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: Roundtable Discussion

Author: Dennis Fixler, John S. Greenlees, David Fenwick, Robin Lowe, Mick Silver

URL: http://www.nber.org/chapters/c9731

Roundtable Discussion

Dennis Fixler, John S. Greenlees, David Fenwick,
Robin Lowe, and Mick Silver

Dennis Fixler: These papers illustrate the way that scanner data can help statistical agencies improve the quality of their consumer price indexes. In particular, the papers highlight the enormous amount of information that is available and simultaneously illustrate the implementation difficulties that must be overcome with the use of scanner data.

It should be noted that these papers, and the other papers at the conference, focus on the scanner data collected at retail outlets. However, scanner data can also be collected at households—indeed, some of the firms providing the retail scanner data also have household-based databases. Given that the goal of using scanner data is to improve the quality of consumer price indexes, an examination of the household scanner data might prove useful.

The underlying issue addressed in these papers is whether scanner data can replace the price quotes collected by statistical agencies. The answer to this broad question is comprised of answers to a set of subsidiary questions. I shall address some of these questions.

1. *What kind of information is available with scanner data that is unavailable with the collected data or vice versa?* Clearly, the scanner data provide a greater number and scope of transactions for a particular product. For the Consumer Price Index (CPI) ready-to-eat breakfast cereal categories examined in the Richardson paper there are fifty-five collected price quotes, whereas the scanner data provide 80 thousand quotes. The attendant reduction in index variation is an obvious advantage of scanner data. However, the greater volume of data evidently does not include product characteristics. The Richardson paper points out that the U.S. CPI collection of product characteristics is much greater than that provided by scanner data.

In addition, the Lowe and Ruscher paper speaks of the need to look at brochures from manufacturers to obtain product characteristic data for the products in the scanner data set.

2. *How do unit values in the scanner data compare with the collected CPI data?* To some extent the answer depends on the time period being examined. In the case of a week in which a retailer does not change the price, the unit value should equal the collected price if the CPI collection occurs on a day within that week. However, it is not possible to use the weekly scanner data because there are too many missing observations; the Bureau of Labor Statistics (BLS) has decided that a month's worth of data must be used. Even at this frequency there is a need for imputing missing values. The result is that there is a considerable difference between the price measure in the scanner data and the transaction price collected by the CPI field economist. Table 2.3 in the Richardson paper shows a difference between the CPI cereal index and the amalgamated scanner index. Table 3.2 in the Fenwick, Ball, Morgan, and Silver paper presents a qualitatively equivalent finding for the United Kingdom. These findings may derive from the difference between transaction prices and unit values or from differences in weights and in the case of the CPI the inclusion of outlets that do not use scanner equipment. It would be interesting to examine the extent to which the difference in the price measures explains the differences in the indexes.

3. *What is the currency of the set of products? Is the age distribution of products in the CPI older than the average age in scanner data?* The advantage of scanner data is that whereas statistical agencies have to choose among different varieties of a given product to price, it is possible with scanner data to collect prices on all of them. Furthermore, the scanner data provide exact information about the sales attributable to the various varieties. Table 3.1 in the Fenwick, et al. paper shows that there can be a considerable difference in the items that are priced and in the number of price quotes collected for each item. These differences lead to some of the scanner data–based indexes moving quite differently from the corresponding component in the U.K. Retail Price Index (RPI).

4. *Are quality adjustments easier with scanner data?* One of the oftenremarked limitations of current quality adjustment practices is that they are too reliant on subjective judgments. Such judgments come into play when a good that is being priced disappears and a replacement is selected. The immediate question is whether the replacement is a comparable (close) or noncomparable substitute. Only in the case of a noncomparable replacement will a quality adjustment be made. At first blush one might think that the Universal Product Code (UPC) based scanner data make such a question relatively easy to answer. Unfortunately that is not the case. Although distinct goods are assigned unique UPC codes, the assignments are made in such a way that the closeness of one number to another does not indicate

that the two goods are close substitutes. In fact, the UPC codes are not useful in making judgments about how products should be grouped.

When a noncomparable replacement is made, the analyst must determine the value of the difference between the new item and the replaced good, and it is here that subjective judgments can enter. However, scanner data do not help reduce this possibility. As mentioned earlier, the Lowe and Ruscher paper discusses how scanner data do not provide a set of recorded product characteristics that is sufficient to meet the needs of the quality adjustment practices. The U.S. CPI program also found that it collects a much more extensive set of product characteristics than is available with scanner data, as stated in the Richardson paper.

