

This PDF is a selection from a published volume from the  
National Bureau of Economic Research

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: Estimating Price Movements for Consumer Durables  
Using Electronic Retail Transactions Data

Author: Robin Lowe, Candace Ruscher

URL: <http://www.nber.org/chapters/c9730>

---

# Estimating Price Movements for Consumer Durables Using Electronic Retail Transactions Data

Robin Lowe and Candace Ruscher

---

## 4.1 Introduction

The emergence of electronic data records<sup>1</sup> that are kept by retailers, recording every transaction, provides new opportunities for price index makers. This paper compares the existing treatment for quality adjustment and sample maintenance for televisions in the Canadian Consumer Price Index (CPI) with what might be feasible using scanner data. Research has documented some of the costs and benefits in the use of scanner data for food items (see Reinsdorf 1996; Dalén 1997; Hawkes 1997; de Haan and Opperdoes 1997a,b; Scobie 1997). However, it stands to be of greater benefit for consumer durable products, especially those based on electronics: audio and visual equipment, cameras, and home computers (Silver 1995; Silver, Ioannides, and Haworth 1997). The discussion in this paper focuses on televisions, but the same approaches can probably apply to other durable 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. The more complete records from scanner data should ensure that a more representative sample of products is monitored.

Robin Lowe and Candace Ruscher are in the Prices Division of Statistics Canada.

The views expressed in this chapter are those only of the authors and not necessarily those of Statistics Canada.

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

In current practice it usually happens that the exact item being observed becomes unavailable. For consumer durables this is almost inevitable because manufacturers routinely modify their products and introduce new ones, making existing ones obsolete. When this happens, a replacement must be made in the sample, and the comparison between the qualities of a replaced item and its replacement often triggers an adjustment to the index measurement. This is where one of the weaknesses of scanner data lies, because 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.

The first part of this 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 need to update the sample more quickly and to improve the quality adjustment procedure. It appears that these two problems are of similar magnitude and that an improvement of the sampling procedure would mitigate the effect of the weaknesses in the quality adjustment methods. The second part of the paper examines the results using scanner data for the period December 1997 to November 1999. Although the database could be used in a limited way—to evaluate quality differences periodically, or to assist sampling and weighting representative products—the attention is given to the use of these data to replace the existing methodology. We find that the breadth of the records identifies clearly different price movements for different sizes of televisions and provides the means to adjust for changes in purchase patterns. We find also, however, that it is difficult to compare qualities of products, even to identify similar or identical products, and that the results are very sensitive to this problem. A number of ways of dealing with it are tried, but none of the ways of handling the data seems entirely satisfactory without detailed examination of the microdata. Doing that examination as well as acquiring the data is costly, so it appears that a practical use of this source will involve using only a part of it, or using it in a limited way that only requires acquisition of the data occasionally.

## **4.2 Consumer Price Index Construction for Televisions**

Televisions exhibit relatively uncommon price behavior, in that 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" color TV—but a television in the 32" to 36" range was added in June 1999. Deviations from the preferred specification are allowed; for example, 19" screens are permitted in the first representa-

tive 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 was replaced by a new variant of the model or because the outlet stopped selling that item altogether. 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 at that outlet. This may be because the manufacturer has stopped making that product or simply because 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 changes are 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, in which the replacement is not a modification of the old item, it is more difficult to value the changes. 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 the selection of 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, a change from mono sound to stereo sound, and other common improvements, so they can make reasonable comparisons in many cases. However, the comparison is complicated when it involves a change of manufacturer. Although it is recognized 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.

For many replacements, the new item is linked into the index to show no

change from the old one. To calculate a price ratio the current price is divided by a reference price. If, for example, the last report for item A was \$20 in September, while the reference price was \$10, and the first report for item B was \$25 in October, the reference price would be changed to \$12.50 so that the ratio remains the same. If the report for item B had been available in September at \$25, the new item could have been substituted into the index calculations then. That method relies on the assumption that the ratio of prices reflects the ratio of qualities. However, because there is not usually a period of overlap, the method of linking from September to October relies on a further assumption, that prices have not changed between the two months. Thus the link is stretched across the two months, and the method is described as stretch-linked in this paper.<sup>2</sup>

