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
Consumer Price Index (CPI) requires a complex set of statistical techniques and samples.
Conceptually, the potential 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 Commodities and Services (C&S) portion of the CPI is
approximately 35,547 outlets with 89,708 price quotations per month.

The outlet response rate for ongoing pricing is 90.1% per month over the time period from
October 2018 to September 2019. The roughly 10% non-response rate in outlets is due to
refusals or outlets being temporarily unavailable for pricing.

The response rate at initiation is 79.1% of eligible outlets. During initiation 10.6% of outlets are
terminated, either because they refuse to participate (2.0%), are ineligible (7.7%), or cannot be
located (0.9%). The following table presents response rates for outlets undergoing CPI initiation
during the two most recent initiation cycles, August 2018 and February 2019:

Type of Response at Initiation	Percent
Data obtained	72.9
Data pending – awaiting central office clearance, temporarily unavailable	16.4
Refusal	2.0
Ineligible – no CPI items available	5.0
Ineligible – out-of-business, out of scope, outlet moved, outlet outside PSU	2.8
Unable to locate	0.9

Following a formula from the Office of Management and Budget (OMB), the 79.1% 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: 79.1% = 72.9 / (72.9 + 16.4 + 2.0 +
0.829). [1]

The table below shows an estimate of an overall response rate, determined by multiplying the
average monthly estimation rate for repricing (October 2018 to September 2019) with the
average initiation rate for the initiation cycles (August 2017 to July 2019). 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.

---

[1] The estimated unable to locate outlets that are eligible (0.829%) was computed by multiplying the
unable to locate outlet rate (0.9%) with the unable to locate units but are eligible rate (92.1%). The
unable to locate units but are eligible rate (92.1%) is based on the percentage of located units that are
eligible: 92.1% = (72.9 + 16.4 + 2.0) / (72.9 + 16.4 + 2.0 + 5.0 + 2.8).

	Outlets (%)	Quotes (%)
Average Initiation Rates *	80.0	80.1
Average Estimation Rates for Repricing **	87.4	78.5
Overall Response Rates ***	69.9	62.9

*   (Aug 2017 to Jul 2019)
**   (Oct 2018 to Sept 2019)
*** (Average estimation rate for repricing multiplied with the average initiation rate)

2. Collection Procedures

2.i. 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 within the geographic areas, (3) the sampling of entry-level items (ELIs) 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 of PSUs by metropolitan/micropolitan status. In the 1998 sample design, four independent variables were used for stratifying the non-self-representing PSUs: normalized (centered and scaled by the range) longitude, the square of normalized longitude, normalized latitude, and percent urban. The initial stratification for the 2018 PSU design was based on the variables average income, average 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 frame is constructed from several sources. The primary source for all food and the majority of the other C&S items prior to 2021 was the Telephone Point of Purchase Survey. [2] For samples selected after 2021, the Consumer Expenditure Survey [3] is the source of outlet sampling frames rather than TPOPS. Data is augmented by information collected by the Quarterly Census of Employment and Wages. [4] The TPOPS or CE Surveys provide coverage for 55% of all consumption expenditures for the CPI-U, as of December 2018. Renter and owner-occupied housing account for 32%. The remaining 13% of consumption expenditures are covered from a variety of sampling frames constructed by BLS or obtained from other sources.

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.

---

[2] TPOPS – OMB Control Number 1220-0044
[3] CE – OMB Control Number 1220-0050
[4] QCEW – OMB Control Number 1220-0012

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.

Changes to the CPI establishment frame started in October 2019. The CPI Program ended collection of the TPOPS in June 2019 and will obtain its retail establishment frame from the same household survey used to obtain the expenditure weights needed to calculate the index. This change to the CE Surveys information will eliminate redundancies and inefficiencies in survey operations, and will result in lower household burden. The CE-Diary will be used as the source for most food and food away from home outlets, while the CE-Interview will be used for the majority of CPI's outlet frames. CPI plans to refine 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.

2.ii. Description of Estimation Methodology

Based on December 2018 CPI-U relative importances, 56% of the CPI is calculated using a Geometric mean formula and 44% is based on the Laspeyres index formula. The Laspeyres portion is composed of Rent (8%), Owners' equivalent rent (24%), and C&S items (12%).

A price index constructed using geometric means more closely approximates a true cost-of-living index than does the Laspeyres, for some items. 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.

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

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

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

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

Where

$_{(i,a)}R^G_{[t;t-1]}$

is the geometric price relative for the item-area combination *(i,a)* from the previous period *(t-1)* to the current period *(t)*;

$_{(i,a)}R^L_{[t;t-1]}$

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 the 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 product 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) in both periods in order to measure price change.

<u>Quote weights</u>.

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

$$W_{j,b} = \frac{AEfg\eta}{BN}$$

where

A

is defined as the proportion of CE expenditures for the ELI relative to the entire item category within the Census region;

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);

4

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;
B	is the proportion of CE expenditures for the ELI relative to the item stratum within the region; and
η	is a nonresponse adjustment factor calculated as the quantity $1 + \dfrac{y}{n-y}$ 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.

In mid-2002, BLS began publishing a Chained Consumer Price Index for All Urban Consumers (C-CPI-U). [i] 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 subsequent revisions. No additional data collection is required to support the publication of the C-CPI-U.

BLS periodically issues a report on its experimental index for the elderly. The CPI for the elderly or CPI-E is calculated monthly and is available on request. It should be emphasized, the 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 CPI-E.

2.iii. Degree of Accuracy 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. [ii]

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.[iii] 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*.

From 1978-2018, BLS collected CPI data in 38 geographic areas across the United States. These areas consisted 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, Syracuse, Burlington, and others), of which 8 were randomly selected to represent the entire set. Within each of the 32 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 current 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 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.

In 2018, BLS introduced a new geographic area sample for the CPI. The new area sample has 23 self-representing areas and 9 Census divisions for a total of 32 geographic areas. The 23 self-representing areas include 21 PSUs whose population is greater than 2.5 million 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 that in other markets. For this reason, the CBSAs in Alaska and Hawaii are treated as separate geographic strata. With 23 self-representing PSUs and nine Census divisions, the new area design will yield 6,752 basic indexes (32 index areas by 211 item strata) for the U.S. all-items CPI.

The estimate of the CPI-U median standard error for 12-month intervals from January 2019 through December 2019 was 0.08% for All Items.

2.iv. 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. 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.

2.v. Use of Periodic Data Collection Cycles

Although BLS publishes monthly estimates of the CPI, prices for about only 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 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 ensure 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, e.g., 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, the CPI will continue pricing the current sample.

4. Testing Plans/Procedures

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 nonsubstantive change to OMB for approval.

5. Statistical Responsibility

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.

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. *The Consumer Price Index: Modernizing Data Collection: APIs and Web Scraping* Fact Sheet
   E. *The Consumer Price Index: Modernizing Data Collection: APIs and Web Scraping* Pamphlet
   F. C&S Introductory Letter from a Regional Commissioner of Labor Statistics

[i] For additional details regarding the C-CPI-U, please see BLS Handbook of Methods, Chapter 17, The Consumer Price Index, pages 28-34. https://www.bls.gov/opub/hom/pdf/homch17.pdf

[ii] 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).

[iii] For additional details regarding variance estimation, please see BLS Handbook of Methods, Chapter 17, The Consumer Price Index, pages 37-39. https://www.bls.gov/opub/hom/pdf/homch17.pdf