Evaluating the First-Stage Sample Design for the 2020 Redesign of the Consumer Expenditures Surveys October 2022

Stephen Ash
Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

Abstract
This paper discusses methods for evaluating the first-stage sample design of complex sample designs. We need these methods because the Bureau of Labor Statistics is in the process of redesigning the 2020 sample design of the Consumer Expenditures Surveys. The research is important because we only have the opportunity to redesign the first stage every ten years. The paper provides expressions for the variance of the collapsed strata estimator and the with replacement variance estimator and uses those expressions to evaluate alternative sample designs for the 2020 redesign of the Consumer Expenditures Surveys.

Key Words: *Two-stage sample design, variance estimation, balanced repeated replication variance estimator, collapsed-strata variance estimator*

1. Introduction

The Consumer Expenditure Survey (CE) collects data on the buying habits of U.S. consumers. The CE data are used to produce estimates of household expenditures in the U.S; it is used by researchers in government, business, labor, and academic analysts; and it is used with the periodic revision of the Consumer Price Index (CPI). The survey is conducted by the U.S. Census Bureau for the Bureau of Labor Statistics and consists of two components: an interview survey in which expenditures of consumer units[1] (CUs) are obtained in four interviews conducted at three-month intervals and a diary survey completed by participating consumer units for two consecutive one-week periods.

Historically, the sample design of CE has been redesigned every ten years. The timing of the redesign is shortly after the decennial census which allows the redesign to make use of the results from the most recent census.

The 2020 redesign will be unique as compared with prior redesigns: it will have two new goals. First, the sample design should produce a sample of PSUs that are geographically close to each other. This new goal is an attempt to control costs: it should help reduce costs because interviewers from nearby PSUs should be able to assist in other PSUs with shorter

[1] The unit from which the CE seeks to collect expenditure data is a "consumer unit." A CU is a group of people living together who are (1) related by blood, marriage, adoption, or some other legal arrangement such as foster children; (2) unrelated but who pool their incomes to make joint expenditure decisions; or (3) is a person living alone or sharing a housing unit with other people but who is financially independent of the other people. Approximately 99 percent of all occupied housing units have one CU, hence the terms "household" and "consumer unit" are often used interchangeably."

travel. Second, the sample design should produce improved state-level estimates for California, Florida, New York, and Texas.

Our research examines the impact of the preliminary 2020 sample design on variances and the estimates of variances. Our specific objectives correspond to the following three sets of analyses:

1) Compare the bias and variances of the 2010 and 2020 sample designs with respect to the estimate of mean total expenditures – the variable of interest for CE.
2) Optimize the CE's Balanced Repeated Replication (BRR) variance estimator with respect to bias and variance.
3) Explore how we can estimate the variance of the CE's BRR variance estimator.

Before we consider the objectives further, we begin by reviewing the sample design of CE in section 2 and variance estimation for CE in section 3. Section 4 describes the simulation, section 5 reviews the analysis of the three objectives using the simulation, and section 6 provides our conclusions.

2. Sample Design of CE

In this background section, we review the aspects of the CE sample design that are needed to understand our research. Since the 2010 and 2020 sample designs will share many features, we begin by describing the 2010 sample design. For the 2020 sample design, we only highlight the new features, instead of repeating the aspects of the sample design that will not change between 2010 and 2020. For a more detailed description of the 2010 sample design, see Neiman et al. (2015). Although the sample design is shared with CPI, CPI is not in the scope of our research.

2.1 2010 Sample Design

CE employs a two-stage sample design. In the first stage, both the Interview and Diary Surveys use the same first-stage sample design. In the first stage, the counties of the U.S. are grouped into Primary Sample Units (PSUs), which are single counties or groups of contiguous counties. In urban areas, PSUs were defined as the Core-Based Statistical Areas (CBSA) as defined by the Office of Management and Budget that use the results of the most recent decennial census. The PSUs were then grouped into strata that are either self-representing (SR) or non-self-representing (NSR).

In 2010, the SR PSUs included the 23 largest metropolitan areas and are selected with certainty. NSR PSUs were grouped at two levels: a high-level stratification and a low-level stratification within the high level. At the high level, NSR PSUs were grouped by urban and rural status (N-size and R-size, respectively) and the nine Census Divisions. The NSR PSUs were further grouped into low-level strata that maximize the homogeneity of a set of variables that are associated with total expenditures. Additionally, the low-level stratification was constrained to produce approximately equal-size strata with respect to the Measure of Size (MOS) within the high-level stratification. For more details on the 2010 stratification, see King, Schilp, and Bergmann (2011).

One PSU was selected within each stratum with probability proportion to the MOS, where the MOS is the most recent decennial census population estimates. The selection procedure maximizes the overlap between the previous and current first-stage samples. Maximizing the overlap reduces costs by reducing the number of 2010 PSUs that need to replace 2000 PSUs which reduces the number of field staff that that need to be replaced.

In the second stage of the sample design, the address frame, maintained by the Census Bureau, was stratified and a sample of addresses was selected with equal probability within each of the first-stage sample PSUs. The allocation of the second-stage sample sizes is completed separately for the Interview and Diary surveys and is generally proportional allocation but with limits on minimum sample sizes. For more details on the 2010 allocation, see Ash et al. (2012).

2.2 2020 First-Stage Sample Design

At the time of our research, we are in the middle of developing the 2020 sample design and not all decisions have been finalized. Therefore, what we describe presently is not final, but we use it with our research.

In addition to all the features described for the 2010 sample design, the 2020 sample design has been asked to do more. The first new goal is to produce a sample of PSUs that are geographically close so that interviewers from different PSUs can assist in other PSUs. To accomplish this goal, the high-level strata will be expanded. The N-size NSR PSUs will be further divided into three substrata: (1) PSUs with population greater than 200,000, (2) PSUs with populations less than 200K and within 20 miles of a Stage 1 sample PSU, and (2R) remote PSUs. Urban Remote PSUs or (2R) will be identified as being 20 miles from any SR PSU or Substrata (1) sample PSU.

NSR Rural PSUs will be further divided into two substrata that will be defined after we select the NSR Urban sample. The PSUs of substrata (3) include PSUs being within 20 miles from any SR PSU or previously selected NSR Urban PSU. The rural remote PSUs or (3R) include PSUs that are outside of 20 miles of any previously selected PSU. Table 1 summaries the expanded urban/rural status for the high-level stratification of the 2020 sample design.

Table 1: Preliminary Urban/Rural Status for the High-Level Stratification of the 2020 Sample Design

High-Level Strata		Definition	
Type of PSU	Substrata	CBSAs with population…	… and...
SR		> 2.8 M	Honolulu and Anchorage
NSR Urban	(1)	in [200 K, 2.8M)	
	(2)	< 200 K	≤ 20 miles from an SR or Stage 1 sample PSU
	(2R) Remote	< 200 K	> 20 miles from an SR or Stage 1 sample PSU
NSR Rural	(3)	Not a CBSA	≤ 20 miles from an SR, Stage 1 or Stage 2 sample PSU
	(3R) Remote	Not a CBSA	> 20 miles from an SR, Stage 1, or Stage 2 sample PSU

The second new goal of the 2020 sample design is to improve state-level estimates for the selected states California, Florida, New York, and Texas. To accomplish this goal, the high-level strata will be defined by the combination of the boundaries of the nine Census Divisions and the four selected states.

Within the high-level strata, PSUs will be further grouped into low-level strata with a data science algorithm. The Expectation-Maximation algorithm will group the PSUs into strata that maximizes the homogeneity of four variables related to total expenditures: percent households (HHs) with income greater than or equal to $75,000, percent of HHs with a computer, percent HHs with a Bachelor degree or more education, and percent of the population urban. The four variables will be generated from the American Community Survey. Although this may change, the algorithm will not employ any constraints for equal size strata as was done in the 2010 redesign,.

