ABSTRACT

Title of dissertation: SAMPLING WEIGHT CALIBRATION WITH ESTIMATED CONTROL TOTALS

Jill A. Dever, Doctor of Philosophy, 2008

Dissertation directed by: Professor Richard Valliant
Joint Program in Survey Methodology

Sample weight calibration, also referred to as calibration estimation, is a widely applied technique in the analysis of survey data. This method borrows strength from a set of auxiliary variables and can produce weighted estimates with smaller mean square errors than those estimators that do not use the calibration adjustments. Poststratification is a well-known calibration method that forces weighted counts within cells generated by cross-classifying the categorical (or categorized) auxiliary variables to equal the corresponding population control totals.

Several assumptions are critical to the theory developed to date for weight calibration. Two assumptions relevant to this research include: (i) the control totals calculated from the population of interest and known without (sampling) error; and (ii) the sample units selected for the survey are taken from a sampling frame that completely covers the population of interest (e.g., no problems with frame undercoverage).

With a few exceptions, research to date generally is conducted as if these assumptions hold, or that any violation does not affect estimation. Our research

directly examines the violation of the two assumptions by evaluating the theoretical and empirical properties of the mean square error for a set of calibration estimators, newly labeled as estimated-control (EC) calibration estimators. Specifically, this dissertation addresses the use of control totals estimated from a relatively small survey to calibrate sample weights for an independent survey suffering from undercoverage and sampling errors. The EC calibration estimators under review in the current work include estimated totals and ratios of two totals, both across all and within certain domains. The ultimate goal of this research is to provide survey statisticians with a sample variance estimator that accounts for the violated assumptions, and has good theoretical and empirical properties.

SAMPLING WEIGHT CALIBRATION WITH ESTIMATED CONTROL TOTALS

by

Jill A. Dever

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Dr. Richard Valliant, Chair/Advisor
Dr. Phillip S. Kott
Dr. Frauke Kreuter
Dr. Stephen Miller
Dr. Paul J. Smith

Table of Contents

List of Tables

List of Figures

Chapter 1

Statement of Work

Sample weight calibration, also referred to as calibration estimation, is a technique widely applied throughout the survey research world. This method borrows strength from a set of auxiliary variables which in general results in weighted estimates with smaller mean square errors (MSE = variance + squared bias) than those which do not use the calibration adjustments. Reduction in MSE is related to the association between the auxiliary and analysis variables. Poststratification is a well-known calibration method that forces weighted *counts* within cells generated by cross-classifying the categorical (or categorized) auxiliary variables to equal the corresponding population counts. These counts are also known as control totals or benchmark controls. If the population cell counts are unavailable, the estimated and true marginal counts are equalized through iterative proportional fitting (i.e., raking) or other regression techniques.

Several assumptions are critical to the theory developed to date for weight calibration. Some of these assumptions are explicitly stated in the literature, while others are more implicit and identified based on the theoretical evaluations presented. The assumptions include, for example: (i) the control totals are calculated from the population and known without sampling or other errors (e.g., measurement); (*ii*) the sample units selected for the survey are taken from a sampling frame that completely covers the population of interest (e.g., no missing population units resulting in undercoverage problems); and (*iii*) the survey requiring calibration does not suffer from nonsampling errors such as nonresponse (e.g., 100 percent response rate) or measurement error (i.e., data values are given and recorded accurately).

With a few exceptions, research to date generally is conducted by assuming that these theoretical requirements hold or that any violation of these assumptions is minimal and does not impact estimation. Our research examines the effects of violating several assumptions on the theoretical properties and empirical MSE estimates of calibrated estimators. Specifically, we address the use of control totals *estimated* from a potentially small survey to calibrate sample weights for a survey suffering from undercoverage and sampling errors. We label this weighting methodology as *estimated-control* (EC) *calibration*. Our methods will be extended at a later date to address the effects of nonresponse and measurement error. Estimated totals and ratios of two totals, both across all and within certain domains, are of particular interest to our research. Ultimately our goal is to develop one or more variance estimators that account for the violated assumptions, thereby translating our theoretical findings into practical applications for survey statisticians. The association between the analysis variables and auxiliary variables is assumed to be adequately modeled through a linear regression (i.e., linear calibration). Other estimators such as regression coefficients and non-linear calibration are reserved for future research.

The results obtained from our current research expand the body of knowledge on weight calibration and are presented in the subsequent chapters. We provide a brief overview of the extensive research conducted to date in Chapter 2 as it relates to our work. Chapter 3 details the scope of our research including notation, assumptions for the theoretical evaluation, and data used to generate the empirical results. We begin in Chapter 4 with the development and evaluation of bias and variance estimators for (overall) estimated population totals. A similar structure is used in Chapter 5 to present findings for ratios of two estimated totals. Domain estimation for both totals and ratios is discussed in Chapter 6. Chapter 6, as do Chapters 4 and 5, begins with theory and then proceeds to a summary of the empirical results from a simulation study. We conclude the dissertation in Chapter 7 with a overall summary of our findings and a map for our future endeavors on weight calibration.

Chapter 2

Historical Perspective

The discussion of calibration estimation below begins with an overview of the extensive literature on traditional weight calibration in which control totals are assumed to be fixed population values. We label this methodology as "traditional" to distinguish it from the weight calibration discussed later in this chapter. Where appropriate, we point to issues reserved for future research in comparison to areas covered by our current work. We focus on a specific set of calibration estimators for population totals and ratios of two totals (Section 2.1), and discuss the theoretical properties of these point estimators in Section 2.2. As reiterated throughout the text, literature related to weight calibration using survey-estimated controls, hereafter referred to as estimated-control (EC) calibration, does exist but is sparse. A discussion of the current techniques for calibration variance estimation follows in Section 2.3. We conclude this chapter with issues related to domain-specific calibration estimation (Section 2.4).

2.1 Calibration Estimators

Calibration estimators, a label first used by Deville & Särndal (1992), identify a class of estimators that borrow strength from auxiliary information to improve the efficiency of survey estimates over more traditional weighting methods such as simple inverse probability weighting. When the G ($G \geq 1$) auxiliary variables are strongly related to a survey outcome (y), the corresponding calibration estimate will be very efficient. However, we can not expect a high level of association between the auxiliary variables and every outcome measured in the survey, so that the efficiency will naturally vary. We briefly compare this efficiency against levels for other estimators in the next section.

Calibration estimators are used in all types of surveys. These include, for example, large U.S. government surveys, such as the Consumer Expenditure Survey (see, e.g., Jayasuriya & Valliant, 1996) and the National Health Interview Survey (National Center for Health Statistics, 2006); surveys of specialized populations, such as the U.S. Department of Defense (DoD) Survey of Health Related Behaviors among Military Personnel (Bray et al., 2003); and a myriad of surveys outside the U.S. including the Canadian Retail Trade Survey (see, e.g., Hidiroglou & Patak, 2006), the Swedish Labour Force Survey (Sweden, 2005), and the British Household Panel Survey (Taylor et al., 2007).

Weight calibration is used to correct survey estimates for sampling frame problems such as undercoverage and to reduce errors associated with sampling and nonresponse (see, e.g., Särndal et al., 1992; Kott, 2006). Undercoverage occurs when the sampling frame fails to contain all units for the population under study (see, e.g., Lessler & Kalsbeek, 1992). For example, estimates from the Behavioral Risk Factor Surveillance System (BRFSS), a nationwide random-digit-dial (RDD) telephone survey conducted by the U.S. Centers for Disease Control and Prevention (CDC), are calibrated (i.e., benchmarked or poststratified) to population counts that include households with and without landline telephone service (Centers for Disease Control and Prevention, 2006). Preliminary results from the 2007 National Health Interview Survey (NHIS) suggest that approximately 15.8 percent of American homes prefer wireless communications and no longer have a landline service (Blumberg & Luke, 2008). If population values are different for the covered and not-covered groups and the proportion not covered is sizeable, then estimates obtained from the BRFSS can have non-trivial levels of error without the use of corrective methods such as calibration. Groves (1989, Section 3.2) provides the following formula for undercoverage error associated with a linear estimator:

$$\theta_c - \theta = \frac{N_{nc}}{N} (\theta_c - \theta_{nc}) \tag{2.1}$$

where θ is the true value for a population of size N; θ_c and θ_{nc} are the population values for covered and not-covered subsets of the population, respectively; and N_{nc}/N is the proportion of the population not covered by the sampling frame.

The calibrated weight w_k is composed of the original design weight π_k^{-1}, the inverse of the sample inclusion probability for the k^{th} unit of observation, multiplied by a calibration-adjustment factor a_k. Traditional weight calibration assumes that the analytic survey (i.e., the survey requiring weight calibration) has no nonresponse. In practice, however, a separate adjustment for nonresponse may also be applied to the design weights. Calibrated weights are historically calculated by minimizing a specified function that measures the distance between w_k and π_k^{-1}. The distance function, $F(w_k, \pi_k^{-1})$, is minimized subject to a set of calibration constraints (or calibration equations) defined as:

$$\mathbf{t}_x = \sum_{k \in s_A} w_k \mathbf{x}_k \tag{2.2}$$

where $\mathbf{t}_x = \sum_{l \in U} \mathbf{x}_l = [t_1, ..., t_G]'$ is the vector of population control (benchmark) totals corresponding to G chosen (auxiliary) survey variables, and $\mathbf{x}$ is a vector of length G containing either analytic survey ($k \in s_A$) or benchmark ($k \in U$) values. The vector $\mathbf{x}$ may include a column of ones ($x = 1$) for constrained estimation of the overall population size, ones and zeros to indicate the presence or absence of a characteristic (e.g., age 18-25 or gender), or larger values (e.g., number of children, or household income). The calibration system (distance function and calibration constraints) results in calibration weights of the form

$$w_k = \pi_k^{-1} \; F^{-1}(\mathbf{x}_k, \boldsymbol{\lambda}, c_k) \tag{2.3}$$

where F^{-1} is the inverse function of $\partial F / \partial w_k$, the first derivative of the distance function taken with respect to the calibrated weight; $\boldsymbol{\lambda}$ is the G-length vector of Lagrange multipliers that satisfies the calibration constraints (2.2) given the design weights π_k^{-1}; and c_k is a value associated with the estimator of choice.

The distance function, $F(w_k, \pi_k^{-1})$, can take multiple forms but is generally chosen from a class of functions that are monotonic and twice-differentiable (Deville & Särndal, 1992). Several of these distance functions are discussed in Deville &

Särndal (1992), Huang & Fuller (1978), and Singh & Mohl (1996). Empirical studies such as those in Singh & Mohl (1996) and Stukel et al. (1996) show that the specific choice of the distance function does not greatly affect either the point or variance estimates, provided that the data are complete (i.e., no missing values). They suggest that the choice of the particular distance function is often more related to personal preference for the resulting estimator or to the structure of the control totals than to an optimality justification.

Deville & Särndal (1992), by contrast, point to potential problems with five commonly used distance functions. For example, the generalized least squares (GLS) distance function

$$\sum_{k \in s_A} (w_k - \pi_k^{-1})^2 / 2c_k\pi_k^{-1}, \tag{2.4}$$

also known as the average or chi-square distance function, generates a closed-form solution to the minimization problem but can result in one or more negative weights. Practitioners consider negative weights to be highly undesirable because they do not have the intuitive interpretation present for inverse-probability weights, i.e., w_k provides the number of units represented in the population by the results collected for the k^{th} sample unit.

To remedy the problem of negative weights, Deville & Särndal (1992) proposed two additional distance functions (Cases 6 and 7 in their article). These distance functions are constrained to produce calibration-adjustment factors ($a_k = w_k \times \pi_k$), referred to as a g-weight in Section 6.5 of Särndal et al. (1992), that fall within a range of values specified by the researcher (e.g., lower bound greater than zero and upper bound less than some extreme value). Calibration with constrained weight-adjustment factors is widely applied through existing software such as a quadratic or optimization programming routine from IMSL used by Isaki et al. (2004); the generalized exponential modeling (GEM) software developed by Folsom & Singh (2000) using SAS® IML; and the calibrate function in the R® language survey library (R Development Core Team, 2005). Even with its popularity, theory to date has been developed under the assumption that the distance function produces nicely behaved weights because bounding complicates the theory. We shall follow this direction with our current research and plan to address constrained a_k's in our future EC calibration work.

Returning to expression (2.4), the c_k's are positive "weights" unrelated to the design weights that are chosen to generate specific types of estimators (Estevao & Särndal, 2000; Lundström & Särndal, 1999; Stukel et al., 1996; Tracy et al., 2003). This property is related to the popularity of the GLS distance function. For example, $c_k = x_k^{-1}$ for a model that relates the outcome variable y to a single auxiliary variable x with $Var_\epsilon(y_k) = \sigma^2 x_k$, and motivates the ratio estimator of a population total

$$\hat{t}_{yRT} = t_x \left(\frac{\hat{t}_{Ay}}{\hat{t}_{Ax}} \right) = \sum_{l \in U} x_l \left(\frac{\sum_{k \in s_A} \pi_k^{-1} y_k}{\sum_{k \in s_A} \pi_k^{-1} x_k} \right).$$

Estimates, as opposed to population values, are identified in formulae in this and subsequent chapters by the "hat" notation. For example, t_x is a population total of x while $\hat{t}_{Ax}$ is the corresponding estimated total from the analytic survey data.

Not all distance functions produce a closed-form solution as with the GLS;

some functions require iterative procedures to solve the calibration system. For example, the raking ratio (or iterative proportional fitting) distance function, defined as $F(w_k, \pi_k^{-1}) = c_k\pi_k^{-1}[w_k\pi_k(ln(w_k\pi_k) - 1) + 1]$, requires iteration techniques to calculate the estimates. This distance function, however, does guarantee positive calibrated weights. Iterative methods are easily applied in practice but complicate the theoretical development of new techniques because a closed-form solution is not available. Therefore, such distance functions have limited use in our current research.

The GLS distance function (2.4) is also referred to as a linear distance function because the resulting inverse function (F^{-1}) is linear only in the auxiliary variables ($\mathbf{x}$). The benefit of such a property is that the resulting calibrated analysis weights are functions only of the auxiliary variables and not any of the outcome variables. In other words, one set of final analysis weights is created instead of requiring weights specific to each variable within a set of key outcome variables. This feature is of particular interest to organizations that produce analysis files for use either by the public or by client agencies. For example, minimizing the GLS distance function subject to the controls in (2.2) with $c_k = 1$ (i.e., $Var_\epsilon(y_k) = \sigma^2$) generates the well-known generalized (linear) regression estimator (GREG). The GREG of a population total is calculated as follows, using the *traditional* calibration assumptions noted in

$$\begin{aligned}\hat{t}_{yTRGR} &= \sum_{k\in s_A} w_k y_k \\ &= \sum_{k\in s_A} a_k \pi_k^{-1} y_k \\ &= \sum_{k\in s_A} \left[1 + \left(\mathbf{t}_x - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{l\in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l'\right)^{-1} \mathbf{x}_k\right] \pi_k^{-1} y_k \end{aligned} \quad (2.5)$$

where the vector of Horvitz-Thompson (HT) auxiliary variable estimates, $\hat{\mathbf{t}}_{Ax} = \sum_{s_A} \pi_k^{-1}\mathbf{x}_k = \left[\hat{t}_{A1}, ..., \hat{t}_{AG}\right]'$, corresponds to the G-length vector of population controls $\mathbf{t}_x$ (Horvitz & Thompson, 1952). Here we see that the calibration-adjustment factor $a_k = 1 + \left(\mathbf{t}_x - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l'\right)^{-1} \mathbf{x}_k$ is a function of the population control total vector ($\mathbf{t}_x$), the vector of estimated totals ($\hat{\mathbf{t}}_{Ax}$), the auxiliary variables ($\mathbf{x}$), and the design weights (π_k^{-1}), but *not* the outcome variable y. Hence, this same set of calibrated weights can be used with any analysis variable.

Generation of estimators by minimizing a distance function is labeled as the *calibration approach*, while another method is referred to as “GREG thinking” or the *regression approach* (Särndal, 2007). With the regression approach, estimators are calculated by way of an *assisting model* that closely represents the relationship between the outcome variable (y) and the auxiliary variables ($\mathbf{x}$). The assisting model is also referred to as the calibration model or the working prediction model by Kott (2006) to distinguish it from other models such as those used to address response propensity. The model is labeled as “assisting” or “working” because we do not assume equivalence with the true (unknown) underlying population model. The size of the residuals measures the effectiveness of the model; the benefits of small residuals are highlighted in Section 2.3. Therefore, $\hat{t}_{yTRGR}$ in (2.5) is equivalently justified as follows using a *linear assisting model* such that $E_{\epsilon}(y_k) = \mathbf{x}_k'\mathbf{B}$ and $Var_{\epsilon}(y_k) = \sigma^2$, where E_{ϵ} and Var_{ϵ} represent the expectation and variance evaluated with respect to the specified working model:

$$\hat{t}_{yTRGR} = \hat{t}_{Ay} + (\mathbf{t}_x - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_A. \tag{2.6}$$

The HT estimator of y is defined as $\hat{t}_{Ay} = \sum_{k\in s_A} \pi_k^{-1} y_k$, a function of the outcome variable and the design weights. The model coefficient vector

$$\hat{\mathbf{B}}_A = \left[\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l'\right]^{-1} \sum_{k\in s_A} \mathbf{x}_k \pi_k^{-1} y_k \tag{2.7}$$

is calculated based on the specification of a working model, $y_k = \mathbf{x}_k'\mathbf{B} + E_k$, and is approximately design unbiased for the corresponding population parameters $\mathbf{B} = \left[\sum_{l\in U} \mathbf{x}_l'\mathbf{x}_l\right]^{-1} \sum_{k\in U} \mathbf{x}_k y_k$ (see, e.g., Result 5.10.1 in Särndal et al., 1992), under an assumption of complete response and no sampling frame error. It is also assumed that the matrix $\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l'$ of dimension G is nonsingular so that the inverse exists. Finally, $a_k = 1+(\mathbf{t}_x - \hat{\mathbf{t}}_{Ax})'(\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l')^{-1}\mathbf{x}_k$ is the calibration-adjustment factor, thus demonstrating the equivalence of (2.5) and (2.6).

