

SUPPORTING STATEMENT  
Consumer Price Index Commodities and Services

**OMB CONTROL NO. 1220-0039**

**B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

**1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.**

Because of the complexity, importance and diversity of its universe, the construction of the Consumer Price Index (CPI) requires a complex set of statistical techniques and samples. Conceptually, the potential universe of price quotes 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. Similarly, the potential universe of outlets for the CPI is the total set of all businesses to whom those prices are paid. The sample for ongoing pricing for the Commodities and Services (C&S) portion of the CPI is approximately 34,000 outlets surveyed monthly with 94,000 price quotes per month.

The average monthly outlet response rate for ongoing pricing was 81.6% per month over the time period from October 2021 to September 2022. The roughly 18% non-response rate in outlets is due to refusals or outlets being temporarily unavailable for pricing.

The response rate at initiation was 75.3% of eligible outlets. During initiation 5.5% of outlets are terminated, either because they refuse to participate (0.5%), are ineligible (4.2%), or cannot be located (0.8%). Table 1 presents cumulative response rates for outlets undergoing CPI initiation during the two most recent 6-month initiation cycles, August 2021 and February 2022:

<b>Table 1: Type of Response at Initiation</b>	<b>Percent</b>
Data obtained	72.2
Data pending – awaiting central office clearance, temporarily unavailable	22.4
Refusal	0.5
Ineligible – no CPI eligible items available	3.1
Ineligible – out-of-business, out of scope, outlet moved, outlet outside PSU	1.1
Unable to locate	0.8

Following a formula from the Office of Management and Budget (OMB) <sup>i</sup>, the 75.3 % response rate at initiation is calculated by dividing the percent of outlets with data obtained by the percent of eligible outlets, estimated by the sum of outlets with data obtained, data pending, refusals, and the estimated unable to locate outlets that are eligible:  $75.3 \% = 72.2 / (72.2 + 22.4 + 0.5 + 0.8)$ .

766).<sup>1</sup>

Table 2 below shows an estimate of an overall response rate, determined by multiplying the average monthly estimation rate for repricing (October 2021 to September 2022) with the average initiation rate for the initiation cycles (February 2018 to February 2022). 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 were used in estimation.

<b>Table 2: Overall Response Rate</b>	<b>Outlets (%)</b>	<b>Quotes (%)</b>
Average Initiation Rates *	76.1	75.8
Average Estimation Rates for Repricing **	77.0	67.6
Overall Response Rates ***	56.4	47.2

\* (February 2018 to February 2022)

\*\* (October 2021 to September 2022)

\*\*\* (Average estimation rate for repricing multiplied with the average initiation rate)

## 2. Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**
- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

### Description of Sampling Methodology

A multi-stage stratified 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 based on the reported expenditures of consumers residing in those geographic areas, (3) the sampling of entry-level items (ELIs)<sup>ii</sup> to be priced in the outlets, and (4) the sampling of unique items from each ELI in each outlet.

BLS selects Primary Sampling Units (PSUs) or geographic areas for pricing. The geographic area definitions used are the same as those used for Core Based Statistical Areas (CBSAs). The sample pricing areas are derived from a stratified design using a selection procedure that provides for the selection of one sample area from each stratum with a control on the distribution

<sup>1</sup> The estimated unable to locate outlets that are eligible (0.766%) was computed by multiplying the unable to locate outlet rate (0.8%) with the unable to locate units but are eligible rate (95.8%). The unable to locate units but are eligible rate (95.8%) is based on the percentage of located units that are eligible:  $95.8\% = (72.2 + 22.4 + 0.5) / (72.2 + 22.4 + 0.5 + 3.1 + 1.1)$ .

of PSUs by metropolitan/micropolitan status. The initial stratification for the 2018 PSU sample design<sup>iii</sup> was based on the variables: median income, median property value, latitude, and longitude.