3. Variances and Variance Estimation for CE

3.1 Variance Estimation for CE

Because there is no unbiased estimator of the variance for sample designs that select one unit per strata, all variance estimators produce a certain amount of bias. CE's solution is to collapse the NSR strata into pseudo strata and then apply balanced repeated replication (BRR) [McCarthy 1966] to the pseudo strata. With SR PSUs, each PSU is partitioned into two approximately equal parts, except very large SR PSUs, which are partitioned into four approximately equal parts; the parts are collapsed into pseudo strata; and then BRR is applied to the pseudo strata. See also Swanson (2017) for more details on variance estimation for CE.

For our research, the estimation of variances for SR strata is out of scope because the current plan is to have the same set of SR PSUs and the estimation of variances for SR strata is a much different problem. See Ash (2014) for more about estimating variances for SR PSUs. We are more interested in the NSR PSUs and the impact of the first-stage stratification and estimators of the variance; therefore, the SR PSUs will be excluded in most of the analyses of our paper.

3.2 CE's use of BRR with NSR PSUs

With CE's use of BRR, there are two choices that impact the bias and variance of the BRR variance estimator. First, we can use the replicate factors described by Judkins (1990) that mitigate the impact of unequal strata within a pseudo stratum. Second, we can minimize the bias and variance with our choice of criteria for collapsing strata into pseudo strata. To minimize the bias and variance, we need expressions of the bias and variance. The key to the expressions of variance and bias of the BRR estimator is understanding that BRR is mathematically equivalent to the collapsed-strata (*cs*) estimator. Then expressions for the bias and variance of the *cs* estimator are also expressions of the bias and variance of the BRR estimator. Before we discuss BRR and the *cs* estimators in more detail, we define some basic notations that will be used throughout the paper.

The two-stage stratified total of some characteristic of interest Y_i is defined as:

$$Y = \sum_{h} \sum_{i \in U_h} Y_i$$

where h is an index on the first-stage strata, i is an index on the first-stage units or PSUs, and U_h is the set of PSUs in first-stage stratum h. We define the PSU total $Y_i = \sum_{k \in U_{hi}} y_k$

where k is an index on the second-stage CUs of U_h, and U_{hi} is the set of second-stage CUs in PSU i of stratum h.

Sometimes the statistic of interest is the ratio of two totals or $R = Y / X$, where X is a different total defined analogous to Y. With CE, the statistic of interest is mean total expenditures and we define y_k as the total expenditures for CU k and x_k as an indicator variable that identifies whether the sample unit is an eligible CU or not: $x_k = 1$ for eligible CU k and $x_k = 0$ for ineligible sample CU k. The two-stage stratified with replacement estimator (*wr*) of Y with a first-stage sample size of n_h is:

$$\hat{Y} = \sum_h \frac{1}{n_h} \sum_{i \in s_h} \frac{\hat{Y}_i}{p_i}$$

where $\hat{Y}_i$ is the estimator of the total of Y_i for PSU i, s_h is the first-stage sample for stratum h, n_h is the first-stage sample size for first-stage stratum h, and p_i is the select one unit probability of selection for PSU i within first-stage stratum h. For CE, we have the special case of $n_h = 1$ for all strata h.

Variance Estimation with Balanced Repeated Replication. Using Fay's method of BRR, as described by Dippo, Fay, and Morganstein (1984) and Judkins (1990), we can express the BRR variance estimator of $\hat{Y}$ as:

$$\hat{v}_{BRR}(\hat{Y}) = \frac{1}{R(\kappa - 1)^2} \sum_r \left(\hat{Y}_r - \hat{Y}\right)^2,$$

where R is the number of replicates, r is the index for the replicates, and κ has values $0 \leq \kappa < 1$. The replicate estimates for each replicate r are defined as
$\hat{Y}_r = \sum_g \left(\hat{Y}_{r,g,h=1} + \hat{Y}_{r,g,h=2}\right)$ and $\hat{Y}_{r,g,h}$ is estimated as:

$$\hat{Y}_{r,g,h} = \sum_{k \in s_{hi}} w_i w_k F_{r,g,h} y_k.$$

where g is the index on the pseudo strata. For CE, the first-stage weight w_i for PSU i is equal to the inverse of the first-stage probability of selection p_i for PSU i or $w_i = p_i^{-1}$. Within each PSU, the second-stage weight is w_k. We define the replicate factors $F_{r,g,h=1}$ as generally as possible. Following Judkins (1990), the replicate factor of replicate r, pseudo strata g, and stratum $h = 1$ is defined as:

$$\mathrm{F}_{r,g,h=1} = 1 + 2a_{rg}(1 - \kappa)P_{g,h=2} \quad (1)$$

and the replicate factor of replicate r, pseudo strata g, and stratum $h = 2$ is defined as:

$$\mathrm{F}_{r,g,h=2} = 1 - 2a_{rg}(1 - \kappa)P_{g,h=1}. \quad (2)$$

We will see later that different values of P_{gh} can be used to minimize the bias and/or variance of the estimator. Also note that $P_{g,h=2} = 1 - P_{g,h=1}$. The covariance generalization of the BRR estimator is:

$$\widehat{cov}_{BRR}(\hat{X}, \hat{Y}) = \frac{1}{R(\kappa - 1)^2} \sum_r (\hat{X}_r - \hat{X})(\hat{Y}_r - \hat{Y}),$$

where $\hat{X}_r$ is defined analogous to $\hat{Y}_r$.

3.3 Collapsed-Strata Estimator

Hansen , Hurwitz, and Madow (1953; chapter 9, section 15) and Wolter (1985; section 2.5) discuss a *cs* variance estimator that collapses multiple strata into a pseudo stratum. We limit our discussion to only collapsing two PSUs – no triples or multiple strata. Therefore, we define our *cs* variance estimator specialized for two collapsed strata and using the generally defined replicate factors (1) and (2) as:

$$\hat{v}_{cs}(\hat{Y}) = 4 \sum_g \left(P_{g2}\hat{Y}_{g1} - P_{g1}\hat{Y}_{g2}\right)^2$$

and covariance generalization as:

$$\widehat{cov}_{cs}(\hat{X}, \hat{Y}) = 4 \sum_g \left(P_{g2}\hat{X}_{g1} - P_{g1}\hat{X}_{g2}\right)\left(P_{g2}\hat{Y}_{g1} - P_{g1}\hat{Y}_{g2}\right).$$

We are interested in the *cs* variance and covariance estimators because they are equivalent to the BRR variance and covariance estimators (Ash 2022b). So, if we know the expectation, bias, and variance of the *cs* estimator, we also know the expectation, bias, and variance of the BRR estimator. The expectation and the bias of the *cs* estimator are:

$$\begin{aligned} E\left(\hat{v}_{cs}(\hat{Y})\right) &= 4 \sum_g \left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)^2 + 4 \sum_g \left(P_{g2}^2 v(\hat{Y}_{g1}) + P_{g1}^2 v(\hat{Y}_{g2})\right) \\ &= v(\hat{Y}) + bias\left(\hat{v}_{cs}(\hat{Y})\right) \end{aligned}$$

and

$$bias\left(\hat{v}_{cs}(\hat{Y})\right) = 4 \sum_g \left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)^2 - \sum_g \left((1 - 4P_{g2}^2)v(\hat{Y}_{g1}) + (1 - 4P_{g1}^2)v(\hat{Y}_{g2})\right).$$

Note that the expectation and bias can be applied to one stage, two stages, or multiple stages as long as the variances $v(\hat{Y}_{g1})$ and $v(\hat{Y}_{g2})$ reflect the same number of stages. The expectation and bias of the *cs* covariance are:

$$\begin{aligned} E\left(\widehat{cov}_{cs}(\hat{Y}, \hat{X})\right) &= 4 \sum_g \left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)\left(P_{g2}X_{g1} - P_{g1}X_{g2}\right) + 4 \sum_g \left(P_{g2}^2 cov(\hat{X}_{g1}, \hat{Y}_{g1}) + P_{g1}^2 cov(\hat{X}_{g2}, \hat{Y}_{g2})\right). \\ &= cov(\hat{Y}, \hat{X}) + bias\left(\widehat{cov}_{cs}(\hat{Y}, \hat{X})\right) \end{aligned}$$

and

$$bias\left(\widehat{cov}_{cs}(\hat{Y}, \hat{X})\right) = 4 \sum_g \left(P_{g2}X_{g1} - P_{g1}X_{g2}\right)\left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right) - \sum_g \left((1 - 4P_{g2}^2)cov(\hat{Y}_{g1}, \hat{X}_{g1}) + (1 - 4P_{g1}^2)cov(\hat{Y}_{g2}, \hat{X}_{g2})\right).$$

