An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI

Ralph Bradley, Bill Cook, Sylvia G. Leaver and Brent R. Moulton[1]

Abstract: The staff of the U.S. Bureau of Labor Statistics has conducted research on a variety of potential uses of electronic point-of-sale data in production of the Consumer Price Index (CPI). This paper summarizes the results of the BLS research. The BLS research has focused on the potential for scanner data to account for product and outlet substitutions in some categories of consumer spending, the ability to produce indexes with greatly reduced sampling error for some item strata, and the ability to ensure a more accurate probability sampling of items with fewer manual steps, a sampling procedure known as "disaggregation."

1. Introduction

The staff of the U.S. Bureau of Labor Statistics (BLS) has been in contact with several private vendors of scanner data since 1993, conducting research with the use of scanner data purchased from these vendors on a variety of potential uses of such electronic point-of-sale data in production of the Consumer Price Index (CPI). This paper will summarize the results of the BLS research.. This research is continuing, and the BLS is currently in the process of purchasing additional test datasets. Although this paper will focus solely on potential uses of scanner data for items sold at supermarkets, BLS staff are also starting to study potential uses of electronic point-of-sale data from other vendors and non supermarket retailers covering consumer electronics, pharmaceuticals, etc. The two major vendors of supermarket scanner data in the U.S., A.C. Nielsen and Information Resources, Inc. (IRI), have both cooperated with BLS staff in these studies. Currently, their clients include mostly the manufacturers of non durable consumer items, and the data is used for their marketing programs.

The BLS research has focused on four of the potential uses of scanner data that would enable the CPI to more accurately measure the cost-of-living (COL) index. These potential improvements are

a. the ability to account for product and outlet substitutions in some categories of consumer spending, the ability to produce indexes with greatly reduced sampling error for some item strata,
b. the ability to identify new products and quality enhancements,
c. and the ability to ensure a more accurate probability sampling of items with fewer manual steps, a sampling procedure known as "disaggregation."

Section 2 of the paper discusses (a) and (b), the analysis of substitution through use of superlative measures at the elementary level, and the reduction of sampling error. Section 3 discusses (d), potential uses of scanner data in item selection or disaggregation. This paper will not cover research on topic (c). Section 4 concludes with a discussion of practical issues that need to be addressed in evaluating the potential use of scanner data.

---

[1] The opinions expressed in this paper are those of the authors and do not represent an official policy of the Bureau of Labor Statistics or the views of other BLS staff. The authors would like to thank Thomas Benson and Tameka Coley for their help in data processing for the production disaggregation study.

2. BLS findings on the potential benefits and disadvantages of using scanner data

There are several potential ways to improve the CPI with the use of scanner data. First, the availability of current expenditure data enables the calculation of superlative indexes whose asymptotic values are closer to a COL index. Second, the outlet and item samples are very large, and this not only reduces sampling error and nonlinear bias, but better allows BLS to allocate a fixed sample across item characteristics and outlets so that the index estimate has minimum variance. Third, scanner data can help identify and incorporate into the CPI new products whose expenditure share of US households is increasing rapidly. Likewise, if a product disappears, scanner data may enhance BLS's ability to find a substitute.

Reinsdorf (1996) used A.C. Nielsen scanner coffee data from the Chicago and Washington areas between December 1992 and December 1994 to compute alternative coffee indexes. During 1994, Brazilian frosts induced a rapid rise in coffee prices. A summary of his results is listed in Table 1 below. For the two year period, Reinsdorf found that for ground coffee, a direct Laspeyres index (with weights revised each month) was 212.3 while the direct Fisher index that allows for substitution was 190.6. If the indexes were chained monthly, the Laspeyres index exhibited enormous upward drift, increasing to 574.7, whereas the chained Fisher index was 183.0. When using prices from the 3rd week, the magnitude of the difference was smaller, the direct Laspeyres was 209.7 compared to 191.2 for the direct Fisher index.

