WITHIN OUTLET ITEM SELECTION TECHNIQUES FOR THE CONSUMER PRICE INDEX REVISION

William L. Weber and Frank R. Lambrecht, Bureau of Labor Statistics

INTRODUCTION

The publication of the revised Consumer Price Index (CPI) by the Bureau of Labor Statistics (BLS) in February, 1978 marked the culmination of the most complex and extensive revision of the CPI since its origination during World War I. Virtually every aspect of the CPI methodology was reviewed, analyzed, and improved as part of the revision effort. Two new indexes were actually produced. One for the urban wage and clerical worker population - corresponding to the unrevised CPI, and one for the all urban population. Ideally each CPI would have as inputs the total set of prices, placed in one-to-one correspondence to the total set of purchases of all members of the index family population. However, the magnitude of these variables and the constraints of time and cost will not permit the achievement of this ideal. Thus, a complex and multi-faceted sampling scheme was designed and employed to produce the revised CPI.

The subject of this paper is the within outlet sampling techniques which were employed to determine the specific items to be priced in each sample outlet. This is the final stage of the CPI sampling scheme. The new techniques employed for within outlet item sample selection constitute one of the areas of major improvement in the CPI methodology. For the first time, multi-stage probability sampling techniques are employed to select the items for pricing within the sample outlets. With this new methodology, each item available in a given sample outlet and included in the general category of items assigned for pricing in that outlet is given a probability of selection proportional to its sales in the outlet. Each brand, variety, size, style, model, etc., of the items available for pricing is given a chance of selection proportional to its importance to the total sales in the outlet. Thus, the selection of the specific items priced for the CPI is keyed to the sales experience of each sample outlet. Since the within outlet sampling methodology does not differ by population, the population distinction will be omitted from the discussion of this portion of the CPI sampling scheme.

Before presenting the within outlet sampling methodology, we will briefly discuss the selection of the general categories of items comprising the CPI market baskets and the selection of the outlet samples of the CPI. This discussion will provide the reader with an understanding of how the within outlet sampling methodology fits into the overall CPI sampling scheme. In addition, a brief discription of the within outlet sampling procedures for the unrevised CPI is provided for comparison with the new methodology.

CONSUMER EXPENDITURES SURVEY AND ENTRY LEVEL ITEM SAMPLING

As in previous revisions of the CPI, the first task was to design and implement a survey of consumer expenditures, income, assets and liabilities This survey, called the Consumer Expenditure Survey (CES) covered the years 1972 and 1973. It was executed for the BLS by the Bureau of the Census. This survey provided the BLS with data to construct the sampling frames from which the sample market baskets of items were selected. The item weights used in the calculation of the indexes were also derived from this data. Eight item sampling frames were constructed, one for each of the two index families in each of four geographic regions.

Each of the eight item sampling frames consisted of approximately 267 item strata. Each item stratum consisted of one or more similar but still broadly defined items. For example, 20051 - Beer and ale away from home, 20052 - Wine away from home and 20053 - Other alcoholic beverages away from home comprised an item stratum. The items within each stratum were called Entry Level Items (ELI's). In the above example, each of the three enumerated items comprise an ELI. Expenditures from the CES were aggregated to the ELI level. Within a stratum each ELI was assigned a probability of selection proportional to the amount of consumer expenditures for the ELI. A single independent PPS sample of ELI's was selected from each item stratum. Eight independent samples of ELI's were selected in each region for each population. These ELI samples comprise the general categories of items assigned for pricing in the outlet samples. Each sample of ELI's was initiated for pricing in a sample of three to four PSU's.

THE OUTLET SAMPLES

The outlet sampling frames from which the sample of outlets were selected for pricing the sample ELI's were constructed from several sources. The outlet sampling frames for all of the food ELI's and most of the commodities and services ELI's were constructed from data obtained from the Point of-Purchase Survey (POPS). The POPS was a household survey conducted by the Bureau of the Census in 1974 for the BLS in the 85 PSU's designated as CPI pricing areas. Respondents were asked to iden tify the amount of their expenditures and the place of their purchase for broad categories of items. A concordance between the ELI's and the POPS categories was defined. Thus, when a particular ELI was selected, a corresponding POPS category was uniquely identified as the outlet sampling frame for that ELI. The outlet sample for each ELI was selected with each outlet's probability of selection proportional to the amount of the expenditures reported for it for the corresponding category on the POPS. The POPS provided outlet sampling frames for about 60 percent of the consumer expenditures included in the CPI. The housing component accounts for approximately 20 percent of consumer expenditures. The remaining 20 percent of consumption expenditures are accounted for by ELI's which we have grouped under the label "Non-POPS."

