On point estimation and variance estimation for mean expenditures in the Consumer Expenditure Survey

Michail Sverchkov, John Eltinge, Lawrence R. Ernst

Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 1950, Washington, DC 20212-0001
Sverchkov.Michael@bls.gov

Abstract

The Consumer Expenditure Survey (CE) is a survey with a multistage design. The first stage of the CE sample includes a set of areas (PSUs) selected from the set of U.S. Core Based Statistical Areas, and CE additionally selects a set of PSUs to represent the rest of the nation. After selecting the original sample of PSUs, a number of selected PSUs were cut from the sample for budgetary reasons. However, the deletion process did not use an explicit probability mechanism in some cases. Consequently, it was of interest to develop weighting-adjustment methods that provide some degree of robustness against potential bias. Mean expenditures are estimated by a Hajek-type ratio estimator based on sample weights calibrated to 23 known demographic population totals. The variances of these estimators are estimated by the Balanced Repeated Replication (BRR) technique. In this paper we investigate how the sample cut in CE influences current estimates of mean expenditures on national level and examine BRR estimates of their variances.

Key words: calibrated estimator, 0-calibrated variance estimator, sample cut selection probability correction, BRR variance estimation, variance strata collapsing, bias of variance estimate

1. Introduction

The Consumer Expenditure Survey (CE) is a nationwide household survey designed by the U.S. Bureau of Labor Statistics (BLS) to find out how Americans spend their money. One of the primary uses of its data is to provide expenditure weights for the Consumer Price Index (CPI). The CPI affects millions of Americans by its use in cost-of-living wage adjustments for many workers, cost-of-living adjustments to Social Security payments, and inflation adjustments to Federal income-tax brackets. CE data are also used to compare expenditure patterns of various segments of the population, such as elderly versus non-elderly people. In addition, the data are being used by the Federal Government in a new experimental poverty index.

Sample Design. The selection of specific households to participate in the CE survey is carried out in multiple stages. The first stage of sampling is defining and selecting a random sample of geographic areas called “primary sampling units” (PSUs) from across the United States. In this stage, the set of counties in the United States is divided into small groups of one or more counties (called PSUs), and a representative sample of them is selected to be in the survey. After the PSUs are defined and selected, the second stage of sampling involves determining the number of households to be visited in each PSU. The CE’s budget allows for a certain number of households to be visited each year nationwide, and, in this stage, that number is allocated across the individual PSUs selected for the survey. The final stage of sampling is selecting specific households to be visited within the PSUs. Households are selected using a systematic selection procedure to ensure that each category of households is well-represented in the survey.

Defining and Selecting the PSUs. In the first stage of sampling, PSUs are defined and selected for the survey. PSUs are counties or groups of counties grouped together into geographic entities called “core-based statistical areas” (CBSAs) by the U.S. Office of Management and Budget (see http://www.census.gov/population/www/metroareas/metrodef.html for the details). CBSAs were defined for use by Federal statistical agencies in collecting data and tabulating statistics.

There are two types of CBSAs, metropolitan and micropolitan. “Metropolitan” CBSAs consist of one or more counties with at least one urban area of 50,000 or more people, while “micropolitan” CBSAs consist of one or more counties centered around an urban area with 10,000-50,000 people. Both include the adjacent counties that have a high degree of social and economic integration with the area's core as measured by commuting ties. Areas outside CBSAs are called "non-CBSA" areas and are mostly rural.

After the PSUs are defined, they are categorized according to their population and region of the country. There are four regions of the country (Northeast, Midwest, South, and West), and four PSU "size classes":
"A" PSUs, which are metropolitan CBSAs with a population over 2 million people
"X" PSUs, which are metropolitan CBSAs with a population between 50,000 and 2 million people
"Y" PSUs, which are micropolitan CBSAs
"Z" PSUs, which are non-CBSA areas, and are often referred to as "rural" PSUs

By definition, the "A" PSUs are "self-representing" and, therefore, have a 100 percent probability of selection in the CE survey. The "X," "Y," and "Z" PSUs are "non-self-representing." The non-self-representing PSUs are grouped together into groups of PSUs (called "strata") according to a 5-variable geographic model whose variables are latitude, longitude, latitude squared, longitude squared, and percent of the population in the PSU who live in an urban area. A typical stratum has approximately 10 PSUs, and all of the PSUs are in the same "region-size class." After the PSUs are grouped into strata, one PSU per stratum is randomly selected with probability proportional to its population. The PSU that is randomly selected represents the whole stratum (see King and Johnson-Herring 2007 for the details).