Reinsdorf had additional findings. Ten percent of the universal product codes (UPCs) for roasted coffee accounted for 75 to 80 percent of expenditures. The rest of the UPCs had small expenditure shares, but their weekly unit sales volume was highly volatile. For these UPCs there would be weeks when many outlets did not record any sales in some weeks. He found that the Fisher Index computed with unit values averaged across the weeks within a month differed from "current market basket" indexes where the price quote came from the 3rd week of the month.

In a separate study, Bradley (1995) compares various types of price indexes using the A.C. Nielsen Academic database. This data source reports weekly unit and dollar sales for four products (ketchup, toilet tissue, milk, and tuna) for ten regions for two years. Unlike the coffee data that is the basis for Reinsdorf's study, there was little price growth in these items over this two year period. Yet, there were still differences between the modified Laspeyres index that is currently computed by BLS and superlative indexes that use current expenditure weights rather than lagged expenditure weights. Table 2 summarizes the results in Bradley (1995).

It seems apparent that the modified Laspeyres results in substitution bias even with items that represent a small fraction of household expenditures, and that have a slow price growth. The average difference between the modified Laspeyres and the Törnqvist across seven one-year periods is .8%.

The two studies mentioned above used the entire scanner dataset, which represents a sample of outlets and essentially the entire population of items (UPCs). The goal was to determine the magnitude of difference between the population values of two different index concepts, a "fixed market basket" index, the modified Laspeyres, and an index that allowed for product substitution, the Fisher. Therefore, the results are of an asymptotic nature, and both studies have results where allowing for item substitution does indeed lower the asymptotic value of the index. However, the databases generated by the vendors are quite large, and consequently to produce an index within required time and cost constraints it might be necessary either to sample items or to perform aggregations. Either route has its advantages and disadvantages.

The major disadvantages with the use of sampling are sampling error, finite sample bias, and item attrition. The first two problems are easily addressed with the use of scanner since they decrease with increases in sample size and it seems possible to choose manageable sizes that will make these two problems immaterial. Even if the target index remains the modified Laspeyres, there will be benefits from the ability to choose larger samples.

However, item attrition is an unresolved issue even when the target index is the modified Laspeyres. This issue is studied in detail in Bradley (1996). Under current BLS methodology, outlets are selected first using a probability proportional to expenditures selection (PPS), and once outlets are selected, an item within the selected outlet is sampled using the same PPS procedure. The target index for an area-stratum with N distinct commodities is a modified Laspeyres Index of the form:

(1) $$I_{T-1,T} = \frac{\sum_{I=1}^{N} \frac{P_{I,T}}{P_{I,0}} E_{I,0}}{\sum_{I=1}^{N} \frac{P_{I,T-1}}{P_{I,0}} E_{I,0}},$$

where $P_{I,S}$ is the price of item I in time S (S=0,T–1,T), and $E_{I,0}$ is base-period expenditure. The base period is denoted as 0. The sample index for the same area-stratum is:

(2) $$\hat{I}_{T-1,T} = \frac{\sum_{I=1}^{N} \frac{P_{I,T}}{P_{I,0}} \chi_{I,0}}{\sum_{I=1}^{N} \frac{P_{I,T-1}}{P_{I,0}} \chi_{I,0}}$$

where $\chi_{I,0}$ is 1 if the item is in the sample and 0 if not. The BLS sampling mechanism is designed so that the probability of selection , i.e. Pr($\chi_{I,0}$ =1), is proportional to $E_{I,0}$. There are times when an item that has been selected for a particular outlet is no longer sold. Under current BLS procedures, in the first month that an item disappears, (2) is computed by leaving the missing item out. In a subsequent collection period, a replacement item in the same outlet is identified and used in the index. (This is referred to as a "substitution.") Although the item might be missing in a particular outlet, it is most often still available in other outlets, and it still might be available to the "representative individual" in a particular area. The current BLS methodology estimates the availability of an item for "the representative individual" by its availability within the outlet that was selected in the sample. When an item disappears, an unresolved question is whether to use the current BLS methodology or to search for the item in another outlet.

When item attrition occurs for good, I, the expected value of $\chi_{I,0}$ is no longer proportional to $E_{I,0}$ if substitution is done within the outlet. However, others argue that it is appropriate to substitute within the same outlet because "the representative individual" will not usually go to another outlet to purchase the missing item. Rather, he would select another product in the same outlet whose characteristics most closely resembled the characteristics of the missing item.

