Experiments with Swedish Scanner Data

Jörgen Dalén

Abstract: Scanner data from supermarkets are able to provide an account of the full universe of prices and quantities for all goods sold. This fact raises a number of issues on index construction. In this report experiments with scanner data for four ”supermarket” item groups are carried out, showing the difference in index outcome from various approaches.

1. Introduction

The problem with elementary aggregation in a CPI is to a high degree due to the fact that, for most products, Statistical Agencies have not so far had access to sales volumes, which could serve as weights for measured price changes for varieties of commodities in outlets. This is now about to change through the rise of so called scanner data, in which all purchases are registered, typically at the cashier’s desk in the outlet. These data thus contain both the price paid for a commodity and the exact number of items purchased at that price.

In Sweden A.C. Nielsen company now has built up a database with scanner data for sample outlets which, it claims, are representative of the universe of supermarkets in Sweden. Item groups included are basically food and other daily necessities (e.g. cleaning articles and articles for personal care) typically sold in supermarkets. In Sweden almost all supermarkets belong to one of three major chains (called ICA, KF and DAGAB).

For experimental purposes Statistics Sweden has been given access to a portion of these data. The purpose of our experiments is to explore the possibilities to utilise these data in monthly CPIs. This paper reports some early results. Reinsdorf (1996) is a first report from a similar U.S. project dealing with data for coffee purchases in Chicago and Washington. Silver (1995) analyses UK scanner data for colour TVs.

2. The experimental data

The experimental data cover 26 outlets of four different sizes according to the Nielsen size classification which are situated all over Sweden. For these outlets weekly *sales volumes* and *prices* of all varieties within four item groups (*fats, detergents, breakfast cereals* and *frozen fish*) are included for all weeks from October 94 to December 95. Prices are quantity weighted averages in case they have varied within a week. Varieties are defined in two slightly different ways: by a verbal *item description* and by a 13-digit *EAN code.* Other variables in the scanner files were: *region*, *chain* (ICA, KF or DAGAB) and *week*. For identification we chose to work with item description (id) rather than EAN code since sometimes different EAN codes were used for the same item description. In Table 1 we give some descriptive statistics concerning the data sets.
Roughly speaking, the number of varieties in each group is around 100. We can compare this figure with 1269, which is the number of coffee varieties reported by Reinsdorf for the U.S. market. (He used a concept for variety similar to EAN code, though, which involves some double-counting.)

The 20-80 rule seems applicable. 20-28% of the varieties account for 80% of the total sales value. Reindorf reported an even skewer distribution for roasted coffee but similar figures for instant coffee.

Table 1. Descriptive statistics for the item group data sets

	Fats	Deter-gents	Breakfast cereals	Frozen fish
# ids Dec 94	58	104	79	133
# ids Dec 95	65	99	80	126
% ids 94 covering 80% of value	28	20	27	25
% ids common to both periods	79/71	61/64	85/84	77/82
% sales value in common in both periods	82/81	74/57	97/97	98/84
% ids common at outlet level, average	76/69	52/54	81/86	72/75
% sales value for items common at outlet level, average	80/77	60/57	94/93	81/78

The last four rows of table 1 give a couple of measures of the degree of matching between the sets of varieties in December 94 vs December 95. For e.g. fats the figures tell the following story: Of the 58 varieties marketed in at least one of the 26 outlets in December 94, 79% were still there a year later and of the 65 existing December 95, 71% were already there a year earlier. Based on a weighting with sales value the same results were 82 and 81 %, respectively. Looking within a single (average) outlet the unweighted forward (backward) rate of matching was 76% (69%) and the last row gives the corresponding sales weighted figures.

The figures in table 1 correspond to monthly attrition rates of 1.3-4% for the four item groups, somewhat more than the 1-2.5% reported by Reinsdorf.

3. Conceptual considerations

The traditional method of item selection used in many countries (partially so in Sweden) is the representative item method. With this non-probabilistic method, the central office, within a certain item group such as *white bread* defines a specification such as ”*tin loaf, sliced, 300-450 g*”. Within this specification a price collector is to choose one variety in an outlet and observe its price at the point of her visits to the outlet. Thereby a sample of observations is accomplished, consisting of prices for different varieties in a set of outlets in successive time periods, and some function of these is used as the estimator of price change.

Actually, the universe of prices which this sample is supposed to represent has three dimensions:
1. Varieties of products within an item group,
2. Outlets within a geographic area (and/or of a certain type/industry) and
3. Points in time within a certain time period, usually a month.

Scanner data have the potential to provide us with the whole universe of prices in all these three dimensions. (In practice there will be limitations in the outlet dimension, since not all outlets will use scanners [for example market stalls]. Also, at present not all outlets co-operate with Nielsen and there is some item under-coverage in certain product groups.)

Now, given complete data on prices and quantities in all these three dimensions, how do we define the true index of a price change?

Dalén (1992) discussed this matter and presented two basic alternatives - a *fixed base index* (where quantities are held constant at some reference period level) and a *unit value index* (a ratio of quantity-weighted prices over the whole universe in both time periods independently). Diewert (1995) continued this discussion and argued for the use of unit values at least over time and perhaps also over outlets in a market area.