Another special case of the traditional GREG estimator, which is well-known and widely applied, is the poststratified estimator. Using the assisting-model approach, these estimators are generated under the group-mean (linear) assisting model defined by $E_\epsilon(y_k) = B_g$ and $Var_\epsilon(y_k) = \sigma_g^2$ for units within each of $g = 1, ..., G$ poststrata. A single auxiliary variable is used in the model which indicates unit membership in the mutually exclusive poststrata. Thus, $B_g = \bar{y}_g$, the average of y in poststratum g. The poststratified estimator of a population total, sometimes referred to as a ratio estimator, is calculated as follows again using the traditional calibration assumptions:

$$\begin{aligned}
\hat{t}_{yTRPS} &= \sum_{g=1}^{G} N_g \, \frac{\hat{t}_{Ayg}}{\hat{N}_{Ag}} \\
&= \sum_{g=1}^{G} N_g \hat{N}_{Ag}^{-1} \left[\sum_{k \in s_A} \delta_{gk} \pi_k^{-1} y_k \right] \\
&= \sum_{k \in s_A} \left[\sum_{g=1}^{G} N_g \hat{N}_{Ag}^{-1} \delta_{gk} \right] \pi_k^{-1} y_k \\
&= \sum_{k \in s_A} \sum_{g=1}^{G} a_k \pi_k^{-1} y_k. \qquad (2.8)
\end{aligned}$$

The number of (true) population units in the g^{th} poststratum is denoted as N_g. The poststratum sizes estimated from the analytic survey $\hat{N}_{Ag}$ are calculated by summing the design weights across primary sampling units (PSUs) and design strata for units within each poststratum, i.e., $\sum_{k \in s_{Ag}} \pi_k^{-1} = \sum_{k \in s_A} \delta_{gk} \pi_k^{-1}$. Though a simplified notation is used, $\hat{t}_{Ayg}$ represents the HT estimated total of y within poststratum g calculated under the analytic survey sampling design. The zero/one variable δ_{gk} identifies members of poststratum g (s_{Ag}) from within the complete sample (s_A). The ratio $\hat{t}_{Ayg}/\hat{N}_{Ag}$ is widely referred to as a combined ratio estimator when the components are calculated by summing across the analytic survey design strata, i.e., the poststrata cross the design strata. The calibration-adjustment weights for $\hat{t}_{yTRPS}$ are calculated as $a_k = \sum_{g=1}^{G} \delta_{gk}(N_g/\hat{N}_{Ag}) = (N_g/\hat{N}_{Ag})$.

Other less prominent forms of the GREG are also found in the literature. The *functional form approach* discussed in Estevao & Särndal (2000), referred to as the *instrumental vector method* in Särndal (2007), generalizes the GREG to include a vector of instrumental variables $\mathbf{z}_k$ in addition to the set of auxiliary variables $\mathbf{x}_k$ (see also Kott, 2006). This method is strictly applied through the regression approach by requiring that the calibrated weights have the form $w_k = \pi_k^{-1}F(\mathbf{z}_k, \mathbf{x}_k)$, where $F()$ is *any* monotonic, twice-differentiable function. Note that the change from the design to the calibrated weights is not minimized as with the original approach proposed by Deville & Särndal (1992). The instrumental variables are incorporated into $\hat{t}_{yTRGR}$, for example, through the assisting-model coefficients $\hat{\mathbf{B}}_A = \left[\sum_{l \in s_A} \pi_l^{-1}\mathbf{z}_l\mathbf{x}_l'\right]^{-1} \sum_{k \in s_A} \mathbf{z}_k \pi_k^{-1} y_k$. Using this method, Estevao & Särndal (2004, Result 8.1) determined the set of optimal instrumental variables which minimizes the asymptotic variance of the calibration estimator calculated for a general sampling design:

$$z_{k(opt)} = \pi_k \sum_{l \in s_A} (\pi_k^{-1}\pi_l^{-1} - \pi_{kl}^{-1})\mathbf{x}_l$$

where π_{kl} is the joint inclusion probability for the k^{th} and l^{th} units in the analytic survey sample. Though minimal variance is always desirable, we choose to focus on more traditional calibration weights within our current research.

The association between the outcome variable y and the auxiliary variables x

may not be best represented through a linear model as with the GREG. Lehtonen & Veijanen (1998) proposed logistic generalized regression estimators (LGREGs) for use with binary outcome variables. The LGREG of a population total, as presented in Duchesne (2003), is calculated as:

$$\hat{t}_{yLGREG} = \sum_{k \in U} \hat{\mu}_k + \sum_{k \in s_A} \pi_k^{-1} \left(y_k - \hat{\mu}_k \right) \tag{2.9}$$

where $\hat{\mu}_k = exp(\mathbf{x}'_k \hat{\mathbf{B}}_A)/(1 + exp(\mathbf{x}'_k \hat{\mathbf{B}}_A))$, the predicted values from the logistic model of $\mathbf{x}$ on y such that $0 < \hat{\mu}_k < 1$ by definition. As alluded to in Särndal (2007), LGREG weights are outcome variable specific which removes the "GREG advantage" of a single set of analysis weights. Additionally, the simulation study results presented by Lehtonen & Veijanen (1998) suggest that the empirical differences for GREG and LGREG estimators do not differ by appreciable levels. Given these two points, we will reserve LGREG estimators for future research.

In our discussions so far, we have emphasized the adjective *traditional* when discussing weight calibration. This is to distinguish it from calibration to *estimated* control totals. In practice, population totals or counts that are unknown are ideally estimated from independent, high-quality surveys with large sample sizes and negligible sampling and non-sampling errors. Because the calibration system requires estimates from more than one survey, we label the *benchmark survey* as the control total source, and the *analytic survey* as the survey requiring calibration. Given the practical issues with weight calibration, we rephrase the estimated total formulae presented previously using notation that is relevant to our research. The GREG of a population total, using control totals from one or more benchmark surveys, is defined as:

$$\hat{t}_{yGREG} = \hat{t}_{Ay} + (\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_A \tag{2.10}$$

using components defined in (2.6), e.g., $\hat{\mathbf{B}}_A = \left[\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l'\right]^{-1} \sum_{k\in s_A} \mathbf{x}_k\pi_k^{-1}y_k$. The only difference from expression (2.6) is to replace the population control-total vector $\mathbf{t}_x$ with a vector produced from the benchmark survey(s), $\mathbf{t}_{Bx}$. We do not use the hat notation for this vector due to the assumption that the control totals are estimated with negligible sampling variance. In other words, the population covariance matrix for $\mathbf{t}_{Bx}$, $Cov\left(\mathbf{t}_{Bx}\right) \equiv \mathbf{V}_B$, is presumed to contain values close enough to zero to support the claim $\mathbf{V}_B \equiv \mathbf{0}$, a matrix of zeros. The calibration-adjustment weights for $\hat{t}_{yGREG}$ are calculated as $a_k = 1 + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l'\right)^{-1} \mathbf{x}_k$. These estimators are generated using either the calibration approach by minimizing the GLS distance function (2.4) subject to the constraints $\mathbf{t}_{Bx} = \sum_{k\in s_A} w_k\mathbf{x}_k$, or the regression approach through the linear model specified for expression (2.6), i.e., $E_\epsilon(y_k) = \mathbf{x}_k'\mathbf{B}$ and $Var_\epsilon(y_k) = \sigma^2$.

The corresponding poststratified estimator of a population total, defined by a slight relaxation of the population control total assumption, is calculated as:

$$\hat{t}_{yPSGR} = \sum_{g=1}^{G} N_{Bg}\, \frac{\hat{t}_{Ayg}}{\hat{N}_{Ag}} \tag{2.11}$$

where N_{Bg} is the benchmark survey count within poststratum g. The remaining terms are defined in expression (2.8). The poststratified estimator can be expressed in matrix notation as:

$$\hat{t}_{yPSGR} = \mathbf{N}'_B \ \hat{\mathbf{N}}_A^{-1} \ \hat{\mathbf{t}}_{Ay}$$

where $\mathbf{N}_B = [N_{B1}, ..., N_{BG}]'$, a G-length vector of benchmark survey totals by post-strata; $\hat{\mathbf{N}}_A$ is a $G \times G$ diagonal matrix with elements equal to the estimated analytic survey poststratum counts $\hat{N}_{Ag} = \sum_{k \in s_A} \delta_{gk} \pi_k^{-1}$ with $\delta_{gk} = 1$ if unit k is a member of the g^{th} poststratum (zero otherwise); and, $\hat{\mathbf{t}}_{Ay} = \left[\hat{t}_{Ay1}, ..., \hat{t}_{AyG}\right]'$ with $\hat{t}_{Ayg} = \sum_{k \in s_A} \delta_{gk} \pi_k^{-1} y_k$. Poststratified estimators are also generated by minimizing the GLS distance function (2.4) given the calibration constraints $N_{Bg} = \sum_{k \in s_A} w_k \delta_{gk}$ for every poststratum g. Using a regression approach, the poststratified estimators are again generated through the group-mean model, i.e., $E_\epsilon(y_k) = B_g$ and $Var_\epsilon(y_k) = \sigma_g^2$.

Functions of GREG-estimated totals are also relevant for the analysis of survey data. The ratio of two GREG totals, of particular interest to our research, is one that approximates a population mean and takes the form

$$\hat{\bar{y}}_{GREG} = \frac{\hat{t}_{yGREG}}{\hat{N}_{GREG}} \tag{2.12}$$

for $\hat{t}_{yGREG}$ defined in (2.10). The estimated population size in the denominator of the ratio is

$$\hat{N}_{GREG} = \hat{N}_A + (\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_{AN} \tag{2.13}$$

where $\hat{N}_A = \sum_{k \in s_A} \pi_k^{-1}$, the population size estimated from the analytic survey data, and

$$\hat{\mathbf{B}}_{AN} = \left(\sum_{l \in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l' \right)^{-1} \hat{\mathbf{t}}_{Ax}, \tag{2.14}$$

the model coefficient vector for the linear assisting model discussed in (2.10) with $y_k = 1$ for all sample units. Under the group-mean model associated with $\hat{t}_{yPSGR}$ (2.11), the ratio becomes

$$\hat{\bar{y}}_{PSGR} = \frac{\hat{t}_{yPSGR}}{\hat{N}_{PSGR}} = \frac{\hat{t}_{yPSGR}}{N_B} \tag{2.15}$$

because $\hat{N}_{PSGR} = \sum_{g=1}^{G} N_{Bg} \hat{N}_{Ag}^{-1} \sum_{k \in s_{Ag}} \pi_k^{-1} = \sum_{g=1}^{G} N_{Bg} = N_B$. The estimators $\hat{\bar{y}}_{GREG}$ and $\hat{\bar{y}}_{PSGR}$ are also known as Hájek estimators (Hájek, 1971; Smith, 1991).

Given that our current research can not address all aspects of weight calibration, we have chosen to focus specifically on the GLS distance function (2.4) due to its ability to generate a closed-form solution for various estimators calculated with weights that are not a function of the outcome variable. Additionally, GREG estimators provide an explicit form to the calibration weights which allows for a direct examination of the theoretical properties (bias and variance) for these widely used estimators. Therefore, the remaining discussion and the research results detailed in the subsequent chapters will deal with GREG estimators of population totals (2.10) and (2.11), and the ratio of two GREG totals.

2.2 Bias of Calibration Estimators

Särndal et al. (1989, 1992) show that the GREG of a population total has many desirable properties such as approximate or asymptotic design-unbiasedness (ADU) and design consistency. The "approximate" label stems from the approximate unbiasedness property of the regression coefficients vector, $\hat{\mathbf{B}}_A$ in (2.10). They also claim that the variance estimator of the GREG is approximately model-unbiased under certain conditions; we save the discussion of variance estimation for Section 2.3. Estevao & Särndal (2002), as well as others, state that the design-unbiased property is only attained if the calibration weights are approximately equal to one. Deville & Särndal (1992) develop a set of conditions under which calibration estimators are asymptotically equivalent to the GREG, and therefore share the desirable properties above. However, Estevao & Särndal (2000) demonstrate that the GREG and the family of calibration estimators are always equivalent only if the assisting model is correctly specified with *all* relevant auxiliary variable covariates, an unlikely condition. Those calibration estimators which are not equivalent to the GREG do not necessarily possess the ADU property (Estevao & Särndal, 2000). The authors, as do others, restricted their examination to the portion of the calibration family that is ADU.

The bias of GREG ratio estimators in comparison are generally assumed to be small such as bias of order $O(n_A^{-1})$ for a simple random sampling (SRS) design of size n_A (see, e.g., Section 7.3.1 of Särndal et al., 1992). The bias in general is a function of the variation in the denominator term plus the association between the numerator and denominator. For example, the expectation of $\hat{\bar{y}}_{TRGR}$ with respect to the analytic survey design (E_A) begins with a second-order Taylor series approximation centered around (t_y, N):

$$\begin{aligned}\hat{\bar{y}}_{TRGR} &= \frac{\hat{t}_{yTRGR}}{\hat{N}_{TRGR}} \\ &\cong \bar{y} + \frac{1}{N}\left(\hat{t}_{yTRGR} - t_y\right) - \frac{t_y}{N^2}\left(\hat{N}_{TRGR} - N\right) \\ &+ \frac{1}{2}\left[0 \times \left(\hat{t}_{yTRGR} - t_y\right)^2 + 2\frac{t_y}{N^3}\left(\hat{N}_{TRGR} - N\right)^2 - \right. \\ &\left. - \; 2\frac{1}{N^2}\left(\hat{t}_{yTRGR} - t_y\right)\left(\hat{N}_{TRGR} - N\right)\right]\end{aligned}$$

with $\hat{t}_{yTRGR}$ defined in (2.5); $\hat{N}_{TRGR}$ calculated by substituting $\mathbf{t}_x$ for $\mathbf{t}_{Bx}$ in $\hat{N}_{GREG}$ (2.13); and $\bar{y} = t_y/N$, the true population mean. The design-based bias is then calculated as:

$$\begin{aligned}Bias\left(\hat{\bar{y}}_{TRGR}\right) &= E_A\left(\hat{\bar{y}}_{TRGR}\right) - \bar{y} \\ &\cong \frac{1}{N}(0) - \frac{t_y}{N^2}(0) + \frac{t_y}{N^3}Var\left(\hat{N}_{TRGR}\right) - \frac{1}{N^2}Cov\left(\hat{t}_{yTRGR},\ \hat{N}_{TRGR}\right) \\ &= \frac{1}{N^2}\left[\bar{y}Var\left(\hat{N}_{TRGR}\right) - Cov\left(\hat{t}_{yTRGR},\ \hat{N}_{TRGR}\right)\right] \qquad (2.16)\end{aligned}$$

where $E_A\left(\hat{t}_{yTRGR}\right) \cong t_y$ and $E_A\left(\hat{N}_{TRGR}\right) \cong N$ as assumed in Särndal et al. (1992, Section 6.6). For large finite populations of size N, the terms $Var\left(\hat{N}_{TRGR}\right)/N^2$ and $Cov\left(\hat{t}_{yTRGR},\ \hat{N}_{TRGR}\right)/N^2$ are negligible, $O\left(n^{-1}\right)$, so that the claim of small bias for traditional calibration appears reasonable. Unfortunately, the absolute value of the bias can change dramatically with the introduction of estimated controls in the numerator and denominator (see Chapter 5).

The theoretical development presented above and in the literature relies on the assumption of negligible errors in the data used to calculate the estimates (e.g., coverage, nonresponse, and measurement). We extend this research by allowing for undercoverage bias in our estimates. For example, HT estimators, such as $\hat{t}_{Ay}$ in (2.10), are known to be design unbiased under perfect survey conditions (i.e., no nonresponse, no frame errors, etc.) and biased otherwise. To account for the possibility of frame undercoverage, we assume that the presence of each population unit on the sampling frame can be modeled as a random event. The expectation of a HT total estimated from an SRS sample of size n_A with 100 percent response but selected from a frame suffering from undercoverage is evaluated below. Here, E_{c_A} and E_A represent the expectations with respect to the frame coverage propensities and the sample selection given the set of units, c_A, covered by the analytic survey sampling frame, respectively:

$$
\begin{aligned}
E(\hat{t}_{Ay}) &= E_{c_A}\left[E_A\left(\sum_{k\in s_A}\pi_k^{-1}y_k \mid c_A\right)\right] \\
&= E_{c_A}\left[E_A\left(\sum_{k\in U}I_{Ak}C_{Ak}\pi_k^{-1}y_k|c_A\right)\right] \\
&= \sum_{k\in U}E_A(I_{Ak}|c_A)E_{c_A}(C_{Ak})\pi_k^{-1}y_k \\
&= \sum_{k\in U}\phi_{Ak}y_k \\
&\equiv t_{Ay} \qquad (2.17)
\end{aligned}
$$

where $C_{Ak}=1$ if the k^{th} unit is listed on the analytic survey sampling frame (zero otherwise) so that $E_{c_A}(C_{Ak})=\phi_{Ak}$, the population propensity for inclusion on the sampling frame; and $I_{Ak}=1$ if the same unit is selected into the sample (zero otherwise), so that $E_A(I_{Ak}\mid c_A)=\pi_k$, the inclusion probability for unit k. Only if $\phi_{Ak} = 1$ for all units in the population (i.e., no undercoverage in the sampling frame) can we claim unbiasedness, $E(\hat{t}_{Ay}) \equiv t_y$. This issue is further developed for EC calibration with complex analytic survey designs in subsequent chapters.