Sample cut. After this sample design was implemented, newly imposed budget constraints forced the CE and CPI to eliminate 11 "X" PSUs from the sample. In addition, seven "A" PSUs were reclassified to the "X" category for publication purposes. However, these seven reclassified PSU's were still truly self-representing for design and estimation purposes. As a result, the sample of PSUs currently used by the CE has 91 PSUs, of which 75 urban PSUs are also used by the CPI (see Ernst, Johnson and Larsen 2007 for the detailed description of the sample cut procedure and the sample weights adjustment for the sample cut).

Mean expenditure estimates. For a given expenditure the national estimate of its mean during a particular time period is defined as a ratio of "calibrated probability weighted" estimate of its total cost in the time period to "calibrated probability weighted" estimate of the population total (Hajek estimate), where "calibrated probability weighted" estimate is a weighted estimate with the weights equal to inverse probability of selection adjusted to sample cut and non-response and calibrated to 23 known population totals. These 23 calibration totals are the total numbers of members within different age, geographical and urban groups. Note also that since the weighted estimate of the population total is calibrated to the known totals, the denominator of the Hajek estimator is constant (see Section 2). Therefore, the national estimate of the mean of a particular expenditure can be viewed as simple calibrated probability weighted estimator divided by a constant. The latter can be not correct for sub-national mean expenditure estimates, such as mean expenditure estimates within specific age or income group or within the giving part of the country.

Variance estimates of the mean expenditure estimators. Variances of the mean expenditure estimators are estimated by Balance Repeated Replication (BRR) technique (see Wolter 1985, Chapter 3). CE uses 0-calibrated BRR estimator, i.e. each replicate estimate is calibrated to the same 23 known totals (see Sverchkov, Dorfman, Ernst and Guciardo 2004). Recall that the CE sample design selects only one PSU from each stratum on the first stage of the selection where classical BRR variance estimator assumes selection of two PSUs from each stratum. Consequently, CE (i) divides A-PSUs into "pair of variance PSUs" such that population totals in the respective variance PSUs are as close as possible, (ii) collapses geographically close X-, Y- and Z-PSUs into pairs of variance PSU's such that population totals in respective variance PSUs are as close as possible, and then apply BRR technique as if the sample design is a stratified sample of two PSUs from variance PSU's from the respective strata. It is well known that (i) decreases the expectation of the variance estimator, and thus may induce negative bias; and (ii) increase the expectation, and thus may induce positive bias (see Wolter 1985, Chapters 2 and 3, Rust and Kalton 1987).

This paper investigates whether the sample cut in CE influences significantly the estimator of mean expenditures and examine whether collapse of the X-, Y-, Z- PSU produces significant bias of 0-calibrated variance BRR estimators.

2. Definitions.

Let H be the number of the original sampling strata, y_{hi} be an expenditure of interest of Consumer Unit (CU) i in the selected PSU in stratum h, N_h be the number of CUs in the selected PSU in stratum h, and $N = \sum_{h=1}^{H} N_h$.

Estimators before the sample cut. Let s_h denote the sample of CUs selected from the selected PSU in stratum h, w_{hi} be inverse probability of selection of the selected unit i in the selected PSU in stratum h before the sample cut adjusted for non-response (non-calibrated sample weights before the sample cut). The calibrated sample weights v_{hi} are defined as the weights that satisfy $\sum_{h=1}^{H}\sum_{i\in s_h} v_{hi}\mathbf{x}_{hi}=\mathbf{t}_{\mathbf{x}}$, $L\le \frac{v_{hi}}{w_{hi}}\le U$, and vector $\mathbf{v}$ minimizes $\sum_{h=1}^{H}\sum_{i\in s_h}(v_{hi}-w_{hi})^2/w_{hi}$, where L and U are predetermined lower and upper bounds for calibration adjustment $\frac{v_{hi}}{w_{hi}}$, $\sum_{h=1}^{H}\sum_{i\in Stratum_h}\mathbf{x}_{hi}=\mathbf{t}_{\mathbf{x}}$ is a vector of the known 23 total numbers of members within different age, geographical and urban groups and $\mathbf{x}_{hi}=(x_{hi,1},...,x_{hi,23})^{\mathbf{T}}$ is respective vector of numbers of members within different age, geographical and urban groups within CU.

