Sampling and Weighting of Commodity and Service Units for the Elementary Level of Computation of the U.S. Consumer Price Index

Mary Lee Fuxa
Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3655,
Washington, D.C. 20212

Abstract
Elementary-level indexes are weighted and aggregated to produce the all-items U.S. Consumer Price Index each month. An elementary-level index is computed for each combination of an item stratum and index area. Given an item stratum and a geographic sample unit representing an index area, the probabilities used to select specific items and the outlets that sell them are not always equal to the desired probabilities that would lead to (near) unbiased estimation of targeted indexes. In this case, the sample is "imbalanced" -- the expected sample proportion for a sample unit is not equal to what it should be. In the estimator for the elementary-level index, weighting factors are used to correct the imbalance. This paper shows the construction of the weighting factors.

Key Words: Price Relative, Economic Weight, Sample Weight

1. Background

The Consumer Price Index (CPI) is a nationwide survey conducted by the U.S. Bureau of Labor Statistics (BLS). It measures the change in prices over time of the goods and services that people buy for day-to-day living. Elementary-level indexes are computed for 211 item categories called "item strata" in 38 geographic areas called "index areas." The elementary-level indexes are aggregated to produce various price indexes, including the all items – all U.S. price index most frequently mentioned in the media.

For operational purposes, the CPI's 211 item strata are divided into two sub-categories, "Housing" and "Commodities and Services" (C&S). Housing has 33% relative importance by weight in the CPI, and C&S has 67%. This paper focuses on the C&S portion of the CPI.

A geographic sample unit representing an index area is referred to as a primary sample unit (PSU). It is a small cluster of counties. Replicate geographic samples are used for the purpose of computing variances.

For C&S, samples of specific items and the outlets that sell them are independently selected for each combination of an item stratum, PSU, and replicate. The samples are updated periodically by re-selecting them on a continuing rotation basis. The prices of the specific items are collected monthly or bimonthly by representatives of the BLS.

2. Purpose of the Study

This paper shows the construction of the economic weight and the sample weight in the estimator for an elementary-level index.

Given an item stratum and an intersection of a PSU and replicate, the economic weights for items and outlets are equal to their consumer expenditure shares. Likewise, the desired probabilities of selection for items and outlets are equal to the consumer expenditure shares. However, the probabilities used for selection are not always equal to the desired probabilities. In this case, the sample is imbalanced because the expected sample proportion for a sample unit is not equal to the desired probability for that unit. The sample weighting factors correct this imbalance.

If the probabilities of selection for the sample units at a particular level of sampling and lower are equal to the desired probabilities, then the weighting factors at that level and lower cancel out into a constant across the sample units. The sample units are self-weighted by their sample proportions which reflect their consumer expenditure shares. If the probabilities of selection are not equal to the desired probabilities, then the weighting factors do not cancel out and are variable. Though the variability of the weighting factors serves to balance the sample, it is a source of sampling error in estimation. This paper evaluates the variability.

The resulting sampling error might be minimal compared to the potential sampling error resulting from a sample design that would be subject to the constraint of using probabilities of selection equal to the desired probabilities. Since the design would be subject to ongoing budget constraints, smaller sample sizes might be required to balance out the cost of the new constraint and sampling error would increase.

3. Method of Sampling

The item strata contain sub-categories called entry-level items (ELIs). An example is the item stratum *Cakes, Cupcakes, and Cookies*. It contains the ELI *Cakes and Cupcakes (Excluding Frozen)* and the ELI *Cookies*. For the first stage of sampling for an item stratum and a PSU-replicate intersection, one or more ELIs are selected with probability based on a measure that is positively correlated with consumer expenditure share. Multiple selections of an ELI are possible.

ELI selection is based on data from the Consumer Expenditure Survey sponsored by the BLS. Data at the level of a region, rather than the level of a PSU-replicate intersection, are used. Thus the probability of selection may not equal the desired probability.

Sampling frames of outlets for a PSU-replicate intersection are obtained from the Telephone Point of Purchase Survey sponsored by the BLS. Each ELI has a corresponding sampling frame of outlets that sell the ELI. A bakery in the food court of a shopping mall is an example of an outlet that sells the ELI *Cookies*. In the second stage of sampling, two or more outlets from a frame are selected with probability based on a measure that is positively correlated with consumer expenditure share. Multiple selections of an outlet are possible.