Each year BLS systematically 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 for each item in each area are constructed from several sources. The primary source for all food and the majority of the other C&S items is the Consumer Expenditure Survey<sup>2</sup>. Data is augmented by information collected by the Quarterly Census of Employment and Wages.<sup>3</sup> The CE-Diary is used as the source for most food and food away from home outlets, while the CE-Interview is used for the majority of CPI's other outlet frames. CPI refines the location and address data reported in the CE by comparing the reported data to establishments in the Quarterly Census of Employment and Wages business registry. The CE Survey provides coverage for most of the C&S Survey. Additional frame sources are listed on the [BLS website](#).

The sampling frames from which the item sample market baskets are derived are constructed using data from the most current two years of the CE Survey, which is an ongoing survey. Each year as the CPI rotates a portion of the outlet sample and the ELIs are resampled too. With data from these surveys assembled into the CPI item classification structure, the CPI selects the sample of ELIs using a stratified random selection procedure with each ELI having a probability of selection proportional to the expenditures reported for it on the CE Survey.

The BLS National Office merges the sample of ELIs with the appropriate sample of outlets. BLS data collectors 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 amount methodology.

## **Description of Estimation Methodology**

To aggregate prices into price indexes, the BLS uses a geometric mean or a Laspeyres index formula. The choice of formula depends on the level to which consumers alter their purchases in response to relative price change (consumer substitution). 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<sup>iv</sup>. For all other items, BLS has determined the geometric mean formula yields the least biased estimate of a cost-of-living index.

### **C&S item stratum indexes using the Laspeyres (arithmetic mean) Formula**

- Lodging at school, excluding board
- Electricity
- Utility (piped) gas service
- Water and sewerage maintenance

---

<sup>2</sup> CE – OMB Control Number 1220-0050

<sup>3</sup> QCEW – OMB Control Number 1220-0012

- State motor vehicle registration and license fees
- Physicians' services
- Dental services
- Services by other medical professionals
- Hospital services
- Nursing homes and adult day services
- Prescription drugs

Price relatives.

The price relative for each basic item-area for C&S using the **Geometric Mean** is based on the formula:

$${}_{(i,a)}R_{[t;t-1]}^G = \prod_{j \in (i,a)} \left[ \frac{P_{j,t}}{P_{j,t-1}} \right]^{\left( \frac{W_{j,b}}{\sum_{j \in (i,a)} W_{j,b}} \right)}$$

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

$${}_{(i,a)}R_{[t;t-1]}^L = \frac{\sum_{j \in (i,a)} \left( \frac{W_{j,b}}{P_{j,b}} \right) P_{j,t}}{\sum_{j \in (i,a)} \left( \frac{W_{j,b}}{P_{j,b}} \right) P_{j,t-1}}$$

where

${}_{(i,a)}R_{[t;t-1]}^G$  is the geometric price relative for the item-area combination  $(i,a)$  from the previous period  $(t-1)$  to the current period  $(t)$ ;

${}_{(i,a)}R_{[t;t-1]}^L$  is the Laspeyres price relative for the item-area combination  $(i,a)$  from the previous period  $(t-1)$  to the current period  $(t)$ ;

$P_{j,t}$  is the price of an item,  $j$ , which is a member of item stratum  $i$ , for which a price quote is being collected in area  $a$ , observed in period  $t$ ;

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

$P_{j,b}$  is an estimate of the item  $j$ 's price in the base period; and

$W_{j,b}$  is item  $j$ 's weight in the base period.

The products and sums in the formulas presented above are taken over all price quotes which are usable for estimation in the item-area combination  $(i,a)$ . It is important that the price of each quote be collected (or estimated<sup>4</sup>) in both periods in order to measure price change.

#### Quote weights.

For each individual quote, the weight, or each quote's share of the average daily expenditure on the ELI in the PSU, is given by  $W_{i,b}$ , which is computed as<sup>4</sup>

$$W_{i,b} = \frac{Efg\eta}{N}$$

where

- $E$  is an estimate of the total daily expenditure for the item category in the PSU by people in the CPI-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 its area design;
- $N$  is the number of quotes planned for collection in the item stratum-PSU, which is also the sum of duplication factors for all sampled quotes in the item stratum-PSU;
- $\eta$  is a nonresponse adjustment factor calculated as the quantity  $1 + \frac{y}{n-y}$  where  $y$  is the sum of duplication factors for uninitiated quotes and  $n$  is the number of quotes in the sample design in the ELI-PSU. This is the ratio of planned quotes to quotes with usable prices in both period  $t$  and period  $t-1$  for the ELI-PSU.