It should be noted that for the purposes of quality adjustment the attention does not have to be limited to data collected at cash registers—the typical source of scanner data. Some companies collect product varieties and prices for different products and sell them to marketing firms for use in planning product development, advertising, and so on. The U.S. CPI program is currently using data from such a company in their effort to develop a quality adjustment protocol for audio products.

5. *Can scanner data provide a better way for selecting items to be priced?* The answer seems to be yes; Fenwick et al. illustrate this point in their work comparing the number of quotes collected by top-selling items and the number of quotes used in the RPI. However, there can be too much data. The Lowe and Ruscher paper speaks of the advantages of reducing the scanner data set to those products that sell best; they show that such a decision requires the determination of a significant contribution to sales level as well as the determination of the relevant time period.

In sum, it does not appear that scanner data can generally replace the prices collected by statistical agencies. However, these papers show that the use of scanner data can help to improve the current computation of the CPI. The main advantage seems to be for item selection and maintaining the currency of the set of items being priced.

John S. Greenlees: The three papers by Lowe and Ruscher; Richardson; and Fenwick, Ball, Morgan, and Silver have presented valuable and sometimes fascinating empirical results.

All the papers highlight the potential benefits of scanner data. Those benefits are considerable. Not necessarily in order of importance, they include more up-to-date samples and weights; the use of superlative formulas, made possible by current-period sales information; and much lower index variance.

It is interesting, by the way, to note how this variance contrast is demonstrated in the papers. At the level of the individual products analyzed, such as breakfast cereal and televisions, the official Consumer Price Index (CPI) samples are so small that their sampling variances can explain even wide observed divergences between the CPI and scanner-based indexes. This

makes it very difficult to determine whether there is an issue of bias in one or the other index.

Each of the papers also highlights the obstacles standing in the way of the use of scanner data. I believe it was John Astin, at a recent meeting of the Conference of European Statisticians, who said that within a few years the statistical agencies in many countries would be using scanner data to produce price indexes. I tend to agree with John, but there is a real question about how we can or will get there from here.

I will proceed by highlighting three issues, or themes, with illustrations from the three papers. The papers take different approaches and make different empirical uses of scanner data, but these issues arise in each. The emphasis will be on how I view these issues from my perspective in the U.S. CPI program.

1. *Sample representativeness.* The first major issue that has to be addressed by any statistical agency is the representativeness of the scanner data samples. This is the major emphasis of the paper by Fenwick et al.; they present very interesting comparisons of product samples between the U.K. Retail Price Index and their scanner data set. There are noticeable differences between the samples in item mix, average prices, and price changes. Given these differences, the authors recommend using scanner data for sample selection, as a resampling trigger, and for reweighting.

The Canadian authors, Lowe and Ruscher, focus their attention on analysis of quality change rather than on sample distribution, but their results nevertheless suggest that representativeness is an important issue. Lowe and Ruscher observe different rates of price change for 20" and 27" TVs, implying that there could be important effects of using the more up-to-date sample distributions taken from scanner data.

Finally, the Richardson paper also notes differences in coverage. Notably, scanner data cover weekends, unlike the CPI samples, but the scanner samples may not cover the universe of geography or outlet type appropriate to the CPI measurement objective. This is reflected in the differences between Richardson's indexes computed with scanner data only and his "amalgamated" indexes that combine scanner and CPI data. The New York cereal scanner data used thus far exclude mass merchandisers and wholesale clubs, so they could differ systematically from the universe of cereal outlets in terms of item mix or price change.

Mention of amalgamation raises another point with respect to index variance. Amalgamation is employed in the U.S. research effort as a means of eliminating potential bias, but it can come with a potentially significant cost in sampling error. A simple spreadsheet calculation can illustrate this point using sample sizes drawn from the Richardson paper for realism. Assume that the scanner index sample is 1,454 times as large as the regular CPI sample but that the underlying data are otherwise similar. Further assume

that in construction of the amalgamated index the scanner index replaces 51/55 of the CPI sample and receives a corresponding share of the weight. In that case the amalgamated index will have a standard error about 27 percent as large as the CPI. Put another way, although the scanner index reduces the standard error by a factor of 38 compared to the CPI, the amalgamated index reduces the standard error by a factor of less than 4.