The same outlet sampling frame often is used for multiple ELIs, and the outlets' probabilities of selection are based on the expenditures consumers make in the outlets for the full set of ELIs. Thus an outlet's probability of selection may not equal the desired probability for any individual ELI in the set.

For a combination of a single selection of an ELI and a single selection of an outlet, a unique item within the outlet is selected with probability based on a measure that is positively correlated with consumer expenditure share. Multiple selections of a unique item are possible. An example of a unique item for the ELI *Cookies* is a 1-lb. bag of chewy-style chocolate-chip cookies with walnuts, of a particular brand name. A single selection of a unique item is referred to as a quote. For each quote, the price of the unique item is collected on a monthly or bimonthly basis.

4. Elementary-Level Index

An index $I_{0,t}$, reflecting price change from month 0 to the current month *t*, is calculated at the item stratum and index-area level by multiplying the index for the previous month by a price relative that measures the price change between the previous month and the current month:

$$I_{0,t} = I_{0,(t-1)} \; R_{(t-1),t} = \prod_{k=1}^{t} R_{(k-1),k} \qquad (1)$$

where: $I_{0,(t-1)}$ = index from 0 to (*t*-1); $R_{(t-1),t}$ = price relative from (*t*-1) to *t*.

Equation 2 shows the geometric mean that is the target price relative for all but a few item strata. For the sake of simplicity, it is shown at the level of a PSU-replicate intersection.

$$R_{(t-1),t} = \prod_{E=1}^{N} \prod_{O=1}^{N_E} \prod_{U=1}^{N_{E,O}} \left(\frac{p_t}{p_{t-1}} \right)_{E,O,U}^{\frac{(v_a)_{E,O,U}}{\sum\sum\sum (v_a)_{E,O,U}}} \qquad (2)$$

where: N = number of ELIs in the item stratum; N_E = number of outlets in the frame corresponding to ELI *E*; $N_{E,O}$ = number of unique items in outlet *O* for ELI *E*; $(p_t / p_{t-1})_{E,O,U}$ = the ratio of the price in month *t* to the price in month *t*-1 for unique item *U* in outlet *O* for ELI *E*; $(v_a)_{E,O,U}$ = consumer expenditure in month *a* for unique item *U* in outlet *O* for ELI *E*, where month *a* is associated with the timing of sample rotation.

5. Probability Framework for Estimation

Given an item stratum and a PSU-replicate intersection, the probabilities of selection for an ELI *E*, an outlet *O*, and a unique item *U*, are based on measures that are positively correlated with their consumer expenditure shares. Equation 3 shows the joint probability that is equal to the product of the conditional probabilities.

$$\mathrm{P}_{E,O,U} \;=\; \Pr(E)\;\Pr(O|E)\;\Pr(U|E,O) \tag{3}$$

The desired probability, $\mathrm{P}^{*}_{E,O,U}$, is equal to the product of the desired conditional probabilities. Equation 4 shows two equivalent formulas for $\mathrm{P}^{*}_{E,O,U}$. The first one shows it calculated directly as the consumer expenditure share portrayed in Equation 2 for a population element *E,O,U*. The second one shows it as a special case of the joint probability shown in Equation 3.

$$\mathrm{P}^{*}_{E,O,U} \;=\; \frac{(v_a)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}$$

$$= \left(\frac{\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}\right)\left(\frac{\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}{\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}\right)\left(\frac{(v_a)_{E,O,U}}{\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U}}\right) \tag{4}$$

Equation 5 shows that, for each population element *E,O,U*, the product of the inverse of the desired probability multiplied by the consumer expenditure is equal to the total consumer expenditure over all population elements:

$$\left(\frac{1}{\mathrm{P}^{*}_{E,O,U}}\right)(v_a)_{E,O,U} \;=\; \sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}(v_a)_{E,O,U} \tag{5}$$

6. Estimator of the Target Price Relative

For an item stratum and a PSU-replicate intersection, Equation 6 shows the current estimator for the target price relative shown in Equation 2.

