of models that are identical or very similar and group their sales together as is done in the last column, above. The results of both approaches when applied to 27” televisions are shown in Table 4.

TABLE 4
Indexes in October-November 1999 for 27” televisions with adjustments for overlap. (December 1997-January 1998 = 100)

	Laspeyres	Paasche	Fisher
With no corrections	78.7	75.4	77.0
Data for first and last period excluded	79.3	76.7	78.3
Identical models grouped together	87.2	83.7	85.4

There are disadvantages of excluding data. One is that the exclusion of genuine sales is regrettable, and as model replacements tend to be grouped at two times of the year the exclusions tend to occur together. A second is that the choice of period to exclude data is arbitrary. Furthermore, the difference it makes is not substantial.

Grouping similar models, however, makes a large difference. The groups were defined by the judgements made by price editors when they were faced with replacements in the course of regular surveying. Only those that were judged to be equal in quality were grouped together. There were only seven groupings over three manufacturers. It is likely that more groupings could be made, but there was no evidence to determine them.

3.2 Reducing the database

One of the hoped for advantages of using scanner data was that it could be analysed without too much examination of micro-data. It is clear that this is not the case. It will be necessary to examine and compare streams of data pertaining to different products.
Except for the corrections described all the data that could be matched from period to period was used. This resulted in excluding very little data, about 1% of the revenue of the first period in each matched pair, and about 1.5% of the second period. If the database can be trimmed of the sales of less important products, the number of products whose quality need be kept track of will be reduced. The results for 27” televisions from taking the products that account for 75% or 90% of sales in each period-to-period comparison are given in Table 5.

TABLE 5
Indexes for 27' televisions, October-November 1999, (December1997-January1998=100), under various selections of subsets

Criteria for selection	Laspeyres	Paasche	Fisher	Number of products included
On reported data				
All sales	78.7	75.4	77.0	99
90% of sales in first period	79.6	77.2	78.4	45
90% of sales in second period	79.7	76.1	77.9	54
75% of sales in first period	83.3	87.5	83.0	37
75% of sales in second period	81.3	77.2	79.2	34
On data with similar products grouped				
All sales	87.2	83.7	85.4	88
90% of sales in first period	89.2	87.5	88.3	37
90% of sales in second period	86.5	85.5	85.5	47
75% of sales in first period	92.5	92.0	92.2	28
75% of sales in second period	90.5	87.6	90.0	27

Most of the savings in the number of products to consider come from cutting the data to the first 90% of sales. The lower number of products for the grouped data is solely due to the replacement of several products by one in each group. The differences when excluding the bottom 10% are small. The differences when including only the top 75% of sales are substantial, and so are the differences according to whether we order the sales according to their value in the first period in each comparison, or the second. Either way the data is filtered, but the filtering matters; the indexes are systematically lower if the second period sales are used. For symmetry we could used the sales of the two periods combined to select data, but the consistent direction of the difference suggests something else at play. Selecting by the second period sales leads to more new products being included, whose prices seem to be falling faster than the older products that are on the way out. We believe by excluding the lesser sales that we are avoiding somewhat the racheting problem when products are being replaced by similar products. However, not all replacements are by similar products, and some may argue that by excluding a quarter of sales, and waiting for new products to establish a large enough share of the market unnecessarily delays the price falls induced by their introduction. The share of the marginal product at the 75% sales total is about 4% in most periods. Most worrying is the fact that the choice of 75% or 90% or any other level is arbitrary, but it affects systematically the index measure.

3.3 Regression approaches

The other way to adjust for the changing mix of quality in different models is to use multiple regression. The method and justification is widely available (see Silver(1999) for an exposition and references). It depends on transforming the description of each product sold into a list of its characteristics, and regressing the price on the characteristics. As manufacturers provide detailed specifications of their products there is a wealth of description available.

The regressions are run for models of the following form:

$$\ln P_i = X_{it}ß_t + u_{it}$$

Where: P_i is the price of model i in time t

X_{it} is a vector of characteristics describing model i in time t

$ß_t$ is a vector of parameters representing the implicit prices of the characteristics in time t

and u_{it} is an error term representing the factors not incorporated in the model.

