SUPPORTING STATEMENT
Consumer Price Index Commodities and Services

B. DESCRIPTION OF INFORMATION COLLECTIONS EMPLOYING STATISTICAL METHODS

1. Universe and Sample Size Summary

Because of the complexity, importance and diversity of its universe, the construction of the CPI requires a complex set of statistical techniques and samples. Conceptually, the potential respondent universe of price quotations for the CPI is the total set of prices, placed in one-to-one correspondence to the total set of purchases of all urban consumers. The sample for ongoing pricing for the C&S portion of the CPI (including outlets selected but not directly contacted, e.g., using the SABRE system to price airfares) is approximately 57,848 outlets with 188,126 price quotations.

The outlet response rate for ongoing pricing is 95.8 percent per month over the time period from October 2009 to September 2010. The roughly four percent non response in outlets is due to refusals (we remove outlet from the sample) or outlets being temporarily unavailable for pricing. The overall response rate at initiation is 87.3 percent (data obtained or pending); that is, 12.7 percent of the outlets are terminated during initiation. The following table presents response rates for outlets undergoing CPI initiation during the two most recent initiation cycles, August 2009 and February 2010:

Type of Response	**Percent**
Data obtained	77.7
Data pending – awaiting central office clearance, temporarily unavailable	9.6
Refusal	2.3
No CPI items available	5.3
Out-of-business, out of scope, outlet moved, outlet outside PSU	3.7
Unable to locate	1.4

The table below reflects an estimate of an overall response rate determined by multiplying the average estimation rate for repricing (October 2009 to September 2010) with the average initiation rate for the initiation cycles (August 2008 to July 2010) The estimation rate for repricing is determined by dividing the number of outlets (or quotes) used in estimation by the number of outlets (or quotes) collected in repricing, in other words, of all the outlets (or quotes) collected in repricing, how many (what percentage of those) were used in estimation.

Type of Rates	**Outlets**	**Quotes**
Average Estimation Rates (Oct 2009 to Sept 2010)	91.2	82.2
Average Initiation Rates (Aug 2008 to Jul 2010)	87.5	82.9
Overall Estimated Response Rates (avg estimation rate for repricing multiplied with the average initiation rate	79.8	68.1

2. Collection Procedures

2.i. Description of Sampling Methodology

A multi-stage sampling process is employed for the CPI. The four main stages of selection are: (1) the sampling of geographic areas, (2) the sampling of outlets within the geographic areas, (3) the sampling of item groups (ELI's) to be priced in the outlets, and (4) the sampling of items from each ELI in each outlet.

(a) BLS selects Primary Sampling Units (PSU's) or geographic areas for pricing. The sample pricing areas were derived from a stratified design using a controlled selection procedure that provided for the selection of one sample area from each stratum with a control on the distribution of PSU's by state. The initial stratification for the PSU design was based on such variables as percent black consumer units, percent fuel oil heated housing units, percent electric heated housing units, mean wage and salary income per consumer unit, and mean interest and dividend income per consumer unit.

(b) Each year BLS systematically re-selects a portion of the sample of outlets and quotes such that over a four-year period most C&S sample outlets have a chance to be replaced. Not only does this re-establish the distribution of the sample, incorporate new outlet construction and reflect shifts in outlet preferences, but it also allows many respondents to rotate out of the sample. Thus, all respondents are not indefinitely retained in the sample.

The outlet sampling frames are constructed from several sources. The primary source for all food and the majority of the other C&S items is the Telephone Point of Purchase Survey (TPOPS). The TPOPS provides coverage for 60.3 percent of all consumption expenditures for the CPI-U, as of December 2006. Renter and owner-occupied housing account for 29.8 percent. The remaining 9.9 percent of consumption expenditures are covered from a variety of sampling frames constructed by BLS or obtained from other sources.

The TPOPS is a computer assisted telephone collection effort, used to identify a universe of outlets from which CPI sample outlets are selected, and is conducted by the Census Bureau for BLS. TPOPS is made up of 214 purchase categories of goods and services, e.g., prescription drugs. Under TPOPS, during each quarter of the year, in rotating groups of PSU/purchase category groups, households are asked to identify the amount of their expenditures and the names and addresses of the outlets where purchases were made. Samples of outlets for pricing are selected from the TPOPS generated frames using a systematic sampling procedure with each outlet having a probability of selection proportional to the expenditures reported for it on the TPOPS.