$$\hat{R}_{(t-1),t} \;=\; \prod_{c=1}^{K}\prod_{(\text{quote}=1)}^{H_c n_c}\left(\frac{P_t}{P_{t-1}}\right)_{c,\text{quote}}^{\left(\frac{(W\,V_a)_{c,\text{quote}}}{\sum_{c=1}^{K}\sum_{(\text{quote}=1)}^{H_c n_c}(W\,V_a)_{c,\text{quote}}}\right)} \tag{6}$$

where: K = number of ELIs in the item stratum that will have received at least one sample hit; H_c = number of sample hits that ELI c will have received, where $0 < H_c \le n$ and n = number of sample hits over all ELIs in the item stratum; n_c = number of outlet sample hits over all outlets for ELI c; quote = a combination of a single sample hit of ELI c and a single sample hit of an outlet for ELI c; $\left(P_t / P_{t-1}\right)_{c,\text{quote}}$ = random variable that is the ratio of the price in month t to the price in month t-1 for the unique item to be sampled for a quote; $W_{c,\text{quote}} = \left(H_c / n\right)\left(1/H_c n_c\right)\left(1/\mathrm{P}_{c,\text{quote}}\right)$, where $\mathrm{P}_{c,\text{ quote}}$ = the joint probability of selection for the ELI and the outlet that will correspond to a quote and for the unique item that will be sampled for it; $\left(V_a\right)_{c,\text{quote}}$ = random variable that is the consumer expenditure in month a for the unique item to be sampled for a quote.

7. Heuristic View of the Exponent

This section illustrates how, in the exponent in Equation 6, the sample weight $W_{c,\text{quote}}$ and the economic weight $\left(V_a\right)_{c,\text{quote}}$ work together. It compares estimation of the price relative for a large realized sample to that for a small realized sample.

With a large realized sample, the sample units corresponding to the set of quotes will be weighted by values that are close to their consumer expenditure share. The estimate of the price relative based on Equation 6 will be close to the target price relative shown in Equation 2. With a small realized sample, the weighting factors in Equation 6 will balance the quotes so that the consumer expenditure shares of the sample units corresponding to the set of quotes will be appropriately reflected.

Given a realized sample, the estimate of the price relative for an item stratum and a PSU-replicate intersection can be expressed as Equation 7:

$$\hat{r}_{(t-1),t} = \prod_{c=1}^{k} \prod_{(\text{quote}=1)}^{h_c n_c} \left(\frac{p_t}{p_{t-1}}\right)_{c,\text{quote}}^{\left(\frac{\left(\frac{1}{n\,n_c}\right)\left(\frac{1}{\mathrm{P}}\,v_a\right)_{c,\text{ quote}}}{\sum_{c=1}^{k}\sum_{(\text{quote}=1)}^{h_c n_c}\left(\frac{1}{n\,n_c}\right)\left(\frac{1}{\mathrm{P}}\,v_a\right)_{c,\text{ quote}}}\right)} \qquad (7)$$

where: k , h_c , $\left(p_t / p_{t-1}\right)_{c,\text{quote}}$, and $\left(v_a\right)_{c,\text{quote}}$ are the sample realizations of random variables K , H_c , $\left(P_t / P_{t-1}\right)_{c,\text{quote}}$, and $\left(V_a\right)_{c,\text{quote}}$; $\mathrm{P}_{c,\text{ quote}}$ = the joint probability of selection for the ELI, outlet, and unique item that define the population element that is the sample realization for a quote.

Equation 7 can be expressed as Equation 8:

$$\hat{r}_{(t-1),t} = \prod_{c=1}^{k} \prod_{(\text{quote}=1)}^{h_c n_c} \left(\frac{p_t}{p_{t-1}}\right)_{c,\text{quote}}^{\left(\frac{1}{n\,n_c}\right)\left(\frac{x_{c,\text{quote}}}{\bar{x}}\right)} \qquad (8)$$

where: $x_{c,\text{quote}} = \left(\frac{1}{\text{P}} \; v_a\right)_{c,\text{ quote}}$; $\bar{x} = \sum_{c=1}^{k} \sum_{(\text{quote}=1)}^{h_c n_c} \left(\frac{1}{n\, n_c}\right)\left(\frac{1}{\text{P}} \; v_a\right)_{c,\text{ quote}}$