The first assumption listed previously for traditional calibration estimation is that the control totals are known without error. Most of the real-world examples presented in this chapter are actually calibration to estimated control totals generated from other surveys instead of calibration to population values as assumed. We coin the term "estimated-control calibration" or "EC calibration" to distinguish from the traditional or fixed-control calibration. The possible exception is with person surveys administered in Scandinavian countries (Denmark, Finland, Iceland, Norway, and Sweden). These countries maintain total population registers including identifying information such as name, address, and personal identity number (e.g., Särndal & Lundström, 2005). Scandinavian surveys calibrating to the population registers may be classified as traditional calibration if one is willing to assume that there are no errors in the register.

Some researchers acknowledge that the controls are taken from benchmark surveys. However, many of these same researchers assume that the mean square error (MSE) associated with the benchmark controls is negligible without completely understanding if or when these errors can be ignored. For example, estimates from the Current Population Survey (CPS) and counts from the Decennial Census (Census) are regularly cited as sources for calibration controls due to their size, extent of the data collected, high levels of accuracy, and perceived low levels of error. The CPS is a source for U.S. labor-force statistics. Data are gathered for the civilian non-institutionalized population, 16 years of age and older, each month through in-person and telephone interviews. Though the CPS design weights are adjusted for undercoverage, Nadimpalli et al. (2004) relate the "negative side" of using CPS estimates for calibration controls to the unknown undercoverage errors between and within households. Weinberg (2006) discusses the biasing impact of undercoverage related to CPS income estimates. The explicit purpose of the Post-Enumeration Survey as summarized in Chao & Tsay (1998) is to estimate the undercount in the U.S. Census by various demographic groups. Stepping away from the coverage issue, West et al. (2005) relate the age rounding (response) errors reported in the Census to data collected from less than knowledgeable proxy respondents. Additionally, both large-scale surveys suffer from nonresponse. Another example focuses on adjustments for differential nonresponse. Researchers calculate a "nonresponse multiplier" (i.e., nonresponse adjustment weight) for non-white respondents to the British Crime Survey by calibrating the design weights to the ethnic group, age, and gender distributions estimated from the British Labour Force Survey (Bolling et al., 2006, Section 7.4). However, the variance estimation discussion seems to indicate that the benchmark controls are treated as population values.

Our last example, potentially with stronger implications, comes from a Web survey with sample members identified through a volunteer (non-random) panel. Terhanian et al. (2000) calibrate the weights for the Web responses to the distribution of characteristics within an RDD telephone survey by assuming the latter to be "relatively free of bias." We can only assume that the benchmark RDD survey discussed here is typical in that it suffers from low levels of response because such information was not provided in the white paper.

On the surface, bias (and variance) implications for calibrating the analytic survey to either the CPS or Census estimates would seem minor in comparison to the Web/RDD example. Särndal et al. (1992) state in Remark 6.4.3, "If erroneous totals are used, the estimator is biased." However, the examination stops there without the much needed information that quantifies the level of bias and impact on MSE and variance estimates. Our research will provide this extension for weight calibration.

2.3 Variance of Calibrated Survey Estimates

An extensive list of references details variance estimation for weight calibration with population control totals. The variance estimation techniques include Taylor (series) linearization, jackknife replication, balanced repeated replication (BRR), bootstrap, and jackknife linearization. We focus specifically on Taylor linearization and jackknife replication in our current research and in the discussion given below (Section 2.3.1). BRR variance estimation has been shown to be consistent for all types of estimators, including non-smooth statistics such as quantiles (Rao & Shao, 1999), and is therefore of particular interest of future work. A few references also exist for the methodology we label as estimated-control (EC) calibration. These sources are briefly reviewed in Section 2.3.2.

2.3.1 Traditional Calibration

Taylor linearization, also known as the delta method, is a well known technique for approximating the mean and variance of linearizable (i.e., differentiable) complex statistics. These statistics include those with one or more random variables such as the regression estimator or the ratio of two estimated totals. Binder (1995) provides a step-by-step description of the linearization approach for several estimators including the GREG under single- and two-phase designs. For example, the poststratified estimator of a population total, $\hat{t}_{yTRPS}$ given in expression (2.8), can be linearly approximated (i.e., linearized) as follows:

$$\begin{aligned} \hat{t}_{yTRPS} &= \sum_{g=1}^{G} N_g \hat{N}_{Ag}^{-1} \hat{t}_{Ayg} \\ &\cong \; t_y + \sum_{g=1}^{G} \left[\frac{\partial \hat{t}_{yTRPS}}{\partial \hat{t}_{Ayg}} |_{t_{yg}} \left(\hat{t}_{Ayg} - t_{yg} \right) + \frac{\partial \hat{t}_{yTRPS}}{\partial \hat{N}_{Ag}} |_{N_g} \left(\hat{N}_{Ag} - N_g \right) \right] \end{aligned} \qquad (2.18)$$

where "$|_{t_{yg}}$" refers to the partial derivatives evaluated at the population parameters. Under some reasonable conditions, the second- and higher-order terms converge in probability to zero at faster rates than the remaining terms, thereby justifying the approximation.

Särndal et al. (1989) developed an approximate linearization population sampling variance (AV) for $\hat{t}_{yGREG}$ (2.10) as a function of population or "census fit" residuals determined from an assisting model — see discussion of the model for expression (2.6). Using notation from Section 6.5 of Särndal et al. (1992), the general form of the approximate linearization variance estimator is calculated as follows:

$$AV_{TS}(\hat{t}_{yGREG}) = \sum_{k\in U}\sum_{l\in U}(\pi_{kl} - \pi_k\pi_l)\left(\frac{E_k}{\pi_k}\right)\left(\frac{E_l}{\pi_l}\right) \tag{2.19}$$

where $E_k = y_k - \mathbf{x}_k'\mathbf{B}$, the assisting model population residual for unit k determined by regressing the outcome y on the auxiliary variables $\mathbf{x}$; $\mathbf{B} = \left(\sum_{l\in U}\mathbf{x}_l\mathbf{x}_l'\right)^{-1}$ $\times \sum_{k\in U}\mathbf{x}_k y_k$, the vector of regression coefficients; π_k is the analytic survey sample inclusion probability for the k^{th} population unit; and π_{kl} is the joint inclusion probability for units k and l. This approximate variance incorporates only the first-order linearization terms and is therefore not an exact estimator. Expression (2.19) is tailored to various types of GREG estimators by choosing an appropriate assisting model which generates different E_k's. For example, an assisting model defined by $E_\epsilon(y_k) = B_g$ and $Var_\epsilon(y_k) = \sigma_g^2$ generates residuals associated with the poststratified estimator $\hat{t}_{yTRPS}$ (2.8).

Särndal et al. (1992) and Stukel et al. (1996), among others, discuss a design-consistent sample variance estimator for expression (2.19) under a general sampling design:

$$var_{TS}(\hat{t}_{yGREG}) = \sum_{k\in s_A}\sum_{l\in s_A}\frac{\pi_{kl} - \pi_k\pi_l}{\pi_{kl}}\left(\frac{a_k e_k}{\pi_k}\right)\left(\frac{a_l e_l}{\pi_l}\right) \tag{2.20}$$

where $e_k = y_k - \mathbf{x}_k'\hat{\mathbf{B}}_A$, the sample estimated residual from the assisting model; $\hat{\mathbf{B}}_A$, the sample-based vector of regression coefficients defined in (2.10) that is assumed to be an approximately unbiased estimator of $\mathbf{B}$; and a_k is the calibration-adjustment factor for unit k also defined for expression (2.10). Särndal et al. (1992) also note that the confidence interval coverage rates associated with (2.20) are near or exactly equal to the specified levels (e.g., 95 percent). For the claim of design-consistency of $var_{TS}(\hat{t}_{yGREG})$ (2.20) to hold, Särndal et al. (1992) require (*i*) the assisting model to be a reasonable representation of the population in that the residuals are small and the "residual variance is small compared to the total variance" of the estimate; (*ii*) a consistent system of calibration equations, i.e., $\sum_{k \in s_A} a_k \pi_k^{-1} \mathbf{x}_k$ is equal to $\mathbf{t}_x$ as specified in (2.2); and (*iii*) $\left(\sum_{l \in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l'\right)^{-1} \sum_{k \in U} \mathbf{x}_k \mathbf{x}_k'$ converges elementwise in probability to one. The first condition leads to the claim that the general form of (2.20) is approximately design-unbiased regardless of the difference between the working and population assisting models because this difference converges to zero in "model probability" (also see Särndal et al., 1989; Deville & Särndal, 1992). Hedlin et al. (2001), however, warn that this condition is not always satisfied making the GREG susceptible to model misspecification and emphasize the importance of model diagnostics to assess model quality.

The difficulties of applying the sample variance estimator (2.20) increase with the complexity of the population estimator. For example, the variance of the post-stratified estimator $\hat{t}_{yPSGR}$ (2.11) uses an approximation similar to $\hat{t}_{yTRPS}$ in (2.18) and requires the estimation of several variance and covariance estimates. The *linear substitute method* eliminates the need for these higher-order estimates (Woodruff,

1971) as shown for $\hat{t}_{yPSGR}$ below:

$$\begin{aligned}
\hat{t}_{yPSGR} - t_y &\cong \sum_{g=1}^{G}\left[\left(\hat{t}_{Ayg} - t_{Ayg}\right) - \frac{t_{Ayg}}{N_{Ag}}\left(\hat{N}_{Ag} - N_{Ag}\right)\right] \\
&= \sum_{g=1}^{G}\left[\hat{t}_{Ayg} - \frac{t_{Ayg}}{N_{Ag}}\hat{N}_{Ag}\right] \\
&= \sum_{g=1}^{G}\left[\sum_{k\in s_A}\pi_k^{-1}\delta_{gk}y_k - \frac{t_{Ayg}}{N_{Ag}}\sum_{k\in s_A}\pi_k^{-1}\delta_{gk}\right] \\
&= \sum_{k\in s_A}\pi_k^{-1}\sum_{g=1}^{G}\delta_{gk}\left(y_k - \frac{t_{Ayg}}{N_{Ag}}\right) \\
&= \sum_{k\in s_A}u_k \qquad (2.21)
\end{aligned}$$

where $\delta_{gk} = 1$ if the k^{th} analytic survey sample unit is a member of the g^{th} poststratum (zero otherwise); and, $E\left(\hat{t}_{Ayg}\right) = t_{Ayg}$ and $E\left(\hat{N}_{Ag}\right) = N_{Ag}$ using the technique demonstrated in expression (2.17). The linear substitute u_k is estimated from the sample data, and $var_{TS}\left(\hat{t}_{yPSGR}\right)$ is estimated from a design-appropriate variance estimator of (2.21). A linear-substitute variance estimator of $\hat{t}_{yGREG}$ described in Stukel et al. (1996) takes the following form for a stratified, multi-stage analytic survey sampling design:

$$var_{LS}(\hat{t}_{yGREG}) = \sum_{h=1}^{H}\frac{m_{Ah}}{m_{Ah}-1}\sum_{i\in s_{Ah}}\left(\sum_{k\in s_{Ahi}}\frac{a_k e_k}{\pi_k} - \frac{1}{m_{Ah}}\sum_{i\in s_{Ah}}\sum_{k\in s_{Ahi}}\frac{a_k e_k}{\pi_k}\right)^2 \qquad (2.22)$$

where h identifies the sampling strata in the analytic survey design ($h = 1, ..., H$); a set s_{Ah} of m_{Ah} PSUs is selected from M_{Ah} within stratum h; and a random sample of units, denoted as s_{Ahi}, is selected within PSU hi. The remaining terms are defined for expression (2.20). Though not mentioned explicitly in their article, (2.22) is an "ultimate cluster" variance formula because the estimate is determined by calculating the variance of design-unbiased PSU-level estimates under an assumption of with-replacement PSU sampling (Kalton, 1979). Stukel et al. (1996), Särndal et al. (1992), and others note that $var_{LS}(\hat{t}_{yGREG})$ in (2.22) and $var_{TS}(\hat{t}_{yGREG})$ in (2.20) are asymptotically equivalent. Linear-substitute variance estimators, however, are computationally easier to use and are therefore included in many software packages (see, e.g., SUDAAN® documentation Research Triangle Institute, 2004).

Following the derivation in (2.21), the population linear substitute for the estimated mean $\hat{\bar{y}}_{TRGR}$ is calculated through a first-order Taylor series approximation as follows:

$$\begin{aligned}
\hat{\bar{y}}_{TRGR} - \bar{y} &\cong \frac{1}{N}\left(\hat{t}_{yTRGR} - t_y\right) - \frac{\bar{y}}{N}\left(\hat{N}_{TRGR} - N\right) \\
&= \frac{1}{N}\left(\hat{t}_{yTRGR} - \bar{y}\hat{N}_{TRGR}\right) \\
&= \sum_{k\in s_A} \pi_k^{-1} a_k \frac{1}{N}\left(y_k - \bar{y}\right) \\
&= \sum_{k\in s_A} u_k
\end{aligned} \tag{2.23}$$

for a_k defined in (2.5), $\hat{t}_{yTRGR} = \sum_{k\in s_A} \pi_k^{-1} a_k y_k$, and $\hat{N}_{TRGR} = \sum_{k\in s_A} \pi_k^{-1} a_k$. As discussed in Särndal et al. (1992, chapter 6), $\hat{t}_{yTRGR}$ and $\hat{N}_{TRGR}$ may be approximated as follows using the residuals from the respective assisting models:

$$\hat{t}_{yTRGR} \cong \sum_{k\in s_A} \pi_k^{-1}\left(y_k - \mathbf{x}_k'\mathbf{B}_A\right) = \sum_{k\in s_A} \pi_k^{-1} E_{Ak}$$

and

$$\hat{N}_{TRGR} \cong \sum_{k \in s_A} \pi_k^{-1} \left(1 - \mathbf{x}_k' \mathbf{B}_{AN}\right) = \sum_{k \in s_A} \pi_k^{-1} E_{ANk}.$$

These approximations require $(\hat{\mathbf{t}}_{Ax} - \mathbf{t}_x)$ to be of order (in probability) population (PSU) size divided by the square root of the sample (PSU) size. Therefore, the linear substitute of $\hat{\bar{y}}_{TRGR}$ may be approximated as

$$u_k \cong \frac{1}{N} \pi_k^{-1} \left(E_{Ak} - \bar{y} E_{ANk}\right)$$

where $E_{Ak} = y_k - \mathbf{x}_k' \mathbf{B}_A$ and $E_{ANk} = 1 - \mathbf{x}_k' \mathbf{B}_{AN}$. The corresponding g-weighted (a_k) sample estimator is used to generate the approximately unbiased linear-substitute sample variance estimator, i.e.,

$$\breve{u}_k = \frac{1}{\hat{N}_{TRGR}} a_k \pi_k^{-1} \left(e_{Ak} - \hat{\bar{y}}_{TRGR} e_{ANk}\right)$$

for the calibration-adjustment factor a_k defined in (2.5); $e_{Ak} = y_k - \mathbf{x}_k' \hat{\mathbf{B}}_A$; and $e_{ANk} = 1 - \mathbf{x}_k' \hat{\mathbf{B}}_{AN}$.

Linearization variance estimation is an option in several software packages designed to analyze survey data. Data files need to contain relevant information such as first-stage strata and PSUs to properly account for the sampling design. For example, the Division of Health Interview Statistics at the National Center for Health Statistics (NCHS) released public-use data files (National Center for Health Statistics, 2006) from the NHIS with such information along with code to produce linearization variance estimates using SUDAAN® (Research Triangle Institute, 2004). However, some organizations choose to withhold the design information from public-use files (PUFs) as an additional step to mask the identity of survey participants (i.e., data confidentiality procedures). For example, the UCLA Center for Health Policy Research (2006) states the following in their weighting and variance estimation document for the California Health Interview Survey (CHIS): "The CHIS PUFs, however, do not include strata information in order to protect data confidentiality and respondent privacy." Therefore, linearization variance estimation is not possible from the CHIS and many other publicly available survey data sets. Instead, replication methods such as jackknife variance estimation are required.

Jackknife variance estimation is a commonly used replication method. Formulae for the jackknife variance are available for single-stage designs, with and without stratification, as well as for more complex designs through an ultimate cluster formulation. The stratified formula is applicable to survey designs with two or more PSUs selected per stratum ($m_{Ah} \geq 2$), and to a wide array of estimates including means, totals, and more complex statistics.

The standard "delete-one" or "delete-a-PSU" jackknife is calculated through the variance of the replicate population estimates, also referred to as the pseudovalues (Wolter, 2007, Chapter 4). The m_A replicate estimates are calculated in the same way as the full sample estimates but require the generation of jackknife weights. The jackknife weights are created by systematically removing one the m_A PSUs from the sample and inflating the design weights for the remaining PSUs within the same stratum by $m_{Ah}/(m_{Ah} - 1)$ to account for the PSU subsampling.

The complete sets of m_A jackknife weights are included on the analysis file and used to generate variance estimates for various types of statistics.

Once the replicate estimates are calculated, statisticians must choose from among a set of formulae to calculate the jackknife variance estimate. The formulae vary based on the centering value. For example, the stratified variance estimator v_4 shown below is centered on the full-sample estimate $(\hat{\theta})$, and is classified as a conservative estimator as discussed in Wolter (2007, Section 4.5):

$$v_4 \equiv var_{JK}\left(\hat{\theta}\right) = \sum_{h=1}^{H} \frac{m_{Ah} - 1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{\theta}_{(hr)} - \hat{\theta}\right)^2 \tag{2.24}$$

where $\hat{\theta}_{(hr)}$ is the replicate estimate calculated after removing the r^{th} PSU from the h^{th} stratum and adjusting the remaining PSUs in stratum h for the loss. Krewski & Rao (1981), in addition to Wolter (2007), discuss other jackknife variance estimators including the less conservative estimator centered on the average of the replicate estimates, i.e., $(\sum_h \sum_r \hat{\theta}_{(hr)})/\sum_h m_{Ah}$, referred to as v_2 in Wolter (2007, Section 4.5). Rust & Rao (1996) demonstrate the unbiasedness of the v_2 variance estimator for a population total and other types of linear estimators. The estimator is also design consistent for nonlinear estimators such as the ratio estimator. The consistency property also holds when the design weights are adjusted for nonresponse and for poststratification but is lost for non-smooth statistics such as quantiles (Yung & Rao, 1996, 2000). The estimators v_2 and v_4 (2.24), however, have been shown to be asymptotically equivalent so that the choice of estimators is related to the sampling design or statistical preference (Krewski & Rao, 1981; Wolter, 2007).