See Ash (2022b) for the derivation of $E\left(\hat{v}_{cs}(\hat{Y})\right)$, $bias\left(\hat{v}_{cs}(\hat{Y})\right)$, $E\left(\widehat{cov}_{cs}(\hat{Y},\hat{X})\right)$, and $bias\left(\widehat{cov}_{cs}(\hat{Y},\hat{X})\right)$. See also Wolter (2007; section 2.5) and Judkins (1990) for alternative expressions of $E\left(\hat{v}_{cs}(\hat{Y})\right)$. The variance of the *cs* variance estimator are:

$$v\left(\hat{v}_{cs}(\hat{Y})\right) = A_1 + B_1 + C_1 + D_1$$

where:

$$A_1 = 16\sum_g \left(P_{g2}^4\left(\mu_{g1}^{(4)} - \sigma_{g1}^4\right) + P_{g1}^4\left(\mu_{g2}^{(4)} - \sigma_{g2}^4\right)\right),$$
$$B_1 = 64\sum_g \left(\left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)^2\left(P_{g2}^2\sigma_{g1}^2 + P_{g1}^2\sigma_{g2}^2\right)\right),$$
$$C_1 = 64\sum_g \left(\left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)\left(P_{g2}^3\mu_{g1}^{(3)} - P_{g1}^3\mu_{g2}^{(3)}\right)\right),$$
$$D_1 = 64\sum_g P_{g1}^2P_{g2}^2\sigma_{g1}^2\sigma_{g2}^2,$$

and where we define:

$$\mu_{gh}^{(a)} = E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^a\right),$$
$$\sigma_{gh}^2 = v\left(\hat{Y}_{gh}\right) = E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^2\right),$$
$$\sigma_{gh}^4 = \left(\sigma_{gh}^2\right)^2.$$

Rust (1984) and Rust and Kalton (1987) have a more general result for collapsing multiple strata, but it does not include P_{gh}. Wolter (2007; p. 53) provides a different expression for the variance for collapsing multiple strata which includes P_{gh}. The variance of the *cs* covariance is:

$$v\left(\widehat{cov}_{cs}(\hat{Y},\hat{X})\right) = A_2 + B_2 + C_2 + D_2,$$

where:

$$A_2 = 16\sum_g \left(P_{g2}^4\left(\phi_{g1}^{(2,2)} - \left(cov\left(\hat{Y}_{g1},\hat{X}_{g1}\right)\right)^2\right) + P_{g1}^4\left(\phi_{g2}^{(2,2)} - \left(cov\left(\hat{Y}_{g2},\hat{X}_{g2}\right)\right)^2\right)\right),$$
$$B_2 = 64\sum_g \left(\left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)\left(P_{g2}X_{g1} - P_{g1}X_{g2}\right)\left(P_{g2}^2 cov\left(\hat{Y}_{g1},\hat{X}_{g1}\right) + P_{g1}^2 cov\left(\hat{Y}_{g2},\hat{X}_{g2}\right)\right)\right),$$
$$C_2 = 64\sum_g \left(\left(P_{g2}Y_{g1} - P_{g1}Y_{g2}\right)\left(P_{g2}X_{g1} - P_{g1}X_{g2}\right)\left(P_{g2}^3\left(\phi_{g1}^{(1,2)} + \phi_{g1}^{(2,1)}\right) - P_{g1}^3\left(\phi_{g2}^{(1,2)} + \phi_{g2}^{(2,1)}\right)\right)\right),$$
$$D_2 = 64\sum_g \left(P_{g2}^2P_{g1}^2 cov\left(\hat{Y}_{g1},\hat{X}_{g1}\right)cov\left(\hat{Y}_{g2},\hat{X}_{g2}\right)\right)^4,$$

and where we define:

$$\phi_{gh}^{(a,b)} = E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^a\left(\hat{X}_{gh} - X_{gh}\right)^b\right),$$
$$cov\left(\hat{Y}_{gh},\hat{X}_{gh}\right) = \phi_{gh}^{(1,1)} = E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)\left(\hat{X}_{gh} - X_{gh}\right)\right).$$

See Ash (2022b) for the derivation of $v\left(\hat{v}_{cs}(\hat{Y})\right)$ and $v\left(\widehat{cov}_{cs}(\hat{Y},\hat{X})\right)$.

Unlike BRR, the *cs* estimator only applies to totals. However, we can apply the *cs* estimator to a ratio estimator with linearization. We linearize the one-stage estimator for a ratio estimator $\hat{R} = \hat{Y}/\hat{X}$ as $\hat{Z}_i = \left(Y_i - \hat{R}X_i\right)/\hat{X}$. For the two-stage estimator, $\hat{Z}_i = \left(\hat{Y}_i - \hat{R}\hat{X}_i\right)/\hat{X}$ for the first stage and $z_k = \left(y_k - \hat{R}x_k\right)/\hat{X}$ for the second stage.

3.4 Choices for the BRR and Collapsed-Strata Estimator

In applications of BRR and *cs* to sample designs with $n_h = 1$, there are two choices that we need to make: (1) how to define P_{gh} and (2) what criteria should we use for collapsing the strata into pseudo strata?

(1) How to choose P_{gh}? The simplest choice is $P_{gh} = 1/2$. This is reasonable because it makes the second term of $bias(\hat{v}_{cs}(\hat{Y}))$ equal to zero. A second option is to choose P_{gh} proportional to the MOS, which reduces the first term of $bias(\hat{v}_{cs}(\hat{Y}))$ when MOS is proportional to the strata totals Y_{gh} (Wolter 2007; p. 52). This can be expressed as $P_{g1} = \text{MOS}_{g1} / \text{MOS}_g$ and $P_{g2} = \text{MOS}_{g2} / \text{MOS}_g$, where MOS_g is the MOS for pseudo stratum g and MOS_{gh} is the MOS for stratum h of pseudo stratum g. Note that $\text{MOS}_g = \text{MOS}_{g1} + \text{MOS}_{g2}$. For CE, the MOS is the population count from the most recent Decennial Census.

The value of P_{g1} that minimizes $bias(\hat{v}_{cs}(\hat{Y}))$ is:

$$P_{g1,min\,bias} = \frac{Y_{g1}Y_g + v\left(\hat{Y}_{g1}\right)}{Y_g^2 + v\left(\hat{Y}_g\right)}$$

and $P_{g2,min\,bias} = 1 - P_{g1,min\,bias}$.

The value of P_{g1} that minimizes $\left(bias(\hat{v}_{cs}(\hat{Y}))\right)^2$ is:

$$P_{g1,min\,bias^2} = \begin{cases} P_{g1,min\,bias} \pm q & r < 0 \\ P_{g1,min\,bias} & r \geq 0 \end{cases}$$

where:

$$q = \frac{\sqrt{4\left(Y_{g1}Y_g + v\left(\hat{Y}_g\right)\right)^2 - \left(Y_g^2 + v\left(\hat{Y}_g\right)\right)\left(4Y_{g1}^2 - \left(v\left(\hat{Y}_g\right) - 4v\left(\hat{Y}_{g1}\right)\right)\right)}}{2\left(Y_g^2 + v\left(\hat{Y}_g\right)\right)}$$

and

$$r = \left(1 - 4P_{g2,min\,bias}^2\right)v\left(\hat{Y}_{g1}\right) + \left(1 - 4P_{g1,min\,bias}^2\right)v\left(\hat{Y}_{g2}\right).$$

Note that $P_{g2,min\,bias^2} = 1 - P_{g1,min\,bias^2}$. When $r < 0$ and there are two solutions, we suggest using the value of $P_{g1,min\,bias} \pm q$ that is closest to ½ because the second term of the bias is equal to zero when $P_{gh} = ½$. See Ash (2022b) for the derivation of $P_{g1,min\,bias}$ and $P_{g1,min\,bias^2}$.