The stretch-link method is the default in index calculations. However, in many cases it cannot be justified and the price evaluator imposes his own judgment. Sometimes the change in the product description, (a minor modification for example), is easy to assess, and the reference price will be adjusted for the value of the change independent of the ratio of prices. In some cases the evaluator may judge the value of the change as zero. Sometimes the ratio of prices may not reflect the ratio of qualities because one or other of the prices is out of line. This could be for any of several reasons. It can easily happen if the price of the old item had been discounted before its disappearance. It can also happen if prices generally are falling, and an old product remains in the sample with an unchanged price because it has become obsolete in the market, although still available. On the other hand, a new item might be introduced into the market at an unusually high price to attract the attention of those who will pay a premium for something new. Or the manufacturer may take the opportunity to adjust his price when he introduces a new product. Because a product usually remains in the sample as long as a price for it can be found, it is the price of the old item that is more likely to be out of line, but any of these conditions could hold. In all of these cases the evaluator must make the best judgment possible of the ratio of qualities. All cases in which the evaluator intercedes are called judgments in this paper. The judgment used might either raise or lower the index, depending on the conditions.

Analysis of the results of replacements in the index during the 1990s shows certain patterns (Lowe 1997). Overall the replacement was linked into the index to show no price change in slightly more than half the cases. There is a difference depending on how different in price the replacement was from the item it replaced. If the price of the new item was more than 10 percent different from the price of the old item, we call it a major change;

2. Some products, including televisions, are not surveyed every month, so the stretching can be for more than one month. Televisions are priced approximately every second month, so most stretch-links are for a two-month period.

otherwise it is a minor change. The figure of 10 percent is arbitrary, but it separates the small modifications from the rest. With that distinction we find that the evaluator used his judgment in over 60 percent of the minor changes. In a significant portion of these the editor judged that quality had not changed at all—that all the nominal price change was pure price change. On the other hand, for major changes, the evaluator only imposed a judgment 45 percent of the time, and even less often—less than one-third of the time—if the replacement was made by a different manufacturer. The single best improvement in quality change evaluation would be a better way to evaluate such large changes.

Given this program of quality assessment, these 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.

#### 4.2.1 Applying Different Assessments of Quality Change

The first scenario replicates current practice.<sup>3</sup> The second scenario uses stretch-links<sup>4</sup> for all replacements—an option that is easy to adopt. The difference between these two shows the effect of the evaluators' judgments. In order to see how these differences came about, we show the impact of large changes and small changes separately. The third scenario stretch-links all large price changes but keeps the existing results for small changes. The fourth scenario is the reverse—adjustments for large price changes are kept, but small changes are stretch-linked. Finally, 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. It is also a measure of the movement through time of the unit values of the items in the sample.

The stretch-link method has been criticized as a default method. Other organizations have different methods for comparing items when replacements take place, and some differ from the stretch-link by the means they use to impute the movement between the two periods that an item and its replacement were in the sample. One method is to omit the records for those periods and estimate the movement on the remaining items that were in the sample in both periods. The difference between these methods amounts to the following. Assume there are ten items equally weighted in the sample, and one is replaced. Using common records only, the index movement between the two periods will be the average of the nine price ratios. Using the stretch-link method the movement will be the average of those nine plus an extra ratio of 1. Thus, if prices generally are rising, the stretch-link method

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

4. In earlier papers a stretch-link has been described sometimes as a splice, but splicing implies that both prices are available in the same period.

**Table 4.1** Percentage Change in TV Index, Canada: June 1990 to November 1997

Scenario	All TVs	20" TV	27" TV	Difference between 27" TV and 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	-26.9	-23.1	-31.4	-8.3
Difference between scenarios 1 and 2	4.5	3.1	6.1	

will tend to keep the index down, and if they are falling it will tend to keep it up. Scenario 1A in table 4.1 shows the result of replacing most stretch-links in our actual practice with the movement of common records instead of stretch-links.<sup>5</sup>

Table 4.1 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.

There are several things to note in table 4.1. 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. In fact, the fastest rate of decline for 20" televisions is less than the slowest for the 27" size. On the other hand, the range of results from different treatments of quality change is only 3.1 percent for 20" televisions and 6.1 percent for 27" televisions. What evidence we have on sales indicates that there has been a substantial shift toward larger sizes, that continues toward larger sizes still, 32" to 35". This spread of price movements suggests that making sure the sample selection is representative is in the same order of importance as choosing the best quality adjustment technique.