The outlet sampling frames for the Non-POPS ELI's were constructed from a variety of sources. Outlets for ELI's such as telephone expenses, public utilities, and public transportation are members of publicly regulated industries. Outlet sampling frames for these items were constructed by consulting the regulatory agencies and industry sources. The Unemployment Insurance file (ES-202) maintained - by the BLS provided the sampling

frames for some other Non-POPS ELI's. Outlet frames for the remaining Non-POPS ELI's were constructed from data provided by other government agencies (HEW, DOT, CAB) and industry sources. As with the POPS ELI's, a concordance was defined between each Non-POPS ELI and the appropriate sampling frame. When a particular Non-POPS ELI was selected for pricing, a corresponding outlet sample was identified. In this way, the sample of ELI's, both POPS and Non-POPS, was merged with the corresponding outlet sample for each ELI pricing area. A computer generated schedule indicating the ELI's to be priced was produced for each CPI outlet. The BLS field representatives entered each sample outlet with this list of the general categories of items to be priced and conducted the final stages of the sampling process.

WITHIN OUTLET ITEM SELECTION FOR THE UNREVISED CPI

For the unrevised CPI, the BLS field representatives enter the sample outlets with a very detailed description or tight specification of each item to be priced. These tight specifications are prepared by the Washington Office staff. They are basically the same for every store in the outlet sample in every CPI pricing area for the unrevised CPI. Within each sample outlet in which the item is to be priced, the BLS field representative attempts to obtain the price for the best-selling item meeting the tight specification. Thus, this procedure has been named the "volume seller" technique. Items which meet the tight specification but are not the top selling item within the outlet of those meeting the tight specification are given no chance of selection. In a few cases the tight specification, provides for variations in what is eligible for pricing, but only within a relatively narrow range. In a few other cases, alternative specifications are eligible for pricing, but only if nothing meeting the preferred tight specification can be found.

Over the years that this process has been in use, a number of shortcomings have become apparent. Since the selection of the tight specifications for pricing is independent of the individual outlets sales experience, items were sometimes selected for pricing which are not as representative of a particular outlets merchandise as is desirable. The definition of national tight specifications has also resulted in the confinement of the items priced for the CPI to a relatively narrow segment of the quality range available in the market. In addition, the out-of-scope rate for a given item was quite high with this procedure. Many sample outlets were lost simply because they did not carry the national tight specification designated for pricing. This large scale substitution of outlets resulted in, essentially, a quota sample of outlets.

With this brief history of the within outlet sampling procedures of the unrevised CPI and the short description of the sampling plan up to item selection for the revised CPI, we are now ready to discuss the within outlet sampling techniques of the revised CPI. These techniques have been named "Disaggregation" by the BLS staff.

DISAGGREGATION

The goals of the within-outlet sampling procedures adopted for the revised CPI are as follows: 1) To key the selection of the items to be priced for the CPI to the sales experience of the individual sample outlet. 2) To obtain, via probability sampling, a sample of items which better represents the distribution of items purchased by CPI families in the market place. This includes the pricing of a broader segment of the quality range of items available in the market place. 3) To increase the likelihood that a selected item can be priced in a given sample outlet, thereby lowering the non-response rate and reducing costs.

In the remainder of this paper we will describe the disaggregation techniques and present some evaluations of how well the above goals were satisfied.

As described earlier, one or more relatively broad categories of items (ELI's) is assigned for pricing in each sample outlet. The ELI's assigned for pricing in a given sample outlet are listed on BLS form 3400, the initiation facesheet. This form is computer generated and contains all the necessary identifying information for the outlet, the ELI's assigned for pricing in the outlet, and the number of quotes to be obtained for each ELI.

Within each outlet the BLS field representative executes the disaggregation procedures for each assigned ELI. The disaggregation process begins with the relatively broad category (ELI) and is narrowed through successive stages of sampling until a single unique item is identified. For example, assume that ELI 09011-Fresh Whole Milk has been assigned for pricing in a given sample outlet The field representative introduces hisself/herself to a respondent in the outlet who is knowledgeable about the outlets milk sales. This might be the manager of the dairy department. After explaining the purpose of the survey, the field representative begins disaggregation for the assigned milk quote.