For the simple case where all $n_c = m$, Equation 8 can be expressed as Equation 9:

$$\hat{r}_{(t-1),t} = \prod_{(\text{quote}=1)}^{nm} \left(\frac{p_t}{p_{t-1}}\right)_{\text{quote}}^{\left(\frac{1}{nm}\right)\left(\frac{x_{\text{quote}}}{\bar{x}}\right)} \qquad (9)$$

where: $\bar{x} = \sum_{(\text{quote}=1)}^{nm} \left(\frac{1}{nm}\right)\left(\frac{1}{\text{P}} \; v_a\right)_{\text{quote}}$

Equation 9, which is a weighted geometric mean over quotes, can be expressed as Equation 10, which is a weighted geometric mean over population elements:

$$\hat{r}_{(t-1),t} = \prod_{E=1}^{N} \prod_{O=1}^{N_E} \prod_{U=1}^{N_{E,O}} \left(\frac{p_t}{p_{t-1}}\right)_{E,O,U}^{\left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{x_{\text{E,O,U}}}{\bar{x}}\right)} \qquad (10)$$

where: $h_{E,O,U}$ = total number of quotes for population element *E,O,U*. It is the realization of the random variable $H_{E,O,U}$, where $0 \le H_{E,O,U} \le H_E m$. Random variable H_E is the number of sample hits of ELI *E*, where $0 \le H_E \le n$. Also, $x_{E,O,U} = \left(\frac{1}{P} \; v_a\right)_{E,O,U}$ and $\bar{x} = \sum_{E=1}^{N} \sum_{O-1}^{N_E} \sum_{U=1}^{N_{E,O}} \left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{1}{P} \; v_a\right)_{E,O,U}$, where: $x_{E,O,U}$ and $\bar{x}$ apply to a population element *E,O,U* rather than a single quote for a population element; $x_{E,O,U}$ and $\bar{x}$ also apply to the unsampled population elements for which $h_{E,O,U} = 0$; $\bar{x}$ is the average value of $x_{E,O,U}$ weighted by its associated sample proportion, $\left(h_{E,O,U} / nm\right)$.

With very large samples, where $n \to \infty$, $m \to \infty$, the variance of the sample proportion for population element *E,O,U* approaches 0: $\text{V}\left(\frac{H_{E,O,U}}{nm}\right) = \frac{\left(\text{P}_{E,O,U}\right)\left(1-\text{P}_{E,O,U}\right)}{nm} \to 0$. The realized sample proportion for population element *E,O,U* is close to the expected value which is equal to the probability of selection: $\left(\frac{h_{E,O,U}}{nm}\right) \approx \text{E}\left(\frac{H_{E,O,U}}{nm}\right) = \text{P}_{E,O,U}$. Equation 11 shows that the exponent in Equation 10 is close to the consumer expenditure share for population element *E,O,U* in the exponent of the target price relative shown in Equation 2:

$$\left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{x_{E,O,U}}{\bar{x}}\right) = \left(\frac{\left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{1}{\mathrm{P}}\ v_a\right)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{1}{\mathrm{P}}\ v_a\right)_{E,O,U}}\right)$$

$$\approx \left(\frac{\left(\mathrm{P}_{E,O,U}\right)\left(\frac{1}{\mathrm{P}}\ v_a\right)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(\mathrm{P}_{E,O,U}\right)\left(\frac{1}{\mathrm{P}}\ v_a\right)_{E,O,U}}\right) = \left(\frac{\left(v_a\right)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(v_a\right)_{E,O,U}}\right) \qquad (11)$$

With a small realized sample, the sample proportion $\left(h_{E,O,U}/nm\right)$ for each population element *E,O,U* most likely does not cancel out with the inverse of the probability of selection. If the number of population elements is large, the sample proportion for many of the elements is equal to 0. Together, the sample proportion and the probability inverse balance out the sample.