Let x_{hi}^{*} be the total number of members within CU(h,i), and define the mean expenditure estimator of national population mean expenditure, $\bar{Y}=N^{-1}\sum_{h=1}^{H}\sum_{i=1}^{N_h} y_{hi}$, as a ratio estimator $\hat{\bar{Y}}=\sum_{h=1}^{H}\sum_{i=s_h} v_{hi}y_{hi}\Big/\sum_{h=1}^{H}\sum_{i=s_h} v_{hi}x_{hi}^{*}$. It is well known that under very general assumptions $\hat{\bar{Y}}$ is consistent for $\bar{Y}$. Note that x_{hi}^{*} is a linear function of $x_{hi,1},...,x_{hi,23}$, so that $\sum_{h=1}^{H}\sum_{i\in s_h} v_{hi}x_{hi}^{*}=N$ and therefore $\hat{\bar{Y}}=N^{-1}\sum_{h=1}^{H}\sum_{i=s_h} v_{hi}y_{hi}$.

Remark. The latter is correct for the estimators on the national level, for estimator in a sub-area *A* of the population, such as mean expenditure estimator within specific age or income group or within the giving part of the country, $\sum_{h=1}^{H}\sum_{\substack{i\in s_h\\ h,i\in A}} v_{hi}x_{hi}^{*}$ is not necessarily equal to the population in area *A* and therefore $\hat{\bar{Y}}$ can be non-linear and the calibration property can be violated in the sub-area. In what follows we consider the estimators on the national level and give possible generalizations for the estimators on sub-national level in the comments.

Estimators after the sample cut. The mean expenditure estimator after the sample cut is defined similarly to the estimator before the sample cut except that the non-calibrated sample weights w_{hi} are replaced by the sample weights corrected to the sample cut, $\tilde{w}_{hi}$,as defined in Ernst, Johnson and Larsen (2007). Thus the calibrated sample weights after the sample cut, $\tilde{v}_{hi}$ satisfy $\sum_{h=1}^{H}\sum_{i\in s_h}\tilde{v}_{hi}\mathbf{x}_{hi}=\mathbf{t}_{\mathbf{x}}$, $L\le\frac{v_{hi}}{w_{hi}}\le U$, and vector $\tilde{\mathbf{v}}$ minimizes $\sum_{h=1}^{H}\sum_{i\in s_h}(v_{hi}-\tilde{w}_{hi})^2/\tilde{w}_{hi}$ and $\hat{\tilde{Y}}=\sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}y_{hi}\Big/\sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}x_{hi}^{*}=N^{-1}\sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}y_{hi}$. Note that unlike the estimators before the sample cut, since $\tilde{w}_{hi}$ is not necessarily equal to the inverse of the actual probability of inclusion into the cut-sample, $\hat{\tilde{Y}}$ is not necessarily consistent for $\bar{Y}$.

3. Testing whether the sample cut induces bias in the mean expenditure estimator on the national level. Similarly to Sverchkov, Dorfman, Ernst, Moerhle, Paben and Ponikowski (2005, Remark 1 and 2), one can show that if $L=-\infty$ and $U=\infty$ then

$$\hat{\hat{Y}} - \hat{\tilde{Y}} = \ N^{-1}\mathbf{t}_{\mathbf{x}}^{T}\left[\sum_{h=1}^{H}\ \sum_{i\in s_h} w_{hi}\mathbf{x}_{h\mathbf{i}}\mathbf{x}_{h\mathbf{i}}^{\mathbf{T}}\right]^{-1}\sum_{h=1}^{H}\ \sum_{i\in s_h}(w_{hi}-\tilde{w}_{hi})\mathbf{x}_{h\mathbf{i}}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}}) \tag{1}$$