It seems to me that this issue of representativeness is absolutely crucial. Scanner data offer sizable payoffs as a means of evaluating and updating CPI item samples. It is equally important, however, to make sure that the scanner samples are representative.

In the U.S. context, I would also note that probability selection is inherent in our CPI for items as well as outlets, so some of the methods and solutions proposed by Fenwick may not be *directly* portable. For example, the use of scanner samples for postweighting or for quota sampling would be difficult, or at least complicated, to combine with our standard procedures for selecting and weighting items.

2. *Quality adjustment.* Quality adjustment is a second key issue with respect to scanner data. As noted, this is the focus of the Lowe and Ruscher paper. Scanner data, because of their volume and because of the way we receive and process them, do not lend themselves readily to the methods for comparison of old and new items that are presented in ordinary CPI samples at the time of item replacement (substitution). That is, the need to select and compare replacements when items disappear from shelves is a well-known problem in CPI construction. The problem may be even more severe for scanner data sets, however.

The purely mechanical processing of scanner data can implicitly treat all replacements using the often-criticized "linking" procedure. Tables 4.3 and 4.4 of the Lowe and Ruscher paper are very striking and suggest how the failure to compare the qualities of entering and exiting models can yield undesirable results. This is true even when the index is constructed with scanner data using monthly weights and the superlative Fisher formula.

A parallel issue is mentioned also by Richardson in his sections on imputation and item definition. It seems that we don't have a full solution for the quality adjustment issue yet, in these papers generally or in the U.S. CPI. More discussion of new and disappearing "items" (Universal Product Codes, or UPCs) in U.S. scanner data is in Ralph Bradley's paper presented at this conference (chap. 11 in this volume).

3. *Cost.* For the most part, the three papers do not discuss cost, but one point made in the Canadian paper is the relatively low marginal cost of CPI data collection—around $10,000 annually for televisions. Similarly, the marginal cost of collecting CPI cereal data in New York is probably only $100–200 per month, extremely small compared to the cost of scanner data. In the near term, therefore, the gains from scanner-based indexes likely will be confined to improved accuracy. Any resource savings will be small until

we learn how to employ scanner data to calculate large sections of the CPI. This is especially true if it will be necessary to maintain a direct-collection activity, either for amalgamation or for quality control purposes.

To conclude, I would like to emphasize again the value of the research presented at this conference. The authors have uncovered some issues, like quality adjustment, that may have been underappreciated when statistical agencies first set out to compute scanner-based indexes. I am sure that we will see much more research like this in the near future. Finally, these papers demonstrate that in addition to direct index construction there are many other important potential applications of scanner data, such as for hedonic regression estimation and for sample and weight evaluation.

David Fenwick: It is clear from studies of scanner data that the full potential to use these data in the construction of consumer price indexes has yet to be realized. I believe the reason for this is twofold: first, the lack of responsiveness of scanner data suppliers to the special needs of those who construct price indexes, and second, the limited knowledge of scanner data and a lack of appreciation of their detailed characteristics by price index statisticians. The former is the case not through a lack of dialogue; indeed, price index statisticians are greatly indebted to the main suppliers of scanner data for their cooperation and encouragement in piloting the use of scanner data for index construction. Rather, the lack of responsiveness is due to the new challenges of exploiting, at a micro level, scanner data that traditionally have been used for analysis only at a macro level. Similarly, the knowledge issue arises not through lack of awareness or effort but more through a growing realization that scanner data are a complex data set and that much greater knowledge is required to exploit for statistical purposes any data set that is essentially collected for nonstatistical—in this case, retail management—purposes. Both general points are well illustrated by the extraordinary effort that has been put into the construction of a price index for wheat in New York by the U.S. Bureau of Labor Statistics and for televisions by Statistics Canada. One is tempted to ask whether such efforts would be economic or feasible if applied to the construction of a national index covering all items or whether the vision is one of selective use of scanner data in those circumstances in which its advantages are particularly strong.