In the special case where the probability of selection $\mathrm{P}_{E,O,U}$ is equal to the desired probability $\mathrm{P}^*_{E,O,U}$ for each population element, the realized sample is inherently balanced. Using Equation 5:

$$\left(\frac{x_{E,O,U}}{\bar{x}}\right) = \left(\frac{\left(\frac{1}{\mathrm{P}^*}\ v_a\right)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(\frac{h_{E,O,U}}{nm}\right)\left(\frac{1}{\mathrm{P}^*}\ v_a\right)_{E,O,U}}\right)$$

$$= \left(\frac{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(v_a\right)_{E,O,U}}{\sum_{E=1}^{N}\sum_{O=1}^{N_E}\sum_{U=1}^{N_{E,O}}\left(v_a\right)_{E,O,U}}\right) = 1 \qquad (12)$$

Looking at this result with respect to Equation 10, $\left(x_{E,O,U}/\bar{x}\right)$ is the extraneous weight factor in the geometric mean used to estimate a replicate-level index. When each $\mathrm{P}_{E,O,U}$ is equal to $\mathrm{P}^*_{E,O,U}$, each $\left(x_{E,O,U}/\bar{x}\right)$ is equal to 1 and each population element is weighted only by its sample proportion $\left(h_{E,O,U}/nm\right)$. The sample proportion reflects the economic weight. It does so because the expected value of the random sample proportion, $\mathrm{E}\left(H_{E,O,U}/nm\right)$, is equal to $\mathrm{P}^*_{E,O,U}$ which is equal to the consumer expenditure share in the target price relative shown in Equation 2. When each $\mathrm{P}_{E,O,U}$ is not equal to $\mathrm{P}^{*}_{E,O,U}$, the factor $\left(x_{E,O,U} / \bar{x}\right)$ corrects the imbalance of the sample proportions.

In like manner, in Equation 9 each $\left(x_{\text{quote}} / \bar{x}\right)$ is equal to 1 when sample selection is based on $\mathrm{P}^{*}_{E,O,U}$. When sample selection is not based on $\mathrm{P}^{*}_{E,O,U}$, the factor $\left(x_{\text{quote}} / \bar{x}\right)$ corrects the sample imbalance.

The advantage of Equation 10 is that it is easier to compare it to the target price relative shown in Equation 2. The advantage of Equation 9 is that, typically, computation is based on a product or summation over quotes.

8. Variability of Exponent Factors

The variability of $\left(x_{c,\text{quote}} / \bar{x}\right)$ in Equation 8 can be used to profile the realized samples, at the level of an item stratum and a PSU-replicate intersection, that support estimation. Given a realized sample, the sample average and sample standard deviation of $\left(x_{c,\text{quote}} / \bar{x}\right)$ can be used to evaluate the variability.

For all samples, the sample average of $\left(x_{c,\text{quote}} / \bar{x}\right)$ is equal to $\left(\bar{x} / \bar{x}\right) = 1$. For a sample that is balanced because $\mathrm{P}_{\mathrm{E,O,U}} = \mathrm{P}^{*}_{E,O,U}$ for all of the population elements, the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ is equal to 1 for each quote and the sample standard deviation is equal to 0. For a sample that is imbalanced because $\mathrm{P}_{\mathrm{E,O,U}} \neq \mathrm{P}^{*}_{E,O,U}$ for some or all of the population elements, the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ is equal to $\left(1 \pm \text{an error}\right)$, and the factor corrects the sample imbalance.

It is important to note that this method of evaluation cannot be used directly to profile the realized samples at the item stratum and index-area level because the factor in this case is $\left(x_{\text{PSU-replicate},\, c,\, \text{quote}} \;\; / \sum_{\text{all PSU-replicates}} \left(\bar{x}\right)_{\text{PSU-replicate}} \right)$.

9. Method of Measurement

The following conditions support the case where the probabilities of selection for an ELI and an outlet are equal to the desired probabilities: (1) the item stratum contains a single ELI; (2) the outlet frame maps to a single ELI. In the estimate of the price relative at the level of an item stratum and a PSU-replicate intersection, the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ in Equation 8 is equal to 1 for all of the quotes in the realized sample.

When condition 1 is not true, the ELI probability of selection and desired probability are not equal to 1.00. Variability of the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ potentially exists since: (a) the ELIs' probabilities of selection are based on data from the Consumer Expenditure (CE) Survey rather than the Telephone Point of Purchase Survey (TPOPS) which is used for outlet selection; (b) the consumer expenditure data from the CE Survey are at the regional level, whereas the consumer expenditure data from the TPOPS are specific to the PSU-replicate intersection.