The respondent may be asked to list all of the brands of fresh whole milk offered for sale by the outlet. The field representative enters this information on a BLS form 3400A, disaggregation worksheet. Three stages of disaggregation may be completed on each worksheet. Probabilities of selection, based on sales of each brand within the outlet, are then assigned to each brand listed. These values are entered on the disaggregation worksheet. Cumulative totals of the probabilities of selection are computed and entered on the worksheet for each brand. A random number is selected from the "random number" table provided on each worksheet. Twenty-five editions of the disaggregation worksheets were produced. Each edition of the worksheet contained an independently generated "random number" table. The BLS field representatives used a different edition of the worksheet in each sample outlet. The first brand for which the cumulative total of the probabilities of selection is greater than or equal to the selected random number is designated for selection Assume that we have selected Brand A.

The respondent is now asked to list all of the types and sizes of containers in which Brand A of fresh whole milk is offered for sale in the outlet This information is listed on the disaggregation worksheet. Probabilities of selection, based on the sales in the outlet of each type and size of container of Barnd A fresh whole milk, are determined and assigned to each type and size of container listed. These values are entered on the disaggregation worksheet. Cumulative totals of the probabilities of selection are computed and

entered on the worksheet for each item listed. The next available random number is then selected from the "random number" table on the worksheet. The first type and size of container for which the cumulative total of the probabilities of selection is greater than or equal to the selected random number is designated for selection. Assume that we have selected a one gallon plastic container.

The field representative determines that the brand, type and size of container uniquely identifies the fresh whole milk sold in the sample outlet. Thus, we have identified the unique store item which will be priced for the CPI. A unique store item is defined to be one for which there are no price differences between it and any other item within the ELI and there are no differences in price determinant characteristics between it and any other item in the ELI. The selected item is described on BLS form 3400B, the ELI checklist. An ELI checklist has been designed by the Washington Office staff of the Bureau for each ELI in the universe. All of the information necessary to identify and price the selected item over time is entered on the ELI checklist. This involves listing all the characteristics of the items which are price determining and/or for which a change in the characteristic may constitute a change in the quality of the item.

As illustrated by the above example, the key element of the disaggregation process is the determination of the probability of selection for each item eligible for selection for a given stage of the process. Four alternative methods have been devised for determining these probabilities. These methods were developed during a one year field test of the disaggregation process. The four methods may be used alone or in combination throughout the various stages of within outlet sampling. The four procedures are: Percent of dollar volume of sales, Ranking on the basis of dollar volume of sales, Shelf space, and equal probability.

PERCENT OF DOLLAR VOLUME OF SALES

The percent of dollar volume of sales is the preferred method for determining the disaggregation probabilities. This method relies on the respondents knowledge of the dollar volume of sales for the items listed for selection for a given stage of the sampling process. After the field representative has listed the items eligible for selection for a given stage of sampling, the respondent is asked to assign a percentage of the total dollar volume of sales for the group of items to each item listed. The respondent may refer to records or give his best estimates for these percents of sales. These percentages are posted in Column C of the disaggregation worksheet (Attachment II). A "running total" of these percentages is posted in Column D of the worksheet. The random number, determining the item selected, is posted in Column E.

RANKING

The ranking procedure is used whenever a respondent cannot or does not want to provide percentages of the dollar volume of sales for the items eligible for selection. Using the dollar volume of sales for each item as the criteria, the respondent is asked to rank the eligible items from most important to least important. Having ranked the items the respondent is asked if he/she can now assign percentages of the dollar volume of sales corresponding to the rankings. If the respondent is unable to determine percentages of sales corresponding to the rankings for all of the items, he may obtain it for some and adjust the remaining percentages accordingly. The field representatives refers to a predetermined table of percentages to obtain the appropriate percentage to assign to each item. This "ranking" table was derived by Washington Office staff by examining actual percentages provided by respondents during the period of field testing of the disaggregation procedures. This table is included for convenient reference on the disaggregation worksheet.