Examples of the way in which scanner data do not currently meet user needs include the following:

1. The limited coverage. For example, in the United States scanner data cover no more than 10 percent of transactions. The use of scanner data is uneconomic if price collectors still need to be sent into shops to collect prices for the remaining 90 percent.
2. The recording of average transaction value as opposed to display price. The former can include sale prices for soiled stock and give-aways

such as free videotapes given to customers who purchase a video recorder. It also includes the effect of outlet substitution that occurs when customers transfer their custom to cheaper outlets, something that should be included in a cost-of-living index but not in a "pure" price index such as the U.K. Retail Price Index.

3. The presentation of data, in particular coding. For example, the coding of shop type is not ideal for statistical purposes, and the European Union Classification of Individual Consumption According to Purpose [COICOP] classification for items in the Harmonised Index of Consumer Prices is not used.

There is clearly a challenge here for scanner data providers. For example, would it be possible for individual shop identifiers to be added to enable the monitoring of transaction values across a fixed sample of shops? Could another identifier be added so that shop-soiled or end-line stock can be excluded? Regardless of the answers to these questions, I also believe there is a challenge to price index statisticians, namely, whether some of the conventions used in the index materially affect the index, thus precluding the use of scanner data. For example, are trends in average transaction value materially different from trends in display prices? Have price statisticians collected the latter in the past simply because the transactions values would have been obtainable only at disproportionate cost and were therefore not a practical option?

The complexity of scanner data and the expertise required to manipulate them, the second point I mention in my opening paragraph, raise interesting questions about partnerships between collectors and users and the role of the former in educating the latter. From a user perspective I would certainly welcome more information from individual suppliers on the strengths and weaknesses of scanner data, their representativity, and relevant considerations that need to be taken into account when compiling a price index. For instance, the work reported in the Office of National Statistics (ONS) paper starts from the assumption that scanner data are representative of the consumer sales and can be used as a benchmark for quality control. Whether the latter assumption is a realistic one has yet to be proven, at least to me.

Against this background, what can reasonably be expected from scanner data over the next few years? At this point I should perhaps apologize if my remarks so far have appeared to be pessimistic. This is rather a reflection of the need I feel to throw out a few challenges. I do not take the view that there is no future in the use of scanner data for the construction of consumer price indexes. This would indeed be perverse, given the amount of time and effort we in ONS have devoted to research this area. Scanner data do have a lot of potential and I take the view that there are a number of uses to which such data can be put without delay. First, the research undertaken by ONS in

conjunction the Cardiff Business School shows that there can be immediate gains for item and outlet selection by using scanner data to impose a form of quota sampling. Second, the use of hedonic regression, although in some instances still problematic for quality adjustment, can provide guidelines on relevant item characteristics to allow for when making forced replacements in connection with old items that have disappeared from the shelves.

Finally, what about the direct use of scanner data for price collection? Whether the potential to do so is universally fulfilled depends, in part, on whether the challenges I have laid down both to the data collector and to the data provider are met. Only history will be able to judge this, but I am clear in my own mind that there is potential for much greater exploitation of this very rich data source, both for construction of consumer price indexes by individual statistical offices and, equally important, for international price comparison exercises, a subject that I regret no time was available to debate.

Robin Lowe: We have seen in the foregoing papers a range of potential uses of scanner data in our programs and some of the difficulties in using them. I would like to address the issue of how the costs of acquiring the data may affect our programs.

We have acquired data through two sources, directly from retailers and (by purchasing it) from market research companies. Because the costs of capturing the data are hidden in our other activities, we have preferred until recently to deal with retailers directly. We can see the data in a more detailed form, so we can do our own editing. There do seem to be different reporting problems with different suppliers—in accuracy and in dealing with unusual sales and returns—so it is useful to be able to contact the original data source directly. Although the data we have bought appear to be plausible, data from different retailers have been combined before they reach us, and although we are sure they have been complied diligently, we have to accept this on faith. There has also been some loss of detail to preserve commercial confidentiality, which limits our analysis. At the moment, with the limited data we have been collecting for research, the data capture costs have been small, but if we were to collect for a large number of commodities more widely, they would become significant. Because the marginal costs of collecting prices for these products in our current surveys are relatively small we would not anticipate any savings in costs by moving to the use of scanner data; the justification for such a move would have to be in the improved indexes that result.