When condition 2 is not true, variability of the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ potentially exists since the outlets' probabilities of selection are based on the consumer expenditures reported for the outlets for the set of ELIs mapping to the TPOPS outlet frame rather than the specific ELI.

The actual variability of the factor $\left(x_{c,\text{quote}} / \bar{x}\right)$ depends on how $x_{c,\text{quote}} = \left(\left(1/\text{P}\right) v_a\right)_{c,\text{ quote}}$ is measured. The consumer expenditure $v_{E,O,U}$ for a population element that is the sample realization for a quote divided by the outlet and unique-item probabilities of selection is formulated and measured as shown in Equation 13:

$$\frac{v_{E,O,U}}{\Pr(O|E)\ \Pr(U|E,O)} = \frac{v_{E,O,U}}{\left(\dfrac{\left(s_O + \sum_{U=1}^{N_{E,O}} v_{E,O,U}\right)}{\sum_{O=1}^{N_E}\left(s_O + \sum_{U=1}^{N_{E,O}} v_{E,O,U}\right)}\right)\left(\dfrac{v_{E,O,U}}{\sum_{U=1}^{N_{E,O}} v_{E,O,U}}\right)} \qquad (13)$$

where: s_O = consumer expenditure in outlet *O* for the other ELIs in the set of ELIs containing ELI *E*.

Equation 13 reduces to Equation 14:

$$\frac{v_{E,O,U}}{\Pr(O|E)\ \Pr(U|E,O)} = \left(\sum_{O=1}^{N_E}\left(s_O + \sum_{U=1}^{N_{E,O}} v_{E,O,U}\right)\right)\left(\frac{\sum_{U=1}^{N_{E,O}} v_{E,O,U}}{\left(s_O + \sum_{U=1}^{N_{E,O}} v_{E,O,U}\right)}\right)$$

$$= \mathsf{E}_E\,\alpha_{E,O} \qquad (14)$$

where: E_E = consumer expenditure for the set of ELIs containing ELI *E* summed over all outlets in the TPOPS outlet frame; $\alpha_{E,O}$ = ELI E 's proportion of the consumer expenditure in outlet *O* for the set of ELIs containing ELI *E*.

For the purpose of practicality and flexibility in construction and estimation of an elementary-level index, all $\alpha_{E,O}$ for ELI *E* and the PSU-replicate intersection are set equal to a constant α_E derived from regional-level data from the CE Survey. Therefore, in Equation 8:

$$x_{c,\text{quote}} \;=\; x_{E,O,U} \;=\; \left(\frac{1}{\mathrm{P}}\; v_a\right)_{E,O,U} \;=\; \frac{\mathsf{E}_E\,\alpha_E}{\Pr(E)} \qquad (15)$$

It is important to note that in Equation 8, $x_{c,\text{quote}}$ is fixed and constant over the quotes within an ELI given a realized sample at the level of an item stratum and a PSU-replicate intersection. However, it is variable across ELIs if the item stratum contains multiple ELIs and the realized sample consists of more than one of them.

10. Empirical Results

This analysis profiles the realized samples at the level of an item stratum and a PSU-replicate intersection that support estimation by evaluating the variability of $\left(x_{c,\text{quote}} / \bar{x}\right)$ in Equation 8. The samples are classified into 2 types, as shown in Table 1:

Table 1: Classification of Samples

Sample Type	*Item Stratum Composition*	*Characteristic of the Sample*
1	Single ELI	$\mathrm{P}_{\mathrm{E,O,U}} = \mathrm{P}^{*}_{E,O,U}$ for all population elements
2	Multiple ELIs	$\mathrm{P}_{\mathrm{E,O,U}} \neq \mathrm{P}^{*}_{E,O,U}$ for some or all population elements

Based on this classification, Table 2 shows by major item group the number of samples at the level of an item stratum and a PSU-replicate intersection that are used in the study. In the Fall of 2010, these samples will be rotated into the sample design for C&S and will represent approximately 1/8 of the total number of samples. Table 2 also shows the number of item strata that the samples in a cell are associated with. There are approximately 15 samples for each item stratum.