Most of the characteristics are dichotomous variables representing the presence of a certain physical attribute or identification with a particular brand. Of the few which are not, the most significant variable is the screen area. A list of the characteristics collected is given in appendix 1.

There is a high degree of correlation among the variables so that the list of significant variables can be reduced considerably, and the ones that remain must be considered as representatives for others. The most satisfactory model, based on relationships over the two-year period, depended on the following variables: size, size squared, whether the brand is Hitachi, Panasonic/JVC, RCA/Sanyo, Samsung, or Sony (Sony was the excluded option, being generally considered to command the highest prices), the incidence of a Picture in Picture feature, with one tuner, or two, the incidence of Surround sound, either a basic or a more sophisticated version, and the number of S-video inputs, front and rear, front inputs being preferred.

Calculating chained Laspeyres, Paasche and Fisher indexes, using as weights the average quantities of the various characteristics each period we get, values of 89.8, 92.2 and the chained Fisher index, 91.0, for October/November 1999, with December 1997/January 1998 =100.

Another index estimate was run using the model derived in Moulton. That study ranged over a longer period and a wider range of products. In our data from one seller there was no variation in home delivery, console or LCD display. However, to the extent that the North American market is unified, at least to the extent of the variety of products available a similar hedonic equation may apply.

The results for that model are, for Laspeyres, Paasche and Fisher, respectively: 90.3, 93.4 and 91.8. Thus the two models produce similar results, with the Moulton model showing less decline as should be expected. The main differences between the two models is that the one derived from this database uses a finer distinction of the varieties of Surround sound, and S-video inputs than the Moulton model could as it had to cover several more years. The other variable in the Moulton model, whether there was a universal remote controller, was also coded to finer distinctions, but this variable proved not to be significant. Consequently, the refinement of these features, which made later products more attractive could be picked up by one regression model but not the other.

These results seem plausible. They are at the high end of the range of results got from matching data, but it is recognised that the intervention in editing it has not been completed. Nevertheless, there are some drawbacks to using this method in regular production. First, the models are derived from examining the data after all the records are available. In running a monthly or bi-monthly index one would want to be able to change the list of significant characteristics at the time the changes occur, rather than afterwards. Second, the high proportion of variance explained, around 90% on monthly data-sets, is misleading. By far the most significant variable is screen size, which explains about 85% of price variation. It is not necessary to group all products together irrespective of size. The scanner data gives us good information on total sales by size, so that if we had separate price indexes for each size category we could produce accurate indexes. However, if we stratify by size it is not nearly so easy to design reliable models. Third, some of the characteristics are given too much significance while other changes cannot be captured. It will be remembered that close to a half of the amount of quality change adjustment in the current procedure occurred because of small modifications to products. Most of these would not register in a regression model. On the other hand, some characteristics that can appear significant in the regression model were valued as insignificant in current production. For example the variable S-video describes the existence of connections for a high-quality feed from another device. In products where this has been added as an update from the previous version this improvement has been regarded as an insignificant quality change. However, for most of this period Sony made two series of televisions for each screen size, an "S' series and a "V" series, one substantially more expensive than the other but with generally similar characteristics. This particular characteristic is one that distinguishes the two series. Fourth, it is a lot of work to collect the characteristics on all the models to be used in the regression.

4.0 Costs

The cost savings in using scanner data are the field expenses currently incurred in collecting data and the processing and editing costs. The in-house costs would be replaced by the costs of collecting and processing the scanner data. The current sample includes three price observations from each of about seventy outlets. Given an average time per store of fifteen minutes and a driving distance of about ten kilometres the annual cost, allowing for checking and overhead, would be in the order of $10,000 a year. Many of the prices are collected from outlets that provide many other prices as well, so unless the visit to the outlet can be avoided, which is possible only if all products were collected by scanner data, this is an overestimate of the field savings.

The collection costs from any individual company are relatively small, and once the processing has been set up to massage the data, collection and checking are straightforward. With a large number of retailers reporting processing costs would not be insignificant, though probably still no more than current processing costs. The major increase in costs would be with the editors. At present, though they have to keep up-to-date with product developments across the whole field, they only have to evaluate between one hundred and two hundred quality changes, many of which are similar – the same model being replaced by another in different locations. To collect and codify thirty or more characteristics for all the models that may be sold, and appear in the scanner data is a job several orders of magnitude higher. So any method that depends on using all scanner data requires an increase in resources or a simpler approach.