SHELF SPACE

When the respondent cannot provide percentages of dollar volume of sales or rankings for the items listed for selection, the shelf space method for estimating relative dollar volumes of sales may be employed. If the amount of shelf space for display of the items eligible for selection is roughly comparable to the dollar volume of sales for those items the field representative may resort to this procedure. The shelf space method takes into account both display space and unit price, whenever possible. The display space times the unit price of the item is used as the approximate proportional equivalent of dollar volume of sales. In this way, percentages obtained by the shelf space method may be used directly in the disaggregation process. If the respondent feels these percentages reflect only the rankings of the items, then ranks should be assigned based on these percentages. New percentages should be determined from the ranking table. This method was primarily used for food items.

EQUAL PROBABILITY

This procedure is used whenever the respondent cannot provide the percentages of dollar volume of sales, cannot rank the relative importance of the eligible units, and the shelf space method cannot be used. While the four methods of assigning probabilities can be used in combination throughout the disaggregation process, the equal probability method may not be used exclusively within a given outlet. With this procedure equal probabilities of selection are assigned to all items eligible for selection for the given stage of sampling. An equal probability table is provided on the disaggregation worksheet for the convenience of the field representatives.

EVALUATIONS

Having described the methodology employed for the within outlet item selection for the revised CPI, we will now present some evaluations of these procedures. These evaluations are not intended to be final or all inclusive. They, do, however, provide some insights into how well the goals of this methodology are being achieved. As indicated earlier, the first goal of the within-outlet procedures was to key the selection of the items to be priced for the CPI to the sales experience of the individual sample outlet. The first three methods of assigning probabilities of selection are directly dependent on the sales experience of the outlet. Only the equal probability method does not key the selection of the items to be priced directly to the sales experience of the sample outlet. The first three procedures were used for selection of the vast majority of items. The

equal probability procedure was used for the selection of less than five percent of the items priced for the revised CPI.

The second goal of the within outlet sampling procedures was to obtain a sample of items which better represents the distribution of items purchased by CPI families in the market place. It was hoped that the sample of items selected for pricing for a given ELI might better span the quality range of items available in the market place. The degree of attainment of this goal is somewhat harder to evaluate. An in depth analysis of the distributions of items and item characteristics selected for pricing for all of the ELI's is not currently available. However, we have examined the items selected for pricing for two ELI's: Beer and ale for use at home and Cigarettes. For each of these ELI's, we have tabulated the distribution of brands of the items priced for the CPI. For the ELI, "Beer and ale at home" eighty different brands were selected for pricing for the all urban population. Seventy-five different brands were selected for pricing for the wage earner and clerical worker populations. For the ELI, "Cigarettes" thirty-eight different brands were selected for pricing for the all urban population and thirty-seven different brands were selected for pricing for the wage and clerical population.

For each ELI, the percentage of the total CPI weight for the ELI represented by each brand was calculated for each CPI population. For "Beer and ale for use at home", the percentages of the CPI weight for each brand were aggregated by manufacturers. Standard errors for these estimates were also computed assuming simple random sampling as $\sigma = \sqrt{\frac{pq}{n}}$, where p is the proportion of the CPI weight, q = 1-p, and n is the number of quotes. Table I compares the percentages of the CPI weight for a given manufacturer with that manufacturer's percentage of total U.S. beer sales for 1975, the initiation year. This table includes the top 17 breweries in the U.S. for 1975. In order to facilitate comparisons the percentages of sales and the percentages of CPI weights have been normalized across the 17 brewers. For the all urban population, for all but one of the manufacturers, the percentages of national sales are within 3σ of the corresponding percentage of the CPI weights. For the wage earner and clerical worker population, for all of the manufacturers, the percentage of national sales are all within 3σ of the corresponding percentage of CPI weights. Although the total sales of beer manufacturers includes sales which are not part of the ELI, i.e., beer purchased for use away from home, we feel that for this ELI the above comparisons indicate that the disaggregation process satisfies our second goal. That is, the disaggregation process has yielded an item sample in which each manufacturer is represented in relatively close accordance with its proportion of purchases by CPI families.

Table II compares the percentages of the CPI weight for "cigarettes" by brand to the percentage of national sales for 1975 by brand. Standard errors were also computed assuming simple random sampling for these estimates. The top twenty brands by sales are included in this table. The percentages of sales and the percentages of CPI weights have been normalized across the twenty brands. For the all urban population, the percentage of sales for fifteen brands are within 3σ of the corresponding percentage of the CPI weight. For the wage earner and clerical worker population the percentage of sales for eighteen brands are within 3σ of the corresponding percentage of the CPI weight. We hypothesize that the more frequent occurence of discrepancies between the percentages of sales and the percentages of the CPI weights are due to marketing characteristics of this ELI. That is, a significant amount of cigarette sales are made by vending machines. The only disaggregation methods applicable to vending machines are the shelf space and equal probability methods. Thus, some discrepancies of the CPI weight might be expected for this ELI. With this constraint in mind, we would conclude that the disaggregation process has reasonably satisfied our second goal for this ELI.