This problem is in fact very difficult and no traditional index formula can take into account all aspects of reality. Index theory works with functions of N prices and quantities at two or more points in time. But in reality both the set of varieties and the set of outlets change over time so that matched comparisons cannot provide the whole answer.

The concept of product quality makes the matter even more complicated. No two varieties are exactly the same, so theoretically not even all prices and quantities are enough to define price change; a set of product characteristics for all distinct varieties that could be entered into some form of hedonic index is also necessary, especially when dealing with durables.

Not only varieties but outlets too could be of different quality due to service level, accessibility, opening hours etc. But at least all time periods should be of the same quality? Not necessarily; if an outlet increases opening hours and for example starts a 24 hour service, the late night hours may be considered to be of a different quality. If, in addition, prices were even different during nights vs. Days, doubts would be cast on the use of unit values over time.

This problem exposition gives an idea of the complexity involved in this issue. Different economic concepts or uses (cost-of-living index, inflation, deflators) may also require different definitions.

4. Experimental indices

After the depressing theoretical background description above, we shall now define the experimental indices that we have worked with so far. They will all be of the matched kind so that the same item descriptions are compared between December 94 and December 95 and the non-matching descriptions are excluded. Varieties (item descriptions) within the item groups will always have fixed weights.

We make four distinctions:

- *Between a direct comparison December - December and a monthly chained comparison in which 12 one-month changes are multiplied*. Direct comparisons are the usual practice in official price indices but monthly chaining has the advantage of needing fewer exclusions and thus making maximum use of the data. It also approaches the concept of a Divisia index where time is continuous and periods are arbitrarily small. On the other hand the path dependence of the chained indices may be considered problematic.

- *Between the Laspeyres and the Fisher index formula*. The Fisher index is a superlative index formula in the sense that it equals or closely approximates a cost-of-living index under a broad range of circumstances - see Diewert (1976) for details. It should therefore, when conditions permit, be preferred to the Laspeyres' index.

- *Between fixed weights for outlets and unit values computed over outlets* (within varieties). Fixed weights for outlets correspond to the economic assumption that differences in price levels between outlets is wholly attributable to differences in quality. Using unit values over outlets, on the other hand, presupposes that all outlets have the same quality. Neither assumption is realistic, the truth is probably somewhere ”in between”. (Note that scanner data of the kind we have acess to are not able to distinguish the so-called outlet bias discussed by the Boskin commission, since the outlet sample is static and does not reflect the opening and closing of outlets in real life.)

- *Between using only prices for the week in which the 15th of a month occurs (the midweek) and using unit values over all weeks (for which most of the days are) within the month*. This corresponds to an issue which has been discussed for some time in Sweden. Unit values over the month is generally the superior concept.

All combinations in these four dimensions, giving 2^4=16 alternatives, are presented in Table 2a-d below, for each of the four item groups. In the table headings the KPI (Swedish CPI) outcomes for the corresponding item groups are also given in their long-term (L) and short-term (K) versions.

Table 2a. Fats. Index outcomes, December 94 - December 95. KPI_L=96.5; KPI_K=98.3

	Direct index		Monthly chained	
	Laspeyres	**Fisher**	**Laspeyres**	**Fisher**
Fixed outlet weights				
Midweek only	98.3	97.5	125.6	98.1
Monthly unit value in outlet	99.5	98.7	112.5	98.3
Unit value over outlets				
Midweek only	95.8	96.0	101.4	94.9
Monthly unit value in outlet	98.0	98.0	101.8	98.7

Table 2b. Detergents. Index outcomes, December 94 - December 95. KPI_L=105.7; KPI_K=106.0

	Direct index		Monthly chained	
	Laspeyres	**Fisher**	**Laspeyres**	**Fisher**
Fixed outlet weights				
Midweek only	107.6	105.8	137.0	103.5
Monthly unit value in outlet	106.4	105.1	122.1	104.1
Unit value over outlets				
Midweek only	105.6	105.1	121.6	105.3
Monthly unit value in outlet	105.3	104.8	110.4	104.0

**Table 2c. Breakfast cereals. Index outcomes, December 94 - December 95.
KPI_L=101.7; KPI_K=101.6 (including 20% snacks)**

	Direct index		Monthly chained	
	Laspeyres	**Fisher**	**Laspeyres**	**Fisher**
Fixed outlet weights				
Midweek only	102.9	101.7	110.3	101.1
Monthly unit value in outlet	102.0	101.1	110.3	100.5
Unit value over outlets				
Midweek only	102.2	101.9	105.2	101.9
Monthly unit value in outlet	101.3	101.1	104.5	101.4

Table 2d: Frozen fish. Index outcomes, December 94 - December 95. KPI_L=98.4; KPI_K=98.3

	Direct index		Monthly chained	
	Laspeyres	**Fisher**	**Laspeyres**	**Fisher**
Fixed outlet weights				
Midweek only	104.0	101.1	151.1	96.7
Monthly unit value in outlet	104.5	102.3	145.5	97.7
Unit value over outlets				
Midweek only	104.0	101.3	134.1	98.4
Monthly unit value in outlet	103.3	102.0	122.2	99.6