5.0 Conclusion

The opportunities provided by the scanner data are extensive, but pose new problems. The scanner data have to be managed carefully. There are challenging questions to answer concerning the choice of data to be included in the calculations, how to group together different products and how to describe them.

One of the hoped-for advantages of using the scanner data was that because of the large amount of information, amounting to a census for the chosen source, little micro-editing would be required. However, this has not happened. It is clear that the assumption underlying any use of matched samples—that relative prices reflect relative qualities—does not hold here. At the very least, replacements that are really continuations of the same product under a different brand name must be recognised. More generally, regression analysis might be applied if the characteristics can be obtained quickly enough. It is essential that products be classified effectively. A basic classification is into size categories, for each of which price measures should be developed.

The database from an individual retailer does not provide enough diversity to estimate regression models for televisions by size category with any degree of degree of reliability, but it may be possible from the market research company database. This will be the next step in the analysis. This database will also be analysed for its behaviour using matched records; because of the broader range of products and outlets, one would expect that the concentration of sales by product would be less. Whether or not the sensitivity to filtering data will be as great remains to be seen, the database provided includes some filtering already.

Because of the cost of collecting full descriptions of all products, it may be more practical to work with a subset. It would be easier to identify the main families of products that account for these sales and monitor their quality changes directly. The same applies to collecting the characteristics of products to be included in regression models. The drawback of this is that that method is a step back towards the position where new varieties of product are not included in the measurement soon enough. Whatever is chosen would be a compromise; at this stage we cannot yet estimate how big a compromise it would be.

References

Dalén, J. (1997), *Experiments with Swedish scanner data*, Proceedings of the Third Meeting of the International Working Group on Price Indices, pp. 163-168, April 16-18, 1997, Voorburg, Statistics Netherlands.

de Haan, J. and E. Opperdoes (1997), *Estimation of the coffee price index using scanner data: simulation of official practices* and *Estimation of the coffee price index using scanner data: the choice of the microindex*, Proceedings of the Third Meeting of the International Working Group on Price Indices, pp. 183-202, April 16-18, 1997, Voorburg, Statistics Netherlands.

Hawkes, W.J. (1997), *Reconciliation of consumer price index trends with corresponding trends in average prices for quasi-homogeneous goods using scanner data*, Proceedings of the Third Meeting of the International Working Group on Price Indices, pp. 145-162, April 16-18, 1997, Voorburg, Statistics Netherlands.

Lowe, R. (1997), *Item selection and quality change in the Canadian CPI*, Proceedings of the Third Meeting of the International Working Group on Price Indices, pp. 275-290, April 16-18, 1997, Voorburg, Statistics Netherlands.

Moulton, B.R., T.J. Lafleur and K.E. Moses (1998) *Research on improved quality adjustment in the CPI: the case of televisions*, Bureau of Labor Statistics, U.S. Department of Labor, Washington D.C.

Reinsdorf, M. (1996), *Constructing basic components for the U.S. CPI from scanner data: a test using data on coffee*, Presented at NBER Conference on Productivity, Cambridge, Massachusetts, Working Paper 277, U.S. Bureau of Statistics.

Scobie, H. (1997), "Use of scanner data – the impact of new goods", Unpublished paper prepared for meeting of the Price Measurement Advisory Committee, May 1997, Ottawa, Statistics Canada.

Silver, M. (1995), "Elementary aggregates, micro-indices and scanner data: some issues in the compilation of consumer price indices", *Review of Income and Wealth*, Series 41, Number 4, December 1995, Cardiff Business School, University of Wales, College of Cardiff.

Silver, M., C. Ioannides, and M. Haworth (1997), *Hedonic quality adjustments for non-comparable items for consumer price indices*, Proceedings of the Third Meeting of the International Working Group on Price Indices, pp. 203-214, April 16-18, 1997, Voorburg, Statistics Netherlands.