Note that both $P_{g1,min\,bias}$ and $P_{g1,min\,bias^2}$ may be difficult to calculate in practice because they require good information about the strata totals Y_{gh} and strata variances $v(\hat{Y}_{gh})$.

(2) What criteria should we use for collapsing strata into pseudo strata? There are two reasonable choices: collapsing strata that are similar with respect to the variable of interest or MOS. Collapsing strata with respect to the variable of interest minimizes the first term of $bias(\hat{v}_{cs}(\hat{Y}))$. This is important because the first term of $bias(\hat{v}_{cs}(\hat{Y}))$ usually contributes the most to the overall bias. It is also reasonable to collapse with respect to MOS because, when done in conjunction with defining P_{gh} proportional to MOS, the second term of $bias(\hat{v}_{cs}(\hat{Y}))$ is zero or close to zero.

If you do not have good information about the strata totals Y_{gh} and strata variances $v(\hat{Y}_{gh})$, the best scenario may be to start with equal-size strata, collapse strata with respect to the variable of interest and use P_{gh} proportional to MOS. Then, the first term of the bias is minimized by collapsing strata with respect to the variable of interest and the second term of the bias is zero or close to zero because $P_{gh} = ½$ due to the equal-size strata.

3.5 Estimating the Variance of the *wr* Variance Estimator

We wanted to go one step further and estimate the variance of BRR with a sample. The *cs* estimator of the previous section is most appropriate; however, it has no unbiased estimator of the variance. For this reason, we took a different route and considered the *wr* estimator because it is appropriate for sample designs with a sample size of $n_h = 1$. However, it has its own drawbacks, as we shall see. Before we discuss those drawbacks, we first define the *wr* estimator for one stage with two different approaches – by assuming independence and assuming a multinomial distribution. We shall see that the two approaches often produce the same or a very similar result for the *wr* estimator; however, we will also see that they sometimes produce different expressions for the same result.

Assuming Independence. We can express the *wr* estimator of the total Y for a single stage as:

$$\hat{Y}_{ind} = \frac{1}{n}\sum_{i\in s}\frac{Y_i}{p_i}.$$

The independence assumption says that the Y_i/p_i are independent random variables with no explicit distribution, but $E(Y_i/p_i) = Y$. We add the subscript *ind* to $\hat{Y}_{ind}$ to indicate that the *wr* estimator uses the independence assumption. The variance of $\hat{Y}_{ind}$ is:

$$v(\hat{Y}_{ind}) = \frac{1}{n}\sum_{i\in U} p_i\left(\frac{Y_i}{p_i} - Y\right)^2.$$

We can estimate $v(\hat{Y}_{ind})$ as:

$$\hat{v}(\hat{Y}_{ind}) = \frac{1}{n(n-1)}\sum_{i\in s}\left(\frac{Y_i}{p_i} - \hat{Y}\right)^2$$

and the variance of $\hat{v}(\hat{Y}_{ind})$ can be expressed as

$$v\left(\hat{v}(\hat{Y}_{ind})\right) = \frac{1}{n^3}\left(\mu_4 - \left(\frac{n-3}{n-1}\right)\sigma^4\right)$$

where:

$$\sigma^2 = \sum_{i \in U} p_i \left(\frac{Y_i}{p_i} - Y\right)^2 \qquad \text{and} \qquad \mu_4 = \sum_{i \in U} p_i \left(\frac{Y_i}{p_i} - Y\right)^4.$$

Hansen, Hurwitz, and Madow (1953) and Valliant and Rust (2010) provide the same expression for $v\left(\hat{v}(\hat{Y}_{ind})\right)$. An unbiased estimator of $v\left(\hat{v}(\hat{Y}_{ind})\right)$ is:

$$\hat{v}\left(\hat{v}(\hat{Y}_{ind})\right) = \frac{1}{n^2(n-2)(n-3)}\left(\hat{\mu}_4 - \frac{(n^2-3)}{n}\,\hat{\sigma}^4\right)$$

where:

$$\hat{\sigma}^2 = \frac{1}{(n-1)}\sum_{i \in s}\left(\frac{Y_i}{p_i} - \hat{Y}\right)^2 \text{ and } \hat{\mu}_4 = \sum_{i \in s}\left(\frac{Y_i}{p_i} - \hat{Y}\right)^4.$$

See Ash (2022a) for the derivation of $v(\hat{Y}_{ind})$, $\hat{v}(\hat{Y}_{ind})$, $v\left(\hat{v}(\hat{Y}_{ind})\right)$, and $\hat{v}\left(\hat{v}(\hat{Y}_{ind})\right)$.

Assuming Multinomial Distribution. We can also express the *wr* estimator of the total *Y* for a single stage as:

$$\hat{Y}_{mult} = \frac{1}{n}\sum_{i \in U} X_i \frac{Y_i}{p_i}.$$

Here we assume that X_i is a multinomial random variable that represents the number of times that a sample unit *i* is selected *wr* with probability p_i. We use the subscript *mult* to indicate that the *wr* estimator uses the multinomial assumption. The expectations of the multinomial distribution that are needed to derive the next set of variances and variance estimators are described by Newcomer, Neerchal, and Morel (2008) and Ouimet (2020). Using the *mult* assumption, it can be shown that $v(\hat{Y}_{mult}) = v(\hat{Y}_{ind})$. An unbiased estimator of $v(\hat{Y}_{mult})$ is:

$$\hat{v}(\hat{Y}_{mult}) = \frac{1}{n(n-1)}\sum_{i \in U} X_i \left(\frac{Y_i}{p_i} - \hat{Y}\right)^2.$$

With the multinomial assumption, the variance of $\hat{v}(\hat{Y}_{mult})$ can be expressed as:

$$v\left(\hat{v}(\hat{Y}_{mult})\right) = \frac{1}{n^3(n-1)}\left[\begin{array}{c} (n-2)\sum\limits_{i \in U}\sum\limits_{\substack{j \in U \\ i<j}} p_i p_j (p_i + p_j)\left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^4 + \sum\limits_{i \in U}\sum\limits_{\substack{j \in U \\ i<j}} p_i p_j \left(\frac{Y_i}{p_i} - \frac{Y}{p_j}\right)^4 \\ +2(n-2)\sum\limits_{i \in U}\sum\limits_{\substack{j \in U \\ i<j<k}}\sum\limits_{k \in U} p_i p_j p_k \left[\left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2\left(\frac{Y_i}{p_i} - \frac{Y_k}{p_k}\right)^2 + \left(\frac{Y_i}{p_i} - \frac{y_j}{p_j}\right)^2\left(\frac{Y_j}{p_j} - \frac{Y_k}{p_k}\right)^2 + \left(\frac{Y_i}{p_i} - \frac{Y_k}{p_k}\right)^2\left(\frac{Y_j}{p_j} - \frac{Y_k}{p_k}\right)^2\right] \\ -2n^2(2n-3)\left(v(\hat{Y}_{mult})\right)^2 \end{array}\right]$$

and an unbiased estimator of $v\left(\hat{v}\left(\hat{Y}_{mult}\right)\right)$ is:

$$\hat{v}\left(\hat{v}\left(\hat{Y}_{mult}\right)\right) = \frac{1}{n^3(n-1)(n-2)(n-3)}\left[\begin{array}{c} \sum_{i \in U}\sum_{\substack{j \in U \\ i<j}} X_i^2 X_j \left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2 + \sum_{i \in U}\sum_{\substack{j \in U \\ i<j}} X_i X_j^2 \left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2 - \sum_{i \in U}\sum_{\substack{j \in U \\ i<j}} X_i X_j \left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^4 \\ +2\sum_{i \in U}\sum_{\substack{j \in U \\ i<j<k}}\sum_{k \in U} X_i X_j X_k \left[\left(\frac{Y_i}{p_i} - \frac{y_j}{p_j}\right)^2\left(\frac{Y_i}{p_i} - \frac{Y_k}{p_k}\right)^2 + \left(\frac{Y_i}{p_i} - \frac{Y_j}{p_j}\right)^2\left(\frac{Y_j}{p_j} - \frac{Y_k}{p_k}\right)^2 + \left(\frac{Y_i}{p_i} - \frac{Y_k}{p_k}\right)^2\left(\frac{Y_j}{p_j} - \frac{Y_k}{p_k}\right)^2\right] \\ -2n^3(n-1)^2(2n-3)\left(\hat{v}\left(\hat{Y}_{mult}\right)\right)^2 \end{array}\right]$$

See Ash (2022a) for the derivation of $v\left(\hat{Y}_{mult}\right)$, $\hat{v}\left(\hat{Y}_{mult}\right)$, $v\left(\hat{v}\left(\hat{Y}_{mult}\right)\right)$, and $\hat{v}\left(\hat{v}\left(\hat{Y}_{mult}\right)\right)$.