Some features of these results are:

1. Laspeyres' indices are markedly higher than Fisher indices as is also noted by Reinsdorf (1996). These differences may be interpreted as the substitution bias of the Laspeyres' index and throw serious doubts on the use of Laspeyres' indices at this level. Monthly chained Laspeyres' indices are a disaster since they accumulate the biases on top of each other (upward drift). It is possible to demonstrate that the chained Laspeyres' bias is approximately proportional to the number of links if the (negative) correlation between price changes and quantity changes and the variances of price and quantity change stay about the same when the time periods are shortened. This appears to be the case here when dividing the year into 12 monthly links. (Indeed, the chained biases are usually even more than 12 times the corresponding unchained ones!) The short term price bouncing phenomenon is the economic explanation.

2. Chained and direct Fisher indices are often quite different. The differences, however, are in both directions. The choice between them is an issue which has to be analysed further. In the case of seasonal goods new kinds of problems will emerge here.

3. Using the midweek vs. the whole month also makes a noticeable difference, in either direction. A problem with using the whole month is that weeks are sometimes divided between months.

4. The actual KPI outcomes, based on two variants of the ”RA” elementary aggregate formula in Dalén (1992) (approximately equal to a geometric mean), are generally in line with the Fisher indices. So there is no indication of any overestimating bias in the present KPI estimates. This also provides some evidence for the approximate unbiasedness of the simple geometric mean as an elementary aggregate formula which agrees with Reinsdorfs findings.

5. Further work

This is a first report on the use of scanner data for CPI purposes. By now, Nielsen Co has provided data for the four item groups and the same outlets for the whole period October 94 to December 96. It is straightforward to replicate the above results for another year (Dec 95-Dec 96), although time has not yet allowed this.

It is of some interest to speculate on future perspectives for the use of scanner data. There are great potentials for improvement of CPIs:

Higher quality. Scanner data can improve CPIs in three ways. Firstly, by allowing the use of superlative and otherwise improved index concepts and formulae like those discussed above. Secondly, by providing data on characteristics and new goods which will make hedonic regression and other more sophisticated methods for adjusting for quality change more tractable alternatives. Thirdly, by greatly increasing sample sizes, especially in the item dimension which, according to European experiences, contributes most to sampling error.

Lower cost. Sharing a common database between outlets, market research firms and statistical agencies is of course highly cost effective. Future negotiations will determine the distribution of these savings between these three parties.

There are also areas which need more research:

The index concept. It is fairly obvious that the best index concept at the lowest aggregation level is some combination of Fisher indices with unit values, see Diewert (1995) for a motivation. Some issues remain, however. i) What is this lowest level? From a cost-of-living index point of view, the consumers’ substitution options should be great within this level but small above. For most item groups this points at a fairly small geographic area being the optimal choice. The propensity of consumers to ”shop around” is greater for durables than for food, however, so this area should not be the same for all item groups. The U.S Bureau of Labor Statistics has declared this to be an important research area and we are eagerly waiting for illuminating results from there. ii) What, exactly, is the best combination of the Fisher and unit values. In particular, what is the best approximation; to consider all shops in a market area to be of the same quality and therefore compute unit values over them or to consider all price differences between them to be exact expressions of the consumers’ perception of quality differentials? In the item dimension, the same issue arises when deciding what items (ids) to aggregate into groups for direct price comparison based on unit values as opposed to keep separate for calculating weighted averages of price change. iii) How should we handle new goods (non-matching items) in a scanner data base? Is the Hicksian shadow price theory really applicable to some of the almost trivial changes in product design that we experience in the market? iv) Should we chain over weeks/months/ neither?

Sampling. Optimal sampling strategies need to be formulated in the new situation. At a first look it seems as if we could use the whole universe of prices without any sampling at all! This is probably not practical, though, given i) the less than 100% market coverage of scanner data, ii) the need to combine items into approximately homogeneous composite items which will require some editing of the data, and iii) the price for and quality of scanner data from the statistical agency's point of view. In the first phase, a reasonable strategy will probably be to seek some combination of scanner data and traditional data for CPIs, even for item groups where scanner data are fairly widespread.

Quality change. Combining not only price and quantity but also quality (characteristics) data, especially for durables, will provide even more challenges. Intensive research and interesting results is apparently carried on in many places but I will abstain from attempts to describe this work in detail.

All in all, analyses of scanner data are bound to improve both the theory and practice of price indexes and economic statistics. Research in this area must be a great priority in the next few years among economic statisticians.

References

Dalén, J. (1992). Computing Elementary Aggregates in the Swedish Consumer Price Index. Journal of Official Statistics, Vol 8, No. 2.

Diewert, W.E. (1976). Exact and Superlative Index Numbers. Journal of Econometrics, 4, 115-145.

Diewert, W.E. (1995). Axiomatic and Economic Approaches to Elementary Price Indexes. Discussion Paper No. 95-01, Department of Economics, University of British Columbia.

Reinsdorf, M. (1996). Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee. Memo, January 11 1996, U.S. Bureau of Labor Statistics.

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, 427-438.