Silver, M. and S. Heravi (1999), *The measurement of quality-adjusted price changes*, Proceedings of the Fifth Meeting of the International Working Group on Price Indices, August 24-26 1999, Reykjavik, Iceland.

Appendix 1: Characteristics used in regression analysis

(Yes/no except where stated)

Brand	
Size	Screen size
Stereo	0=MONO, 1=MTS
DBX	
SAP	
Comb Filter	1 = 2 line, 2=2 line digital, 3=3 line digital, 4=3D y/c, 5=advanced digital
Picture in picture	1 = 1 tuner, 2=2 tuner, 3= twin view (side by side)
Surround Sound	with built in speakers 1=basic 2= 3D, SRS, Dolby pro logic 3=advanced digital or DTS
Internal speaker on/off	
External speaker option	1 = for 4 speaker surround sound
Type of Speakers	0=side 1=front, 2=dome 3= vertical side firing
Number of speakers	number
Audio Output	
Notch Filter	
A/V program outputs	number
Audio inputs Rear	number
Audio Inputs Front	number
RF Inputs	number
S-Video rear	number
S-Video front	number
Fixed audio output	
Variable audio output	
Remote control	1= tv only 2=basic universal 3=hometheatre universal. 4=joystick universal
On screen programming	1=basic menu 2=icon or rolling icon 3=bitmat with pulldowns or presets
Channel guard	
Game and video guard	
TV Lock	
Sleep timer	1=limited preset times, 2=15 to 20 min intervals up to max 3=program for any time
On/off timer	1=yes,2=prog incl channel(eg 2 events) 3= same as 2 but more times and days
Closed caption	
Auto Channel programming	
Channel labelling	
Volume correction	1= manual, set by channel 2=auto for ads and channels
Picture/colour correction	
Headphone jack	
Favourite channel	1= yes 2= pop up screen 3= preview with PIP 4=view and hear with PIP & multiple screens
Xds	
Commercial skip	
Scan velocity modulation	
V-CHIP	

Component video in Wireless headphones

Appendix 2: Regression results

A) Model derived from original data

Estimates of parameters

Per	Const	Sam	Hit	Pan/J	RCA/	Size	Size^2	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR
1	5.188		-0.288	-0.208	-0.315	0.035	6.97E-04	-0.027	0.081	-0.100	0.135	0.195	-0.060
2	5.322		-0.313	-0.170	-0.313	0.018	1.10E-03	0.031	0.148	-0.135	0.179	0.151	-0.068
3	5.177		-0.299	-0.199	-0.288	0.031	7.59E-04	-0.036	-0.060	-0.015	0.031	0.239	0.068
4	5.079	-0.173	-0.386	-0.098	-0.355	0.025	1.02E-03	0.031	0.005	0.008	0.230	0.081	-0.095
5	4.579	-0.194	-0.096	-0.013	-0.254	0.064	1.60E-04	0.046	-0.044	0.035	0.262	0.067	0.084
6	5.191	-0.033	-0.226	0.076	-0.287	0.002	1.57E-03	0.025	-0.042	0.073	0.293	-0.060	-0.007
7	4.968	-0.081	-0.288	0.017	-0.330	0.029	9.40E-04	0.093	-0.011	0.136	0.223	-0.022	0.036
8	4.736	-0.025	-0.213	0.183	-0.128	0.039	6.23E-04	0.041	0.011	0.173	0.285	-0.010	0.070
9	5.657	-0.397	-0.272	-0.158	-0.455	-0.001	1.20E-03	0.087	0.037	-0.006	0.087	0.156	0.133
10	5.881	-0.303	-0.252	-0.106	-0.353	-0.029	1.81E-03	0.033	0.060	0.065	0.018	0.155	0.175
11	5.411	-0.287	-0.336	-0.138	-0.542	0.009	1.17E-03	0.074	0.089	0.113	0.012	0.077	0.241
12	5.908	-0.281	-0.304	-0.082	-0.350	-0.034	1.87E-03	0.061	-0.018	0.156	0.117	0.133	0.238