Jackknife variance estimation for single-stage surveys and those with a large number of PSUs can be problematic in two ways: (*i*) additional time is required to produce and check the jackknife weights, and (*ii*) the analysis file size increases with the inclusion of a large set of jackknife weights. The "delete-a-group" technique is the same as described above except for the deletion of a *group* of PSUs instead of one PSU for each replicate. Valliant et al. (2008) warn that jackknife variance estimate for a population total under this method can result in a severely overestimated variance when the groups do not contain an *equal* number of PSUs. They suggest, as does Kott (2001), a revised variance estimator to account for this problem. Due to the complexity of issue, we postpone for now an examination of the "delete-a-group" jackknife for EC-calibrated estimators.

Linearization variance estimators involving assisting-model residuals, such as the estimator in expression (2.22), usually account only for the last (random) adjustment applied to the weights, e.g., calibration. This is in contrast to accounting for all random weight adjustments (e.g., unknown eligibility, nonresponse, etc.). Replication may remedy this problem by explicitly accounting for all adjustments applied to the design weights. Valliant (1993), for example, showed that jackknife variance estimators are consistent for two-stage sampling design *only* if the post-stratification adjustments are newly applied for each replicate. However, Rao and Shao (1992) showed that re-imputing missing data within each replicate does not give a consistent variance estimate

Variance estimation in the literature also extends to *multi-phase sample designs*. Techniques used in the development of these variance estimators are useful to our research. A multi-phase design is defined as a survey where subsequent-phase units are subsampled from the same type of units identified in a previous phase. Hence, the phase-specific samples are not independent. For example, a design that includes the recontact of a nonrespondent subsample to improve response rates is known as a "two-phase design with a nonresponse follow-up." In contrast, a multi-stage design contains units from differing levels at each stage such as schools within say the third stage and students within schools at the fourth stage of sampling. Note that a nonresponse follow-up phase is dependent on the result of the previous phase(s), which has implications for point and variance estimation. This is not a problem for EC calibration when the control totals are obtained from an independent survey.

Fuller (2004) provides a lengthy reference list in the development of variance estimation for two-phase designs beginning with Rao (1973). Särndal et al. (1992, Result 9.7.1), Binder (1995), Axelson (2000), Fuller (2000), Estevao & Särndal (2002) and others specifically address linearization variance. Replication variance estimation is presented in Fuller (1998) and later expanded by Kim & Sitter (2003). Fuller (2004) extends his own work for regression estimators by developing a two-phase variance formula and providing the relevant asymptotic theory to demonstrate consistency. The theory requires a relatively large phase-two sample and replicates created using the phase-one sample. The basis for the two-phase derivations comes from results for the unconditional expectation and variance of a general estimator (see, e.g., Casella & Berger, 2002, Theorems 4.4.3 and 4.4.7):

$$E(\hat{\theta}) = E_b[E_a(\hat{\theta}|b)]$$

$$Var(\hat{\theta}) = E_b[Var_a(\hat{\theta}|b)] + Var_b[E_a(\hat{\theta}|b)]$$

where, under a two-phase design, subscript a may denote the second-phase sample conditioned on the first-phase results (subscript b). A similar procedure is needed for the development of EC calibration where the benchmark and analytic surveys are associated with the first- and second-phase notation.

2.3.2 Estimated-Control Calibration

Variance estimation for EC calibration is not a new concept; a few articles propose methods to account for the estimated controls. For example, Isaki et al. (2004) applied a delete-one jackknife variance estimator developed by Fuller (1998) for two-phase designs to account for estimated control totals. An overview of Fuller's variance estimator is as follows: take a spectral (eigenvalue) decomposition of the covariance matrix for the vector of G benchmark controls, develop benchmark adjustments as a function of the resulting eigenvalues and eigenvectors, and add the adjustments to the benchmark controls to create a set of replicate controls. Thus, either a benchmark analysis file is needed to calculate the covariance matrix, or the statistician is forced to use only publicly-available benchmark information. A randomly chosen subset of the m_A replicates ($m_A \geq G$) is then calibrated to G replicate controls where $m_A = \sum_h m_{Ah}$, the total number of PSUs in the sample. The resulting variance estimator is shown to be an approximately unbiased estimator of the population sampling variance and to contain components for the variation within the analytic and benchmark surveys, both as desired. A more extensive simulation study is needed to empirically demonstrate the theoretical findings. Also, the methodology does not address coverage error in either of the sampling frames.

Nadimpalli et al. (2004) calibrate weights for the 2003 National Survey of Parents and Youth (NSPY) to the number of U.S. households with children ages 9–18 estimated from the Current Population Survey (CPS) using a ratio-raking replicate algorithm (www.census.gov/cps). The U.S. Bureau of Labor Statistics conducts the CPS to obtain labor force characteristics for the population ages 16 years and older. They note, however, that the calibration controls change depending on the month of CPS data used in the calculation. The focus of their paper (not of particular interest here) is to evaluate several models for smoothing the monthly estimates to develop a single set of stable marginal control totals by domain such as region of the U.S. The authors were unable to estimate the complete covariance matrix ($\hat{\mathbf{V}}_B$) for $\hat{\mathbf{t}}_{Bx}$ from, for example, a public-use file, and therefore had to assume independent benchmark estimates. To account for the random nature of the CPS controls in the NSPY variance estimates, they assume that the marginal control totals are approximately normally distributed and incorporate a standard normal random variable, $N(0, 1)$, into the equation for a replicate control total. Griffiths (2007) also applies a method similar to Nadimpalli et al. (2004) for calibration of Arbitron data to "stochastic population controls" which again requires the assumption of independent control totals. Unlike Fuller (1998), which specifically focuses on the development of an EC-calibration procedure, the Nadimpalli et al. (2004) paper provides an application of a proposed method. Their research requires additional theoretical work to understand the bias associated with their method.

The American Time Use Survey (ATUS), conducted jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (www.bls.gov/tus), produces estimates of how people in the U.S. spend their time by various demographic characteristics. Samples are selected from CPS responding households that have completed the last in a series of interviews related to unemployment. The ATUS design weights are equivalent to the 161 CPS BRR final weights after adjusting for ATUS subsampling (Tupek 2004). The CPS BRR weights include components for the inclusion probabilities, nonresponse, poststratification of household-level weights to Census counts, raking of person-level weights to Census projections, and seasonal variation (Current Population Survey 2002). Details are lacking in the documentation on the methods used to account specifically for the Census projections. Additional factors are applied to the BRR weights based on the results from the ATUS such as adjustments for nonresponse, day of the week that the interview was conducted, and calibration to CPS microdata. A replication variance estimate should therefore account for the variation from both the analytic survey (ATUS) and the benchmark surveys (CPS and Census); however, published theory or analytic results to support this claim has not been located. Unlike the Fuller (2004) and Nadimpalli et al. (2004) methods, we reserve the ATUS (two-phase) methodology for future research because we have chosen to focus on studies with independent analytic and benchmark surveys.

Renssen & Nieuwenbroek (1997) develop an *adjusted general regression estimator* (AGREG) that calibrates weights from two independent surveys to population controls $\mathbf{t}_x$ and/or to controls estimated from either one or a combination of the surveys $\hat{\mathbf{t}}_z$. The estimated controls, using common variables $\mathbf{z}$ from both surveys, may be calculated with a general composite estimator of the form

$$\hat{\mathbf{t}}_z = \hat{\mathbf{P}}_v \hat{\mathbf{t}}_{z1} + \left(\mathbf{1} - \hat{\mathbf{P}}_v\right) \hat{\mathbf{t}}_{A_z 2}$$

where $\hat{\mathbf{t}}_{z1}$ is a vector of GREG estimates from the first survey sample calibrated to the population controls $\mathbf{t}_x$; $\hat{\mathbf{t}}_{z2}$ is the corresponding vector of GREG estimates from the second sample; and $\hat{\mathbf{P}}_v = var\left(\hat{\mathbf{t}}_{z1}\right)\left[var\left(\hat{\mathbf{t}}_{z1}\right) + var\left(\hat{\mathbf{t}}_{z2}\right)\right]^{-1}$, a matrix containing the proportion of the total variance associated with the first survey for each common variable. The matrix $\hat{\mathbf{P}}_v$ could also be set to a matrix of zeros or ones if estimates from one survey in comparison with the other are believed to be unusable. They suggest that large questionnaires may divided into smaller instruments (see, e.g., matrix sampling in Gonzalez & Eltinge, 2007) with key common variables, administered to independent samples to maximize response, and combined through the use of AGREGs. Their approximate population linearization sampling variance estimator accounts for the variation in the outcome and auxiliary variable estimates but not for the estimation of the composite factor $\hat{P}_v$. They compare the variance for estimates from a Dutch household survey using various sets of values of $\hat{\mathbf{P}}_v$ but do not provide further empirical evidence through a simulation study. We intend to expand on our current work in the future to address studies that involve, for example, matrix sampling with multiple independent subsamples but reserve our current work for only a single survey requiring calibration.

Another example of calibration for two-phase studies, of interest but not considered in our current research, focuses specifically on surveys with less than desired response rates. All sample cases are released in the first phase of this design, and a subsample of nonrespondents is recontacted in the second phase with a data collection mode different than the one used in prior contacts (i.e., nonresponse follow-up). Singh et al. (2003) expand upon the idea of *dual-frame calibration* developed by Singh & Wu (1996, 2003) by applying this estimator to two-phase designs with a nonresponse follow-up. The methodology requires the creation of two analysis files with nonresponse-adjusted design weights — one file contains only phase-one respondents, and the second file includes respondents from both phases. Using an algorithm that simultaneously satisfies the constraints, the estimates for each file are calibrated to the population control totals, while the difference between a set of estimates calculated from each file are calibrated to zero. Estimates from the two files are combined through a composite estimator in such a way that minimizes the variation in the calibrated weights (i.e., unequal weighting effects). Some theory is given in their proceedings paper with mixed results from the analysis of one survey of U.S. military veterans. Thus, additional work is needed to fully develop this methodology. Singh et al. (2004) implement this methodology to examine a new response rate calculation for studies with a nonresponse follow-up using the same example data.

In our final example, we focus on the *regression composite estimator* developed by Singh (1996) for rotating panel surveys. At each time period, the sample contains new cases (birth panel) and cases with previously collected data (overlap panel). The regression equations contain two components: (*i*) current time-point estimates for the birth and overlap panels are calibrated to the corresponding set of population controls; and (*ii*) estimated controls from the overlap sample are calculated from the previous round and used to calibrate prior-round estimates using the combined birth and overlap panel data. Fuller & Rao (2001) expand on this work by incorporating composite estimation to smooth the combined estimates from the birth and overlap panels for use in the regression. We have chosen to examine a single survey within our current research, instead of panel surveys, and therefore reserve this work for future consideration.

2.4 Domain Estimation

Domain or subpopulation estimation is critical to survey research. Surveys are generally designed to produce estimates within a set of domains with specified levels of precision. Two such examples, taken from U.S. surveys, are: (*i*) poverty rates can be compared across domains such as U.S. region and race/ethnicity by analyzing current CPS data; and (*ii*) estimated rates of illicit drug use for young adults aged 18 to 24 in the U.S. are produced from *National Survey on Drug Use and Health* data (SAMHSA, 2007).

Domains can be classified into two categories — design and analytic. *Design domains* are included in the sampling design either explicitly as strata or implicitly by constraining the expected number or level of precision. *Analytic domains* are identified only during the analysis phase of the study and may result in small domain sample size. Hidiroglou & Patak (2004) call these primary and secondary domains, respectively. Domain estimation can reduce the degrees of freedom for statistical tests and confidence intervals below levels defined for the full sample if domain members are not contained in all design strata and PSUs. The number of degrees of freedom is (roughly) the maximum of either the number of PSUs minus the number of strata ($m_A - H$), or the total number of replicates (m_A). Survey inference relies on large samples (i.e., degrees of freedom) along with the central limit theorem developed for finite populations (Krewski & Rao, 1981) under which point estimates will be approximately normally distributed. Korn & Graubard (1999) recommend reducing the degrees of freedom to the numbers of PSUs and strata which contain domain members. Therefore, the degrees of freedom can be managed for design domains but not analytic domains.

Much of the survey theory underlying traditional domain estimation assumes sufficient sample size regardless of the type of domain (see, e.g., Särndal et al., 1992; Rao, 1997; Théberge, 1999; Lohr, 1999; Korn & Graubard, 1999; Chambers & Skinner, 2003). This assumption is important to the development of an EC calibration theory for domains. Small area estimation techniques (e.g., Rao, 2003; Lohr & Prasad, 2003) are reserved for small domains, and are therefore excluded from consideration for our current work.

Overall sample estimators, such as the Horvitz-Thompson or Hájek estimators, are specialized for domain estimation by including a domain indicator variable. An estimate of a domain mean for a stratified design has the following form where δ_{dhk} is a binary variable to indicate membership in domain d:

$$\hat{\bar{y}}_{Ad} = \frac{\hat{t}_{yAd}}{\hat{N}_{Ad}} = \frac{\sum_{h=1}^{H}\sum_{k=1}^{m_{Ah}} \delta_{dhk}\pi_{hk}^{-1} y_{hk}}{\sum_{h=1}^{H}\sum_{k=1}^{m_{Ah}} \delta_{dhk}\pi_{hk}^{-1}}. \tag{2.25}$$

The form of the GREG estimator for domains is not so straight forward. Researchers may assume that a working model defined for each domain, i.e., $E_{\epsilon}(y_k) = \mathbf{x}'_{dk}\mathbf{B}_{dd}$, results in more efficient estimators than those generated from an overall model. This leads to the following GREG estimator of a domain total for a non-specific sampling design:

$$\hat{t}_{yddGREG} = \hat{t}_{Ayd} + \left(\mathbf{t}_{Bdx} - \hat{\mathbf{t}}_{Adx}\right)' \hat{\mathbf{B}}_{Add} \tag{2.26}$$

where $\hat{t}_{Ayd} = \sum_{k\in s_A} \delta_{dk}\pi_k^{-1} y_k$, the estimated total of y within domain d using the analytic survey data; $\hat{\mathbf{t}}_{Adx} = \sum_{k\in s_A} \delta_{dk}\pi_k^{-1}\mathbf{x}_k$, the G-length vector of analytic survey auxiliary values for domain d; $\mathbf{t}_{Bdx}$, the corresponding vector of domain-specific benchmark controls; and

$$\hat{\mathbf{B}}_{Add} = \left[\sum_{l\in s_A} \delta_{dl}\pi_l^{-1}\mathbf{x}_l\mathbf{x}'_l\right]^{-1} \sum_{k\in s_A} \delta_{dk}\pi_k^{-1}\mathbf{x}_k y_k,$$

the model coefficient vector that is a function of the domain indicator in both the numerator and denominator. By decomposing (2.26), we see that the calibration-adjustment factor,

$$a_k = 1 + \left(\mathbf{t}_{Bdx} - \hat{\mathbf{t}}_{Adx}\right)' \left[\sum_{l \in s_A} \delta_{dl} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l'\right]^{-1} \mathbf{x}_k,$$

is a function of the domain indicator both in the denominator of $\hat{\mathbf{B}}_{Add}$, and in the auxiliary vectors $\mathbf{t}_{Bdx}$ and $\hat{\mathbf{t}}_{Adx}$. Even though Hidiroglou & Patak (2004) show the benefits of using domain-specific auxiliary variables (e.g., $\hat{\mathbf{t}}_{Adx}$), we have chosen to exclude this estimator from our current research because of our desire to create one set of analysis weights.

Another GREG estimator of a domain total which does satisfy the "one overall set of analysis weights" criterion is specified under a domain-specific assisting model that incorporates information from non-domain units, namely, $E_\epsilon(y_k) = \mathbf{x}_k' \mathbf{B}_d$ and $Var_\epsilon(y_k) = \sigma^2$. The resulting estimator is expressed as:

$$\hat{t}_{ydGREG} = \hat{t}_{Ayd} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_{Ad} \tag{2.27}$$

where $\hat{t}_{Ayd}$ is defined following (2.26), and $\hat{\mathbf{t}}_{Ax}$ and $\mathbf{t}_{Bx}$ are the vector of auxiliary values used in $\hat{t}_{yGREG}$ (2.10) that are not domain specific. The working model coefficient vector, $\hat{\mathbf{B}}_{Ad}$, incorporates the domain indicator only in the numerator term as seen below:

$$\hat{\mathbf{B}}_{Ad} = \left[\sum_{l \in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l'\right]^{-1} \sum_{k \in s_A} \delta_{dk} \pi_k^{-1} \mathbf{x}_k y_k. \tag{2.28}$$

We demonstrate the creation of a single set of generalized analysis weights with the following derivation:

$$
\begin{aligned}
\hat{t}_{ydGREG} &= \hat{t}_{Ayd} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_{Ad} \\
&= \sum_{k \in s_A} \left[1 + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{l \in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l' \right)^{-1} \mathbf{x}_k \right] \pi_k^{-1} \delta_{dk} y_k \\
&= \sum_{k \in s_A} a_k \pi_k^{-1} \delta_{dk} y_k
\end{aligned}
$$

where a_k equals the calibration-adjustment factor shown for $\hat{t}_{yGREG}$ (2.10). In addition to the advantages listed for a single-set of analysis weights, $\hat{t}_{ydGREG}$ (2.27) also has good theoretical properties. Estevao & Särndal (2004), for example, demonstrate the theoretical and empirical advantages of $\hat{t}_{ydGREG}$ over $\hat{t}_{yddGREG}$ (2.26), as well as a GREG domain estimator that "borrows strength" from non-domain cases through an overall model coefficient vector $\hat{\mathbf{B}}_A$ (2.10). This later domain estimator is constructed, unintuitively, as $\hat{t}_{Ayd} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A$. With domains that cut across the calibration groups, such as strata equivalent to the calibration groups but not the domains, the estimator is calculated as the sum of within-stratum values.