Using the A.C. Nielsen data base, Bradley (1996) found that the current BLS methodology of searching for a replacement within the same store after the first month of disappearance and the alternative approach of immediately searching in other outlets produce different estimates. Bradley runs simulations on a sample size of 12 items. He computes the population target index (1) and then the sample estimate in (2) based on the two methodologies. His results are listed on Table 3 below. The first column identifies the product and market area, and the index under the store restricted substitution method and the unrestricted index. The index labeled "Target" is from (2). Not only are the means different but the mean squared errors are different. The omission of the missing item in the first month's price relative, and the use of a substitute within the same store adds more variance to the current BLS estimator.

Another approach in making the large scanner data base more manageable is to perform aggregations across outlets. Reinsdorf (1996) did this and derived different index values than he did when he only took the third week of every month. (See Table 1.) Aggregation introduces new sources of bias. If the target index is a modified Laspeyres, aggregation will bias the expenditure weights because the price averages that are substituted into the Laspeyres have been weighted by current expenditures rather than base period expenditures. If one uses unit values in a Törnqvist, bias can result from the condition that the mean of a function that is nonlinear in prices will not be equal to the function of the mean.

Figures 1 and 2 demonstrate the difference between a non aggregated Törnqvist index and an aggregated Törnqvist Index and. We used tuna sales in 1992 from the A.C. Nielsen Academic Database. First, we computed a week to week Törnqvist and then a month to month Törnqvist using unit average prices over a month. Figure 1 compares the Törnqvist index computed week to week and then month to month. Figure 2 compares the chained indexes. It is clear that the two indexes are not equal and even if the entire data set is used the aggregated index does not converge to the true index.

The bias caused by aggregation is the result of the variation within the aggregated group. In the above example, the variance of an item within each month created a bias when prices were aggregated by month. However, when weekly prices are aggregated by chain the bias is less because the price variance within many chains is equal to zero. For each week, we aggregated prices by chain and computed both a weekly and a chained weekly index. (Please note that we define "chain" as a cluster of outlets that belong to the same firm, and we define a "chained index" as the product of an index over time.) The results are depicted in Figures 3 and 4. Figure 3 compares the weekly index and the weekly index for prices aggregated by chain. We can see that the bias is far less than the index that used prices that are aggregated over months. Figure 4 makes the same comparisons for the chained index.

It seems apparent that aggregation can effectively reduce data set sizes, and results in little bias if the variances within the aggregating groups are small. We cannot conclude that aggregating by chain will always produce immaterial bias, because we cannot be assured that all price variances within chains are close to zero.

An additional benefit from scanner data is the ability to compute accurate variances by product characteristics so that the sample can be allocated within the area-stratum as efficiently as possible. Currently, BLS uses sample replicates to compute variances on an area-stratum level so that limited samples can be efficiently allocated across area strata. Scanner data may allow for more efficient allocation across characteristics within an area-stratum. A fixed sample size is allocated across characteristics according to the characteristic's expenditure weighted standard deviation. If one characteristic's expenditure weighted standard deviation is twice another's then it has twice the sample size. A potential drawback to this allocation technique is that it assumes a constant variance over time and this might not hold.

In Bradley (1996b), a second set of simulations were run where the index was computed using a sample allocation scheme where a fixed sample was allocated across characteristics proportional to its expenditure weighted standard deviation and then the index was computed without such a scheme. The results are listed in Table 4. To keep the experiment simple, the optimal sample allocation scheme used only two clusters. Cluster 1 contains items whose size was below the median size and Cluster 2 contains all other items. The table lists the variance from the index that was simulated from a pure PPS method, and another index that was simulated by allocating the sample by the optimal share using the weighted standard deviation. Additionally the table lists both the expenditure share and the optimal allocation.