(c) The sampling frames from which the item sample market baskets are derived are constructed using data from the most current two years of the Continuing Consumer Expenditure Survey, which is an ongoing survey. Each year as we rotate a portion of the outlet sample we also resample the ELIs. With data from these surveys assembled into the CPI item classification structure, we select the sample of ELI's using a stratified random selection procedure with each ELI having a probability of selection proportional to the expenditures reported for it on the Continuing Consumer Expenditure Surveys.

(d) The BLS Washington Office merges the sample of ELIs with the appropriate sample of outlets. BLS field representatives then initiate the new outlets and select the specific unique items to be priced within each ELI by following an outlet based multistage probability proportional to sales methodology.

2.ii. Description of Estimation Methodology

A price index constructed using geometric means more closely approximates a true cost-of-living index than does the Laspeyres. This occurs because the geometric means formula, unlike the Laspeyres formula, implicitly assumes that product substitution takes place when relative prices change. The geometric means formula assumes that relative expenditures are kept constant over time.

The Laspeyres index formula in concept simply measures the change in the weighted arithmetic mean of prices. As a fixed-weight index, the Laspeyres formula assumes that consumers do not change the amount of each item purchased as relative prices change. In reality, as relative prices change, consumers do switch toward items whose prices become relatively lower.

Based on December 2009 CPI-U relative importances, 61 percent of the CPI is calculated using a Geometric mean formula and 39 percent is based on the Laspeyres index formula. (The Laspeyres portion is composed of Rent, 6 percent, Owners' equivalent rent, 24 percent, and C&S items, 9 percent) Also note that C&S items accounts for 70 percent of the CPI-U weight.

All C&S stratum indexes are calculated using a geometric formula, except for those listed below. Demand elasticity studies led BLS to conclude that the Laspeyres index formula would yield the least biased measure of price change for these items.

C&S Components retaining the Laspeyres (arithmetic mean) Formula		
Lodging while at school, excluding board;	State and local registration, license and motor vehicle property tax;	Physicians' services;
Dental services; Services by other medical professionals;	Hospital services;	Nursing homes; Adult daycare.
Electricity; Utility piped gas service;	Residential water and sewerage maintenance;	

Price relatives.
The price relative for each basic item-area for commodities and services (C&S) using the **geometric mean**-is based on the formula:

$$ {}_{a,i}R^{G}_{[t;t-1]} = \prod_{j \epsilon a,i} \left(\frac{P_{j,t}}{P_{j,t-1}} \right)^{\left(W_{j,POPS} / \sum_{k \epsilon a,i} W_{k,POPS} \right)} $$

The price relative for each basic item-area for commodities and services (C&S) using **Laspeyres**-is based on the formula:

$$ {}_{a,i}R^{L}_{[t;t-1]} = \frac{\sum_{j \epsilon a,i} \left(W_{j,POPS} / P_{j,POPS} \right) P_{j,t}}{\sum_{j \epsilon a,i} \left(W_{j,POPS} / P_{j,POPS} \right) P_{j,t-1}} $$

${}_{a,i}R^{G}_{[t;t-1]}$ and ${}_{a,i}R^{L}_{[t;t-1]}$, are, respectively, the geometric and Laspeyres price relatives for area-item combination, a,i, from the previous period, t-1 (either 1 month or 2 months ago), to the current month, t;

$P_{j,t}$ is the price of the jth observed item in month t for area-item combination a,i;

$P_{j,t-1}$ is the price of the same item in time t-1;

$P_{j,POPS}$ is an estimate of the item j's price in the sampling period when its POPS was conducted; and

$W_{j,POPS}$ is item j's weight in the POPS, defined in detail below

The product in the geomeans formula and sums in the Laspeyres formula are taken over all useable quotes in area-item combination *a, i*. It is important that the price of each quote be collected (or estimated) in both months in order to measure price change.

Quote weights.
For each individual observation, the weight *Wj,POPS* is computed as:

$$Wj,POPS = \alpha\ E\ f\ g\ b\ /\ M\ B$$

where

α is the proportion of the total dollar volume of sales for the ELI relative to the entire POPS category within the outlet (called the outlet's *percent of POPS* for the ELI);

E is an estimate of the total daily expenditure for the POPS category in the PSU half-sample by people in the U population (called the *basic weight*);

f is a *duplication factor* that accounts for any special subsampling of outlets and quotes;

g is a *geographic factor* used to account for differences in the index area's coverage when the CPI is changing from an area design based on an old decennial census to a design based on a more recent census;

b is the number of times the ELI was selected to represent the item stratum, divided by the total selections for the item stratum, in the PSU half-sample;

M is the number of quotes with usable prices in both months *t-1* and *t* for the ELI-PSU half-sample; and

B is the proportion of the item stratum's expenditure accounted for by the ELI in the region.