#### Index calculation.

When aggregating together price relatives above the elementary index level, the Laspeyres formula is used exclusively implying no substitution across elementary index cells in the CPI-U and CPI-W.

In mid-2002, BLS began publishing a Chained Consumer Price Index for All Urban Consumers (C-CPI-U).<sup>vi</sup> The C-CPI-U is a monthly-chained index that uses a Tornqvist formula, a weighted geometric mean formula which uses the average of expenditure shares in consecutive periods as weights, 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 two consecutive 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

---

<sup>4</sup> This formula is undergoing revision in the Handbook of Methods and should soon match the formula given here.

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 subsequent revisions. No additional data collection is required to support the publication of the C-CPI-U.

BLS issues a report on its research index for the elderly. The CPI for the elderly or R-CPI-E is calculated monthly and is available on the CPI website, alongside other CPI research series. It should be emphasized, the R-CPI-E is merely a reweighting of the CPI basic indexes using expenditure weights from households headed by someone 62 years of age or older. No additional data collection is required to support the publication of the R-CPI-E.

### **Degree of Accuracy and Precision Required**

Section 2 of Title 29, Chapter 1, Subchapter 1, United States Code 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 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 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.<sup>vii</sup>

Since 1978, the CPI's sample design has accomplished variance estimation by using two or more independent samples of items and outlets in each geographic area.<sup>viii</sup> 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 32 geographic areas across the United States. These areas consisted of 23 *self-representing* areas and 9 *non-self-representing* areas. The 23 self-representing areas include 21 PSUs whose population is greater than 2.5 million (such as Boston, Chicago, or San Francisco) and 2 additional units - Anchorage, AK, and Honolulu, HI. Anchorage represents all CBSAs in Alaska, and Honolulu represents all CBSAs in Hawaii. These CBSAs are unique because the locations of both states make price change in their markets geographically isolated from price change in other markets. For this reason, the CBSAs in Alaska and Hawaii are treated as separate geographic strata. 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, Syracuse, and others), of which 2 were randomly selected to represent the entire set. Within each of the 32 areas, price data are collected for 243 item categories called *basic items*<sup>5</sup>. Together the 243 basic items cover all consumer purchases. Examples of basic items are bananas, women's dresses, and electricity.

Multiplying the number of current areas by the number of basic items gives 7,776 (= 32 x 243) different area and item combinations for which price indexes need to be calculated. Separate price indexes are calculated for each one of these 7,776 area and item combinations. After all

---

<sup>5</sup> The set of all goods and services purchased by consumers is divided into 211 categories called item strata: 209 Commodities and Services item strata, plus 2 housing item strata. The number of basic items used for the calculation of aggregate indexes is larger than this, at 243, because the entry level item (ELI) level is used for the calculation of basic cells for health insurance retained earnings (item code SEME) rather than the higher item stratum level. This results in 7,776 (32 x 243) item-area combinations.

7,776 of these *basic-level* indexes are calculated, they are aggregated to form *higher-level* indexes, using expenditure estimates from the CE 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 January 2022 through December 2022 was 0.12% for all items.

### **Unusual Problems Requiring Specialized Sampling Procedures**

We do not have any unusual problems requiring specialized sampling procedures.

### **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 category 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, such that the overall sample is representative of consumer purchases throughout the month. Assigning pricing periods also ensures there is a full month between pricings for each monthly priced outlet or a full two months between pricings for bi-monthly collected outlets.

### **Use of Periodic Data Collection Cycles**

Although BLS publishes monthly estimates of the CPI, prices for only about 59% of the total covered expenditures are collected monthly in all sampling areas. Of the 59% priced monthly, 32% reflects rent and owners' equivalent rent and 27% C&S items.