In all but one item-area, the optimal allocation scheme outperformed the PPS scheme. The one exception might have occurred because the assumption of a constant variance over time was violated.
Another issue for scanner data is that often an item is on a store shelf, but it has no recorded purchases. Currently, there is no method to determine if an item is unrecorded because the retailer decided to discontinue selling the product or if consumers decided not to buy it. These two different scenarios have different impacts on the COL. When we did the comparison of an aggregated monthly Törnqvist and the weekly Törnqvist we needed observed prices for each outlet-UPC combination for each week. However, as already mentioned for many outlet-UPC combinations there was not a complete set of observable prices. The true Törnqvist index needs observed prices for items that did not sell within a particular month, and since we cannot access this, we cannot get a precise Törnqvist Index.

Scanner data might assist BLS in testing for the presence of time substitution. For non-perishable items, households may store inventories as a means to lower its COL. From inspection of tuna sales in the A.C. Nielsen data base, it seems that when a brand of tuna goes on sale, purchases can sometime increase over one hundred fold. The theory of the COL index that underlies the superlative index is based on an assumption that the purchase and the consumption occurred in the same week. However, this might be too restrictive an assumption. Households may store some of these purchases for later consumption. Current COL theory has not provided a methodology to detect inventory use and then to incorporate this use into the COL. Bradley and Verdon (1997) have shown for a two period model how inventories can affect the COL. Work is currently focusing on the ability to estimate the latent inventory accumulation variables.

3. Disaggregation using electronically scanned data

Disaggregation is the process used by CPI data collectors to select store-specific items that will be used to calculate measures of price change for the CPI. Prior to a data collector's store visit, broad categories of goods and services called Entry Level Items (ELIs) are assigned to outlets for initiation. Cereal, Cola Drinks, and Frozen Prepared Meals are examples of Food At Home ELIs. During the store visit, the data collector attempts to enlist the assistance of establishment personnel who can provide percent of sales data broken out by item characteristics that are listed on an ELI Checklist. Packaging and brand are examples of the item characteristics. By summing the percentages for individual item characteristics and applying a random number, the data collector selects unique items with probability proportionate to the sales (PPS) experience of the outlet. When percent of sales data cannot be obtained, the data collector asks the store respondent to rank sales for the item characteristics. Failing that, the field collector estimates sales percentages by shelf space methods or by equal probability.

Because food prices are volatile, and, because once having traveled to the establishment they are relatively inexpensive to collect, the number of food price quotations in grocery stores tends to be large. Similarly, the burden imposed on store respondents by asking them to supply sales data is large. One might suspect that grocery stores may not record sales data according to the item characteristics denoted on the ELI Checklist or that respondents do not have time to provide the percent of sales data that are needed.

Currently available data bases containing electronically scanned product information may be important alternative sources of the sales data that are needed to select items for many food ELIs. Because these data bases contain expenditures accumulated by Uniform Product Code (UPC), it may be possible to eliminate all or part of the costly, multi-stage item characteristic approach used now. In addition, it may be possible to obtain more accurate percent of sales data and reduce respondent burden. These considerations are motivating factors in a disaggregation study now in progress. Since disaggregation is conceptually simple, the potential payoff for the disaggregation research may occur sooner than for the index calculation research.

The CPI disaggregation test is confined to items sold in supermarkets for several reasons. Standardized UPCs are used extensively for many categories of supermarket items.[2] In addition, A.C. Nielsen, and, Information Resources, Inc. (IRI), the two major vendors of data used by the supermarket industry, are providing support for the study. However, problems in fitting vendor data into the CPI will make it necessary to approach disaggregation somewhat differently than is presently done in the CPI. Store addresses, for example, that have been selected for the CPI outlet sample may not be members of the scanner data bases. Also, scanned data are collected by the vendors in supermarkets and large grocery stores, but do not exist or are not collected in some other types of businesses in the CPI outlet sample.

One component of the disaggregation feasibility study will be an assessment of current disaggregation procedures and their effectiveness. Tabulations have been constructed from disaggregation records from the commodities and services sample rotation performed in 1995. The method of item selection for each step of disaggregation has been tabulated for all quotes initiated in outlets with more than 20 quotes in that sample rotation, as well as for all quotes in 6 selected sample cities. Examination of these tabulations indicates high respondent participation in the initial 3 stages of disaggregation in the large sample volume outlets; the use of PPS or ranking sampling in these cases for food-at-home quotes exceeded 98%, with PPS alone exceeding 58%. This is remarkably different for food-at-home in the sample of 6 cities, where equal probability sampling was used in the first stage of sampling 4.8% of the time and had grown to 13% by the third stage of sampling. These results are shown in Figures 5a-5h and 6a-6h.