where w_{hi} are the same inverse probability weights considered above, and $\hat{\mathbf{B}} = [\sum_{h=1}^{H}\ \sum_{i\in s_h} w_{hi}\mathbf{x}_{hi}\mathbf{x}_{hi}^{T}]^{-1}\sum_{h=1}^{H}\ \sum_{i\in s_h} w_{hi}\mathbf{x}_{hi}y_{hi}$ is weighted least square estimate of the regression coefficient (From Result 5 of Deville and Särndal 1992 it follows that the same remains correct asymptotically for reasonably bounded *L* and *U*). Note that $\sum_{h=1}^{H}\ \sum_{i\in s_h} w_{hi}\mathbf{x}_{h\mathbf{i}}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}}) = 0$ since $\hat{\mathbf{B}}$ is the weighted least square regression coefficient, and therefore (1) is equivalent to

$$\hat{\hat{Y}} - \hat{\tilde{Y}} = \ N^{-1}\mathbf{t}_{\mathbf{x}}^{T}\left[\sum_{h=1}^{H}\ \sum_{i\in s_h} w_{hi}\mathbf{x}_{h\mathbf{i}}\mathbf{x}_{h\mathbf{i}}^{\mathbf{T}}\right]^{-1}\left[\sum_{h=1}^{H}\ \sum_{i\in s_h}\mathbf{z}_{h\mathbf{i}}\mathbf{z}_{h\mathbf{i}}^{\mathbf{T}}\right]\left[\sum_{h=1}^{H}\ \sum_{i\in s_h}\mathbf{z}_{h\mathbf{i}}\mathbf{z}_{h\mathbf{i}}^{\mathbf{T}}\right]^{-1}\sum_{h=1}^{H}\ \sum_{i\in s_h}\mathbf{z}_{h\mathbf{i}}z_{hi}^{*} \tag{2}$$

where $\mathbf{z}_{hi} = \mathbf{x}_{h\mathbf{i}}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}})$, $z_{hi}^{*} = \tilde{w}_{hi}$ or $\mathbf{z}_{hi} = \mathbf{x}_{h\mathbf{i}}$, $z_{hi}^{*} = \tilde{w}_{hi}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}})$. In what follows we use the latter, $\mathbf{z}_{hi} = \mathbf{x}_{h\mathbf{i}}$, $z_{hi}^{*} = \tilde{w}_{hi}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}})$. Therefore the difference between the calibrated estimators before and after the sample cut, $\hat{\hat{Y}} - \hat{\tilde{Y}}$, is insignificant if ordinary least squares regression coefficient $\mathbf{A} = \left[\sum_{h=1}^{H}\ \sum_{i\in s_h}\mathbf{z}_{h\mathbf{i}}\mathbf{z}_{h\mathbf{i}}^{\mathbf{T}}\right]^{-1}\sum_{h=1}^{H}\ \sum_{i\in s_h}\mathbf{z}_{h\mathbf{i}}z_{hi}^{*}$ is insignificant, the latter can be checked for example by F-test by use of SAS PROC REG. The sufficient condition for use of F-test is that z_{hi}^{*} -s in the sample are independent and normally distributed approximately. Since only one PSU is selected into the sample, the first stage of the selection does not induce dependence. Recall that the second stage of selection is systematic sample with relatively small sampling fractions. Pfeffermann et al. (1998) show by simulation studies that distribution of sampled measurements under systematic sampling design are independent asymptotically where asymptotics means that population size increases when the sample size held fixed. We applied Shapiro-Wilk test for normality to each $z_{hi}^{*} = \tilde{w}_{hi}(y_{hi}-\mathbf{x}_{hi}^{\mathbf{T}}\hat{\mathbf{B}})$, y_{hi} corresponds to a particular representative expenditure in CE 2005 Diary Survey and the normality hypothesis was not rejected for any of them on 0.1 significance level. Finally we applied F-test to CE 2005 Diary Survey and $H_0 : \mathbf{A} = \mathbf{0}$ were not rejected for all representative CEs on 0.1 significance level. Therefore we decided that the sample cut does not induce any significant bias to the national estimates of mean expenditures.