Sam= Samsung, Hit = Hitachi, Pan/J= Panasonic or JVC, RCA/= RCA or Sanyo

t-stats

Per	Const	Sam	Hit	Pan/J	RCA/	Size	Size^2	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR
1	15.60		-2.39	-2.33	-3.47	1.20	1.07	-0.20	0.67	-0.91	1.18	1.72	-0.36
2	21.00		-3.22	-2.53	-4.37	0.84	2.31	0.31	1.61	-1.49	2.24	1.86	-0.58
3	15.30		-2.31	-2.41	-3.17	1.11	1.25	-0.30	-0.50	-0.13	0.31	2.31	0.48
4	17.40	-1.16	-3.13	-1.38	-4.30	1.03	2.04	0.34	0.05	0.09	2.97	0.96	-0.71
5	12.02	-1.14	-0.69	-0.13	-1.98	2.01	0.24	0.36	-0.32	0.27	2.34	0.60	0.46
6	10.88	-0.49	-1.68	0.15	-2.43	0.75	1.17	0.73	-0.08	0.89	1.80	-0.15	0.15
7	11.44	-0.16	-1.38	1.66	-0.98	1.17	0.92	0.35	0.08	1.34	2.59	-0.09	0.44
8	19.91	-4.23	-1.91	-2.17	-5.42	-0.04	2.51	0.97	0.38	-0.07	0.94	1.91	1.08
9	21.11	-3.23	-1.81	-1.51	-4.03	-1.32	4.13	0.42	0.71	0.84	0.19	2.00	1.51
10	20.72	-3.10	-2.40	-2.04	-6.29	0.43	2.81	0.89	1.15	1.38	0.13	0.90	2.08
11	22.87	-3.27	-2.20	-1.40	-4.16	-1.62	4.60	0.72	-0.24	2.08	1.08	1.59	2.03
12	22.87	-3.27	-2.20	-1.40	-4.16	-1.62	4.60	0.72	-0.24	2.08	1.08	1.59	2.03

Period	1	2	3	4	5	6	7	8	9	10	11	12
Adj R^2	.952	.963	.928	.932	.885	.910	.877	.872	.887	910	.927	.925
N	37	45	48	67	61	61	53	59	79	83	81	78

Weights for physical characteristics

Period	Size	Size^2	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR
1	24.32	628.52	0.03	0.15	0.43	0.11	0.34	0.07
2	25.10	665.90	0.05	0.22	0.46	0.15	0.38	0.09
3	24.55	638.61	0.06	0.16	0.35	0.12	0.31	0.07
4	24.36	635.97	0.07	0.14	0.36	0.11	0.31	0.03
5	25.03	665.14	0.09	0.15	0.43	0.22	0.38	0.04
6	25.35	682.07	0.07	0.14	0.46	0.21	0.45	0.04
7	24.71	653.69	0.05	0.14	0.45	0.21	0.43	0.06
8	24.84	655.31	0.06	0.10	0.37	0.16	0.37	0.04
9	24.75	650.22	0.08	0.11	0.42	0.16	0.46	0.02
10	24.66	645.99	0.07	0.11	0.39	0.15	0.49	0.05
11	24.70	647.26	0.07	0.13	0.39	0.16	0.53	0.08
12	25.09	666.52	0.09	0.15	0.40	0.17	0.58	0.11

Appendix 2 (cont.)