Second, the impact of the judgments is 4.5 percent overall. When we look at the relative impact of large and small changes in scenarios 3 and 4 in table 4.1 we see that large and small changes had about equal effect. It was to be expected that the net impact of assessing small price changes was to lower the index, because these are mainly small modifications at little or no cost, but it was not obvious that the judgments on the larger price changes would

5. Those cases in which 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.

have the same impact. Since the judgments on large changes tended to lower the index, it must be suspected that the index would have been lower still if the large percentage of changes that were linked to show no price change had been replaced by some kind of judgment. Both this and the lack of adjustment in the sample for televisions of different sizes suggest that any error in this index is likely to be upward.

Third, 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 onward) 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.

Finally, the computational practice of keeping the linked observation in the sample for the month in which it is linked has a fairly significant impact. Scenario 1A shows that the drag on indexes by stretch-linking was about 1 percent over the period, whereas the impact of quality adjustment was 4.5 percent. 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.

### **4.3 Calculations Based on Scanner Data**

#### **4.3.1 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 behavior 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 data-



**Table 4.2** Chained Bimonthly Indexes, October–November 1999, All Sales (December 1997–January 1998 = 1)

Size	Laspeyres	Paasche	Fisher
13"	0.748	0.785	0.750
20"	0.767	0.772	0.770
27"	0.787	0.754	0.770
31"	0.751	0.710	0.730
35"	0.718	0.691	0.705
All	0.764	0.738	0.745

base. 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 it carries 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, one can obtain the characteristics of each.<sup>6</sup> 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.

Because 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<sup>7</sup> indexes over the twelve two-month periods from December 1997–January 1998 to October–November 1999 are given in table 4.2.

As expected, except for 20" televisions, the larger the screen size, the greater the rate of decline. All of the index declines are substantial. One reason, which biases the measures downward, is the effect on the indexes of individual models at the beginning and the end of their market life. Typically prices fall 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. This behavior did not appear in the conventional surveys, partly because models at the beginning of their lives were rarely included in the sample, and partly because a direct comparison of the

6. The market research data provided have 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 to January, February to March, and so on.

7. If the quantities sold and average prices in successive periods are  $q$ ,  $q'$  and  $p$ ,  $p'$  respectively, the Laspeyres price index is  $\sum qp'/\sum qp$ , the Paasche index is  $\sum q'p'/\sum q'p$ , and the Fisher index is the geometric average of the two,  $(\sum qp' \cdot \sum q'p'/\sum qp \cdot \sum q'p)^{1/2}$ . Indexes for longer periods are calculated by chaining the period-to-period indexes.

price of models that disappeared, and their successors, usually produced an adjustment.

For many models, sales are small during these periods, so they should 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. Table 4.3 illustrates.

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 because 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, and they all show a sustained decline.

The last column in table 4.3 is based on recognizing 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. This index shows some price drops during the overlap period, but by the end of the period shown the index is higher than in February 1997, as it should be because the identical model is selling for close to \$900, compared to about \$850 a year earlier.

If these prices had been collected and used in our conventional survey, model 1 would have been replaced by model 2 some time in the period. The evaluator should have judged that because the two models were of equal quality, all the change was price change, and a price increase would have been shown, from \$828 to \$883 if the replacement was made in August, \$778 to \$884 if in September, \$732 to \$863 if in October, or (most likely) from \$697 to \$852 if in November, when a price for the old model was no longer available. Depending on when the replacement was made, the index would differ through the replacement period, but in any case, it would finish the same by December.

Clearly, to avoid this ratcheting effect, one would like to treat these models as identical, but it is not very easy. Although the data for the two models are shown together in table 4.3, they would not appear in such an easy way to compare in the set of scanner data. The only identifying information is a unique code number to identify each model, and some description, which is not standardized. One has to know that these two models are identical ahead of time to treat them as such. Furthermore, pairs of models may not be identical, but very similar, and the same phenomenon of falling prices during the overlap may occur to lower the index. We need to identify and exclude that effect. One possibility is to exclude sales at the beginning and end of market runs, when their price movements may be abnormal.