Index calculation.
When aggregating together price relatives above the elementary index level, the Laspeyres formula is used exclusively implying no substitution across strata in the CPI.

In mid-2002, BLS began publishing a Chained Consumer Price Index for All Urban Consumers (C-CPI-U). [1] The C-CPI-U is a monthly-chained index that uses a Tornqvist formula to aggregate indexes. This index is designed to be a closer approximation to a "cost-of-living index" than the present measures. By utilizing expenditure data in adjoining periods, it reflects consumer substitution across item categories in response to relative prices. The use of expenditure data for both a base period and the current period to average price change across item categories distinguishes the C-CPI-U from the existing CPI measures. Expenditure data required for the C-CPI-U calculations are available only with a lag. Thus, the C-CPI-U, unlike the CPI-U and CPI-W, is issued first in preliminary form and then subject to two subsequent revisions. No additional data collection is required to support the publication of the C-CPI-U.

2.iii. Degree of Accuracy Required

The statute mandating the CPI does not specify a required precision or accuracy for the index. BLS requires that the precision of the CPI be maximized given the total cost constraint imposed by the authorized spending level. BLS developed an allocation model to examine relative efficiencies of various alternative sample designs. The objective of the allocation process is to

[1] For additional details regarding the C-CPI-U, please see BLS Handbook of Methods, Chapter 17, The Consumer Price Index, pages 33-38. http://stats.bls.gov/opub/hom/homch17_a.htm

determine values for all sample design parameters which will minimize the variance of price change for the CPI at the U.S. level subject to the total cost constraint of the CPI budget. The model uses a variance function to project the variance of price change given a set of sample design parameters. It also has a cost function to project the annual cost given a set of values for the sample design parameters. A non-linear programming technique is used to determine the set of values for the sample design parameters which minimizes the variance of price change given a cost constraint. [2]

Starting in 1978, the CPI's sample design has accommodated variance estimation by using two or more independent samples of items and outlets in each geographic area.[3] This allows two or more statistically independent estimates of the index to be made. The independent samples are called *replicates*, and the set of all observed prices is called the *full sample*.

BLS collects CPI data in 38 geographic areas across the United States. These areas consist of 31 *self representing* areas and 7 *non-self-representing* areas. Self-representing areas are large metropolitan areas, such as the Boston, St. Louis, and San Francisco metropolitan areas. Non-self-representing areas are collections of smaller metropolitan areas. For example, one non-self-representing area is a collection of 32 small metropolitan areas in the Northeast region (Buffalo, Hartford, Providence, Bangor, and others), of which 8 were randomly selected to represent the entire set. Within each of the 38 areas, price data are collected for 211 item categories called *item strata*. Together the 211 item strata cover all consumer purchases. Examples of item strata are bananas, women's dresses, and electricity.

Multiplying the number of areas by the number of item strata gives 8,018 (= 38 x 211) different area and item combinations for which price indexes need to be calculated. Separate price indexes are calculated for each one of these 8,018 area and item combinations. After all 8,018 of these *basic-level* indexes are calculated, they are aggregated to form *higher-level* indexes, using expenditure estimates from the Consumer Expenditure Survey as their weights. Examples of higher-level geographic areas are the four regions (Northeast, Midwest, South, and West); and examples of higher-level item categories are the eight major groups (food & beverages, housing, apparel, transportation, medical care, education and communication, recreation, and other goods and services). The highest level of geographic aggregation is the U.S. city average, and the highest level of item aggregation is all items. Variances are computed with a Stratified Random Groups Method, in which variances are computed separately for certain subsets of areas and items and are then combined to produce the variance of the entire area and item combination. Subsets of items are formed by the intersection of the item category with each of the eight major groups.

The estimate of the CPI-U median standard error for 12-month intervals from December 2008 through December 2009 was 0.09 for All Items.