The GREG ratio estimator of a population mean within domain d is defined as follows:

$$
\hat{\bar{y}}_{dGREG} = \frac{\hat{t}_{ydGREG}}{\hat{N}_{dGREG}}, \tag{2.29}
$$

a function of $\hat{t}_{ydGREG}$ defined in expression (2.27);

$$
\hat{N}_{dGREG} = \hat{N}_{Ad} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_{ANd} \tag{2.30}
$$

defined in the same way as $\hat{t}_{ydGREG}$ but with $y_k = 1$ (i.e., $\hat{N}_{Ayd} = \sum_{k \in s_A} \delta_{dk} \pi_k^{-1}$);

and

$$\hat{\mathbf{B}}_{ANd} = \left[\sum_{l \in s_A} \pi_l^{-1} \mathbf{x}_l \mathbf{x}_l' \right]^{-1} \sum_{k \in s_A} \delta_{dk} \pi_k^{-1} \mathbf{x}_k. \tag{2.31}$$

The GREG estimator for a domain total (2.27) and a domain mean (2.29) are specialized for poststratification by expressing the assisting-model coefficient vector in terms of a domain group-mean model: $\hat{\mathbf{B}}_{Ad} \equiv \hat{\bar{\mathbf{Y}}}_{Ad} = \left[\hat{\bar{y}}_{Ad1}, \ ..., \ \hat{\bar{y}}_{AdG} \right]$ with $\hat{\bar{y}}_{Adg} = \hat{t}_{Aydg} / \hat{N}_{Ag}$, $\hat{t}_{Aydg} = \sum_{k \in s_A} \delta_{gk} \delta_{dk} \pi_k^{-1} y_k$, and $\hat{N}_{Ag} = \sum_{k \in s_A} \delta_{gk} \pi_k^{-1}$; and $\hat{\mathbf{B}}_{ANd} = \hat{\bar{\mathbf{Y}}}_{ANd} = \left[\hat{N}_{Ad1} / \hat{N}_{A1}, \ ..., \ \hat{N}_{AdG} / \hat{N}_{AG} \right]$ with $\hat{N}_{Adg} = \sum_{k \in s_A} \delta_{gk} \delta_{dk} \pi_k^{-1}$. Thus,

$$\hat{t}_{ydPSGR} = \sum_{g=1}^{G} N_{Bg} \left(\frac{\hat{t}_{Aydg}}{\hat{N}_{Ag}} \right)$$

and

$$\hat{\bar{y}}_{dPSGR} = \frac{\hat{t}_{ydPSGR}}{\hat{N}_{dPSGR}} = \frac{\sum_{g=1}^{G} N_{Bg} \left(\hat{t}_{Aydg} / \hat{N}_{Ag} \right)}{\sum_{g=1}^{G} N_{Bg} \left(\hat{N}_{Adg} / \hat{N}_{Ag} \right)}.$$

Some researchers choose to subset the analysis data to the domain of interest before calculating the population estimates. This technique works for point estimates such as means and totals. However, removal of units outside the domain of interest may inappropriately reduce the size of the variance estimate leading to confidence intervals that cover at less than the nominal rate and hypothesis tests with erroneously inflated Type I errors. Discussions on this point and other issues related to domain estimation may be found in sources such as Särndal et al. (1992), Lohr (1999), and Research Triangle Institute (see, e.g., SUDAAN® documentation 2004). For example, as shown in Example 10.3.1 of Särndal et al. (1992), the sample variance estimator of $\hat{t}_{Ayd}$ given for expression (2.26) is a function of the full sample size (n) in addition to the domain sample size (n_d) under an SRS design:

$$var\left(\hat{t}_{Ayd}\right) = N^2 \frac{(1-f)}{n} \left[\frac{(n_d - 1)\hat{S}^2_{yd} + n_d q_d \hat{\bar{y}}^2_d}{N-1} \right]$$

where $\hat{\bar{y}}_d = n_d^{-1} \sum_{k \in s_A} \delta_{dk} y_k$; $p_d = n_d/n$; $q_d = (1 - p_d)$; and $\hat{S}^2_{yd} = (n_d - 1)^{-1} \sum_{k \in s_A} \delta_{dk} \left(y_k - \hat{\bar{y}}_d\right)^2$. Note that this variance is a function of the variation within the domain ($\hat{S}^2_{yd}$), the average value of y within the domain ($\hat{\bar{y}}_d$), as well as the size of the domain (n_d).

Following the "domain indicator" approach, we rephrase the linear substitute sample variance estimator for $\hat{t}_{yGREG}$ (2.22) for the estimated domain total as follows:

$$var_{LS}(\hat{t}_{ydGREG}) = \sum_{h=1}^{H} c_h^{-2} \sum_{i=1}^{m_{Ah}} \left(\sum_{k \in s_{Ahi}} \frac{a_{hik} e_{dhik}}{\pi_{hik}} - \frac{1}{m_{Ah}} \sum_{i=1}^{m_{Ah}} \sum_{k \in s_{Ahi}} \frac{a_{hik} e_{dhik}}{\pi_{hik}} \right)^2 \quad (2.32)$$

where $c_h^{-2} = (m_{Ah} - 1)/m_{Ah}$, a function of the total number of analytic survey PSUs in stratum h (m_{Ah}); a_{hik} equals the calibration-adjustment factor a_k defined for $\hat{t}_{yGREG}$ (2.10); and, $e_{dhik} = \delta_{dhik} \left(y_{hik} - \mathbf{x}'_{hik} \hat{\mathbf{B}}_{Ad} \right)$. Note that if *all* units within stratum h are excluded from domain d (i.e., $\delta_{dhik} = 0$ for all $i, k \in s_{Ah}$), then these units do not contribute to the overall variance and can be safely removed from the analysis file, but not otherwise.

The v_4 jackknife variance estimator for a domain estimator is derived using a similar technique:

$$var_{JK}\left(\hat{\theta}_d\right) = \sum_{h=1}^{H} \frac{m_{Ah} - 1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{\theta}_{(hdr)} - \hat{\theta}_d\right)^2 \tag{2.33}$$

where $\hat{\theta}_{(hdr)}$ is the replicate domain estimate calculated after removing the r^{th} PSU from the sample and adjusting the remaining PSUs in stratum h for the loss, and $\hat{\theta}_d$ is the full-sample domain estimate. Note that even if $\hat{\theta}_{(hdr)} = 0$, the replicate contributes the value $\hat{\theta}_d^2$ to the overall variance.

The domain-specific variance estimators discussed in this section rely solely on the traditional calibration assumptions. The limited amount of EC-calibration research conducted to date (see Section 2.3.2) addresses overall population estimates and not estimation within a particular domain. The work detailed in Chapter 6 will combine domain estimation with EC-calibration estimators to fill this research gap.

Chapter 3

Scope of the Research

Chapter 1 provides an overview of our research while Chapter 2 contains a discussion of past and current literature on traditional and estimated-control (EC) calibration. In this chapter, we detail the scope of our research contained within the dissertation. Some material provided previously is repeated here for completeness. The assumptions made for the target population and the analytic survey are provided in Section 3.1. The conditions associated with the benchmark survey, from which the control totals are estimated, are highlighted in Section 3.2. We discuss issues related to the particular calibration technique used in our research (Section 3.3), in addition to factors affecting the quality of the analytic and benchmark survey sampling frames (Section 3.4). Section 3.5 is reserved for the assumptions related to domain estimation. Because a large number of survey estimators could be addressed, we identify the particular point and variance estimators examined in this body of work within Section 3.6. Additional assumptions required specifically for the theoretical understanding of EC calibration are identified in Section 3.7. The remaining section (Section 3.8) contains information associated with data used in our empirical simulation studies.

3.1 Analytic Survey Assumptions

Consider a large, finite population U of size N. We assume that this population can be divided into H ($H \geq 2$) mutually exclusive groups indexed by h. Within the h^{th} stratum, the population may be (conceptually) classified into M_h mutually exclusive clusters, each indexed by i, for a total of $M = \sum_{h=1}^{H} M_h$ clusters. The hi^{th} cluster contains a total of N_{hi} units, where $N = \sum_{h=1}^{H} N_h = \sum_{h=1}^{H} \left(\sum_{i=1}^{M_h} N_{hi} \right)$. The groups and clusters are classified as strata and primary sampling units (PSUs), respectively, for the sampling designs developed to estimate the relevant population parameters from U.

Estimates for the finite population U are calculated from survey data collected under a multi-stage, stratified sampling design. This survey is labeled as the *analytic survey* with random sample s_A. For the analytic survey design, m_{Ah} ($m_{Ah} \geq 2$) PSUs, each indexed by i, are selected *with replacement* (*WR*) from a total of M_{Ah} PSUs within the h^{th} stratum. Assuming *WR* sampling of PSUs is a common theoretical device to simplify derivations (e.g., see Krewski & Rao, 1981). Although most samples are selected without replacement, *WR* results provide a practical guidance on the performance of different procedures. The analytic survey sampling frame may have imperfections at each stage of sampling. Sampling frames suffering from undercoverage, i.e., not all population units are accessible from this source, are of particular interest to our research. Therefore, we say that $M_{Ah} \leq M_h$, where the subscript A denotes the analytic survey and M_{Ah} is the number of PSUs available for sampling on the analytic survey frame. The hi^{th} PSU inclusion probability in the WR design is $\pi_{hi} = 1 - (1 - \pi_{hi(1)})^{m_{Ah}}$, where $\pi_{hi(1)}$ is the single-draw inclusion probability. However, we assume that π_{hi} can be approximated by $m_{Ah}\pi_{hi(1)}$ by requiring $m_{Ah} \geq 2$ and $\pi_{hi(1)}$ to be sufficiently small (see, e.g., Särndal et al., 1992, Section 2.9).

Analytic units, units which provide survey response data, are selected in the last stage of the analytic survey design. Särndal et al. (1992, Section 4.1) call these *ultimate sampling units*. Once the sample PSUs are identified, the analytic units may be selected after more than one subsequent stage of sampling. A sample of n_{Ahi} units ($n_{Ahi} \geq 2$) is randomly selected from a total of N_{Ahi} ($N_{Ahi} \leq N_{hi}$) units within PSU hi. The units, indexed by k, are assumed to be selected with a method that results in unbiased estimates of a PSU total for various analysis variables. This assumption, in addition to a WR sample of PSUs, allows the use of "ultimate cluster" variance formulae (Kalton, 1979) in our research. Thus, our notation may be simplified to a two-stage design without loss of generality to multi-stage designs.

The unit-level design weight is represented as the inverse of the unconditional inclusion probability π_{hik}^{-1} for unit k within the hi^{th} PSU. Given our assumption that π_{hi}, is sufficiently approximated by $m_{Ah}\pi_{hi(1)}$, we say that π_{hik} is sufficiently approximated by $m_{Ah}\pi_{hi(1)}\pi_{k|hi}$, where $\pi_{k|hi}$ is the k^{th} inclusion probability given the selection of PSU hi. Note, however, that the point estimators we study will be formulated as "p-expanded with-replacement" (pwr) estimators (Särndal et al., 1992, Section 2.9). The pwr estimators are described in Section 3.6 and do not require that the PSU selection probabilities be approximated as above.

Data for a total of n_A units ($n_A = \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} n_{Ahi}$) is obtained for the analyses. This implies a 100 percent participation rate for the analytic survey. Even though unrealistic in practice, this assumption will facilitate development of the EC calibration theory before inclusion of the nonresponse mechanism in later work. Thus, prior to calibration, the population estimates are calculated using only the design weights. We additionally assume that data are collected without non-sampling errors.

3.2 Benchmark Survey Assumptions

We label the survey requiring calibration as the analytic survey and the source of the control totals under EC calibration as the *benchmark survey.* In practice, more than one benchmark survey may be tapped for control total estimates, though covariances among variables collected from different surveys may be difficult to estimate. However, we will assume only one benchmark survey to simplify the theoretical development and assume that the covariance matrix for the control totals ($\mathbf{V}_B$) can be estimated from the benchmark analysis file. We make no explicit specifications for the benchmark survey design though a stratified, multi-stage design would be a reasonable assumption. As with the analytic survey, we allow for potential errors in the benchmark survey sampling frame from which the random sample, s_B, of size n_B is selected. Hence, the subscript B is used to identify design elements and estimated values associated with the benchmark survey.

Control totals and the benchmark covariance matrix for the control totals are estimated from benchmark survey data using analysis weights, $\{w_l\}_{l=1}^{n_B}$, and formulae that properly account for the sample design. The analysis weights are functions of the design weights and any additional adjustment factors including nonresponse and calibration. The precision of the benchmark estimates reflect the random adjustments in as much as the final analysis weights allow (see the discussion of replicate weights in Section 2.3.1).

3.3 Calibration Procedure

Using the assisting-model approach of Särndal (2007), weight calibration can be classified as linear or nonlinear based on the type of model used to explain the association of auxiliary variables ($\mathbf{x}$) with an outcome (y). Weights generated through nonlinear calibration, such as those required for $\hat{t}_{yLGREG}$ (2.9), are a function of the outcome variable of interest. In other words, nonlinear calibration results in one set of analysis weights for *each* variable within a set of key measures. This trait adds to the unpopularity of calibration estimators such as the logistic generalized regression estimator (LGREG) proposed by Duchesne (2003). Linear calibration produces one set of analysis weights used to generate, for example, generalized regression estimators (GREG) and the specialized GREG known as the poststratified estimator. Both GREG estimators are widely used throughout survey research. Therefore, in our current research we choose to address linear calibration to generate parameter estimates that are functions of GREG estimated totals. Raking ratio (iterated) estimators are also excluded from our research because they do not have a closed-form solution to the calibration equations.

We additionally assume that calibration is implemented with estimates obtained in the last stage of the analytic survey sampling design. For example, in an area household survey, we assume calibration is implemented only for person-level estimates and not simultaneously for person- and household-level estimates (see, e.g., Estevao & Särndal, 2002; Ash, 2003). Additional work remains for EC calibration administered for multiple stages (and phases) of a design.

3.4 Sampling Frame Coverage Errors

A few additional comments are needed regarding the *sampling frames* for the analytic and benchmark surveys. Sampling frames are rarely considered to be perfect representations of the population. Frames may fail to contain all of the population units resulting in an undercoverage error. For example, a source for landline telephone numbers will miss cell-phone only households, as well as those without any telephone service.

Frames may also contain additional units referred to as overcoverage error. These sources include units that are not members of the target population (i.e., ineligibles) and units that are listed more than once (i.e., multiplicities). For example, samples selected from RDD telephone lists will likely contain inoperative numbers, in addition to multiple numbers linked to a single household.

Overcoverage error primarily reduces the number of analysis cases given that the erroneous units can be identified before the weights are finalized. The sample size, n_A using our notation, can be inflated for this potential loss of sample cases.

Undercoverage error, by contrast, can bias the estimated parameters, especially in population totals — see, for example, the discussion related to expression (2.1). Without access to a more complete sampling frame, researchers must rely on methods such as calibration to minimize the bias. Therefore, we choose to focus only on sampling frames that suffer from undercoverage errors.

With this undercoverage error, we say that the analytic and benchmark surveys estimate parameters from their "covered" populations — U_A of size N_A and U_B of size N_B, respectively. The population sizes, N_A and N_B, are assumed to be large which implies that the undercoverage errors are *not* so severe as to claim, for example, $N_A \ll N$ (i.e., N_A is significantly smaller than the complete population size N).

A coverage indicator is used to identify population units contained in the sampling frames: $C_{Ahik} = 1$ if the k^{th} population unit in PSU hi is accessible from the sampling frame used for the analytic survey (zero otherwise). We assume that the event c_A that determines the inclusion of the unit on the frame (i.e., $C_{Ahik} = 1$) is random and independent among the population units. This allows the use of a Bernoulli distribution to say $E_{c_A}(C_{Ahik}) = \phi_{Ahik}$ and $Var_{c_A}(C_{Ahik}) = \phi_{Ahik}(1 - \phi_{Ahik})$, where E_{c_A} and Var_{c_A} are the expectation and variance taken with respect to the coverage mechanism. The benchmark survey coverage indicator, C_{Bl} $(l \in s_B)$, and coverage process, c_B, are similarly defined.

3.5 Domain Estimation

Within the large, finite population U discussed in Section 3.1, let U_d represent the set of population units within domain d of size N_d. The domain members are identified through a binary variable denoted as δ_{dhik} where $\delta_{dhik} = 1$ if unit k in PSU i within stratum h is a member of domain d and $\delta_{dhik} = 0$ otherwise. We assume that δ_{dhik} is a fixed value and therefore, does contribute to the variance of the point estimator. The population domain size is determined by summing the domain indicators, i.e., $N_d = \sum_{hik \in U} \delta_{dhik}$. The domains may span the H design strata and need not be represented within each stratum nor within each PSU in a particular stratum. Thus, we denote the number of population PSUs containing a least one domain member as M_d.

