Source Selection: Selecting and Evaluating America's Expenditures October 2010

Barry Steinberg* Brett Creech Mary Lynn Schmidt Patrick Falwell

Any opinions expressed in this paper are those of the authors, and do not constitute policy of the Bureau of Labor Statistics.

Abstract

The Consumer Expenditure Survey data are obtained from one of two distinct instruments, the Diary Survey and the Interview Survey. The Diary Survey is designed to capture ALL expenses incurred over a two-week period while the Interview Survey generally captures items that can be expected to be recalled for a period of 3-months or longer. The decision as to which survey to use for each Universal Classification Code (UCC) is crucial to the ongoing estimation of expenditures. Over time, the survey questions can change as do the way people respond so it is very important to adapt by identifying which survey provides the best data for all overlapping UCC's. This paper presents a multi-step quantitative method for comparing expenditure data between both surveys so that the best estimate of the mean annual expenditures per household can be obtained.

Key words: Source, Universal Classification Code (UCC)

I. Background

The Consumer Expenditure Survey (CE) is a nationwide household survey conducted by the U.S. Bureau of Labor Statistics (BLS) to find out how Americans spend their money. The CE consists of two components, each with its own questionnaire and sample: the Interview Survey, and the Diary survey. The data are collected for the BLS by the U.S. Census Bureau. There is a significant amount of overlap in the information collected by the two surveys. When expenditure information is collected about a particular item category in both surveys, a decision needs to be made about which source of information is the more reliable one for publication purposes. Data from only one of the surveys are used for each item. This paper describes the process used by the CE program to select the better source of data for use in its official published expenditure estimates.

In the Interview Survey, consumer units[1] (CUs) are visited once every three months over a period of 13 months, the survey collects expenditures on items that respondents can reasonably recall for a period of 3 months or longer, such as furniture or vehicle purchases, and expenses that occur on a regular basis, such as rent, utility payments, and insurance premiums. In the Diary Survey, CUs report expenditures in two consecutive 1-week diaries. The diaries specialize in capturing expenditures on small

[1] [1] A consumer unit is defined as members of a household related by blood, marriage, adoption, or other legal arrangement; a single person living alone or sharing a household with others but who is financially independent; or two or more persons living together who share responsibility for at least 2 out of 3 major types of expenses – food, housing, and other expenses.

***Steinberg.Barry@BLS.gov - U.S. Bureau of Labor Statistics 2 Mass. Ave., NE, Washington, DC 20212**

frequently purchased items that may be difficult for respondents to recall, such as detailed food items and beverages.

Prior to 1980, the CE was conducted at about 10-year intervals. However, since 1980 it has been conducted as an ongoing survey. From 1980 to 1983, CE data were published separately for the Diary and Interview surveys, but beginning in 1984, selected data were chosen from the each survey and combined to produce publication tables. Such integrated data from the BLS Diary and Interview Surveys provide a complete accounting of consumer expenditures and income that neither survey alone is designed to do. For example, the Diary Survey does not collect data on expenditures for overnight travel or information on reimbursements, whereas the Interview Survey does.

Examples of expenditures for which reimbursements are not collected in the Diary Survey are medical care; automobile repair; and construction, repairs, alterations, and maintenance of property. The Interview Survey does not collect detailed food expenditures, or expenditures for housekeeping supplies, personal care products, and nonprescription drugs. These are collected uniquely in the Diary Survey.

For items that are unique to one survey or the other, the choice of which survey to use as the source of data is obvious. There is overlap in coverage between the surveys where this paper briefly describes the methods employed to select the appropriate survey source for published survey estimates of consumer expenditures.

II. Past Methods of Source Selection

Expenditure items in the current CE are identified using the Universal Classification Codes (UCC) system. A UCC is a six-digit code that classifies reported expenditures at the most detailed level. Selection of survey source for UCCs common to both the Diary and Interview surveys was first conducted for tabulations of 1984-86 data through an Estimated Mean Square Error (MSE) method that used 1982-84 data. This method added the variance of the CE data to the squared difference between the mean of the CE data and the Personal Consumption Expenditure (PCE) estimate produced by the Bureau of Economic Analysis (BEA) to produce an estimate of the mean squared error for each CE source. The source with the smaller MSE was chosen. The method of source selection was changed in 1997 and used CE data from 1993-1995. A Coefficient of Variance (CV) was computed for each source, and the source with the smaller CV was selected.