Table 2: Profile of Samples in the Study

	Number of Samples			*Number of Item Strata*		
Major Group	*Type 1*	*Type 2*	*Total*	*Type 1*	*Type 2*	*Total*
Apparel	136	121	257	8	8	16
Education*	144	33	177	9	2	11
Food	859	118	977	55	7	62
Housing**	195	198	393	12	14	26
Medical	98	17	115	7	1	8
Recreation	149	173	322	9	11	20
Transportation	88	47	135	6	4	10
Other	99	59	158	7	4	11
All	1768	766	2534	113	51	164
Percent	70	30	100	69	31	100

* *Education and Communication*
** *Housing (Excluding Shelter)*

For the samples in a cell in Table 2, Table 3 shows the average interval for $\left(x_{c,\text{quote}} / \bar{x}\right)$ based on 1 standard deviation. It also shows the average number of designated quotes per sample. The results show that the samples tend to be quite small.

Table 3: Variability of Exponent Factors that Correct Sample Imbalance

Major Group	*Average Interval of* $\left(x_{c,\text{quote}} / \bar{x}\right)$ *for the Samples in a Cell Based on 1 Standard Deviation* Type 1	Type 2	All	*Average Number of Designated Quotes per Sample in a Cell* Type 1	Type 2	All
Apparel	1.00 ± 0.00	1.00 ± 0.18	1.00 ± 0.08	6.6	9.0	7.8
Education	1.00 ± 0.00	1.00 ± 0.07	1.00 ± 0.01	5.5	3.5	5.1
Food	1.00 ± 0.00	1.00 ± 0.15	1.00 ± 0.02	6.2	5.5	6.1
Housing	1.00 ± 0.00	1.00 ± 0.15	1.00 ± 0.07	6.1	5.6	5.9
Medical	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	21.3	9.6	19.6
Recreation	1.00 ± 0.00	1.00 ± 0.23	1.00 ± 0.12	5.5	5.2	5.3
Transportation	1.00 ± 0.00	1.00 ± 0.17	1.00 ± 0.06	7.8	8.6	8.1
Other	1.00 ± 0.00	1.00 ± 0.26	1.00 ± 0.10	5.9	4.1	5.2
All	1.00 ± 0.00	1.00 ± 0.17	1.00 ± 0.05	7.0	6.1	6.7

Table 2 shows that 70% of the samples in the study are Type 1. As expected for Type 1 samples, Table 3 shows that there is no variability in $\left(x_{c,\text{quote}} / \bar{x}\right)$ since the samples are inherently balanced to the extent allowed by the defined method of measurement of $x_{c,\text{quote}}$.

For the 30% of the samples that are Type 2, Table 3 shows that the major group "Other" has the highest degree of variability in $\left(x_{c,\text{quote}} / \bar{x}\right)$ with a sample standard deviation equal to 0.26. It is followed by "Recreation" with a sample standard deviation equal to 0.23. The major group "Medical" shows a sample standard deviation equal to 0.00. All of the samples in this cell represent the case where a single outlet frame maps to all of the ELIs in an item stratum and only those ELIs. This characteristic in combination with the measurement technique that all $\alpha_{E,O}$ for an ELI and a PSU-replicate intersection are set equal to a constant α_E results in the lack of variability.

11. Conclusion

This paper has explained the weighting factors used at the quote level to correct for sample imbalance. It has profiled the samples supporting estimation by evaluating the variability of these factors. The variability contributes to the sampling error of elementary-level indexes.

Acknowledgements

I would like to thank Ronald Johnson and David Swanson for their continued interest and helpful suggestions.

References

Bureau of Labor Statistics (2008). *BLS Handbook of Methods*, Chapter 17, pp. 21-22.

Casady, Robert J., Alan H. Dorfman, and Richard L. Valliant (1996). "Sample Design and Estimation for the CPI," *BLS Note*.

Fuxa, Mary Lee (1998). “Estimating Price Relatives for the U.S. Consumer Price Index when Priced Items are Restratified,” *Proceedings of the Section on Business and Economic Statistics*, American Statistical Association.

Swanson, David C. (1994, revised 2003). “The Origin of C&S Quote Weights in the CPI,” *BLS Note*.