COMMENT. Consider now the mean expenditure estimator for a sub-area *A*, and let $(y_{hi}^{A}, x_{hi}^{*A}) = (y_{hi}, x_{hi}^{*})$ if unit (h,i) belongs to area *A* and $(0,0)$ otherwise. Then the mean expenditure estimator before the sample cut can be written as $\hat{\hat{Y}}^{A} = \left(N \Big/ \sum_{h=1}^{H}\ \sum_{i=s_h} v_{hi}x_{hi}^{*A} \right) N^{-1}\sum_{h=1}^{H}\ \sum_{i=s_h} v_{hi}y_{hi}^{A}$ and similarly for the estimator after the sample cut, $\hat{\tilde{Y}}^{A} = \left(N \Big/ \sum_{h=1}^{H}\ \sum_{i=s_h} \tilde{v}_{hi}x_{hi}^{*A} \right) N^{-1}\sum_{h=1}^{H}\ \sum_{i=s_h} \tilde{v}_{hi}y_{hi}^{A}$. Therefore, if the difference between $\sum_{h=1}^{H}\ \sum_{i=s_h} v_{hi}x_{hi}^{*A}$ and $\sum_{h=1}^{H}\ \sum_{i=s_h} v_{hi}\tilde{x}_{hi}^{*A}$ is small, which is often the case, the above test procedure can be applied to the estimators in the sub-area *A* by substituting y_{hi}^{A} in place of y_{hi} .

4. Testing whether CE BRR variance estimators based on collapsed variance PSUs overestimate insignificantly the true variances. For the simplicity assume that BRR variance strata are defined by pairs of PSUs, {g1,g2; g=1,…,G}, (for each PSU(gi) corresponds original stratum ST(gi) from wich this PSU was selected), 2G is a number of collapsed original strata, 2G < H = total number of original strata. Generalization of the following results to the case of grouping more than two PSUs in the variance strata is straightforward. As it is shown in Wolter (1985, Chapters 2 and 3), Rust and Kalton (1987), the positive expected bias of BRR variance estimator induced by collapsing of PSUs is equal to

$$Bias[\hat{V}(\hat{\hat{Y}})] = E[\hat{V}(\hat{\hat{Y}}) - V(\hat{\hat{Y}})] = \frac{1}{N^2}\sum_{g=1}^{G}(t_{y,g1} - t_{y,g2})^2 \ , \tag{3}$$

where $t_{y,gj}$ is the total expenditure in PSU(gj), j=1,2. Recall that CE uses 0-calibrated BRR variance estimator, i.e. each replicate estimate, similar to the original one, is calibrated to the same 23 known totals, and the sums of the calibrated sample weights, v_{hi}, (over the cut sample or the replicate cut samples) are equal to the total population, which implies that $\hat{V}(\hat{\hat{Y}}) = \hat{V}(E)$ and $V(\hat{\hat{Y}}) = V(E)$, where $E = \sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}[y_{hi} - (\mathbf{x}_{hi}^{\mathbf{T}}, \tilde{v}_{hi})\mathbf{c}] \Big/ \sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}x_{hi}^{*} = N^{-1}\sum_{h=1}^{H}\sum_{i=s_h}\tilde{v}_{hi}[y_{hi} - (\mathbf{x}_{hi}^{\mathbf{T}}, \tilde{v}_{hi})\mathbf{c}]$ for any vector $\mathbf{c}$, and therefore

$$Bias[\hat{V}(\hat{\hat{Y}})] = \frac{1}{N^2}\sum_{g=1}^{G}(t_{E,g1} - t_{E,g2})^2 \ . \tag{4}$$

Let $\mathbf{u}_{hi} = (u_{1,hi},...,u_{G,hi})$ be a vector of variance strata indicators, $u_{j,hi} = 1$ if CU belongs to the variance stratum {j1,j2} and 0 otherwise, and $\mathbf{l}_{hi} = (l_{1,hi},...,l_{H/2,hi})$ where, $l_{j,hi} = 1$ if CU belongs to the original stratum ST{j1} and 0 otherwise, and consider a linear regression model on the population,

$$e_{hi} = \mathbf{u}_{hi}^{\mathbf{T}}\mathbf{d} + \mathbf{l}_{hi}^{\mathbf{T}}\mathbf{a} + \varepsilon_{hi} \ , \tag{5}$$

where $e_{hi} = y_{hi} - (\mathbf{x}_{hi}^{\mathbf{T}}, \tilde{v}_{hi})\mathbf{c}$.