Because we allow for undercoverage error, analytic survey domain estimates are associated only with the population parameter for those domain members listed on the sampling frame, i.e., U_{Ad} of size N_{Ad}. We assume that N_{Ad}, as well as the domain sample size $n_{Ad} = \sum_{hik \in s_A} \delta_{dhik}$, are sufficiently large so that small area estimation techniques are not required. We additionally assume that coverage mechanism (see the discussion of C_{Ahik} in Section 3.4) is independent of the (fixed) domain indicator. The domain sample units are contained within a total of $m_A d$ sample PSUs.

Calibration weights for domain estimation can take several forms depending on the level of information available from the benchmark survey (Estevao & Särndal, 2004). For example, benchmark control totals may be published for gender by age group but not by race/ethnicity, a domain of interest. Given that the relevant benchmark domain control totals exist, researchers could be faced with creating one set of weights for each key domain possibly in addition to an "overall estimation" set of weights. Keeping with our desire for one set of weights and the possibility that domain control totals are not available, we will use the same set of weights for both the overall and domain-specific population estimates. In other words, the calibrated weights for the domain estimates are functions of the overall auxiliary totals from the analytic and benchmark surveys.

3.6 Study Estimators

Through linear calibration, we will construct GREGs and poststratified (GREG) estimators to address *totals* and *ratios of two totals* for *all* population units and for a *domain* within the population. More complex point estimators, such as regression coefficients and quantiles, are reserved for future research.

Given the specified analytic survey sampling design (see Section 3.1), the estimator used to calculate the estimated population totals is the so-called "p-expanded with-replacement" (*pwr*) estimator discussed in Särndal et al. (1992, Section 2.9). This estimator is also known as the Hansen-Hurwitz estimator (Hansen & Hurwitz, 1943). For example, this *pwr* estimator for the total of y, using the analytic survey notation and suppressing the "pwr" subscript from the Särndal et al. (1992)

notation, is expressed as

$$\begin{aligned}
\hat{t}_{Ay} &= \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{m_{Ah}} \frac{\hat{t}_{Ayhi}}{\pi_{hi(1)}} \\
&= \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{y_{hik}}{\pi_{hi(1)}\pi_{k|hi}} \\
&= \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{y_{hik}}{\pi_{hik}} \qquad (3.1)
\end{aligned}$$

where $\hat{t}_{Ayhi} = \sum_{k=1}^{n_{Ahi}} \pi_{k|hi}^{-1} y_{hik}$. As discussed in Result 2.9.1 of Särndal et al. (1992), this *pwr* estimator is unbiased for the corresponding population total in *WR* PSU sampling. Note that we may simplify the complexity of estimator formulae throughout the remaining chapters when brevity is appropriate. For example, $\hat{t}_{Ay}$ may also be expressed as $\sum_{hik \in s_A} \pi_{hik}^{-1} y_{hik}$.

Taylor linearization and jackknife variance estimation techniques, either newly developed or extracted from the literature, are included in our development of EC calibration. Balanced repeated replication (BRR) variance estimation is needed to address the estimation of population quantiles but will not be covered here. Additional details related to the chosen variance estimators is provided in Section 2.3.

3.7 Assumptions for Asymptotic Theory

As discussed previously, we focus on stratified, multi-stage analytic survey sampling designs, where m_{Ah} PSUs ($m_{Ah} \geq 2$) are selected *with replacement* from within H design strata. The inclusion probability for PSU hi is assumed to be sufficiently approximated by $m_{Ah}\pi_{hi(1)}$. Additional assumptions are required to facilitate the development of asymptotic theory for our research:

- As $m_A = \sum_{h \in s_A} m_{Ah} \to \infty$ and $M = \sum_{h \in U} M_h \to \infty$, $\max_h \left(\frac{M_h}{M}\right)\left(\frac{m_A}{m_{Ah}}\right) = O(1)$. This assumption addresses two cases: (i) a fixed number of strata each containing a large number of PSUs, and (ii) a large number of strata each with a limited number of PSUs.

- The mean per population PSU $(\hat{t}/M)$ and the mean per population unit $(\hat{t}/N)$ are bounded in probability, where $\hat{t}$ is an unspecified sample total calculated from either the analytic or benchmark survey data. This allows statements such as $\hat{t}_{Ay}$ (3.1) is $O_P(M)$.

- The size of the analytic and benchmark PSU samples (m_A and m_B) are sufficiently large to support the claim that $E[f(\hat{\theta}_A, \hat{\theta}_B)] \cong f[E(\hat{\theta}_A), E(\hat{\theta}_B)]$, where $\hat{\theta}_A$ and $\hat{\theta}_B$ are the population estimators of interest from the analytic and benchmark surveys, respectively, and f is a differentiable function.

3.8 Data Source

Theory presented without empirical results to support the development is incomplete. We include the discussion of simulation studies conducted in R® (Lumley, 2005; R Development Core Team, 2005), and the corresponding results in the subsequent research chapters. The simulation population used in our research is a random subset of the 2003 National Health Interview Survey (NHIS) public-use file containing records for 21,664 U.S. residents. These records are contained within $H = 25$

design strata; $M_h = 6$ PSUs are associated with each stratum. Units within the sampling frame, from which the analytic survey samples are selected, are randomly chosen from the simulation population with varying degrees of undercoverage by age group and gender. Benchmark control-total covariance matrices are calculated from the complete NHIS public-use data file (92,148 records). Additional details on the simulation study are provided beginning in Section 4.5.1.

Chapter 4

Estimated Population Totals

4.1 Introduction

Särndal et al. (1992) among others demonstrate the good theoretical properties of the *traditional* generalized regression estimator (GREG) including asymptotic unbiasedness. We add to the literature in this chapter by detailing the theoretical properties of the GREG of a population total under estimated-control (EC) calibration. The form of the estimated-control generalized regression estimator (EC-GREG) of a population total is described in Section 4.2 using notation from Chapters 2 and 3. The specialized EC-GREG known as the estimated-control poststratified estimator (EC-PSGR) is also discussed. We examine factors that effect bias of the EC-GREGs in estimating the population total in Section 4.3. An evaluation of a set of sample variance estimators is discussed in Section 4.4, thereby allowing a complete picture of the mean square error properties of these new estimators. Our theoretical findings are validated with a simulation study in Section 4.5. We conclude the chapter with a summary of our research findings in Section 4.6.

4.2 Point Estimators

The EC-GREG of a population total, denoted as $\hat{t}_{yR}$ in our research, is obtained by replacing the vector of (presumed) population totals in the GREG, $\mathbf{t}_{Bx}$ in expression (2.10), with values estimated from a benchmark survey, $\hat{\mathbf{t}}_{Bx}$. Even though the values within the two vectors are the same, the "hat" notation in the latter vector identifies the set of control total estimates with a non-zero (sampling) covariance matrix, i.e., $var\left(\hat{\mathbf{t}}_{Bx}\right) = \hat{\mathbf{V}}_B$ versus $var\left(\mathbf{t}_{Bx}\right) = \mathbf{0}_G$, a G-length vector of zeroes. The use of differing notation becomes apparent in our discussion of variance estimators in Section 4.4. The formula for $\hat{t}_{yR}$ is explicitly expressed as:

$$\begin{aligned}
\hat{t}_{yR} &= \hat{t}_{Ay} + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_A \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \left[1 + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})' \left(\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{l=1}^{n_{Ahi}} \pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right)^{-1} \mathbf{x}_{hik}\right] \pi_{hik}^{-1} y_{hik} \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1} y_{hik} \; . \qquad (4.1)
\end{aligned}$$

We repeat the definition of terms in the formula above to facilitate the discussion in this chapter. The *pwr* estimator of the outcome variable y is defined as $\hat{t}_{Ay} = \sum_{hik\in s_A} \pi_{hik}^{-1} y_{hik}$, a function of the outcome variable values and the design weights, π_{hik}^{-1}, from the analytic survey sample s_A. We assume that π_{hik} for the k^{th} unit in PSU i within stratum h is reasonably approximated by $m_{Ah}\pi_{hi(1)}\pi_{k|hi}$, where (i) m_{Ah} out of M_{Ah} PSUs are selected with replacement within stratum h with a single-draw selection probability π_{hi1}, and (ii) n_{Ahi} out of N_{Ahi} units are selected with conditional probabilities $\pi_{k|hi}$. Additional details on the assumptions of the analytic survey design are provided in Section 3.1. The vector of *pwr* estimators for the auxiliary ($\mathbf{x}$) variables is similarly calculated as $\hat{\mathbf{t}}_{Ax} = \sum_{hik \in s_A} \pi_{hik}^{-1} \mathbf{x}_{hik}$ for the vector $\mathbf{x}_{hik}$ of size G. The corresponding G-length vector of auxiliary variable totals estimated from the benchmark survey is calculated as $\hat{\mathbf{t}}_{Bx} = \sum_{l \in s_B} w_l \mathbf{x}_l$, where w_l denotes the benchmark analysis weight (i.e., design weight adjusted for issues such as nonresponse) for the l^{th} sample unit in the benchmark survey sample s_B. We make no explicit statement about the estimates in $\hat{\mathbf{t}}_{Bx}$ because no assumptions have been made for the benchmark survey sampling design. The model-coefficient vector

$$\hat{\mathbf{B}}_A = \left[\sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{l=1}^{n_{Ahi}} \pi_{hil}^{-1} \mathbf{x}_{hil} \mathbf{x}_{hil}' \right]^{-1} \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} \mathbf{x}_{hik} y_{hik}, \tag{4.2}$$

also used in the calculation of $\hat{t}_{yGREG}$ in (2.10), is calculated based on the specification of a working population model, $y_{hik} = \mathbf{x}_{hik}' \mathbf{B} + E_{hik}$. Finally, $a_{hik} = 1 + \left[\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax} \right]' \left(\sum_{hil \in s_A} \pi_{hil}^{-1} \mathbf{x}_{hil} \mathbf{x}_{hil}' \right)^{-1} \mathbf{x}_{hik}$ is the calibration adjustment factor also referred to as a *g-weight* by Särndal et al. (1992) in Section 6.5.

Similarly, the estimated-control poststratified estimator (EC-PSGR) of a population total, $\hat{t}_{yP}$, is produced by replacing the population counts within the G poststrata $\left(\{N_{Bg}\}_{g=1}^{G} \right)$ with estimated benchmark survey counts $\left(\left\{ \hat{N}_{Bg} \right\}_{g=1}^{G} \right)$ in the formula for $\hat{t}_{yPSGR}$ shown in expression (2.11). The resulting estimator, a special case of the estimator in (4.1), has the form

$$\begin{aligned}
\hat{t}_{yP} &= \hat{\mathbf{N}}_B'\hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay} = \hat{\mathbf{N}}_B\hat{\bar{\mathbf{Y}}}_A \\
&= \sum_{g=1}^{G} \hat{N}_{Bg}\ \hat{N}_{Ag}^{-1}\ \hat{t}_{Ayg} \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \left[\sum_{g=1}^{G} \hat{N}_{Bg}\ \hat{N}_{Ag}^{-1}\ \delta_{ghik}\right] \pi_{hik}^{-1} y_{hik} \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1} y_{hik}. \qquad (4.3)
\end{aligned}$$

The G-length vector of poststratum counts estimated from the benchmark survey is denoted by $\hat{\mathbf{N}}_B = [\hat{N}_{B1}, ..., \hat{N}_{BG}]'$ where $\hat{N}_{Bg} = \sum_{l\in s_B} \delta_{gl} w_l$, the sum of the benchmark analysis weights for the units within poststratum g. The term $\delta_{gl} = 1$ if the l^{th} benchmark unit is a member of the g^{th} poststratum ($l \in s_{Bg}$), otherwise $\delta_{gl} = 0$. The poststratum sizes estimated from the analytic survey $\left(\hat{N}_{A1}, ..., \hat{N}_{AG}\right)$ are calculated by summing the design weights π_{hik}^{-1} across PSUs and design strata within each poststratum, i.e., $\hat{N}_{Ag} = \sum_{hik\in s_A} \delta_{ghik}\pi_{hik}^{-1}$ where $\delta_{ghik} = 1$ if $k \in s_{Ag}$ (zero otherwise). The estimated counts $\hat{N}_{Ag}$ are contained within a G-dimensional diagonal matrix, $\hat{\mathbf{N}}_A$. The term $\hat{t}_{Ayg} = \sum_{hik\in s_A} \delta_{ghik}\pi_{hik}^{-1} y_{hik}$ represents the *pwr* estimator of the total for y within the g^{th} poststratum and populates the G-length vector $\hat{\mathbf{t}}_{Ay}$. The vector $\hat{\bar{\mathbf{Y}}}_A = \hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay} = \left[\left(\hat{t}_{Ay1}/\hat{N}_{A1}\right),\ , ...,\ \left(\hat{t}_{AyG}/\hat{N}_{AG}\right)\right]'$ $= \left(\hat{\bar{y}}_{A1},\ ...,\ \hat{\bar{y}}_{AG}\right)'$. The calibration-adjustment factor for the k^{th} value in the $\hat{t}_{yP}$ calculation is $a_{hik} = \sum_{g=1}^{G}(\hat{N}_{Bg}/\hat{N}_{Ag})\delta_{ghik} = (\hat{N}_{Bg}/\hat{N}_{Ag})$ because $\delta_{ghik} = 1$ for only one (mutually exclusive) poststratum g.

In practice, the sample point estimates calculated under either the traditional or the estimated-control assumption will be numerically equal, though conceptually different. Numerical differences, however, do occur in the components of the mean square error (MSE) — namely squared bias and variance.

4.3 Bias of Point Estimators

The bias of an estimator $\hat{\theta}$ is evaluated as $Bias(\hat{\theta}) = E(\hat{\theta}) - t_y$ when only the randomness associated with the survey design is considered. Following terms discussed in Särndal et al. (1992), this bias is labeled as the model-assisted randomization (or design-based) bias where the "model" is the assisting model chosen to produce the estimator of interest.

A Taylor linearization is used to approximate the expectation of any nonlinear estimator, such as the EC-GREG totals studied in this chapter. We assume that samples in analytic and benchmark surveys are sufficiently large to facilitate the approximation — see the theoretical assumptions discussed in Section 3.7. In the case of $\hat{t}_{yR}$ in (4.1) with a first-order linearization approximation, we have

$$\begin{aligned} E(\hat{t}_{yR}) &= E(\hat{t}_{Ay}) + E[(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_A] \\ &= E(\hat{t}_{Ay}) + [E(\hat{\mathbf{t}}_{Bx}) - E(\hat{\mathbf{t}}_{Ax})]'E(\hat{\mathbf{B}}_A) \\ &\quad + O\left(\max\left[M/\sqrt{m_A}, M/\sqrt{m_B}\right]\right) \\ &\cong E(\hat{t}_{Ay}) + [E(\hat{\mathbf{t}}_{Bx}) - E(\hat{\mathbf{t}}_{Ax})]'E(\hat{\mathbf{B}}_A) \qquad (4.4) \end{aligned}$$

where m_A is the number of PSUs selected under the analytic survey design and m_B is the number of PSUs selected under the benchmark survey design. The result above is obtained by evaluating the expectations with respect to four random mechanisms using the formula for an unconditional expectation $E(\hat{t}) = E_b[E_a(\hat{t}|b)]$ (see e.g., Casella & Berger, 2002, pp. 164, Theorem 4.4.3). The mechanisms include the analytic survey sample design (E_A), the benchmark survey sample design (E_B), and the population coverage propensities for their respective sampling frames (E_{c_A} and E_{c_B}). The unconditional expectation of the model-coefficient vector $\hat{\mathbf{B}}_A$ in (4.2) is defined as follows using a first-order approximation:

$$\begin{aligned}
E(\hat{\mathbf{B}}_A) &= E_{c_A}\left[E_A\left(\hat{\mathbf{B}}_A|c_A\right)\right] \\
&\cong \left[E_{c_A}\left(E_A\left(\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{l=1}^{n_{Ahi}}\pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right)\right)\right]^{-1} \\
&\quad \times E_{c_A}\left(E_A\left(\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\pi_{hik}^{-1}\mathbf{x}_{hik}y_{hik}\right)\right) \\
&\cong \left[\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{l=1}^{N_{hi}}E_{c_A}\left(C_{Ahil}\right)\mathbf{x}_{hil}\mathbf{x}_{hil}'\right]^{-1} \\
&\quad \times \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}}E_{c_A}\left(C_{Ahik}\right)\mathbf{x}_{hik}y_{hik} \\
&= \left[\sum_{hil\in U}\phi_{Ahil}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right]^{-1}\sum_{hik\in U}\phi_{Ahik}\mathbf{x}_{hik}y_{hik} \equiv \mathbf{B}_A \qquad (4.5)
\end{aligned}$$

where $\sum_{hik\in U}$ represents the sum over the design strata (h), PSUs (i), and units (k) within the complete population (U); and $C_{Ahik} = 1$ indicates that the k^{th} population unit ($k \in U$) is listed on the analytic sampling frame ($C_{Ahik} = 0$ otherwise) such that $E_{c_A}(C_{Ahik}) = \phi_{Ahik}$. Note that we use the subscript A in $E(\hat{\mathbf{B}}_A) = \mathbf{B}_A$ to associate the population model-coefficient vector with any subset of the population covered by the analytic survey sampling frame, i.e., $U_A \subseteq U$. Some researchers implicitly assume an average coverage rate across the frame ($\phi_{Ahik} \equiv \phi_A$) so that the claim of unbiasedness holds, i.e., $\mathbf{B}_A \equiv \mathbf{B}$. However, it is more common for the coverage rates to differ across groups of units (U.S. Census Bureau, 2002). For our research, we will not make these assumptions and instead will allow for a difference between $\mathbf{B}_A$ and $\mathbf{B}$ (i.e., coverage errors may exist in the benchmark survey).

