U.S. BUREAU OF LABOR STATISTICS

An Overview of the State-Level Weighting Procedure for the Consumer Expenditure Survey

Susan King and Taylor J. Wilson

March 2018

Consumer Expenditure Surveys Program

Introduction

This document presents an overview of the process used to create state-level weights for the Consumer Expenditure Survey (CE) public-use microdata. Every household in the sample has a weight indicating the number of households in the population it represents, and this document describes a process in which those weights are modified to make them represent the state in which the households are located rather than the region of the country in which they are located. The process modifies their base weights along with their nonresponse and calibration adjustments that are used to generate their final weights, which are then used to make statistically valid state-level expenditure estimates. State-level weights can only be provided for states with a sufficient sample size and representative primary sampling units (PSUs). The goal of providing state-level weights is to support state-level data analysis by public-use microdata users.

Background

The CE survey is a nationally representative household survey that collects expenditure data from a sample of households in the U.S. The survey program produces expenditure estimates using the data collected from those households. The CE sample design is a two-step process in which a random sample of geographic areas is selected from the U.S., and then a random sample of households is selected inside those selected areas. The geographic areas are small clusters of counties called PSUs, and the process is designed to produce unbiased expenditure estimates at the national, U.S. Census Region, and U.S. Census Division level, but not at the state level.

The U.S. Census Bureau partitions the U.S. into four geographic "regions" (Northeast, Midwest, South, and West), and it partitions each region into two geographic "divisions" (except the South which is partitioned into three divisions). Then a random sample of PSUs is drawn from each division, and a random sample of households is drawn from each PSU. Prior to 2017 CE's weights were calibrated to the four Census region population totals, but beginning in 2017 they are calibrated to the nine Census division population totals. This allows statistically valid expenditure estimates to be made at the division, region, and U.S. levels, but not at geographic levels below divisions and regions (such as the state level). The gap at the state level is filled by calculating base weights from the sample of households in the

existing PSUs in a state and calibrating their weights to their state population totals. The state-level weights are statistically valid only for state-level estimates. They are not statistically valid for lower geographic areas, such as counties within a given state, or higher geographic areas such as divisions, regions, or the nation.

Calculating the State-Level Weights – Overview

The state-level weights are computed in a three-step process that starts with *base weights*, which are the number of households in a state's population that a household in the sample represents. It is typically around 18,000 which means that one out of every 18,000 households in a state is selected for the survey. Then the base weights are increased by multiplying them by a nonresponse adjustment factor to account for households that were selected for the survey but did not participate in it. And finally, the weights are multiplied by a calibration adjustment factor to make the weights of all the respondent households in the state's sample add up to the state's true population. The final weights are typically around 24,000, which means that approximately one out of every 24,000 households in a state participates in the survey.

Calculating the State-level Base Weight

Both the national-level and state-level base weights are calculated using a probability of selection and the "Within PSU Sampling Interval." The two procedures differ in the calculation of the probability of selection. For state-level estimates, U.S. Census tracts replace PSUs for determining the probability of selection. Census tracts are small contiguous areas within a county or county equivalent that are relatively permanent from Census to Census. Tracts are designed to have approximately 4,000 people but they can range from 1,200 to 8,000 people. Census tracts vary in geographical size, depending on population density, and range from a few blocks in densely populated areas to hundreds of square miles in sparsely populated areas.

Census tracts for a state are stratified using the same algorithm for stratifying PSUs. For state estimates, the number of clusters is set to five, corresponding to the number of income quintiles. Only two clustering variables are used for state estimates: median household property value and median household income, which correlate with expenditures and are calculated for each Census tract from five-year American Community Survey's (ACS) estimates. The census tracts, like the PSUs in the national estimate, within each cluster should be homogeneous with respect to median household property value and median income.

The data set with cluster group assignment is merged with the CE data set using the variables county and Census tract so that every CU on the CE data set has a cluster assignment. The total population of the Census tracts in each cluster is summed and the total population of the CE interviews (respondents and Type "A" nonrespondents) is also summed. Thus the probability of selection for each cluster is:

$$Probability\ of\ Selection = \frac{Sum\ of\ CE's\ Tract\ Population\ in\ a\ Cluster}{Sum\ of\ Census\ Tract\ Population\ in\ a\ Cluster}$$