**Table 4.3 Monthly Chained Index Calculations When One TV Model Replaces an Identical One in the Market**

1997-98	Model 1		Model 2		Indexes (February 1997 = 1)				Based on Combined Average Price
	Units Sold	Average Price (\$)	Units Sold	Average Price (\$)	Laspeyres	Paasche	Fisher	Combined Average Price	
					of Models 1 and 2 (\$)				
February	91	846			846	1	1	1	1
March	99	850			850	1.005	1.005	1.005	1.005
April	66	850			850	1.004	1.004	1.004	1.004
May	73	845			845	0.999	0.999	0.999	0.999
June	68	844			844	0.998	0.998	0.998	0.998
July	53	828			828	0.979	0.979	0.979	0.979
August	85	778	15	883	794	0.920	0.920	0.920	0.938
September	63	732	73	874	808	0.873	0.891	0.882	0.955
October	17	697	79	863	833	0.849	0.875	0.862	0.985
November			87	852	852	0.838	0.864	0.851	1.007
December			114	845	845	0.832	0.857	0.844	0.999
January			68	884	884	0.870	0.896	0.883	1.045
February			55	904	904	0.889	0.917	0.903	1.068

**Table 4.4** Indexes in October–November 1999 for 27" Televisions with Adjustments for Overlap (December 1997–January 1998 = 1)

	Laspeyres	Paasche	Fisher
With no corrections	0.787	0.754	0.770
Data for first and last period excluded	0.793	0.767	0.783
Identical models grouped together	0.872	0.837	0.854

Another is to identify families of models that are identical or very similar and group their sales together. The results of both approaches when applied to 27" televisions are shown in table 4.4.

There are disadvantages of excluding data. One is that the exclusion of genuine sales is regrettable, and because 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 judgments 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.

#### 4.3.2 Reducing the Database

One of the hoped-for advantages of using scanner data was that it could be analyzed without too much examination of microdata. It is clear that this is not the case. It will be necessary to examine and compare streams of data pertaining to different products.

In the initial analysis, all data that could be compared from one period to the next were used. This resulted in the exclusion of very little data. Models that had no sales in the second period in each comparison accounted for only about 1 percent of the revenue of each first period. Models that first appeared in the second period in each comparison accounted for about 1.5 percent of revenue in the second period. We do not, perhaps, need all the data. If the database can be trimmed of the sales of less important products, the number of products whose quality must be kept track of will be reduced. The results for 27" televisions from taking the products that account for the first 75 percent or 90 percent of sales in each period-to-period comparison are given in table 4.5.

To choose the data to be used in each period-to-period comparison, the products that could be matched were listed by revenue in descending order in each month. Then records were included by going down the list until the required percentage of sales, 75 percent or 90 percent, was reached.

The column "Number of Products Included" shows the number of differ-

**Table 4.5** Indexes for 27" Televisions, October–November 1999 under Various Selections of Subsets (December 1997–January 1998 = 1)

Criteria for Selection	Laspeyres	Paasche	Fisher	Number of Products Included
On reported data				
All sales	0.787	0.754	0.770	99
90% of sales in first period	0.796	0.772	0.784	45
90% of sales in second period	0.797	0.761	0.779	54
75% of sales in first period	0.833	0.875	0.830	37
75% of sales in second period	0.813	0.772	0.792	34
On data with similar products grouped				
All sales	0.872	0.837	0.854	88
90% of sales in first period	0.892	0.875	0.883	37
90% of sales in second period	0.865	0.855	0.855	47
75% of sales in first period	0.925	0.920	0.922	28
75% of sales in second period	0.905	0.876	0.900	27

ent models included over the whole period from December 1997 to November 1999. In any matched pair of periods the number of products included is much less. The lower number of products for the grouped data is solely due to the replacement of several products by one in each group. Most of the saving in the number of products to consider comes from cutting the data to the first 90 percent of sales. The differences in the index calculation when excluding the bottom 10 percent are small.

The differences when including only the top 75 percent of sales are substantial, and it also matters whether we filter the data according to their value in the first period in each comparison or the second. For symmetry we could use 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 the inclusion of more new products, whose prices seem to be falling faster than those of the older products that are on the way out. We believe that by excluding the lesser sales we are avoiding somewhat the ratcheting problem when products are being replaced by similar products. However, not all replacements are by similar products, and some may argue that 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 percent sales total is about 4 percent in most periods. Most worrying is the fact that the choice of 75 percent or 90 percent or any other level is arbitrary, but it affects systematically the index measure.

#### 4.3.3 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 are widely

available (see Silver and Heravi 1999 for an exposition and references). This method depends on transforming the description of each product sold into a list of its characteristics and regressing the price on the characteristics. Because manufacturers provide detailed specifications of their products, there is a wealth of description available.