The third goal of the disaggregation procedures was to increase the likelihood that a selected item can be priced in a given sample outlet thereby lowering the non-response rate. Response rates are currently available only for those ELI's whose outlet samples were selected from the frames generated from the Point-of-Purchase Survey. These ELI's account for approximately 85 percent of the CPI items for which the disaggregation procedure was employed. For these ELI's, approximately 82 percent of the outlets assigned for pricing were successfully initiated via the disaggregation procedures. This high response rate cannot be attributed to the disaggregation procedures alone. The POPS survey also contributes significantly to this success. We can, however, conclude that the combination of these two innovations, disaggregation and POPS, have enabled us to achieve our goal of significantly lowering the non-response rate.

It is clear that the above evaluations are not comprehensive enough to completely judge the within outlet sampling procedures. However, they do provide some initial insights and a basis for further studies. Analysis of the distributions of items and item characteristics for all of the ELI's for which the disaggregation procedures were employed are desired. These studies would reveal both the strengths and weaknesses of the disaggregation process vis-a-vis each ELI.

TABLE I. Distribution of CPI weights by manufacturer for ELI 20011 - Beer and ale at home.

Manufacturer No.	% of National Sales 1975*	% of CPI Weight-U Pop.	Standard error U - Population	% of CPI Weight-W Pop.	Standard error W - Population
No. 1	25.1	28.0	1.8	26.2	1.8
No. 2	16.6	14.9	1.5	16.5	1.5
No. 3	11.1	9.1	1.2	9.0	1.2
No. 4	9.1	8.8	1.2	9.4	1.2
No. 5	8.5	6.7	1.0	5.5	0.9
No. 6	4.2	2.3	0.6	2.8	0.7
No. 7	4.0	3.5	0.8	2.6	0.7
No. 8	3.7	4.1	0.8	5.1	0.9
No. 9	3.5	4.3	0.8	4.7	0.9
No. 10	3.3	3.5	0.8	2.7	0.7
No. 11	3.2	5.7	1.0	5.3	0.9
No. 12	2.4	3.2	0.7	3.6	0.8
No. 13	1.6	0.7	0.3	2.3	0.6
No. 14	1.4	1.8	0.5	2.3	0.6
No. 15	1.0	1.0	0.4	1.1	0.4
No. 16	0.7	1.3	0.5	0.4	0.4
No. 17	0.6	1.1	0.4	0.5	0.3

* Source: Advertising Age, January 26, 1976

TABLE II. Distribution of CPI weights by brand for ELI 63011 - Cigarettes

Brand Number	% of National Sales 1975*	% of CPI Weight-U Pop.	Standard Error U - Population	% of CPI Weight-W Pop.	Standard Error W - Population
No. 1	16.4	17.7	1.7	16.8	1.7
No. 2	16.4	19.1	1.8	18.9	1.7
No. 3	11.1	12.3	1.5	9.1	1.3
No. 4	9.1	8.4	1.2	7.2	1.1
No. 5	9.0	9.0	1.3	10.5	1.3
No. 6	4.9	5.4	1.0	7.5	1.2
No. 7	4.8	4.6	0.9	4.0	0.9
No. 8	4.7	5.8	1.0	5.1	1.0
No. 9	3.3	4.3	0.9	5.6	1.0
No. 10	2.9	0.6	0.3	1.6	0.6
No. 11	2.5	1.1	0.5	0.5	0.3
No. 12	2.2	0.7	0.4	1.0	0.4
No. 13	2.0	2.3	0.7	3.4	0.8
No. 14	1.8	0.1	0.1	1.1	0.5
No. 15	1.7	2.9	0.6	1.8	0.6
No. 16	1.7	0.3	0.2	0.7	0.4
No. 17	1.7	2.1	0.4	2.8	0.7
No. 18	1.6	0.2	0.2	0.3	0.2
No. 19	1.2	1.4	0.5	0.7	0.4
No. 20	1.0	1.7	0.6	1.4	0.5

* Source: Advertising Age, November 22, 1976