B) Using Moulton model

Estimates of parameters

Per	Const	Sam	Hit	Pan/J	RCA/	Size	$Size^2$	PIP1	PIP2+	SSound	Svideo	Remote
1	5.06		-0.18	-0.12	-0.22	0.04	4.35E-04	0.06	0.03	0.00	0.20	0.06
2	5.36		-0.20	-0.10	-0.22	0.01	1.20E-03	0.07	0.12	-0.04	0.16	0.10
3	5.16		-0.26	-0.15	-0.23	0.03	7.50E-04	0.00	-0.07	0.04	0.20	0.06
4	5.17	-0.13	-0.26	-0.03	-0.27	0.01	1.16E-03	0.05	-0.02	0.10	0.14	0.15
5	4.49	-0.24	0.00	-0.04	-0.30	0.08	-9.0E-05	0.11	0.01	0.06	0.17	-0.08
6	5.47	-0.13	-0.11	0.06	-0.28	-0.02	1.86E-03	0.06	-0.02	0.06	0.10	0.19
7	5.02	-0.15	-0.13	0.01	-0.32	0.03	8.45E-04	0.12	-0.02	0.12	0.12	0.08
8	4.98	-0.13	-0.09	0.10	-0.19	0.02	8.55E-04	0.11	0.13	0.14	0.10	0.06
9	5.76	-0.40	-0.23	-0.15	-0.43	-0.01	1.33E-03	0.08	0.05	0.02	0.18	0.07
10	5.96	-0.32	-0.23	-0.11	-0.35	-0.04	1.91E-03	0.03	0.07	0.06	0.17	0.05
11	5.59	-0.34	-0.32	-0.18	-0.55	-0.01	1.36E-03	0.06	0.12	0.07	0.15	0.08
12	5.97	-0.32	-0.20	-0.11	-0.36	-0.03	1.88E-03	0.01	0.02	0.11	0.24	-0.05

Ssound = any surround sound, Svideo = any Svideo connection, Remote = universal remote

t-Stats

Per	Const	Sam	Hit	Pan/J	RCA/	Size	$Size^2$	PIP1	PIP2+	SSound	Svideo	Remote
1	14.78		-1.75	-1.68	-2.37	1.30	0.66	0.49	0.26	0.02	2.34	0.55
2	19.33		-2.19	-1.58	-3.02	0.29	2.32	0.74	1.26	-0.48	2.33	1.08
3	14.73		-2.24	-1.95	-2.40	0.93	1.20	0.03	-0.55	0.45	2.33	0.49
4	14.69	-0.86	-2.15	-0.44	-3.07	0.31	1.90	0.56	-0.16	1.12	1.83	1.27
5	9.94	-1.43	0.00	-0.39	-2.12	1.84	-0.11	0.80	0.05	0.47	1.58	-0.47
6	12.84	-0.96	-0.88	0.75	-2.44	-0.61	2.54	0.50	-0.14	0.51	1.10	1.33
7	9.68	-0.95	-0.80	0.09	-2.19	0.55	0.93	0.90	-0.10	0.82	0.96	0.50
8	10.26	-0.87	-0.57	0.99	-1.36	0.54	1.07	0.93	0.97	1.19	0.93	0.40
9	18.85	-4.49	-1.72	-2.23	-5.16	-0.42	2.63	0.90	0.53	0.25	2.52	0.74
10	21.21	-3.68	-1.82	-1.92	-4.18	-1.54	4.24	0.41	0.89	0.86	2.73	0.67
11	20.53	-3.74	-2.48	-2.79	-6.41	-0.22	3.09	0.73	1.56	0.92	2.29	0.90
12	22.18	-3.76	-1.70	-2.05	-4.20	-1.55	4.38	0.17	0.31	1.57	3.91	-0.62

Period	1	2	3	4	5	6	7	8	9	10	11	12
Adj R^2	.949	.958	.928	.923	.875	.895	.870	.855	.888	.912	.926	.923
N	37	45	48	67	61	61	53	59	79	83	81	78

Weights for physical characteristics

Period	Size	$Size^2$	PIP1	PIP2+	Ssound	Svideo	Remote
1	24.32	628.52	0.03	0.15	0.43	0.41	0.73
2	25.10	665.90	0.05	0.22	0.46	0.47	0.76
3	24.55	638.61	0.06	0.16	0.35	0.39	0.68
4	24.36	635.97	0.07	0.14	0.36	0.34	0.71
5	25.03	665.14	0.09	0.15	0.41	0.42	0.78
6	25.35	682.07	0.07	0.14	0.43	0.49	0.83
7	24.71	653.69	0.05	0.14	0.41	0.48	0.80
8	24.84	655.31	0.06	0.10	0.33	0.41	0.81
9	24.75	650.22	0.08	0.11	0.39	0.48	0.73
10	24.66	645.99	0.07	0.11	0.38	0.54	0.72
11	24.70	647.26	0.07	0.13	0.39	0.61	0.72
12	25.09	666.52	0.09	0.15	0.40	0.69	0.74