The regressions are run for models of the following form:

$$\ln P_i = \mathbf{X}_{it}\boldsymbol{\beta}_t + u_{it}$$

where  $P_i$  is the price of model  $i$  in time  $t$ ,  $\mathbf{X}_{it}$  is a vector of characteristics describing model  $i$  in time  $t$ ,  $\boldsymbol{\beta}_t$  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 that are not, the most significant variable is the screen area. A list of the characteristics collected is given in appendix A.

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.

Although the selection of significant characteristics was based on the two-year period, we can run a regression for each period separately. We then have a vector of implicit prices  $\boldsymbol{\beta}_t$  (including a constant term) for each period. We can also calculate the average value for each characteristic in each period; for a continuous variable such as screen size it is the average screen size of all models sold, and for a dichotomous characteristic, such as picture-in-picture with two tuners, it is the percentage of models sold that had that that particular characteristic. Call this vector  $\bar{\mathbf{X}}_t$ . (The incidence of the constant is 1.) Then we can estimate a price for period  $t$  using the average characteristics and implicit prices for period  $t$  as  $\exp(\bar{\mathbf{X}}_t\boldsymbol{\beta}_t)$ . We can also estimate a price for period  $t + 1$  using the prices for period  $t + 1$  and the characteristics for period  $t$ . We can call this ratio of estimated prices a Laspeyres price index for  $t + 1$  related to  $t$ . Similarly, we can calculate a Paasche index using the average characteristics in period  $t + 1$ . The implicit prices and the average values of physical characteristics are shown in appendix B.

Calculating Paasche and chained Laspeyres indexes on the implicit

prices and characteristics weights, comparable to those from conventional price and quantity data, we get values of 0.898 and 0.922, respectively, and a chained Fisher index of 0.910, for October/November 1999, with December 1997/January 1998 = 1.

Another index estimate was run using the model derived in Moulton, Lafleur, and Moses (1998). 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: 0.903, 0.934, and 0.918. Thus the two models produce similar results, with the Moulton model showing less decline, as should be expected. The main difference 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 because it had to cover several more years. The other variable in the Moulton model, whether there was a universal remote control, 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 recognized 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 by examining the data after all the records are available. In running a monthly or bimonthly 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 percent on bimonthly data sets, is misleading. By far the most significant variable is screen size, which explains about 85 percent of price variation. It is not necessary to group all products together irrespective of size. The scanner data give 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, whereas other changes cannot be captured. It will be remembered that close to 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 to which this has been added as an update from the previous ver-

sion 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.

#### 4.4 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 kilometers, the annual cost, allowing for checking and overhead, would be on 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, although they would probably still be no more than current processing costs. The major increase in costs would be with the editors. At present, although they have to keep up—to date with product developments across the whole field, they only have to evaluate between 100 and 200 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. Thus, any method that depends on using all scanner data requires an increase in resources or a simpler approach.

#### 4.5 Conclusion

The opportunities provided by the scanner data are extensive, but they 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 recognized. 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 separately.

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 analyzed for its behavior 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 toward 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. In *Proceedings of the third meeting of the International Working Group on Price Indices*, ed. Bert M. Balk, 163–68. Voorburg, the Netherlands: Statistics Netherlands.
- de Haan, J., and E. Opperdoes. 1997a. 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*, ed. Bert M. Balk, 191–202. Voorburg, the Netherlands: Statistics Netherlands.
- . 1997b. Estimation of the coffee price index using scanner data: Simulation of official practices. *Proceedings of the third meeting of the International Working Group on Price Indices*, ed. Bert M. Balk, 183–90. Voorburg, the Netherlands: Statistics Netherlands. Available at [file://www4.statcan.ca/secure/english/ottawagroup/toc3.htm].
- 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 In-*

- dices*, ed. Bert M. Balk, 145–62. Voorburg: Statistics Netherlands. Available at [file://www4.statcan.ca/secure/english/ottawagroup/toc3.htm].
- 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*, ed. Bert M. Balk, 275–90. Voorburg: Statistics Netherlands. Available at [file://www4.statcan.ca/secure/english/ottawagroup/toc3.htm].
- Moulton, B. R., T. J. Lafleur, and K. E. Moses. 1998. *Research on improved quality adjustment in the CPI: The case of televisions*. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.
- Reinsdorf, M. 1996. *Constructing basic components for the U.S. CPI from scanner data: A test using data on coffee*. Paper presented at NBER Conference on Productivity. 17 July, Cambridge, Massachusetts. Working Paper no. 277. Washington, D.C.: U.S. Bureau of Labor Statistics.
- Scobie, H. 1997. Use of scanner data: The impact of new goods. Paper prepared for meeting of the Price Measurement Advisory Committee. 20–21 May, Ottawa, 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* 41 (4): 427–38.
- 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*, ed. Rosmundor Gudnason. Reykjavik, Iceland: Statistics Iceland. Available at [file://www4.statcan.ca/secure/english/ottawagroup/ottawa/list.htm].
- 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*, ed. Bert M. Balk, 203–14. Voorburg: Statistics Netherlands.