In 2001, meetings were conducted at the request of the Consumer Price Index (CPI) group to look at differences in source selection between the CE and the CPI using 1999 data. At that time fewer than 15 UCCs had different sources between the two programs. It was recommended that the CPI adopt the CE source decision in all cases with greater than 50 reports of expenditures at the UCC level. The CPI looks at Urban and Regional consumer units for their source selection while the CE uses all Consumer Units in their calculations. After discussions on the differences, fewer than 5 UCCs were sourced differently. Subsequently, when new expenditure items and UCCs were introduced in 2005, source selection was coordinated so that the CE and the CPI were in agreement on the newly introduced UCCs.

In 2007, a team was formed with the task of reviewing the previous methods of source selection and developing a new quantitative methodology for selecting expenditure data from the two surveys.

III. CE Coverage vs. CPI Coverage

As mentioned above, the population coverage of the CE differs from that of the CPI. The CPI is calculated for urban CUs whereas the CE uses all CUs (urban and rural) in their calculations. Definitions of components also differ between the CE and the CPI. For example, homeownership is treated differently in the two surveys: actual expenditures of homeownership are reported in the CE, whereas the CPI uses a rental equivalence approach that estimates the change in the cost of obtaining, in the rental marketplace, services equivalent to those provided by owner-occupied homes.

The CE publishes expenditures for items such as medical care and auto repair net of reimbursements by health insurance and vehicle insurance. Reimbursements for such expenditures are asked for in the Interview Survey and are used in calculating actual out of pocket expenses. The Diary Survey does not collect reimbursement data so expenditures for these items are necessarily taken from the Interview Survey. There are also transportation UCCs that are derived from information exclusively in the Interview Survey. For example, for a new car purchase, the value of any trade-in vehicles is deducted from the purchase price to calculate the net out-of-pocket expense for the new car.

IV. Other Issues Affecting Source Selection

A small number of UCCs are excluded from the source selection process due to reporting issues. For one UCC, there was a concern that some Diary expenditure reports had been misclassified. In other cases, the number of Diary reports directly by respondents was so small that it disqualified the Diary Survey as a source. While the source selection methodology generally evaluates sources based on the number of expenditure reports over a given year, there are some items included in the chained C-CPI-U price index for which a sufficient number of monthly expenditure reports are required. In order to compare the monthly expenditure counts from both surveys, the Diary Survey's monthly counts have to be adjusted upward to account for the Interview Survey's longer recall period and larger sample size. During the 2007 source selection process, the source for four UCCs was based on a monthly comparison of adjusted Diary Survey counts to Interview Survey counts.

V. Source Selection Methodology

Developing the source selection methodology involved a number of steps. Before calculating means and variances, expenditure data are top coded at the 99^{th} percentile and bottom coded at the 1^{st} percentile to minimize the impact of outliers. Bottom coding is a form of censoring the data and is performed by applying the value of the 1^{st} percentile to replace the very small values and conversely, top coding applies the value at the 99^{th} percentile to replace the very large values for each UCC. If sample size is less than 100 counts then the Interquartile Range will be used to more accurately top code and bottom code. This will represent the distance between the 25^{th} and 75^{th} percentiles and you will multiply this value by a factor to get the 1^{st} and 99^{th} percentiles which are approximately +/- 2.33 standard deviations from the median.

Next, the counts (each represents a reported expenditure for that UCC) and Z-Scores (defined below) are weighted for the three most recent collection years using the following scheme, which places greater emphasis on the more recent collection years:

1st collection year (oldest) by 1/6 (For 2007, 2004 data are used).
2nd collection year (middle) by 2/6 (For 2007, 2005 data are used).
3rd collection year (most recent) by 3/6 (For 2007, 2006 data are used).