2.iv. Special Sampling -- Sampling of Time

The outlet samples of each PSU are divided into three pricing periods. Each outlet is designated for pricing during a specified period of the month. Therefore, a given item is priced at different times in different outlets in order to average out possible systematic differences between one time period of the month and another and to ensure that there is a full month between pricing for each monthly priced outlet or a full two months between pricings for bi-monthly collected outlets.

---

[2] For a complete description of the allocation process, see: Jacobson, Shawn, Leaver, Sylvia G. and Swanson, David C. (1998), "Choosing a Variance Computation Method for The Revised Consumer Price Index," Proceedings of the Business and Economics Statistical Section, American Statistical Association, 131-136, and Swanson, David C., (1999),

[3] For additional details regarding variance estimation, please see BLS Handbook of Methods, Chapter 17, The Consumer Price Index, pages 41-42. http://stats.bls.gov/opub/hom/homch17_a.htm

2.v. Use of Periodic Data Collection Cycles
Although BLS publishes monthly estimates of the CPI, prices for about 58 percent of the total covered expenditures are not collected monthly in all sampling areas. Of the 58 percent not priced monthly, 30 percent reflects rent and owner's equivalent rent and 28 percent C&S items.

Regarding just the C&S portion (70 percent) of the total CPI expenditure weight, 29 percent is collected monthly and 41 percent is collected bimonthly. The monthly priced C&S items include Food at home, Lodging away from home, Tenants insurance, Household fuels, Motor fuels, Motor vehicle parts, equipment and fees, Recreational reading materials, Education, Postage and delivery, Telephone services, and Tobacco products. (Note, in the three largest consolidated areas, New York, Chicago and Los Angeles all sampled items are priced monthly.) Other commodities and services are priced bi-monthly ("even" cycle--February, April, June, August, October and December or "odd" cycle--January, March, May, July, September and November.)

3. Methods of Maximizing Response

BLS utilizes several techniques to insure that adequate sample sizes are maintained for estimating the CPI. Initial sample sizes are larger than the desired sample sizes to cover initial non-responses, i.e., out-of-business, out-of-scope, refusal, sample items not available, and unable to locate. In rare circumstances, if the sample of outlets is deemed insufficient, we continue pricing the current sample. Additionally, the CPI is exploring methods to improve its sample maintenance system that would allow us to offset sample loss in the primary sample. Benefits include improved sample sufficiency and adequacy of CPI indexes, and a reduction in variances.

4. Testing Plans/Procedures

The CPI is not currently planning any procedural or methods tests requiring OMB approval.

Applicable Performance Goal:
Improve the effectiveness of information and analysis on price changes by improving the timeliness, accuracy, and relevance of the CPI. The performance goal supports directly BLS Strategic Goal 2, Product Improvement and DOL Goal 5, Produce Timely and Accurate Data on the Economic Conditions of Workers and their Families. The focus of these goals is on producing statistics and improving the quality of the economic data, not only in terms of statistical validity, but also in terms of geographic detail, industry detail, demographic detail, and subject matter. Achieving the objective of reducing the variance of the CPI (BLS Project 2.2.5) is contingent on funding in the FY 2011 budget. This proposal will increase the number of CPI commodity and services price quotes collected by 50 percent, improving the accuracy of each published index and the overall quality of CPI data.
In the past, the program has met its performance goal, including its relevancy, timeliness, accuracy, improvement, and efficiency targets. The Administration's assessment, which rated BLS as "effective" in 2005, found that the BLS fills a unique role in the development of national labor-related statistics, and that its design for data collection and reporting is sound.

Small Sample Bias
BLS research shows that a small sample of price changes used to calculate the basic components of the index will yield an upward bias in the inflation estimator of 0.1 percentage points or more per year. Reducing this bias requires an increase in sample through additional data collection and/or through pooling sample across basic components or areas. BLS has requested funding to continue to study and address this source of bias.

5. Statistical Responsibility

W. John Layng, Assistant Commissioner, Division of Consumer Prices and Price Indexes, Office of Prices and Living Conditions of BLS (telephone: 202-691-6950) is the CPI program manager and has overall responsibility for the CPI. Bill Johnson, Chief of the Survey and Analysis Research Branch of the Price Statistical Methods Division of the Office of Prices and Living Conditions (Telephone 202-691-6912) has reviewed and approved the statistical methodology for the survey design. BLS's Office of Field Operation will collect all data. The Division of Consumer Price Computer Systems of the Office of Technology and Survey Processing of BLS will process the data.