A similar distinction between large volume outlets and all outlets in the distribution of disaggregation methods held for the entire set of tabulated quotes initiated in that rotation; here the percentage of instances in which equal probability sampling was used in the first three steps was greater in non-food items. In the initial 3 stages of disaggregation in the large sample volume outlets, the use of PPS or ranking sampling in these cases for all quotes exceeded 89%, with PPS alone exceeding 53%. This contrasts with the sample of all quotes for 6 cities, where equal probability sampling was used in the first stage of sampling 10.9% of the time and had grown to 19.5% by the third stage of sampling. Further examination of quote-level disaggregation records has confirmed that the majority of quotes disaggregated using equal probability sampling at the first stage were in other than food at home expenditure classes, with the largest number in any one class being in food-away-from-home item strata.

The second component of the study will be a simulation of disaggregation using scanned data from the A.C. Nielsen and IRI data bases. To circumvent the store address problem, we anticipate using aggregate sales information for chains or selling markets to estimate item expenditure shares for individual stores in specific markets. Programming that will support a comprehensive test of electronically scanned data is being developed presently. These tests and the associated analysis are expected to take one to two years.

[2] Important exceptions are “random” (variable) weight items like meats, fresh produce, and some bakery and dairy products.

4. Conclusions

With respect to the use of scanner data in monthly index calculations, a number of unresolved issues remain. It seems apparent that the use of scanner data has the potential to provide BLS with an opportunity to derive an index that better estimates the true COL. However, before scanner data might be incorporated into the CPI many decisions must be made.

As of December 1996, supermarket items (food, housekeeping supplies, toiletries, OTC health products, tobacco) only account for 13.0% of the total weight in the CPI-U and these items that are covered by scanner data represent a smaller percentage since some outlets do not have scanners. Items that are weighed at the check-out (e.g., fresh meat, produce, bulk items, deli and bakery items) may not be included fully in vendor provided scanner datasets. Although scanner and other point-of-sale data are available for items other than food at home, coverage often tends to be incomplete.

Therefore, traditional sampling methods will probably need to continue for most strata even if scanner data can be used for some strata. Since traditional methods do not allow for the collection of current expenditure information much of the index will still rely on base-period expenditure weights. If superlative indexes are computed with scanner data, decisions must be made on the integration of these superlative indexes with indexes that use base-period weights.

Geographical coverage is also incomplete. For example, data from Alaska or Hawaii may be unavailable from one or another vendor, so that the current collection methods might need to be continued in order to produce the Honolulu and Anchorage indexes.

Currently, the vendors do not use PPS in selecting the outlets in their sample. BLS relies on PPS to select outlets. Therefore, a weighting mechanism will need to be devised that will allow for the construction of an index that is consistent with BLS's PPS procedures.

The vendors that provide these data purchase it from retail outlets, and must work with a variety of retail systems. Consequently, purchasing data for use in the monthly CPI is likely to be fairly expensive. Although BLS would potentially experience cost savings associated with reduced data collection in supermarkets, presently the supermarket data are among the least costly of the CPI price data collected by BLS. It is not clear at this point whether the use of scanner data represents any net cost savings to the agency.

Issues of timeliness still need to be resolved. The CPI is produced very quickly, and it is not yet known whether scanner data could be delivered, verified, and processed, within the deadlines required for monthly production and publication of the CPI.

Finally, any reliance on outside vendors creates some concerns about reliability. In a worst case, if the vendors could not deliver the data within the schedules required by the BLS, it potentially could lead to a failure to produce an important monthly index. Additional scenarios might involve the attrition of a chain from the vendor's sample. Thus, BLS would need to implement a system of reliable backups if these events occurred.