If a new UCC was created within the most recent two years or if there was a change in the collection instrument that caused a significant difference between the means in the years before and after the instrument change, then the two most recent years of data are analyzed. Counts and Z-Scores are weighted with more emphasis given to the most recent collection year:

1st collection year (oldest) by 2/5 (For 2007, 2005 data are used).
2nd collection year (most recent) by 3/5 (For 2007, 2006 data are used).

VI. Source Selection Decision Criteria

Definitions of the statistical terms used in the analysis are:

1) UCC Mean - the unconditional weighted annual average expenditure for the CPI-U population using the adjusted full sample weight.

2) UCC Z-Score = $$\frac{(\overline{X}_I - \overline{X}_D) - (\mu_I - \mu_D)}{\sqrt{\sigma^2_{\overline{X}_I} + \sigma^2_{\overline{X}_D}}}$$

$\overline{X}_I$ - Annual Mean for that UCC from Interview Survey

$\overline{X}_D$ - Annual Mean for that UCC from Diary Survey

μ_I - Annual Population Mean for that UCC

μ_D - Annual Population Mean for that UCC

$\sigma^2_{\overline{X}_I}$ - Annual Variance for that UCC from Interview Survey

$\sigma^2_{\overline{X}_D}$ - Annual Variance for that UCC from Diary Survey

With the null hypothesis that the team tested, H_0: $\mu_I = \mu_D$ or $\mu_I - \mu_D = 0$, the Z-Score represents the test of equality between the two weighted source means. The numerator is the difference between the sample means $(\overline{X}_I - \overline{X}_D) - 0.$ and the denominator is the standard deviation of that difference. It is assumed that the two surveys are statistically independent of each other. The Balanced Repeated Replications (BRR)[2] method of variance estimation with 44 replicates is used to calculate each source variance.

[2] Balanced Repeated Replication (BRR) is a method of variance estimation used for sample survey statistics when the complexity of a survey's sample design prevents standard classical variance estimation techniques from being used. BRR belongs to a class of variance estimation techniques that use *replications*. The basic idea behind replication is to select sub-samples of the collected data repeatedly from the full sample, and then calculate the statistic of interest from both the full sample and from each sub-sample. These sub-samples are called *replicates*. The difference between the replicate estimates and the full sample estimate are then used to estimate the variance of the full sample statistic.

In order to determine which source to select for each UCC, the following decision criteria are used:

Criterion #1: Counts Sufficiency Criterion. For each UCC and each survey, the number of CUs with at least one expenditure is counted for each of the three most recent data collection years. This yields six counts for each UCC: three yearly counts for the Interview Survey and three yearly counts for the Diary Survey. These counts are used to ensure that a sufficient amount of data is available to make source selection decisions. A sufficient amount of data exists when the count for each of the three years is not less than 60. A count of 60 is chosen as the number of observations needed to be considered credible.

- If both surveys have sufficient data, then proceed to Criterion #2.
- If both surveys lack sufficient data, then keep the original source.
- If one survey has sufficient data, but the other has insufficient data, then a weighted average of the three yearly counts for the survey having an insufficient amount of data is computed: $n^* = (3/6)n_{t-1} + (2/6)n_{t-2} + (1/6)n_{t-3}$.

If the weighted average n^* from the insufficient survey is greater than or equal to 60, then proceed to Criterion #2.

If the weighted average n^* from the insufficient survey is still less than 60, then use the survey with sufficient data for the source.

Criterion #2: Statistical Significance Criterion. If the absolute value of the weighted Z-Score, $z^* = (3/6)z_{t-1} + (2/6)z_{t-2} + (1/6)z_{t-3}$, is greater than or equal to 1.645 then select the source based on the following:

If the weighted Z-Score is greater than or equal to 1.645, then the Interview Survey is selected as the source.

If the weighted Z-Score is less than or equal to -1.645, then the Diary Survey is selected as the source.

Criterion #3: Statistical Non-Significance Criterion. If the weighted Z-Score is between -1.000 and 1.000, then the current source will continue to be used.