In the following study we define e_{hi} as the residuals of OLS regression, $y_{hi} = (\mathbf{x}_{hi}^{\mathbf{T}}, \tilde{v}_{hi})\mathbf{c} + e_{hi}$, y-s against x-s, although as it was noted earlier, any constant $\mathbf{c}$ can be used, for example, one can use probability weighted linear regression coefficient. If (5) is a true population model then, by (3) and (4),

$$Bias[\hat{V}(\hat{\hat{Y}})] = \frac{1}{N^2}\sum_{g=1}^{G}(t_{E,g1} - t_{E,g2})^2 = \frac{1}{N^2}\sum_{g=1}^{G}[(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\mathbf{d} + \mathbf{t_{l,g1}^{T}a}]^2 \ , \tag{6}$$

where $\mathbf{t_{u,g1}} = \mathbf{t_{l,g1}}$ is total population in the original strata ST(g1) and $\mathbf{t_{l,g2}}$ is total population in the original strata ST(g2). Recall that CE collapses only geographically close X-, Y- and Z-PSUs into pairs of variance PSU's such that population totals in respective variance PSUs are as close as possible, so we can expect that $\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}}$ is relatively small with respect to the standard error of the mean expenditure estimator. That is why we can assume that $\frac{1}{N^2}\sum_{g=1}^{G}[(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\mathbf{d}]^2$ does not induce essential bias, and therefore, the main part of the bias (if it is significant) is $\frac{1}{N^2}\sum_{g=1}^{G}[\mathbf{t_{l,g1}^{T}a}]^2$. The significance of $\frac{1}{N^2}\sum_{g=1}^{G}[\mathbf{t_{l,g1}^{T}a}]^2$ can be checked by testing whether the regression coefficient $\mathbf{a}$ in model (5) is significant which again can be checked by F-test. We applied classical regression F-test to model (5) (for 157 expenditures with reported number of non-zero expenditures greater than 300, again, as in previous section, we applied Shapiro-Wilk test to the residuals of model $e_{hi} = \mathbf{u}_{hi}^{\mathbf{T}}\mathbf{d} + error$ and for most considered expenditures the Normality hypothesis were not rejected). The results of the F-test are as follows: for 83 of the expenditures $H_0 : \mathbf{a} = \mathbf{0}$ were not rejected on 0.1 significance level, for 106 CEs H_0 were not rejected on 0.01 significance level, and for 17 CE's the hypothesis were rejected on 0.01 significance level.

COMMENT. The above testing procedure assumes essentially the calibration property that can be violated for the estimators in sub-areas of the nation. We suggest the following similar testing procedure that does not require the calibration property and thus can be applied for sub-national estimators. Consider the following regression model, $y_{hi}^{A} = \delta_{hi}^{A}\mathbf{u}_{hi}^{\mathbf{T}}\tilde{\mathbf{d}} + \delta_{hi}^{A}\mathbf{l}_{hi}^{\mathbf{T}}\tilde{\mathbf{a}} + e\tilde{rr}or$, where δ_{hi}^{A} =1 if unit (h,i) belongs to sub-area *A* and 0 otherwise. If the vector coefficient $\tilde{\mathbf{a}}$ is not significant in this model (which can be tested similarly to the above by F-test) then the differences between the totals in the pairs of PSUs that formed the variance strata are insignificant and therefore $Bias[\hat{V}(\hat{\hat{Y}})] = \frac{1}{N^2}\sum_{g=1}^{G}(t_{y,g1} - t_{y,g2})^2$ is not essential.