Despite these serious issues and concerns, the staff of the BLS is continuing to evaluate the potential. The potential benefits are sufficiently great, that development of prototype scanner indexes and continued research on disaggregation are important priorities of the agency.

References

Bradley R, “The Use of Scanner Data as a Means to Reduce the Mean Squared Error in the CPI,” mimeo, U.S. Bureau of Labor Statistics, December 1995.

Bradley R. “Finite Sample Issues in the CPI,” mimeo, U.S. Bureau of Labor Statistics, July 1996.

Bradley, R. “Potential Benefits from the Use of Scanner Data in the Construction of the CPI” *1996 ASA Proceedings* , 1997.

Bradley, R., and Verdon S. “Time Substitution Bias in the CPI,” mimeo, U.S. Bureau of Labor Statistics, February 1997.

Reinsdorf M. “Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee,” BLS Working Paper 277, April 1996.

U.S. Department of Labor, Bureau of Labor Statistics, *BLS Handbook of Methods*, Washington, DC: U.S. Government Printing Office, Chapter 19.

Figure 1
Comparison of Monthly Aggegated and Weekly Indexes

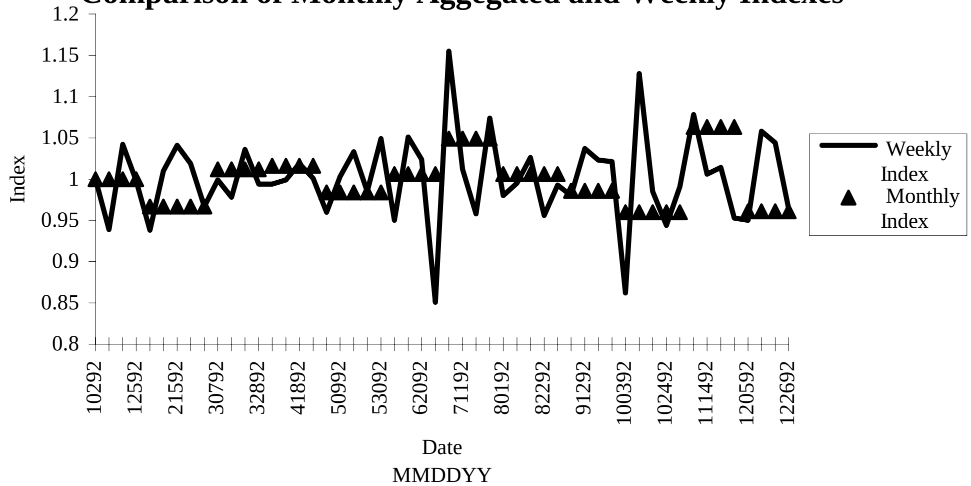

Figure 2
Comparisons of Chained Indexes
for Weekly Index and Monthly Aggregated Index