Regarding just the C&S portion (68%) of the total CPI expenditure weight, 27% is collected monthly and 41% is collected bi-monthly. 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 areas, New York, Chicago, and Los Angeles all sampled items are priced monthly.) Other C&S items are priced bi-monthly ("even" cycle-February, April, June, August, October, and December or "odd" cycle-January, March, May, July, September, and November.)

Some item categories are priced less frequently than our normal monthly or bimonthly cycles. Categories such as College Textbooks, Elementary and High School Books and Supplies, College Tuition and Fixed Fees, Elementary and High School Tuition and Fixed Fees, and Lodging while at School. These categories experience less frequent price change supporting a reduction from monthly and bimonthly data collection while helping to alleviate unnecessary respondent burden and to improve the efficiency of data collection resources.

**3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for**

**intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.**

BLS utilizes several techniques to ensure that adequate sample sizes are maintained for estimating the CPI. Initial outlet sample sizes are larger than the desired sample sizes to cover initial non-responses, e.g., out-of-business, out-of-scope, refusal, sample items not available, and unable to locate. Different modes of collection are also utilized, such as telephone, email, video, and web collection at the request of participants. The use of corporate datasets is another option to encourage participation. In rare circumstances, if a new sample of outlets is deemed insufficient, the CPI will continue pricing the current sample.

**4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.**

Periodically, the CPI may test a new procedure or method to determine its validity. Prior to testing of any new questions CPI will submit a non substantive change to OMB for approval.

**5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze person(s) who will actually collect and/or analyze the information for the agency.**

Rob Cage, Assistant Commissioner, Division of Consumer Prices and Price Indexes, Office of Prices and Living Conditions of BLS (Telephone: 202-691-6959) is the CPI program manager and has overall responsibility for the CPI.

William Johnson, Chief of the Survey Research and Analysis Branch of the Price Statistical Methods Division of the Office of Prices and Living Conditions (Telephone: 202-691-6921) has reviewed and approved the statistical methodology for the survey design.

The Regional Offices of the Division of Price Programs, Office of Field Operations of the BLS has primary responsibility for the collection of Consumer Price Index data.

The Division of Consumer Prices and Price Indexes, Office of Prices and Living Conditions of the BLS has primary responsibility for the analysis of Consumer Price Index data.

OMB Supporting Statement Attachments:

- A. Screen Shots-Collection Instrument for C&S
- B. *The Consumer Price Index Commodities and Service Survey: Questions & Answers Pamphlet*
- C. *The Consumer Price Index Commodities and Service Survey: Questions & Answers Pamphlet (Spanish version)*
- D. C&S Introductory Letter from an Assistant Regional Commissioner of Labor Statistics



Consumer Price Index Commodities and Services

OMB Control Number 1220-0039

OMB Expiration Date: 9/30/2023

- E. The Consumer Price Index: Modernizing Data Collection: APIs and Web Scraping Fact Sheet
- F. The Consumer Price Index: Modernizing Data Collection: APIs and Web Scraping Pamphlet
- G. The Consumer Price Index: Modernizing Data Collection: Corporate Data Fact Sheet
- H. Shuttle Form example: MC011 Physicians' Services

<sup>i</sup> For additional information, see [OMB Standards and Guidelines for Statistical Surveys](#), guideline 3.2.2

<sup>ii</sup> For additional information about ELIs and item strata, see [CPI Item Aggregation](#)

<sup>iii</sup> For additional information, see [The 2018 revision of the Consumer Price Index geographic sample](#)

<sup>iv</sup> For more information, please see [Incorporating a geometric mean formula into the CPI](#)

<sup>v</sup> For more information, please see the CPI Handbook of Methods at

<https://www.bls.gov/opub/hom/cpi/calculation.htm#imputation>

<sup>vi</sup> For additional details regarding the C-CPI-U, please see the CPI Handbook of Methods at

<https://www.bls.gov/opub/hom/cpi/calculation.htm#final-c-cpi-u-upper-level-aggregation-formula>

<sup>vii</sup> 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).

<sup>viii</sup> For additional details regarding variance estimation, please see the CPI Handbook of Methods at

<https://www.bls.gov/opub/hom/cpi/calculation.htm#sources-of-error>