Cost considerations make it unlikely that we would move to using scanner data for everyday items like food, personal care items, and household supplies. These are relatively simple and cheap to survey, and there are not significant quality change problems; the main weakness in these surveys is a relative lack of diversity in the sample, but it does not appear to be serious compared to other products.

At the detailed level, we have a continuous problem in keeping our sample relevant. I suspect that if we did the kind of analysis shown in the U.K. paper—comparing the distribution of sales by model with the distribution of our sample—we would get similar results. There is inertia built into our sampling method so that the sample will always lag the latest distribution of sales. As we do not have a regular program of resampling, and we have no point-of-purchase survey to assist in item selection, using scanner data to postweight the sample or to guide sample selection is an attractive option. However, purchasing data occasionally for this purpose would be a straightforward extra cost because we would not be saving any of our current production costs. This is why we are presently concentrating on methods to replace completely our current surveys for certain commodities. At the moment, because of the kinds of statistical problems we have outlined here, we cannot justify doing that. However, it would be a pity if we had to ignore such a rich source of data, so I hope we will solve these problems and start working the use of scanner data into some durable goods before long.

Mick Silver: The papers by Richardson and by Fenwick, Ball, Morgan, and Silver are most welcome for their detailed matching of scanner against CPI data. Scanner data have many advantages, including their extensive coverage of transactions. Richardson notes the substantial reduction in standard errors: a factor of $\sqrt{80{,}000/55} = 38.1$. However, the fifty-five price quotes are outlet display prices against which there will be a large number of transactions. If, for example, only 100 transactions take place on average at that price in each outlet, the factor reduces to $\sqrt{80{,}000/5{,}500} = 3.81$.

In CPI practice, quality changes are controlled for by the price collector's comparing prices of "like" with "like." A problem arises when a variety becomes unavailable and only noncomparable replacements are available. Lowe and Ruscher provide an extensive account of the results from different imputation methods. They also compare the results with those from two different hedonic regression specifications that use scanner data as an alternative to CPI practice. The account is rigorous and extensive, so I extend the discussion here in a related area: that is, the use of hedonic regressions and scanner data to supplement current CPI practice to better serve, via an integrated approach, the needs of quality adjustment *and representativity.*

Price collectors collect prices on a variety of goods until they become unavailable. They hold on to the product variety until it dies: when few people purchase it, when it has unusual price changes, and when its quality features are quite different from newer models. The quality adjustment for these unpopular, unrepresentative varieties is thus more difficult. Endemic in the fixed basket and matched model approach is that CPIs become increasingly unrepresentative and quality adjustments more difficult. An integrated approach to quality adjustment and representativity, first, requires that switching between "old" and "new" varieties take place before a variety

dies. Scanner data clearly show that models of, for example, television sets of different vintages coexist in the market. Scanner data or other market information may be available to identify when sales of the existing variety are falling and those of a newer one are picking up, so that price collectors can switch earlier. An integrated approach would, second, use scanner data on market sales and the results of hedonic regressions in the period immediately prior to item selection, to guide the initial selection of varieties. The hedonic regressions will help inform price collectors as to which brands and characteristics are salient in explaining price variation, thus ensuring their sample has a mix of such factors. The scanner sales data should help inform price collectors whether these brands and characteristics have a substantial presence in the market. Guidelines or even quota controls based on such information will help ensure representativity from the outset. However, under the integrated approach the selection of varieties is predicated on the hedonic characteristics. Consequently, any subsequent quality adjustment using hedonic regressions should be more satisfactory.

This use of hedonic regressions to supplement CPI practice includes, for example, the prediction of the replacement's prices in the base period by inserting its characteristics into a hedonic regression equation for that period. Silver and Heravi in this volume outline the use of hedonic regressions for predicting "missing" prices. Our concern is that the hedonic equations used for such adjustments relate closely to the samples of items being collected. For example, assume only Sony and Hitachi TVs exist in the market and surround sound is the only other explanatory variable in a hedonic regression. A subsequent quality adjustment from a "Sony" to a "Sony with surround sound" using the hedonic approach is easier because the sample was set up with this in mind. However, if a model with a new characteristic—say, a widescreen TV—is the only available replacement, then either a new regression on current data or an alternative imputation approach, as outlined in Lowe and Ruscher, is required. The integration of item selection and representativity with quality adjustment is not of course a panacea, but a recognition of a real dilemma and the need for a wider strategy.