We admit that the expressions for $\hat{Y}_{ind}$ and $\hat{Y}_{mult}$, and the variance estimators $\hat{v}\left(\hat{Y}_{ind}\right)$ and $\hat{v}\left(\hat{Y}_{mult}\right)$ are essentially equivalent. The duplicate sample units of the *ind* assumption (sample units selected multiple times) are represented by X_i in the *mult* assumption. However, the variance and variance estimators of $\hat{v}\left(\hat{Y}_{ind}\right)$ and $\hat{v}\left(\hat{Y}_{mult}\right)$ are completely different expressions that produce the same values. Raj (1968; p. 120) also notes that the two assumptions result in two different variances and variance estimators when used in the first-stage of a two-stage sample design.

In our notation, we will sometimes include the sample size in our notation. For example, $v\left(\hat{Y}_{mult,n=4}\right)$ refers to the *wr* variance using the multinomial assumption with $n = 4$ and $\hat{v}\left(\hat{Y}_{ind,n=4}\right)$ refers to the *wr* variance estimator using the independence assumption with $n = 4$.

Although $\hat{v}\left(\hat{v}\left(\hat{Y}_{ind}\right)\right)$ and $\hat{v}\left(\hat{v}\left(\hat{Y}_{mult}\right)\right)$ are unbiased estimators of the *wr* variance estimator, they have the drawback that they both require $n \geq 4$. To work around this, we can collapse four original strata into a pseudo stratum. The analysis of the next section considers whether this method produces a safe overestimate of $v\left(\hat{v}_{BRR}\left(\hat{Y}\right)\right)$. We say an estimator produces a safe overestimate because it is "conservative in stating the sampling error of the estimate." (Hansen, Hurwitz, and Madow 1953; p. 439)

4. Description of the Simulation

The only CE data we have is for the 2010 sample PSUs; we have not observed the value of total expenditures for every CU of every PSU in the U.S. Therefore, without a complete universe, we cannot speak definitively about the variances of alternative sample designs. Our solution is to produce a "plausible" or "reasonable" universe using the 2010 sample and simulation. Since our interest is with the variances, the universe does not need to make 100% accurate estimates of total expenditures. What the universe needs to do is to make reasonable variances for making relative comparisons of the variances arising from the alternative sample designs. We are interested in the impact of the first-stage stratification on the variances of estimates and the how well we can estimate those variances. We are not trying to replicate 100% accurate estimates of expenditures.

For the first step of the development of our plausible universe, we applied the SAS PROC SERVERITY ®[2] to the CE data to find the best distribution to represent total expenditures. From this analysis, we determined that a gamma distribution best fit mean total expenditures, given the distributions that SAS considered.

Next, we needed the parameters of the gamma distribution for every PSU in the U.S. To get those values, we modeled PSU-level mean total expenditures $\hat{R}_i$ and its second-stage standard error $\widehat{se}_2(\hat{R}_i) = \left(\hat{v}_2(\hat{R}_i)\right)^{1/2}$, using the observed 2010 sample PSUs. Our regression models were simple and included variables known for all PSUs in the U.S.

Figures 2 and 3 show the fit of the predicted values with the actual values for both the estimate of mean total expenditures and the standard errors of mean total expenditures. Each point in Figures 2 and 3 represents a NSR sample PSU from the 2010 sample design. We generally considered these predicted values as reasonable.

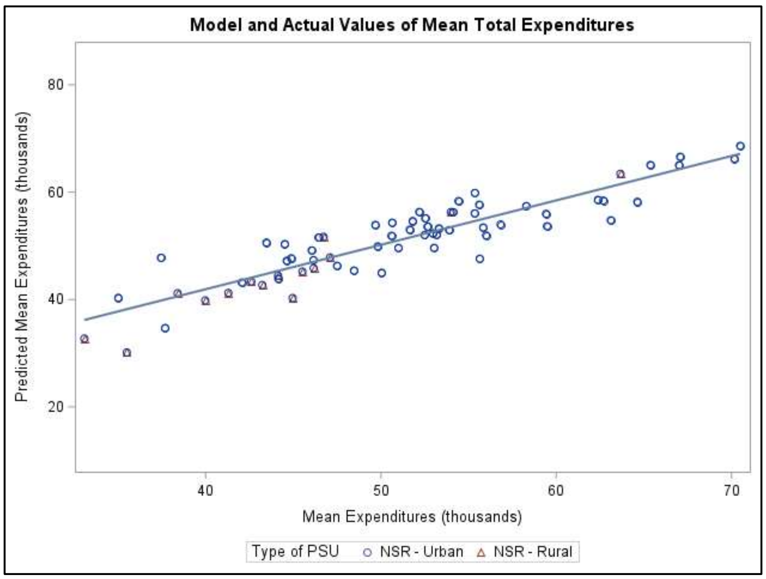

Figure 2: Model and Actual Values for Mean Total Expenditures.

[2] SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

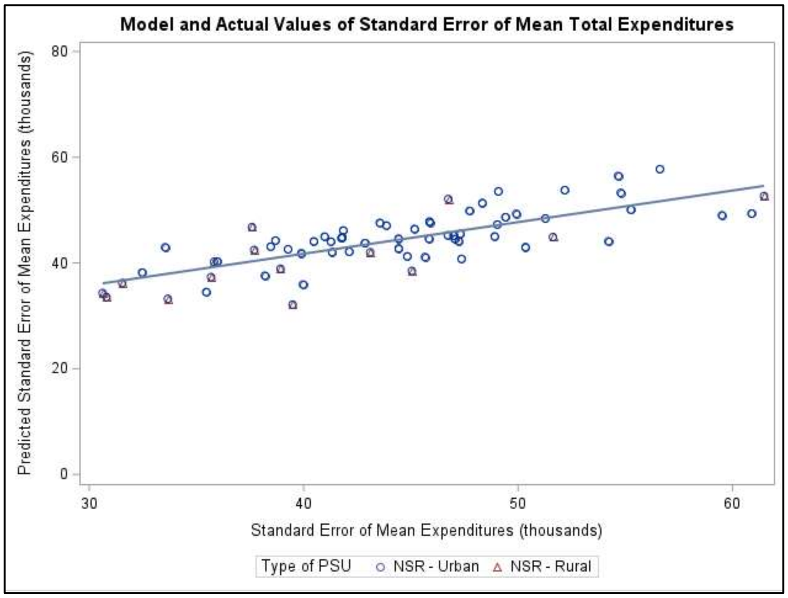

Figure 3: Model and Actual Values for Standard Error of Mean Total Expenditures.

With the predicted values, we defined the shape and scale parameters of the gamma distribution as $shape_i = \left(\hat{R}_i / \widehat{se}_2(\hat{R}_i)\right)^2$ and $scale_i = \left(\widehat{se}_2(\hat{R}_i)\right)^2 / \hat{R}_i$ and used them to randomly generated a universe of 3.5M CUs where the universe size N_i of each PSU was equal to the maximum of either the actual population size or 2,500. A weighting adjustment was calculated and applied to totals to adjust for the difference between 2,500 and N_i. To make the simulation more like actual survey data, we used Bernoulli sampling to change 10% of the sample into ineligible units. Since Bernoulli sampling produces a random number of units with a mean of 10%, the sample sizes of completed interviews are random within the PSUs.