Following the technique in (4.5) for the remaining terms in (4.4), we say that $E(\hat{t}_{Ay}) \cong t_{Ay} = \sum_{hik \in U} \phi_{Ahik} y_{hik}$, and $E(\hat{\mathbf{t}}_{Ax}) \cong \mathbf{t}_{Ax} = \sum_{hik \in U} \phi_{Ahik} \mathbf{x}_{hik}$. The expectation of the benchmark control total vector is equal to $\mathbf{t}_{Bx} = \sum_{hik \in U} \phi_{Bhik} \mathbf{x}_{hik}$ where $C_{Bhik} = 1$ identifies the population units listed on the benchmark survey frame such that $E(C_{Bhik}) = \phi_{Bhik}$ ($C_{Bhik} = 0$ otherwise). Therefore,

$$E(\hat{t}_{yR}) \cong t_{Ay} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})' \mathbf{B}_A. \tag{4.6}$$

The calibration model underlying $\hat{t}_{yR}$ is $y_{hik} = \mathbf{x}'_{hik} \mathbf{B} + E_{hik}$, where we assume the model errors ($\mathbf{E}$) are distributed with mean zero and common variance σ^2. Continuing with the calculation of the design-based bias, we obtain the following expression:

$$\begin{aligned}
Bias(\hat{t}_{yR}) &\cong t_{Ay} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B}_A - t_y \\
&= [t_{Ay} - \mathbf{t}'_{Ax}\mathbf{B} - t_y + \mathbf{t}'_x\mathbf{B}] - (\mathbf{t}_x - \mathbf{t}_{Ax})'\mathbf{B} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\left(\mathbf{B}_A - \mathbf{B}\right) \\
&\quad + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B} \\
&= \sum_{hik \in U} [\phi_{Ahik}(y_{hik} - \mathbf{x}'_{hik}\mathbf{B}) - (y_{hik} - \mathbf{x}'_{hik}\mathbf{B})] - (\mathbf{t}_x - \mathbf{t}_{Ax})'\mathbf{B} \\
&\quad + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_A - \mathbf{B}) + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B} \\
&= \sum_{hik \in U} (E_{hik} - \bar{E})(\phi_{Ahik} - \bar{\phi}_A) + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_A - \mathbf{B}) \\
&\quad - N\bar{E}(1 - \bar{\phi}_A) + (\mathbf{t}_{Bx} - \mathbf{t}_x)'\mathbf{B} \\
&= NC_{AE\phi} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_A - \mathbf{B}) - N\bar{E}(1 - \bar{\phi}_A) \\
&\quad + (\mathbf{t}_{Bx} - \mathbf{t}_x)'\mathbf{B} \qquad (4.7)
\end{aligned}$$

where $E_{hik} = y_{hik} - \mathbf{x}'_{hik}\mathbf{B}$, the population-level assisting model residual; $\bar{\phi}_A$ is the average coverage rate for the analytic survey sampling frame; and $C_{AE\phi} = \sum_{hik \in U}(E_{hik} - \bar{E})(\phi_{Ahik} - \bar{\phi}_A)/N$, the covariance between the coverage rates and the assisting model residuals. The four bias components in (4.7) each can be eliminated under the following conditions. (i) If the auxiliary variables are correlated with the outcome variable y and with the coverage mechanism, and the working model is sufficiently close to the population model, then the random variation left unexplained by the model (in theory) should be uncorrelated with the coverage propensities, i.e., $C_{AE\phi} \cong 0$. Under this scenario, the first bias component $NC_{AE\phi}$ is approximately zero. Note that $C_{AE\phi}$ can also be written as $N\left[C_{Ay\phi} - C'_{A\mathbf{x}\phi}\mathbf{B}\right]$ with

$NC_{Ay\phi} = \sum_{hik \in U}(y_{hik} - \bar{y})(\phi_{Ahik} - \bar{\phi}_A)$ and $NC'_{A\mathbf{x}\phi}\mathbf{B} = \sum_{hik \in U}(\mathbf{x}_{hik} - \bar{\mathbf{x}})(\phi_{Ahik} - \bar{\phi}_A)$. As a result, $C_{AE\phi}$ will also be zero if the coverage probabilities in the analytic survey are uncorrelated with both the outcome and the auxiliary variables. (*ii*) If the coverage for the analytic and benchmark surveys is the same, then $\mathbf{t}_{Bx} = \mathbf{t}_{Ax}$ so that the second bias component $(\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_A - \mathbf{B})$ disappears. Likewise, if the slope $\mathbf{B}_A$ from the universe covered by the analytic survey is the same as that of the full universe, the second term vanishes. (*iii*) If the design matrix contains a column of ones (intercept) so that the overall estimated population size is included as an auxiliary variable, then by definition $\bar{E} = 0$ and the third bias component is eliminated. (*iv*) Finally, if $\mathbf{t}_{Bx} = \mathbf{t}_x$, as with traditional calibration, the last component is zero. Therefore, the estimator $\hat{t}_{yR}$ will be asymptotically design unbiased only if *all* these conditions are satisfied; an unlikely event especially with EC calibration. Generally, the bias will be order $O\left(\max\left[M/\sqrt{m_A}, M/\sqrt{m_B}\right]\right)$.

The last term in (4.7), $(\mathbf{t}_{Bx} - \mathbf{t}_x)$, can be further decomposed into $NC_{Bx\phi} - (1 - \bar{\phi}_B)\mathbf{t}_x$, where $C_{Bx\phi}$ is the vector of covariance terms between the auxiliary variables and the coverage propensities for the benchmark survey(s), and $\bar{\phi}_B$ is the average benchmark coverage rate. If $\mathbf{t}_{Bx} \neq \mathbf{t}_x$, the bias component could be reduced by choosing auxiliary variables from the benchmark survey with high coverage rates.

The model-assisted randomization bias for $\hat{t}_{yP}$, an EC-GREG estimator of a total under a group-mean assisting model, can be derived from expression (4.7) and has the following form:

$$Bias(\hat{t}_{yP}) \cong \sum_{g=1}^{G} N_{Bg} C_{Ay\phi g} \frac{1}{\bar{\phi}_{Ag}} + \sum_{g=1}^{G} t_{yg} \left(\frac{N_{Bg}}{N_g} - 1 \right) \tag{4.8}$$

where N_g is the population size within poststratum g; $C_{Ay\phi g} = N_g^{-1} \sum_{k \in U_g} (y_{hik} - \bar{y}_g)(\phi_{Ahik} - \bar{\phi}_{Ag})$, the population covariance between the outcome variable and the coverage rates within poststratum g; $\bar{y}_g = t_{yg}/N_g$, the g^{th} poststratum mean of y; and $\bar{\phi}_{Ag} = N_{Ag}/N_g$, the average coverage rate within the poststratum under the analytic survey design. If the benchmark survey does not cover the target population correctly, so that $N_{Bg} \neq N_g$, then the first bias component, $t_{yg}(N_{Bg}/N_g - 1)$, will be either positive (overestimate) or negative (underestimate) depending on the magnitude of the bias. This component will be strictly negative if the benchmark survey suffers undercoverage, and can accumulate across the poststrata to a sizeable negative bias depending on the magnitude of the outcome variable. The second component, which is dependent on the particular outcome variable under examination, may also be negative if large y values are more likely to be excluded from a sampling frame.

Components of $Bias(\hat{t}_{yP})$ are zero only under certain conditions. (i) If $N_{Bg} = N_g$ for all g (i.e., no coverage errors in the benchmark survey), then the bias is dependent only on the association between the outcome variable and the coverage probabilities, $C_{Ay\phi g}$. The value of $Bias(\hat{t}_{yP})$ then reduces to the formula provided in equation (2) of Kim et al. (2007) for the traditional poststratified estimator, $\hat{t}_{PS}$. (ii) If the coverage probabilities are constant within each poststratum (i.e., $\phi_{Ahik} = \bar{\phi}_{Ag}$,

$k \in U_g$ for all g), then the first bias component is zero. Only if *both* conditions are satisfied can we say that the $\hat{t}_{yP}$ is approximately unbiased. Some might argue that a "perfect" combination of poststrata could be formed such that the positive and negative bias components cancel; however, we believe this likelihood to be so rare as to be virtually impossible.

For some estimators, the contribution of the squared bias to the total MSE is small relative to the variance. Many researchers will claim (approximate) unbiasedness based on weight adjustments that reduce bias to negligible levels because the "true" levels of bias are generally not available. We next focus on what is for many the primary component of the MSE, i.e., the variance.

4.4 Variance Estimation

Variance estimators have been developed for traditional weight calibration and are available in software designed to analyze survey data, e.g., R® (R Development Core Team, 2005), SAS® (SAS Institute Inc., 2004), Stata® (StataCorp, 2005), SUDAAN® (Research Triangle Institute, 2004), and WesVar® (Westat, 2000). However, limited theoretical work has been completed on variance estimation for EC calibration, and to our knowledge, the associated software is non-existent.

Five EC variance estimators that account for the variation in the benchmark control totals are presented in the following sections. They include two linearization estimators, and three delete-one jackknife variance estimators. With the delete-one jackknife, replicates are created by sequentially deleting one PSU and adjusting the weights for the remaining PSUs within the corresponding design stratum. This results in a total of $m_A = \sum_{h=1}^{H} m_{Ah}$ replicates calculated by summing the number of analytic-survey PSUs per stratum (m_{Ah}) across the design strata (h=1,...,H). We also compare the theoretical properties of the variance estimators.

An effective variance estimator will reproduce the corresponding population sampling variance in expectation (i.e., asymptotically unbiased estimator). The population sampling variance of the EC-GREG total of y, $\hat{t}_{yR}$, is classified as an approximate (or asymptotic) variance because of the approximate form of the regression estimator used in the derivation. The approximation is derived by first rewriting the estimator in terms of the corresponding GREG estimator (2.10):

$$
\begin{aligned}
\hat{t}_{yR} &= \hat{t}_{Ay} + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A \\
&= \left[\hat{t}_{Ay} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A\right] + \left(\hat{\mathbf{t}}_{Bx} - \mathbf{t}_{Bx}\right)' \hat{\mathbf{B}}_A \\
&= \hat{t}_{yGREG} + \left(\hat{\mathbf{t}}_{Bx} - \mathbf{t}_{Bx}\right)' \left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) + \hat{\mathbf{t}}'_{Bx}\mathbf{B}_A - \mathbf{t}'_{Bx}\mathbf{B}_A \\
&= \hat{t}_{yGREG} + O_P\left(M/\sqrt{m_B}\right) O_P\left(m_A^{-1/2}\right) + \hat{\mathbf{t}}'_{Bx}\mathbf{B}_A - \mathbf{t}'_{Bx}\mathbf{B}_A \\
&\cong \hat{t}_{yGREG} + \hat{\mathbf{t}}'_{Bx}\mathbf{B}_A - \mathbf{t}'_{Bx}\mathbf{B}_A
\end{aligned} \tag{4.9}
$$

where $\mathbf{B}_A$ is the (design) expected value of $\hat{\mathbf{B}}_A$, the vector of sample regression coefficients defined in (4.1), such that $\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) = O_P\left(m_A^{-1/2}\right)$; $\hat{\mathbf{t}}'_{Bx}\mathbf{B}_A = O_P(M)$, by the assumptions discussed in Section 3.7, is a function of the vector of estimated benchmark control totals $\hat{\mathbf{t}}_{Bx}$ with M equal to the total number of PSUs in the population; and $\mathbf{t}'_{Bx}\mathbf{B}_A$ is a constant, i.e., $O(M)$. The estimator $\hat{t}_{yGREG} = O_P(M)$

can be written as a function of the population assisting-model residuals:

$$\begin{aligned}
\hat{t}_{yGREG} &= \hat{t}_{Ay} + \left(\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_A \\
&= \hat{t}_{Ay} - \hat{\mathbf{t}}'_{Ax}\mathbf{B}_A - \hat{\mathbf{t}}'_{Ax}\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) + \mathbf{t}'_{Bx}\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) + \mathbf{t}'_{Bx}\mathbf{B}_A \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1}\left(y_{hik} - \mathbf{x}'_{hik}\mathbf{B}_A\right) - \left(\hat{\mathbf{t}}_{Ax} - \mathbf{t}_{Ax}\right)'\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) \\
&\quad -\mathbf{t}'_{Ax}\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) + \mathbf{t}'_{Bx}\left(\hat{\mathbf{B}}_A - \mathbf{B}_A\right) + \mathbf{t}'_{Bx}\mathbf{B}_A \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} E_{Ahik} + O_P\left(M/\sqrt{m_A}\right) + \mathbf{t}'_{Bx}\mathbf{B}_A \\
&\cong \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} E_{Ahik} + \mathbf{t}'_{Bx}\mathbf{B}_A \qquad (4.10)
\end{aligned}$$

where the analytic survey population residuals are calculated as $E_{Ahik} = y_{hik} - \mathbf{x}'_{hik}\mathbf{B}_A$ such that $\sum_{hik \in s_A} \pi_{hik}^{-1} E_{Ahik} = O_P(M)$; $\left(\hat{\mathbf{t}}_{Ax} - \mathbf{t}_{Ax}\right) = O_P\left(M/\sqrt{m_A}\right)$; and $\mathbf{t}'_{Bx}\mathbf{B}_A = O(M)$. Combining the two approximations in (4.9) and (4.10) and noting that the $\mathbf{t}'_{Bx}\mathbf{B}_A$ terms cancel, we express $\hat{t}_{yR}$ as

$$\hat{t}_{yR} \cong \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} E_{Ahik} + \hat{\mathbf{t}}'_{Bx}\mathbf{B}_A. \qquad (4.11)$$

The population sampling variance of $\hat{t}_{yR}$ is generally evaluated with respect to the analytic (A) and benchmark (B) survey designs as well as the coverage mechanisms associated with the respective sampling frames (c_A and c_B). However, for our purposes we will assume that the benchmark survey has only coverage bias (i.e., no detectable variation in coverage). The unconditional population sampling variance is determined by evaluating the following variance components created by recursively administering the unconditional variance formula given in, e.g., Casella & Berger (2002, Theorem 4.4.7):

$$\begin{aligned}
Var(\hat{t}_{yR}) &= E_B\left[Var_{c_A,A}(\hat{t}_{yR} \mid B)\right] + Var_B\left[E_{c_A,A}(\hat{t}_{yR} \mid B)\right] \\
&= E_B\left[E_{c_A}\left\{Var_A(\hat{t}_{yR} \mid c_A, B)\ |B\right\}\right] \\
&\quad + E_B\left[Var_{c_A}\left\{E_A(\hat{t}_{yR} \mid c_A, B)|\ B\right\}\right] \\
&\quad + Var_B\left[E_{c_A}\left\{E_A(\hat{t}_{yR} \mid c_A, B)|\ B\right\}\right] \\
&\equiv V_1 + V_2 + V_3. \qquad (4.12)
\end{aligned}$$

For the purpose of variance computation, we assume the analytic survey sample is generated from a complex, multi-stage design with m_{Ah} ($m_{Ah} \geq 2$) PSUs selected *with replacement* from within each of H design strata, and a without-replacement sample of n_{Ahi} units selected from PSU hi. A complete discussion of analytic survey sampling design assumptions is provided in Section 3.1. The *pwr* estimator of the residual total in stratum h for this design is

$$\begin{aligned}
\hat{t}_{AEh} = \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{E_{Ahik}}{\pi_{hik}} &= \frac{1}{m_{Ah}} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{E_{Ahik}}{\pi_{hi(1)}\pi_{k|hi}} \\
&\equiv \frac{1}{m_{Ah}} \sum_{i=1}^{m_{Ah}} \frac{\hat{t}_{AEhi}}{\pi_{hi(1)}}
\end{aligned}$$

where $\hat{t}_{AEhi} = \sum_{k=1}^{n_{Ahi}} E_{Ahik}/\pi_{k|hi}$ for E_{Ahik} defined in (4.10). To evaluate the expectation of $\hat{t}_{yR}$ with respect to the coverage mechanism and the sampling design associated with the analytic survey, as well as the benchmark survey design, note that each

$\hat{t}_{AEhi}/\pi_{hi(1)}$ is a one-PSU estimate of the stratum total t_{AEh} $(= \sum_{i=1}^{M_h} \sum_{k=1}^{N_{hi}} E_{Ahik})$. It follows that

$$\begin{aligned}
E\left(\hat{t}_{yR}\right) &= E_B\left(E_{c_A}\left[E_A\left(\hat{t}_{yR} \mid c_A\right)\right]\right) \\
&\cong E_{c_A}\left[E_A\left(\sum_{h=1}^{H} \hat{t}_{AEh} \mid c_A\right)\right] + E_B\left[\hat{\mathbf{t}}'_{Bx}\mathbf{B}_A\right] \\
&\cong \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{k=1}^{M_{hi}} E_{c_A}\left(C_{Ahik}\right) E_{Ahik} + E_B\left[\hat{\mathbf{t}}'_{Bx}\right]\mathbf{B}_A \\
&= \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{k=1}^{M_{hi}} \phi_{Ahik} E_{Ahik} + \mathbf{t}'_{Bx}\mathbf{B}_A \qquad (4.13)
\end{aligned}$$

where $C_{Ahik} = 1$ indicates that the k^{th} unit within PSU hi is listed on (i.e., covered by) the analytic survey sampling frame (zero otherwise) with $E_{c_A}\left(C_{Ahik}\right) = \phi_{Ahik}$.