Criterion #4: Borderline Statistical Significance Criterion. If the weighted Z-Score is less than -1.000 and greater than -1.645 or greater than 1.000 and less than 1.645, the following method is used to select the source:

If all three Z-Scores are greater than 1.000, then the Interview Survey is selected as the source.

If all three Z-Scores are less than -1.000, then the Diary Survey is selected as the source.

In all other scenarios, the source remains the same.

VII. 2007 Data Results from the Source Selection Process

Table 1 lists the number of overlapping UCCs researched for source selection and the number of UCCs that changed sources. Of the 222 overlap UCCs that were researched, 22 changed sources. Eighteen changed due to having high Z-Scores and 4 changed due to the Diary Survey failing the counts criterion, thereby switching to the Interview Survey. There were a total of 9 UCCs that had observations fewer than 60 in both surveys; therefore the source stayed the same. A total of 75 UCCs had observations in the Diary Survey that failed the counts criterion, thereby using the Interview Survey as the source. Only one UCC had observations in the Interview survey that failed the counts criterion. There were a total of 94 UCCs that had high Z-Scores (their absolute values were greater than or equal to 1.645), 27 UCCs with low Z-Scores (their absolute values were less than 1.000) and 16 UCCs with Z-Scores between 1.000 and 1.645 or between -1.000 and -1.645.

Table 1: Number of Overlap UCCs Researched for Source Selection	**UCCs Researched**	**UCCs Changing Source**
UCCs for any year with observations less than 60 in both surveys	9	
UCCs in which the observations in the Diary survey fail the counts criterion	75	4
UCCs in which the observations in the Interview survey fail the counts criterion	1	
UCCs with Z-Scores between -1.00 and 1.00	27	
UCCs with Z-Scores between +1.00 and +1.645 Or UCCs with Z-Scores between -1.00 and -1.645	16	
UCCs with Z-Scores of -1.645 or less	68	17
UCCs with Z-Scores of 1.645 or greater	26	1
Total	**222**	22

VIII. 2009 Results

The source selection process developed in 2007 was used to evaluate 2006 to 2008 data for the development of 2009 estimates. Upon receiving a list of new UCCs, the source selection program was run on all UCCs having data in both the Interview and Diary surveys. The three year period from 2006 through 2008 was used for most UCCs, and 2007 and 2008 data for the newer UCCs. After calculating the expenditure counts and Z-Scores, the procedures identified 6 UCCs for which the applicable source selection criterion had changed. After further evaluation the source was changed for only two of them.

The source for two of these 6 UCC's had their source changed for 2007, and it was decided not to change again in 2009 to avoid an undue switching between sources

every two years. Of the remaining four UCC's, it was decided to keep the Interview Survey as the source due to concerns about the quality and consistency of the Diary Survey estimates.

The two UCC's did change source from the Diary Survey to the Interview Survey are as follows:

480214 Vehicle Audio Equipment *failed the expenditure count criterion*

520541 Tolls or Electronic Toll Passes *Z-Score strongly favored use of the Interview Survey*

IX. Summary

Source selection is the process of choosing the better survey to use in CE's official published expenditure estimates. It is a multi-step approach performed every two years for every overlap UCC by comparing expenditure data from both the Interview and Diary surveys using a counts criteria and weighted Z-Score approach to determine the best source. The program uses the previous three years of data when available, giving more weight to the most recent years. For new UCC's, only two years of data are used. The data are adjusted for outliers in both the Interview and Diary Surveys. A number of criteria are tested to determine which source to select. The first criterion assesses the number of unweighted consumer units making an expenditure for each UCC in each survey, and may eliminate a source where an insufficient number of CU's report. The next criterion chooses the source that provides the larger overall expenditure per UCC. The means of reported expenditures, weighted by year, are compared from each survey using a standard z-score and in essence, the statistically larger mean is chosen.

The source selection process will continue to be evaluated every 2 years. In addition to evaluating the process, the CE will be looking for ways to accelerate the process by methods such as automation. Future research will also be performed to adapt the process for changes in the survey instruments, collection methodology and processing.