We generated a full universe of CUs because we want to examine the two-stage variance and the variance of the first-stage only (a census within the second stage).

The 2010 and 2020 estimates, produced for our analyses, will use the 2010 PSU definitions, the 2010 MOS, and the same simulated population to facilitate comparisons. We know that the 2020 PSU definitions and MOS will be updated for the 2020 sample design, but we use the 2010 values because we are interested in the impact of the stratification and not the impact of the PSU definitions or the MOS. Additionally, we made one simulated population; our results are not averaged over multiple simulated populations.

We define the simulation expectation and variance for a general estimator $\hat{\theta}$ respectively as $E_{sim}(\hat{\theta}) = n_{sim}^{-1} \sum_s \hat{\theta}_s$, and $v_{sim}(\hat{\theta}) = n_{sim}^{-1} \sum_s \left(\hat{\theta}_s - E_{sim}(\hat{\theta})\right)^2$. Note that $n_{sim} = \sum_s 1$, $se_{sim}(\hat{\theta}) = \sqrt{v_{sim}(\hat{\theta})}$, and $\sum_s$ refers to the sum over the simulated samples.

4. Simulation Results

This section reviews the results of our simulation analysis in terms of the three original objectives. All of the results are for the estimator of mean total expenditures $\hat{R} = \hat{Y}/\hat{X}$, where $\hat{Y}$ is the estimator of total expenditures and $\hat{X}$ is the estimator of the population of interest. Since the estimator $\hat{R}$ is a ratio of two estimators, we linearized $\hat{R}$ when calculating the *cs* and *wr* variances. Linearizing was not used with BRR. The results of section 4.1 use both the SR and NSR PSUs and sections 4.2 and 4.3 only describe the variances of the NSR PSUs. All of the estimates are national, except in Section 4.1, where we provide national and Census Division estimates. We allocated the overall sample size of 12,246 to the first-stage strata in the same way for the 2010 and 2020 sample designs. With the sample sizes, we selected an equal probability simple random sample without replacement (*srswor*) sample of CUs. Within PSU variances and variance estimators reflected the *srswor* sample design. The simulation estimates used 20,000 simulated samples.

The calculation of the 2020 estimates included an adjustment for the remote PSUs excluded from the sample design. We increased the sample weight of the non remote PSUs so that they accounted for the remote PSUs. The adjustment was weighted by the MOS and calculated within the combination of Census Division and the four selected states.

4.1 Comparison of the Variance and Biases of the 2010 and 2020 Sample Designs

Per our first objective, we compare the bias and variance of the 2010 and 2020 sample designs with respect to estimates of expenditures. Table 2 shows the resulting bias of the 2010 and 2020 first-stage sample designs.

Table 2: Bias of 2010 and 2020 Estimates of Mean Total Expenditures

Sample Design	Census Region	Census Division	Mean Total Expenditures (R)	$bias(\hat{R})$	Relative bias (%)
2010	U.S.			0	0.00
2020	U.S.		58,204	15	0.03
	Northeast	New England	67,260	0	0.00
		Middle Atlantic	58,144	-38	-0.07
	Midwest	East North Central	58,966	56	0.09
		West North Central	63,994	-33	-0.05
	South	South Atlantic	50,608	-17	-0.03
		East South Central	48,410	-180	-0.37
		West South Central	47,285	180	0.38
	West	Mountain	57,852	210	0.36
		Pacific	59,157	20	0.03

In Table 2, our estimate of mean total expenditures for 2010 is unbiased because no PSUs were excluded. Because the 2020 does excluded remote PSUs, the estimates of mean total expenditures, even with our adjustment for the remote PSUs, exhibit a small bias. In relative terms, the bias for national and Census Division-level estimates of expenditures are all small – less than 1 percent of the estimate.

Table 3: Comparisons of 2010 and 2020 Standard Errors for Mean Total Expenditures

Statistic	Census Region	Census Division	First-Stage 2010	First-Stage 2020	Ratio (2020/2010)	Both Stages 2010	Both Stages 2020	Ratio (2020/2010)
$se_{wr1}(\hat{R})$	U.S.		614	632	1.05	736	740	1.04
	Northeast	New England	2,278	2,267	1.00	2,976	2,940	0.99
		Middle Atlantic	2,099	1,971	0.94	2,426	2,298	0.95
	Midwest	East North Central	1,281	1,592	1.24	1,652	1,881	1.14
		West North Central	1,625	1,881	1.16	2,061	2,200	1.07
	South	South Atlantic	1,279	1,241	0.97	1.557	1,503	0.97
		East South Central	1,974	2,114	1.07	2,356	2,460	1.04
		West South Central	1,132	1,711	1.51	1,566	1,985	1.27
	West	Mountain	2,770	3,550	1.28	3,184	3,815	1.20
		Pacific	2,938	1,759	0.60	3,215	2,156	0.67
$se_{sim}(\hat{R})$	U.S.		616	630	1.02	735	754	1.03
	Northeast	New England	2,257	2,266	1.00	2,969	2,940	0.99
		Middle Atlantic	2,116	1,963	0.93	2,436	2,314	0.95
	Midwest	East North Central	1,286	1,594	1.24	1,649	1,912	1.16
		West North Central	1,625	1,878	1.16	2,054	2,329	1.13
	South	South Atlantic	1,277	1,241	0.97	1,546	1,520	0.98
		East South Central	1,972	2,113	1.07	2,351	2,472	1.05
		West South Central	1,134	1,707	1.51	1,558	2,023	1.30
	West	Mountain	2,767	3,617	1.31	3,180	4,012	1.26
		Pacific	2,977	1,746	0.59	3,227	2,155	0.67

Table 3 compares the variances of the 2010 and 2020 sample designs both nationally and by Census Division. We see that the national variance for 2020 sample design is slightly larger than the 2010 sample design. This is expected since the 2020 sample design had more high-level strata and therefore, less low-level strata than the 2010 sample design. Some of the Census Division variances increased and some decreased between 2010 to 2020 due to the less even distribution of sample PSUs to the Census Divisions and selected states.

4.2 How to Optimize the BRR Variance Estimator

For the second objective, we consider how to optimize the 2010 and 2020 CE variance estimator with respect to bias and variance. We do this by examining the bias and variances of the *cs* and BRR variance estimators for different choices of P_{gh} and different methods for collapsing strata into pseudo strata. Table 4 summarizes the results in terms of the standard errors – comparing the simulated values $se_{sim}(\hat{R})$, the direct calculation of the standard errors $se_{wr1}(\hat{R})$, and the expectation of the *cs* estimator $E\left(\hat{v}_{cs}(\hat{R})\right)$.