Both the national-level and state-level base weight equations use the same "Within PSU Sampling Interval." The national sample of 12,000 addresses is allocated to PSUs proportional to the number of people that the PSU represents. The "Within PSU Sampling Interval" is the number of addresses on the sampling frame in the PSU divided by the number of sample addresses allocated to the PSU and is calculated by the U.S. Census Bureau. In other words, if all of the addresses were ordered, the interval

between selected households is the "Within PSU Sampling Interval." Thus the state-level base weight equation is:

$$State\ Base\ Weight = \frac{1}{Probability\ of\ selection} \times (Within\ PSU\ Sampling\ Interval)$$

Using the Weighting Control Factor

The weighting control factor adjust for subsampling in the field, which occurs when the data collector visits a particular address and discovers multiple housing units where only one housing unit was expected. The weighting control factor is generally 1. Therefore, the new weight is:

CONTWT= State Base Weight x Weighting Control Factor

Nonresponse Adjustment Process

The nonresponse adjustment factor adjusts for interviews that cannot be conducted in occupied housing units due to a consumer's refusal to participate in the survey or the inability to contact anyone at the housing unit despite repeated attempts.

For state estimates, the sample CUs for a quarter are partitioned into cells using the following variables: the American Community Survey's (ACS) median household income for each CU's Census tract, CU size, and number of contact attempts. There are three income classes (bottom 10 percent, middle 80 percent, and upper 10 percent), four contact attempt classes, and four consumer unit size classes. The procedure is run separately for each income class creating a 4x4 matrix (16 distinct cells) with the other two variables. Using the formula below, the nonresponse adjustment factor is calculated for each cell, but when there is an insufficient number of CUs in a cell, the factor is adjusted using a hierarchical cell collapsing procedure with the consumer unit size and number of contact attempt variables. Again, the nonresponse adjustment factor is calculated for each cell. The nonresponse adjustment is calculated for each cell using the following formula:

$$NONINTAD_k = \frac{\sum_{cu \in I_k} CONTWT_{cu} + \sum_{cu \in NI_k} CONTWT_{cu}}{\sum_{cu \in I_k} CONTWT_{cu}}$$

where,

NONINTAD = nonresponse adjustment

CONTWT = control factor state base weight

k = cell number

cu = consumer unit

I_k = set of interviewed consumer units in cell=*k*

NI_k = set of noninterviewed consumer units in cell=*k*

Finally, this ratio is multiplied by the CONTWT to adjust for nonresponse and create the STAGE1WT.

STAGE1WT = State Base Weight x Weighting Control Factor x NONINTAD

Calibration Process

The final weight adjustment is calibration and uses only respondent interviews. STAGE1WT is calibrated quarterly to nine "known" population totals to account for frame undercoverage. CE uses Lagrange multipliers to select the weights that minimize the amount of change made to the STAGE1WTs so that the calibrated weights multiplied by the number of CU members in each of the demographic groups sum to the population totals. The nine population totals are from the Current Population Survey (CPS) and are: the total number of households in the state; the total number of homeowning households in the state and the total number of people in seven age categories.

There are infinitely many set of weights that sum to the nine "known" population totals. The algorithm sets limits on the amount the STAGE1WTs can change, between 0.5 and 4. , and the final state weights are:

STATE_WGT = State Base Weight x Weighting Control Factor x NONINTAD x Calibration Factor

Presently, the state-level weighting process is in an experimental phase. State-level weights are being made available to Consumer Expenditure Survey microdata users to gauge interest and usefulness. For further details, questions, or comments on the procedures or on the weights themselves, please contact the Division of Consumer Expenditure Surveys at CEXINFO@bls.gov. For more detail on CE's national-level estimates, see this document's reference section.

References

[1] Neiman, D, King, S., Swanson, D., Ash, S., Enriquez, J., and Rosenbaum, J. (2015). Review of the 2010 Sample Redesign of the Consumer Expenditure Survey. Proceedings of the American Statistical Association, Section on Survey Methodology, 2522-2536.

[2] King, S. (2017). Estimating State Expenditures for New Jersey, Florida, and California. https://www.bls.gov/cex/ws17-estimating-state-expenditures-nj-fl-ca.pdf

[3] King, S. (2017). Calculating State Weights for the Consumer Expenditure Survey. Proceedings of the American Statistical Association.

[4] Bureau of Labor Statistics, U.S. Department of Labor (2017). U.S. Bureau of Labor Statistics Handbook of Methods, Consumer Expenditures and Income https://www.bls.gov/opub/hom/cex/pdf/cex.pdf

[5] Bureau of Labor Statistics, U.S. Department of Labor (2017). Consumer Expenditure Public Use Microdata Documentation (2017). https://www.bls.gov/cex/pumd_doc.htm