4. Additional analysis of the expected bias of collapse BRR variance estimator. The above tests make possible to check whether the expected bias is significant or not. Note that the use of F-test is based on a number of assumptions like independence and normality (at least asymptotically) of the residuals of respective regression models and although in our study we did not get any evidence of strong violation of these assumptions we know that some violations are presented. For example, the variable of interest, the expenditure, is truncated at 0 and therefore even residuals of the model (5) can not have pure central distribution and thus some violation of normality is inevitable. We also neglected by $(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\mathbf{d}$ **-** part of (6) assuming that the differences between the totals in the variance strata, $\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}}$, are relatively small without checking how practically small are they. Therefore it would be of interest to suggest some practical statistics, such as an upper estimate of possible relative expected bias, that have simple practical interpretation. The latter can be obtained from model (5) and representation of the bias (6). Actually, since the totals $\mathbf{t_{u,g1}} = \mathbf{t_{l,g1}}$ and $\mathbf{t_{l,g2}}$ are known, one can construct a naive estimate of possible bias as

$$\hat{B}[\hat{V}(\hat{\hat{Y}})] = \frac{1}{N^2}\sum_{g=1}^{G}[(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\hat{\mathbf{d}} + \mathbf{t_{l,g1}^{T}}\hat{\mathbf{a}}]^2 \ , \tag{7}$$

with $\hat{\mathbf{d}}$ and $\hat{\mathbf{a}}$ be consistent estimators of $\mathbf{d}$ and $\mathbf{a}$. Note that this naïve estimator overestimates the true bias,

$$E\{\hat{B}[\hat{V}(\hat{\hat{Y}})]\} \cong \frac{1}{N^2}\sum_{g=1}^{G}[(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\mathbf{d} + \mathbf{t_{l,g1}^{T}}\mathbf{a}]^2 + \frac{1}{N^2}\sum_{g=1}^{G}V[(\mathbf{t_{u,g1}} - \mathbf{t_{u,g2}})^{\mathbf{T}}\hat{\mathbf{d}} + \mathbf{t_{l,g1}^{T}}\hat{\mathbf{a}}], \tag{8}$$

and therefore (7) is conservative estimate. Finally, the relative bias can be estimated by $\hat{B}[\hat{V}(\hat{\hat{Y}})]/\hat{V}(\hat{\hat{Y}})$.

The following two tables summarize Relative Bias statistics and H_0 rejection for a set of questionable items.

Table 1. Estimates of Relative Bias in percents, RelBias= $\hat{B}[\hat{V}(\hat{\hat{Y}})]/\hat{V}(\hat{\hat{Y}})\times 100$, and F-test p-values for 17 expenditures for which H_0 were rejected on 0.1 significance level (n=number of reported non-zero expenditures).

Item	RelBias (%)	p-value	N
Rice	8.6%	0.0001	753
Fresh fish and shellfish	5.6%	0.0002	765
White bread	12.0%	0.0007	3620
Roasted coffee	16.0%	0.0007	1250
Fresh fruit juice	12.0%	0.001	767
Infant dresses, outerwear	7.1%	0.001	311
Instant and freeze dried coffee	12.2%	0.002	1009
Bananas	11.0%	0.003	3142
Snacks and nonalcoholic beverages at fast food	3.6%	0.003	5060
Other laundry cleaning products	13.7%	0.003	3008
Beer and Ale	6.4%	0.004	1622
Natural gas	13.7%	0.005	732
Tomatoes	10.8%	0.006	1936
Potato chips and other snacks	6.5%	0.007	5271
School supplies, etc.	17.3%	0.008	1249
Electricity	15.0%	0.009	1491
Ready-to-eat and cooked cereals	10.4%	0.0096	3652

Table 2. Estimates of Relative Bias in percents, RelBias= $\hat{B}[\hat{V}(\hat{\hat{Y}})]/\hat{V}(\hat{\hat{Y}})\times 100$, and F-test p-values for 29 expenditures for which RelBias is greater than 10% (n=number of reported non-zero expenditures).