Table 4: 2010 and 2020 Standard Errors for National Estimates of Mean Total Expenditures

Statistic	P_{gh}	k	Pseudo strata collapsing	Standard Errors (SEs)				Bias Ratio of Ses			
				First-Stage Only		Two Stages		First-Stage Only		Two Stages	
				2010	2020	2010	2020	2010	2020	2010	2020
$se_{sim}(\hat{R})$				608	634	732	755				
$se_{wr1}(\hat{R})$				614	632	736	740				
$\sqrt{E\left(\hat{v}_{cs}(\hat{R})\right)}$	½	0	Current 2010	796		893		1.31		1.21	
			Expenditures	752	903	854	991	1.24	1.40	1.16	1.30
			MOS	858	1,064	949	1,140	1.41	1.65	1.29	1.49
	MOS	½	Current 2010	793		890		1.30		1.21	n/a
			Expenditures	745	537	848	618	1.23	0.83	1.15	0.81
			MOS	848	949	940	1,024	1.39	1.47	1.28	1.34
	min bias[2]	½	Current 2010	707		827		1.16		1.12	
			Expenditures	644	666	781	765	1.06	1.03	1.06	1.00
			MOS	741	873	865	963	1.22	1.35	1.18	1.26
$\sqrt{E_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)}$	½	0	Current 2010	792		889		1.30		1.21	
			Expenditures	754	924	856	1,012	1.24	1.46	1.17	1.34
			MOS	860	1,103	951	1,178	1.41	1.74	1.30	1.56
	MOS	½	Current 2010	793		890		1.30		1.22	
			Expenditures	744	546	847	626	1.22	0.86	1.16	1.28
			MOS	849	1,076	940	1,147	1.40	1.70	1.28	1.52
	min bias[2]	½	Current 2010	714		834		1.17		1.14	
			Expenditures	654	684	792	779	1.08	1.08	1.08	1.03
			MOS	737	892	858	979	1.21	1.41	1.17	1.30

The values of $se_{wr1}(\hat{R})$ and $se_{sim}(\hat{R})$ are in the row highlighted in yellow in Table 4 and are in close agreement. Either can be considered the target or known values, but we use $v_{sim}(\hat{R})$ as the base of the bias ratios of the standard errors in the last four columns of Table 4.

The variance estimation for the 2010 sample design uses the choices that are highlighted in the blue in Table 4. The choices include choosing P_{gh} = ½ and collapsing strata using the "Current 2010" method. We see that the variance estimator for the 2010 sample design produces a safe overestimate: the first-stage and two-stage variance estimates are 1.30 and 1.20 times larger than the actual variances. The "Current 2010" method of collapsing strata was a mix of collapsing strata with respect to both expenditures and MOS with more weight given to expenditures. We did not apply the "Current 2010" method to the 2020 sample design because we were not able to develop a general rule based on what was done in 2010.

Before this research, we would have produced replicate factors for the 2020 sample design by choosing P_{gh} proportional to the MOS and collapsing strata with respect to expenditures. These values are in the row highlighted in red in Table 4. We see that this combination provides an underestimate of the variance by a factor of 0.83 for the first stage only and 0.81 for two stages. This large underestimate of the variance is generally unacceptable.

To understand how BRR and *cs* can underestimate the variance, consider one of the pseudo strata that contributed a negative value to the 2020 estimate for the combination of P_{gh} proportional to the MOS and collapsing pseudo strata by expenditures in Table 4. Figure 1 plots $bias(\hat{v}_{cs}(\hat{Y}))$ and the two terms of the $bias(\hat{v}_{cs}(\hat{Y}))$ on the vertical axis and varying value of P_{g1} from 0 to 1 on the horizontal axis.

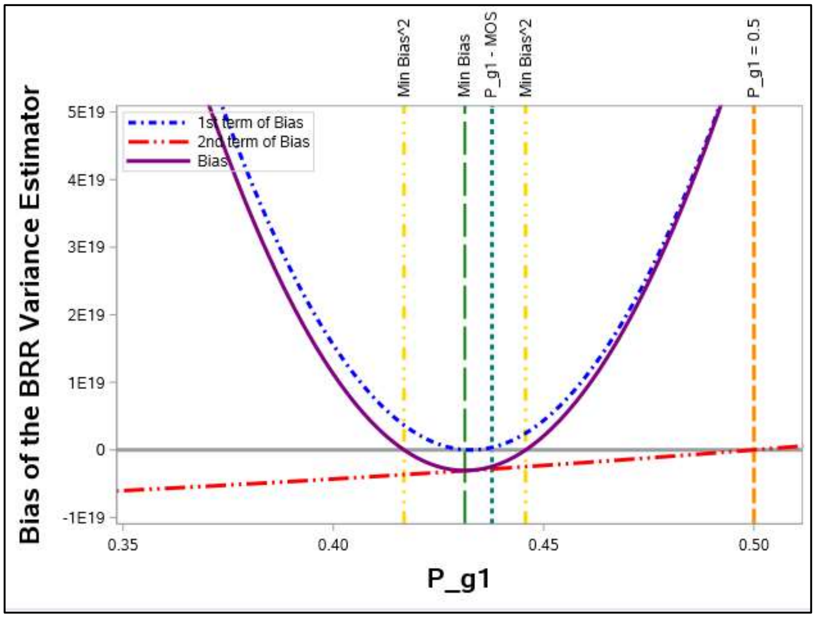

Figure 1. Bias of the Collapsed Strata Variance Estimator – One Pseudo Strata

When we focus on the bias at it is minimum in Figure 1, we see that the negative value of the second term of the bias has the impact of making the overall bias negative where the first term of the bias is minimum. When we choose P_{gh} proportional to MOS, we are trying to approximate the P_{gh} that minimizes first term of the bias (when Y_{gh} is proportional to MOS_{gh}), but this is where the bias is negative and therefore contributes to the underestimate of the variance.

We note that producing an underestimate is less likely if the sample design has equal sized strata and we choose P_{gh} proportional to MOS. In that case, P_{gh} is equal to ½ or close to it which makes the second term of the bias equal to zero or close to zero. The opposite case of very unequal strata can exacerbate the underestimate. This happens because the second term of the $bias(\hat{v}_{cs}(\hat{Y}))$ goes to $\left(v(\hat{Y}_{g1}) - 3v(\hat{Y}_{g2})\right)$ as $P_{g1}^2 \to 0$ or it goes to $\left(-3v(\hat{Y}_{g1}) + v(\hat{Y}_{g2})\right)$ as $P_{g1}^2 \to 1$.

The best choices for the 2020 replicate factors are collapsing strata with respect to expenditures and choosing P_{gh} to minimize $\left(bias(\hat{v}_{cs}(\hat{Y}))\right)^2$. With these choices for 2020, there is no bias with the two-stage variance and the first-stage variance is only 1.03 times larger than the actual variance. In Table 4, the results for this choice are in the row that is highlighted in green.

Next, we examine the variances of the *cs* and BRR estimators for different choices of P_{gh} and choices for collapsing strata into pseudo strata. Table 5 summarizes the results of calculating the standard errors directly in terms of $v\left(\hat{v}_{cs}(\hat{R})\right)$ and calculating the variances via simulation.

Table 5: Comparisons of 2010 and 2020 Standard Errors for Mean Total Expenditures $(\hat{R})$

Sample Design	P_{gh}	k	Pseudo Strata Collapsing	First-Stage Only		Two Stages	
				$se_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)$	$se\left(\hat{v}_{cs}(\hat{R})\right)$	$se_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)$	$se\left(\hat{v}_{cs}(\hat{R})\right)$
2010	½	0	Current 2010	174,129	162,715	225,753	210,051
			Expenditures	173,219	169,229	216,834	211,429
			MOS	182,275	175,952	231,301	224,473
	MOS	½	Current 2010	167,204	162,517	214,257	209,339
			Expenditures	176,463	172,724	218,703	213,959
			MOS	179,393	174,323	227,749	222,338
	min bias^2	½	Current 2010	155,843	149,620	203,944	196,410
			Expenditures	146,274	138,422	196,113	184,708
			MOS	160,179	156,811	211,767	208,211
2020	½	0	Expenditures	225,900	242,083	278,707	292,337
			MOS	343,106	341,131	413,266	410,887
	MOS	½	Expenditures	88,937	93,430	118,775	122,584
			MOS	370,191	346,847	441,825	411,640
	min bias^2	½	Expenditures	168,883	183,033	208,772	220,943
			MOS	284,597	278,731	358,950	353,891

For Table 5, we were able to calculate the expectations $\mu_{gh}^{(a)} = E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^a\right)$ directly with "First-Stage Only" because it is a straightforward calculation:

$$E\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^a\right) = \sum_{i \in U_{gh}} p_{ghi} \left(\frac{Y_{ghi}}{p_{ghi}} - Y_{gh}\right)^a$$

However, for "Two Stages," this expectation becomes much more complicated with $a > 2$. Our solution for Table 5 was to simulate the expectations over the two stages of the sample design. Then the simulated expectations $E_{sim}\left(\left(\hat{Y}_{gh} - Y_{gh}\right)^a\right)$ were used as $\mu_{gh}^{(a)}$ in our calculation of $se\left(\hat{v}_{cs}(\hat{R})\right)$. This is different than $se_{sim}\left(\hat{v}_{cs}(\hat{R})\right)$ which is the variance of $\hat{v}_{cs}(\hat{R})$ via simulation.