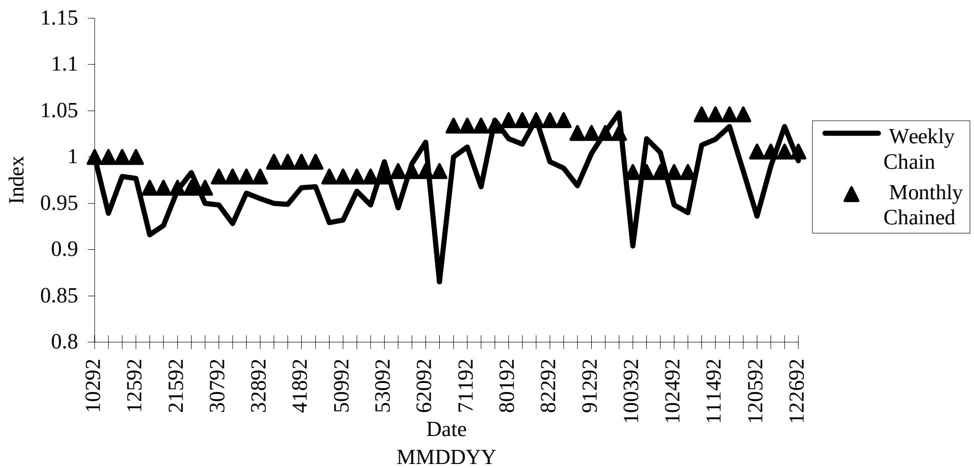

Figure 3
Weekly Index for Individual Outlets and Aggregated by Chain

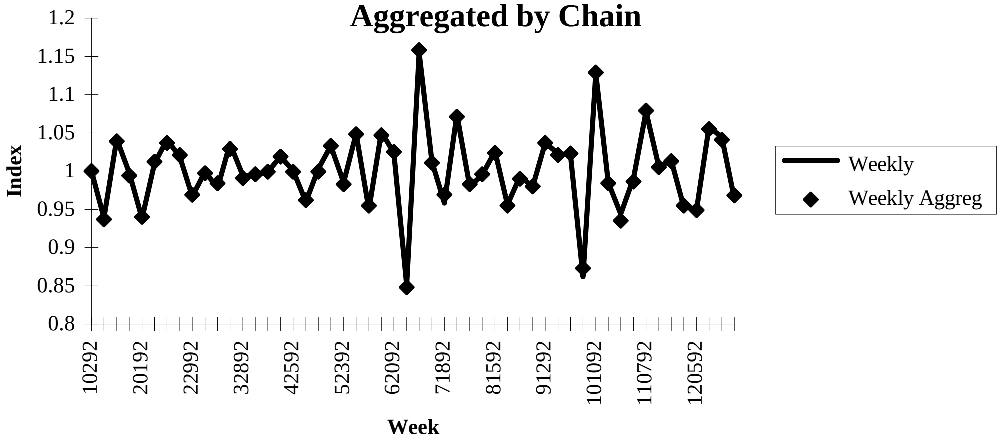

Figure 4
Chained Indexes for Weekly Index by Outlet and Aggregated by Chain

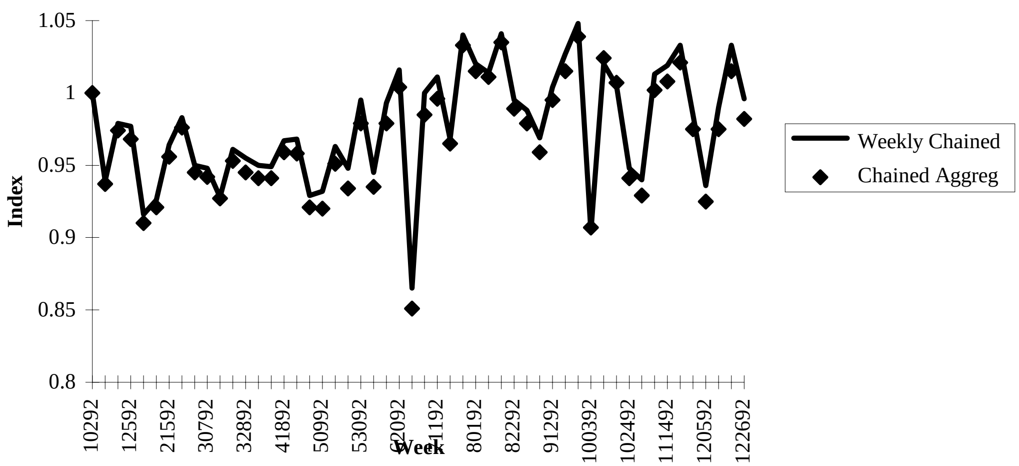

Table 1. Alternative indexes from coffee data

Periods Compared	Price Concept	Index Description	All Types of Coffee	Ground Coffee
12/92 - 12/94	Monthly	All months in market basket	176.7	187.4
"	Unit	Direct Laspeyres, market basket from 12/92	195.9	212.3
"	Value	Chained Laspeyres, market baskets from 12/92 to 11/94	471.1	574.7
"		Direct Fisher, market baskets from 12/92 and 12/94	179.3	190.6
"		Chained Fisher, market baskets from 12/92 to 12/94	171.5	183.0
"	Price for	All months in market basket	176.0	187.1
"	3rd Week	Direct Laspeyres, market basket from 12/92	193.6	209.7
"	of Month	Direct Fisher, market baskets from 12/92 and 12/94	179.4	191.2