## Appendix A

### *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, surround sound, Dolby pro logic, 3 = advanced digital or digital theater sound (DTS)
Internal speaker on/off	

External speaker option	1 = for four-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 = home theater 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 maximum, 3 = program for any time
On/off timer	1 = yes, 2 = programming includes channel (e.g., two events), 3 = same as (2) but more times and days
Closed caption	
Auto channel programming	
Channel labeling	
Volume correction	1 = manual, set by channel; 2 = auto for ads and channels
Picture/color correction	
Headphone jack	
Favorite channel	1 = yes, 2 = pop-up screen, 3 = preview with picture-in-picture (PIP), 4 = view and hear with PIP and multiple screens
Xds	
Commercial skip	
Scan velocity modulation	
V-chip	
Component video in	
Wireless headphones	

# Appendix B

**Table 4B.1 Regression Results: Model Derived from Original Data**

Period	Constant	Sam	Hit	Pan/J	RCA/	Size	Size <sup>2</sup>	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR	Adj R <sup>2</sup>	N
<i>Estimates of Parameters</i>															
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		
<i>t-Statistics</i>															
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	.952	37
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	.963	45
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	.928	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	.932	67
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	.885	61
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	.910	61
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	.887	53
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	.872	59
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	.887	79
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	.910	83

(continued)

**Table 4B.1** (continued)

Period	Constant	Sam	Hit	Pan/J	RCA/	Size	Size <sup>2</sup>	PIP1	PIP2+	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR	Adj R <sup>2</sup>	N
11	22.87	-3.27	-2.20	-1.40	-4.16	-1.62	4.60	0.72	-0.24	0.72	-0.24	2.08	1.08	1.59	2.03	.927	81
12	22.87	-3.27	-2.20	-1.40	-4.16	-1.62	4.60	0.72	-0.24	0.72	-0.24	2.08	1.08	1.59	2.03	.925	78
Period	Size	Size <sup>2</sup>	PIP1	PIP2+	Soun1	Soun2	SvidF	SvidR	Adj R <sup>2</sup>	N							
<i>Weights for Physical Characteristics</i>																	
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									

*Notes:* Sam = Samsung; Hit = Hitachi; Pan/J = Panasonic or JVC; RCA/ = RCA or Sanyo; Size = screen size in inches; PIP1 = picture-in-picture with one tuner; PIP2+ = picture-in-picture with two or more tuners; Soun1 = surround sound coded 1 (see Appendix A); Soun2 = surround sound coded 2 (see Appendix A); SvidF = number of S-type connections at front; SvidR = number of S-type connections at rear.

**Table 4B.2** Using Moulton Model

Period	Constant	Sam	Hit	Pan/J	RCA/	Size	Size <sup>2</sup>	PIP1	PIP2+	SSound	Svideo	Remote	Adj. R <sup>2</sup>	N
<i>Estimates of Parameters</i>														
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		
<i>t-Statistics</i>														
1	14.78		-1.75	-1.68	-2.37	1.30	0.66	0.49	0.26	0.02	2.34	0.55	.949	37
2	19.33		-2.19	-1.58	-3.02	0.29	2.32	0.74	1.26	-0.48	2.33	1.08	.958	45
3	14.73		-2.24	-1.95	-2.40	0.93	1.20	0.03	-0.55	0.45	2.33	0.49	.928	48
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	.923	67
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	.875	61
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	.895	61
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	.870	53
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	.855	59
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	.888	79
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	.912	83
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	.923	78

(continued)

**Table 4B.2** (continued)

Period	Size	Size2	PIP1	PIP2+	Ssound	Svideo	Remote
<i>Weights for Physical Characteristics</i>							
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

*Notes:* Ssound = any surround sound; Svideo = any Svideo connection; Remote = universal remote; all other abbreviations, see notes to table 4B.1.