Comparing the estimates $se_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)$ and $se\left(\hat{v}_{cs}(\hat{R})\right)$ of Table 5, we conclude that that the expression for $v\left(\hat{v}_{cs}(\hat{R})\right)$ is reasonable when applied to both one or two stages since $se\left(\hat{v}_{cs}(\hat{R})\right)$ agrees with the simulated values $se_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)$. Per the comparisons of the sample designs, the combination of choosing P_{gh} to minimize the bias^2 and collapsing strata with respect to expenditures produced the smallest variance for 2010. With 2020, the smallest variance was produced by the combination of choosing P_{gh} proportional to the MOS and collapsing strata with respect to expenditures. However, this combination underestimated the variance in section 4.2. The second-best combination for 2020 is choosing P_{gh} to minimize the bias^2 and collapsing strata with respect to expenditures– the best combination in section 4.2 for 2020 with respect to bias.

4.3 Estimating the Variance of the CE Variance

With our third objective, we consider variance estimators of the CE two-stage variance. We compare the simulation variance of $\hat{v}_{BRR}(\hat{R})$, which we consider the known variance, with the estimators of the variances $\hat{v}\left(\hat{v}(\hat{R}_{ind,n=4})\right)$ and $\hat{v}\left(\hat{v}(\hat{R}_{mult,n=4})\right)$. The simulation variance of $\hat{v}_{cs}(\hat{R})$ and the *ind* and *mult* variances $v\left(\hat{v}(\hat{R}_{ind,n=4})\right)$ and $v\left(\hat{v}(\hat{R}_{mult,n=4})\right)$ are also included in Table 6 for reference.

The BRR and *cs* results of Table 6 reflect the variance estimation for the 2010 sample design: choosing P_{gh} = ½ and collapsing strata with the "Current 2010" method. To apply the *ind* and *mult* variances and estimators of the variance, we collapsed the "Current 2010" pseudo strata into a pseudo stratum of four original stratum with respect to expenditures. We did not include the 2020 sample design because the adjustment for the remote PSUs complicated it. Table 6 summarizes the results of the simulation.

Table 6: 2010 Standard Errors for Variance Estimator of Mean Total Expenditures

Statistic	One Stage	Two Stages
$se_{sim}\left(\hat{v}_{BRR}(\hat{R})\right)$	174,129	225,753
$se_{sim}\left(\hat{v}_{cs}(\hat{R})\right)$	168,163	215,886
$se\left(\hat{v}(\hat{R}_{ind,n=4})\right)$	185,708	
$se\left(\hat{v}(\hat{R}_{mult,n=4})\right)$	185,708	
$\sqrt{E_{sim}\left(\hat{v}\left(\hat{v}(\hat{R}_{ind,n=4})\right)\right)}$	203,168	235,874
$\sqrt{E_{sim}\left(\hat{v}\left(\hat{v}(\hat{R}_{mult,n=4})\right)\right)}$	203,168	235,874

When we applied $\hat{v}\left(\hat{v}(\hat{R}_{ind,=4})\right)$ and $\hat{v}\left(\hat{v}(\hat{R}_{mult,n=4})\right)$ to a two-stage sample design within the simulation, we replaced the PSU totals Y_i with the simulation values of $\hat{Y}_i$.

Table 6 suggests that the variance estimators $\hat{v}\left(\hat{v}(\hat{R}_{ind,n=4})\right)$ and $\hat{v}\left(\hat{v}(\hat{R}_{mult,n=4})\right)$ provide a safe overestimate of the variance of the BRR and *cs* variance estimators.

5. Conclusions

Per our first objective, the 2010 and 2020 sample designs are not much different with respect to both bias and variance. Although the 2020 will not be unbiased, the bias is small and the exclusion of remote PSUs should provide considerable cost savings.

For our second objective, setting up BRR for variance estimation requires two important choices: how to choose P_{gh} and how to collapse strata into pseudo strata. However, there is no simple answer. In our example, choosing P_{gh} that minimizes $\left(bias(\hat{v}_{cs}(\hat{Y}))\right)^2$ is the best choice in terms of bias and either the best or second best with respect to variance.

However, calculating P_{gh} that minimizes the $\left(bias(\hat{v}_{cs}(\hat{Y}))\right)^2$ requires good information about the strata totals Y_{gh} and strata variances $v(\hat{Y}_{gh})$ and this may be difficult to obtain in practice.

We also learned that having or not having equal-size strata in the sample design can have an important impact on variance estimation. When the strata do not have equal size and we choose P_{gh} as proportional to the MOS, it can result in an unacceptable underestimate of the variance with BRR.

The good news is that we have expressions for the bias and variance of BRR that can be used to evaluate our choices. Additionally, we showed how the prior survey data can be used to produce a plausible universe which we can use to calculate the bias and variance of different variance estimators.

With our last objective, we have offered expressions for the variance and variance estimator of the *wr* variance estimator. A small example using our simulation showed that applying the one stage estimator of the variance of the *wr* variance to a two-stage sample design can produce reasonable estimates.

References

Ash, S. (2014). Using successive difference replication for estimating variances, JSM, Survey Methodology, 40, 47-59.

Ash, S. (2022a). Bureau of Labor Statistic memorandum entitled "Results for the With Replacement Variance Estimator," forthcoming.

Ash, S. (2022b). Bureau of Labor Statistic memorandum entitled "Results for the Collapsed Strata Variance Estimator," forthcoming.

Ash, S., Dumbacher, B., Swanson, D., and Reyes-Morales, S. (2012). Allocation of Sample for the 2010 Redesign of the Consumer Expenditure Survey, JSM, Section on Survey Research Methods.

Dippo, C.S., Fay, R.E., Morganstein, D.H. (1994). Computing Variances for Complex Survey Designs with Replication, JSM, Section on Survey Research Methods, 489-494.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Volume 2. John Wiley & Sons, New York, NY.

Judkins, D.R. (1990). Fay's method for Variance Estimation, Journal of Official Statistics, 6, 223-239.

King, S.L., Schilp, J., and Bergmann, E. (2011). Assigning PSUs to a Stratification PSU, JSM, Section on Survey Research Methods.

McCarthy, P.J. (1966). Replication: An Approach to the Analysis of Data from Complex Surveys. Vital and Health Statistics, 79-1269, National Center for Health Statistics, Washington D.C.

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, JSM, Section on Survey Research Methods.

Newcomer, J.T., Neerchal, N.K. and Morel, J.G. (2008). Computation of Higher Order Moments from Two Multinomial Overdispersion Likelihood Models, Department of Mathematics and Statistics, University of Maryland: Baltimore MD, USA, 2008;

p. 1-11. Available online: http://www.math.umbc.edu/~kogan/technical_papers/2008/Newcomer_Nagaraj_Morel.pdf (accessed November 6, 2020).

Ouimet, F. (2020). Explicit formula for the joint third and fourth central moments of the multinomial distribution, arXiv:2006.09059.

Raj, D. (1968). *Sampling Theory*, McGraw Hill.

Rust, K.F. (1984). Techniques for Estimating Variances for Sample Surveys, thesis University of Michigan.

Rust, K. and Kalton, G. (1987). Strategies for Collapsing Strata for Variance Estimation, Journal of Official Statistics, 3, 69-81.

Swanson, D. (2017). Consumer Expenditures Surveys Program Series, "Standard Errors in the 2016 Consumer Expenditure Survey," September 2017. https://www.bls.gov/cex/research_papers/pdf/ce-standard-error-article.pdf

Valliant, R. and Rust, K.F. (2010). Degrees of Freedom Approximations and Rules-of-Thumb, Journal of Official Statistics, 24, 585-602.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. Spring-Verlag.