Table 2. Results from the A.C. Nielsen Academic database for 4 products and ten regions

Periods Compared	Modified Laspeyres	Modified Geomeans	Direct Laspeyres	Direct Geomeans	Fishers	Törnqvist
6/92-6/93	1.031	1.025	1.048	1.038	1.028	1.028
7/92-7/93	1.037	1.034	1.060	1.048	1.033	1.032
8/92-8/93	1.035	1.029	1.055	1.043	1.025	1.025
9/92-9/93	1.026	1.019	1.035	1.027	1.013	1.014
10/92-10/93	1.033	1.030	1.053	1.039	1.028	1.025
11/92-11/93	1.034	1.029	1.036	1.028	1.020	1.021
12/92-12/93	1.027	1.024	1.036	1.028	1.021	1.021

Table 3. Results of simulation comparing a store restricted index to an unrestricted index

Market Area-Product	Index 6/92-12/93	Simulation Result-Population Target	Simulation MSE
Market 01- Ketchup			
Store Restricted	1.0475	-0.0357	0.0025
Unrestricted	1.0847	0.0015	0.0019
Target	1.0832		
Market 02- Ketchup			
Store Restricted	1.048	0.0274	0.0032
Unrestricted	1.0212	0.0006	0.0012
Target	1.0206		
Market 03- Ketchup			
Store Restricted	0.9829	0.0052	0.0054
Unrestricted	0.9797	0.002	0.0032
Target	0.9777		
Market 01- Milk			
Store Restricted	1.0486	-0.0018	0.0012
Unrestricted	1.0494	-0.001	0.0001
Target	1.0504		
Market 02- Milk			
Store Restricted	0.9929	0.0118	0.0008
Unrestricted	0.9806	-0.0005	0.0001
Target	0.9811		
Market 03- Milk			
Store Restricted	1.1376	0.0037	0.0006
Unrestricted	1.1345	0.0006	0.0002
Target	1.1339		
Market 01- Toilet Tiss.			
Store Restricted	1.0317	-0.0148	0.0031
Unrestricted	1.0495	0.003	0.0018
Target	1.0465		
Market 02- Toilet Tiss.			
Store Restricted	1.0362	-0.0057	0.0047
Unrestricted	1.0427	0.0008	0.001
Target	1.0419		
Market 03- Toilet Tiss.			
Store Restricted	0.9571	-0.0218	0.0025
Unrestricted	0.9813	0.0024	0.0006
Target	0.9789		
Market 01- Tuna			
Store Restricted	1.0339	-0.0394	0.0043
Unrestricted	1.0728	-0.0005	0.0015
Target	1.0733		
Market 02- Tuna			
Store Restricted	1.0361	0.0059	0.0049
Unrestricted	1.031	0.0008	0.0011
Target	1.0302		
Market 03- Tuna			
Store Restricted	0.9183	-0.0026	0.0007
Unrestricted	0.9166	-0.00009	0.0017
Target	0.9192		

Table 4. A Comparison of variances between a pure pps and a targeted selection

Product-Market	PPS Variance	Optimal Variance	Cluster 1 Exp. Share	Cluster 1 Opt. Share	Cluster 2 Exp. Share	Cluster 2 Opt. Share
Ketchup-01	.00289	.00235	.406	.333	.594	.667
Ketchup-02	.00222	.00205	.600	.911	.400	.089
Ketchup-03	.00412	.00396	.524	.583	.476	.417
Tuna-01	.00145	.00135	.075	.089	.925	.911
Tuna-02	.00106	.00098	.110	.089	.911	.890
Tuna-03	.00070	.00036	.183	.166	.817	.834
Toi. Tiss.-01	.00409	.00297	.360	.089	.640	.991
Toi.-Tiss-02	.00706	.00587	.677	.178	.333	.822
Toi.-Tiss-03	.00456	.00254	.306	.089	.694	.911
Milk-01	.00011	.00010	.602	.416	.318	.584
Milk-02	.00014	.00013	.760	.416	.240	.584
Milk-03	.00023	.00034	.791	.667	.209	.333