Name	RelBias (%)	p-value	N
Lawn and garden supplies	27.1%	0.025	1076
(*) School supplies, etc.	17.3%	0.008	1249
(*) Roasted coffee	16.0%	0.0007	1250
Sausage	15.3%	0.014	1112
Peanut butter	15.3%	0.130	656
(*) Electricity	15.0%	0.009	1491
Misc. auto repair, servicing	15.0%	0.037	516
Bologna, liverwurst, salami	14.7%	0.097	1153
Jams, preserves, other sweets	14.3%	0.148	1407
Frankfurters	13.9%	0.224	1098
(*) Other laundry cleaning products	13.7%	0.003	3008
(*) Natural gas	13.7%	0.005	732
Other steak	13.1%	0.023	900
Frozen and refrigerated bakery products	12.7%	0.052	424
Closet and storage items	12.6%	0.142	301
Crackers	12.5%	0.012	1890
(*) Instant and freeze dried coffee	12.2%	0.002	1009
Sugar	12.2%	0.313	1080
Women's hosiery	12.0%	0.044	393
(*) White bread	12.0%	0.0007	3620
(*) Fresh fruit juice	12.0%	0.001	767
Fresh and frozen chicken parts	11.8%	0.024	2129
Other poultry	11.5%	0.127	687
Canned miscellaneous vegetables	11.4%	0.117	1789
(*) Bananas	11.0%	0.003	3142
Men's underwear	11.0%	0.044	309
Butter	10.8%	0.067	966
(*) Tomatoes	10.8%	0.006	1936
(*) Ready-to-eat and cooked cereals	10.4%	0.0096	3652

(*) – if RelBias greater than 10% and H_0 were rejected on 0.1 significance level.

Conclusions.

1. For most of the expenditures both statistics, estimates of the relative bias and p-values of F-test statistics, do not indicate essential potentional overestimation of the variances by BRR estimator based on collapsed variance PSUs.

2. At least for 11 items marked by (*) in Table 2 and “Lawn and garden supplies” item one can expect overestimation of the variances by 10 - 20%, although as it was noted earlier, it can be result of conservatism of the estimator define by (7), see (8) for the details.

3. Note that we consider the items with not less than 300 non-zero expenditures reported. Although for items with less reported expenditures BRR variance estimator could be approximately unbiased we do not expect that it would be stable enough since a lot of variance PSUs contain in this case only 0-reported expenditures, therefore we do not consider these items in the present paper.

4. The above analysis can be repeated for the estimators in the sub-areas (following the above Comments 1 and 2), although one has keep in mind that the number of non-zero reported items for sub-area could be to small for use BRR-type or other repeated sample type estimators.

References

Deville, J. C. and Särndal, C. E. (1992), Calibration Estimators in Survey Sampling, *JASA*, v. 87, No. 418, pp. 376-382.

Ernst, L. R., Johnson, W. H., and Larson, William E. (2007), Reduction of Sample Areas in the Consumer Price Index and Consumer Expenditure Survey Designs, Proceedings of the Joint Statistical Meetings

King, S. L., and Johnson-Herring, S. A. (2007), Selecting a Sample of Households for the Consumer Expenditure Survey. The Design and Selection of the Survey's Sample, *Consumer Expenditure Survey Anthology.*

Pfeffermann, D., Krieger, A. M. and Rinott, Y. (1998), "Parametric distributions of complex survey data under informative probability sampling," *Statistica Sinica* 8, 1087-1114.

Rust, K., and Kalton, G. (1987) Strategies for Collapsing Strata for Variance Estimation, Journal of Official Statistics, 3, No. 1, pp. 69-81

Sverchkov, M., Dorfman, A. H., Ernst, L. R., and Guciardo, C. J. (2004), Zero-Calibrated Variance Estimators., *2004 JSM Meetings, Proceedings of the Section on Survey Methods Research*

Sverchkov, M., Dorfman, A. H., Ernst, L. R., Moerhle, T. G., Paben, S. P. and Ponikowski, C. H. (2005) On Non-response Adjustment via Calibration, *2005 JSM Meetings, Proceedings of the Section on Survey Methods Research*

Wolter, K. M. (1985). *Introduction to variance estimation*, Springer-Verlag, New York

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.