The term $V_1 \equiv E_B\left[E_{c_A}\left\{Var_A(\hat{t}_{yR} \mid c_A, B) \mid B\right\}\right]$ in expression (4.12) is evaluated using Result 4.5.1 in Särndal et al. (1992) within each of the H analytic survey design strata in addition to the work presented in expression (4.13):

$$\begin{aligned}
V_1 &\equiv E_B\left[E_{c_A}\left\{Var_A(\hat{t}_{yR} \mid c_A, B) \mid B\right\}\right] = E_{c_A}\left[Var_A(\hat{t}_{yR} \mid c_A, B)\right] \\
&\cong E_{c_A}\left[\sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \pi_{hi(1)} \left(\frac{t_{AEhi}}{\pi_{hi(1)}} - t_{AEh}\right)^2 + \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \frac{V_{Ahi}}{\pi_{hi(1)}}\right] \\
&= \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \pi_{hi(1)} \left(\frac{t_{AEhi}}{\pi_{hi(1)}} - t_{AEh}\right)^2 + \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \frac{E_{c_A}\left(V_{Ahi}\right)}{\pi_{hi(1)}} \qquad (4.14)
\end{aligned}$$

where $t_{AEhi} = \sum_{k=1}^{N_{hi}} \phi_{Ahik} E_{Ahik}$; $t_{AEh} = \sum_{i=1}^{M_h} t_{AEhi}$; and V_{Ahi} is the within-PSU population sampling variance for E_{Ahik}. The within-PSU variance, V_{Ahi}, is determined by the design used to select the second-stage units for the analytic survey. For example, a simple random sample (SRS) of n_{Ahi} out of N_{Ahi} second-stage units within PSU hi results in the following formula:

$$V_{Ahi} = N_{Ahi}\left(\frac{N_{Ahi}}{n_{Ahi}} - 1\right)\sum_{i\in U_{Ahi}} \frac{\left(E_{Ahik} - \bar{E}_{Ahi}\right)^2}{N_{Ahi} - 1}$$

where $\bar{E}_{Ahi} = N_{Ahi}^{-1}\sum_{k\in U_{Ahi}} E_{Ahik}$, the average assisting-model residual in PSU hi within the analytic frame population U_{Ahi}. Taking the expectation with respect to the analytic frame coverage mechanism (c_A), we have

$$E_{c_A}\left(V_{Ahi}\right) \cong E_{c_A}\left(N_{Ahi}\right)\left(\frac{E_{c_A}\left(N_{Ahi}\right)}{n_{Ahi}} - 1\right)\frac{E_{c_A}\left(\sum_{k\in U_{Ahi}}\left(E_{Ahik} - \bar{E}_{Ahi}\right)^2\right)}{\left(E_{c_A}\left(N_{Ahi}\right) - 1\right)}$$

with $E_{c_A}\left(N_{Ahi}\right) = E_{c_A}\left(\sum_{k=1}^{N_{hi}} C_{Ahik}\right) = \sum_{k=1}^{N_{hi}} \phi_{Ahik} = N_{Ahi}^{*}$, the expected PSU size covered by *all* analytic survey sampling frames. The remaining variance component above evaluates to the following expression:

$$\begin{aligned} E_{c_A}\left(\sum_{k\in U_{Ahi}}\left(E_{Ahik} - \bar{E}_{Ahi}\right)^2\right) &= E_{c_A}\left(\sum_{k\in U_{Ahi}} E_{Ahik}^2 - N_{Ahi}\bar{E}_{Ahi}^2\right) \\ &\cong \sum_{k=1}^{N_{hi}} \phi_{Ahik}E_{Ahik}^2 - \frac{\left(\sum_{k=1}^{N_{hi}} \phi_{Ahik}E_{Ahik}\right)^2}{\sum_{k=1}^{N_{hi}} \phi_{Ahik}} \\ &= \sum_{k=1}^{N_{hi}} \phi_{Ahik}\left(E_{Ahik} - \breve{\bar{E}}_{Ahi}\right)^2 \end{aligned}$$

for $\breve{\bar{E}}_{Ahi} = \left(\sum_{k=1}^{N_{hi}} \phi_{Ahik}E_{Ahik}\right) / \left(\sum_{k=1}^{N_{hi}} \phi_{Ahik}\right)$. Thus, the V_1 variance component is associated only with the variance of the analytic survey design, that is, traditional calibration where the benchmark estimates are assumed to be fixed so that $V_1 \equiv AV\left(\hat{t}_{yGREG}\right)$.

Focusing on the coverage errors in the analytic survey, the second variance component in (4.12) is evaluated as follows:

$$\begin{aligned} V_2 &\equiv E_B\left[Var_{c_A}\left\{E_A(\hat{t}_{yR} \mid c_A, B) \middle| B\right\}\right] \\ &= Var_{c_A}\left[E_A(\hat{t}_{yR} \mid c_A, B) \middle| B\right] \\ &\cong Var_{c_A}\left[\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} C_{Ahik}E_{Ahik}\right] \\ &= \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\left(1-\phi_{Ahik}\right)E_{Ahik}^2 \end{aligned} \tag{4.15}$$

under the assumption that $C_{Ahik} \sim Bernoulli(\phi_{Ahik})$ as discussed in Section 3.4. This component is by definition positive and inflates the variance for the analytic survey frame undercoverage errors.

The last variance component in (4.12) addresses the variability of the benchmark control totals and evaluates to the following expression:

$$\begin{aligned} V_3 &\equiv Var_B\left[E_{c_A}\left\{E_A(\hat{t}_{yR} \mid c_A, B) \middle| B\right\}\right] \\ &\cong Var_B\left[\hat{\mathbf{t}}'_{Bx}\mathbf{B}_A\right] \\ &= \mathbf{B}'_A\mathbf{V}_B\mathbf{B}_A \end{aligned}$$

where $\mathbf{V}_B = Var_B\left(\hat{\mathbf{t}}_{Bx}\right)$, the population sampling covariance matrix associated with the vector of estimated control totals. Therefore, after combining the component approximations, we say that the asymptotic population sampling variance (AV) of

$\hat{t}_{yR}$ is:

$$
\begin{aligned}
AV(\hat{t}_{yR}) &\equiv V_1 + V_2 + V_3 \\
&= AV\left(\hat{t}_{yGREG}\right) \\
&\quad + \sum_{h=1}^{H} \sum_{i=1}^{M_h} \sum_{k=1}^{N_{hi}} \phi_{Ahik} \left(1 - \phi_{Ahik}\right) E_{Ahik}^2 \\
&\quad + \mathbf{B}_A' \mathbf{V}_B \mathbf{B}_A. \qquad (4.16)
\end{aligned}
$$

The relative influence of the three components in (4.16) on the overall variance is best examined through their convergence rates after dividing $AV(\hat{t}_{yR})$ by M^2 to ensure the quantity is bounded. The first and third terms in (4.16), under standard conditions (see Rao & Wu, 1985), are $O\left(m_A^{-1}\right)$ and $O\left(m_B^{-1}\right)$, respectively. The term V_2 is of a lower order, $O\left(M^{-1}\right)$. Thus, the sizes of the PSU samples in the analytic and benchmark surveys are the prime determinants for the level of the asymptotic variance. Note that the notation "$AV(\hat{t}_{yR}) =$" in (4.16) is the same as "$Var\left(\hat{t}_{yR}\right) \cong$" in (4.12) and follows the naming convention adopted by Särndal et al. (1992). We use the AV notation in the remainder of the document.

The results for the *EC poststratified estimator* (EC-PSGR), a special case of the regression estimator discussed above, are derived by first detailing the first-order

Taylor approximation of $\hat{t}_{yP}$ (4.3):

$$\begin{aligned}
\hat{t}_{yP} &\cong t_{yP} + \sum_{g=1}^{G} \left\{ \bar{Y}_{Ag} \left(\hat{N}_{Bg} - N_{Bg} \right) + N_{Bg} N_{Ag}^{-1} \left(\hat{t}_{Ayg} - t_{Ayg} \right) \right. \\
&\qquad \left. - N_{Bg} N_{Ag}^{-1} \bar{Y}_{Ag} \left(\hat{N}_{Ag} - N_{Ag} \right) \right\} \\
&= t_{yP} + \sum_{g=1}^{G} \left\{ \bar{Y}_{Ag} \left(\hat{N}_{Bg} - N_{Bg} \right) + N_{Bg} N_{Ag}^{-1} \left(\hat{t}_{Ayg} - \bar{Y}_{Ag} \hat{N}_{Ag} \right) \right\} \\
&= t_{yP} + \left(\hat{\mathbf{N}}_B - \mathbf{N}_B \right)' \bar{\mathbf{Y}}_A + \left(\hat{\mathbf{t}}_{Ay} - \hat{\mathbf{N}}_A \bar{\mathbf{Y}}_A \right)' \mathbf{N}_A^{-1} \mathbf{N}_B, \qquad (4.17)
\end{aligned}$$

where $t_{yP} = \sum_{g=1}^{G} N_{Bg} N_{Ag}^{-1} t_{Ayg}$; $\hat{\mathbf{N}}_B = [\hat{N}_{B1}, ..., \hat{N}_{BG}]'$, the vector of G poststratum counts estimated from the benchmark survey; $\mathbf{N}_B$ is the vector of true counts from the benchmark sampling frame population; $\hat{\bar{\mathbf{Y}}}_A = \hat{\mathbf{N}}_A^{-1} \hat{\mathbf{t}}_{Ay} = [\hat{\bar{y}}_{A1}, ..., \hat{\bar{y}}_{AG}]'$, the G-length vector of model coefficients under the group-mean assisting model for the analytic survey; $\bar{\mathbf{Y}}_A = [t_{Ay1}/N_{A1}, \ ..., \ t_{Ay1}/N_{AG}]'$, the population equivalent to $\hat{\bar{\mathbf{Y}}}_A$; $\hat{\mathbf{t}}_{Ay} = [\hat{t}_{Ay1}, ..., \hat{t}_{AyG}]'$, the vector of total y within each of the G poststrata; and $\hat{\mathbf{N}}_A$ is a diagonal matrix of dimension G with elements equal to the analytic survey poststratum estimates, i.e., $\hat{N}_{A1}, ..., \hat{N}_{AG}$. Note that the first-order approximation to $\hat{\bar{\mathbf{Y}}}_A$ is given as:

$$\begin{aligned}
\hat{\bar{\mathbf{Y}}}_A &\cong \bar{\mathbf{Y}}_A + \mathbf{N}_A^{-1} \left(\hat{\mathbf{t}}_{Ay} - \mathbf{t}_{Ay} \right) - \left(\hat{\mathbf{N}}_A - \mathbf{N}_A \right) \mathbf{N}_A^{-1} \bar{\mathbf{Y}}_A \\
&= \bar{\mathbf{Y}}_A + \mathbf{N}_A^{-1} \left(\hat{\mathbf{t}}_{Ay} - \hat{\mathbf{N}}_A \bar{\mathbf{Y}}_A \right).
\end{aligned}$$

Next, note that $\hat{\mathbf{t}}_{Ay} - \hat{\mathbf{N}}_A \bar{\mathbf{Y}}_A = \hat{\mathbf{N}}_A \left(\hat{\bar{\mathbf{Y}}}_A - \bar{\mathbf{Y}}_A \right)$ and $\mathbf{N}_A^{-1} \hat{\mathbf{N}}_A = \mathbf{I}_G + O_P \left(m_A^{-1/2} \right)$, where $\mathbf{I}_G$ is a G-dimensional identity matrix. Using this and (4.17), we have

$$\begin{aligned} AV\left(\hat{t}_{yP}\right) &= \mathbf{N}_B' E\left[\left[\mathbf{N}_A^{-1}\left(\hat{\mathbf{t}}_{Ay} - \hat{\mathbf{N}}_A\bar{\mathbf{Y}}_A\right)\right]\left[\left(\hat{\mathbf{t}}_{Ay} - \hat{\mathbf{N}}_A\bar{\mathbf{Y}}_A\right)'\mathbf{N}_A^{-1}\right]\right]\mathbf{N}_B \\ &+ 2\bar{\mathbf{Y}}_A' E\left[\left(\hat{\mathbf{N}}_B - \mathbf{N}_B\right)\left(\hat{\bar{\mathbf{Y}}}_A - \bar{\mathbf{Y}}_A\right)'\right] E\left[\hat{\mathbf{N}}_A\right]\mathbf{N}_A^{-1}\mathbf{N}_B \\ &+ \bar{\mathbf{Y}}_A' E\left[\left(\hat{\mathbf{N}}_B - \mathbf{N}_B\right)\left(\hat{\mathbf{N}}_B - \mathbf{N}_B\right)'\right]\bar{\mathbf{Y}}_A \\ &= \mathbf{N}_B' Var\left(\hat{\bar{\mathbf{Y}}}_A\right)\mathbf{N}_B + 2\bar{\mathbf{Y}}_A' Cov\left(\hat{\mathbf{N}}_B, \hat{\bar{\mathbf{Y}}}_A\right)\mathbf{N}_B + \bar{\mathbf{Y}}_A' Var\left(\hat{\mathbf{N}}_B\right)\bar{\mathbf{Y}}_A \\ &= \mathbf{N}_B' Var\left(\hat{\bar{\mathbf{Y}}}_A\right)\mathbf{N}_B + \bar{\mathbf{Y}}_A' Var\left(\hat{\mathbf{N}}_B\right)\bar{\mathbf{Y}}_A. \end{aligned} \tag{4.18}$$

where the covariance term, $Cov\left(\hat{\mathbf{N}}_B, \hat{\bar{\mathbf{Y}}}_A\right)$, is zero because we assume that the benchmark and analytic surveys are independent.

The asymptotic population sampling variance for the EC poststratified total follows the expression (4.16). The V_1 and V_2 variance components are obtained by evaluating the unconditional variance of $\mathbf{N}_B' Var\left(\hat{\bar{\mathbf{Y}}}_A\right)\mathbf{N}_B$ by averaging over the coverage mechanism (c_A) and design (A) for the analytic survey. Therefore, V_1 for the EC-PSGR estimator is approximately equal to $\mathbf{N}_B' E_{c_A}\left(\mathbf{V}_A\right)\mathbf{N}_B = AV\left(\hat{t}_{yPSGR}\right)$, where $\mathbf{V}_A = Var_A(\hat{\bar{\mathbf{Y}}}_A) \cong \mathbf{D}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\mathbf{D}'$ with

$$\mathbf{D} = \begin{bmatrix} \frac{\partial \bar{y}_{A1}}{\partial t_{y1}} & \frac{\partial \bar{y}_{A1}}{\partial t_{y2}} & \cdots & \frac{\partial \bar{y}_{A1}}{\partial N_G} \\ \frac{\partial \bar{y}_{A2}}{\partial t_{y1}} & \frac{\partial \bar{y}_{A2}}{\partial t_{y2}} & \cdots & \frac{\partial \bar{y}_{A2}}{\partial N_G} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \bar{y}_{AG}}{\partial t_{y1}} & \frac{\partial \bar{y}_{AG}}{\partial t_{y2}} & \cdots & \frac{\partial \bar{y}_{AG}}{\partial N_G} \end{bmatrix} = \left[diag\left(\left\{\frac{1}{N_{Ag}}\right\}_{g=1}^{G}\right), diag\left(\left\{\frac{-\bar{y}_{Ag}}{N_{Ag}}\right\}_{g=1}^{G}\right)\right]$$

and

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \sigma_{(\hat{t}_{A1},\hat{t}_{A1})} & \sigma_{(\hat{t}_{A1},\hat{t}_{A2})} & \cdots & \sigma_{(\hat{t}_{A1},\hat{N}_{AG})} \\ \sigma_{(\hat{t}_{A2},\hat{t}_{A1})} & \sigma_{(\hat{t}_{A2},\hat{t}_{A2})} & \cdots & \sigma_{(\hat{t}_{A2},\hat{N}_{AG})} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{(\hat{N}_{A(G-1)},\hat{t}_{A1})} & \sigma_{(\hat{N}_{A(G-1)},\hat{t}_{A2})} & \cdots & \sigma_{(\hat{N}_{A(G-1)},\hat{N}_{AG})} \\ \sigma_{(\hat{N}_{AG},\hat{t}_{A1})} & \sigma_{(\hat{N}_{AG},\hat{t}_{A2})} & \cdots & \sigma_{(\hat{N}_{AG},\hat{N}_{AG})} \end{bmatrix}.$$

The V_2 variance component is defined as in (4.15) with $E_{Ahik} = y_{hik} - \bar{y}_{Ag}$ for unit k within poststratum g and $\bar{y}_{Ag} = t_{Ayg}/N_{Ag}$. The remaining component is defined as $V_3 \cong \bar{\mathbf{Y}}_A' \mathbf{V}_B \bar{\mathbf{Y}}_A$ with $\mathbf{V}_B \equiv Var_B\left(\hat{\mathbf{N}}_B\right)$, the population sample covariance matrix specified under the benchmark design, by noting $\mathbf{B}_A \equiv \bar{\mathbf{Y}}_A$ under poststratification. Therefore, the approximate variance accounting for the estimates from the analytic and benchmark surveys, as well as the analytic survey coverage mechanism, is equal to

$$\begin{aligned} AV(\hat{t}_{yP}) &= AV\left(\hat{t}_{yPSGR}\right) \\ &\quad + \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \delta_{ghik}\phi_{Ahik}\left(1-\phi_{Ahik}\right)E_{Ahik}^2 \\ &\quad + \bar{\mathbf{Y}}_A'\mathbf{V}_B\bar{\mathbf{Y}}_A \end{aligned} \tag{4.19}$$

with $E_{Ahik} = y_{hik} - \bar{y}_g$ and $\bar{y}_g = t_{yg}/N_g$ discussed for expression (4.8).

Krewski & Rao (1981), Rao & Wu (1985), and others demonstrated the asymptotic consistency of the linearization and jackknife variance estimators for nonlinear functions. However, this examination needs to be extended to the EC calibration — this very research begins here. We discuss the set of EC sample variance estimators for the population sampling variance below identified for our research. The sample estimators are derived by substituting (approximately) unbiased sample estimates for the corresponding population parameters. We begin with an evaluation of traditional calibration variance estimators that do not account for the variability in the estimated controls.

4.4.1 Linearization Variance Estimation for Traditional Calibration

A variety of variance estimators have been developed for traditional weight calibration. These include linearization, balanced repeated replication, jackknife (replication), jackknife linearization, and bootstrap. With all of these methods, the controls are assumed to be fixed and the coverage error does not exist. Therefore, the positive variance components in (4.16) associated with the variability in the benchmark controls and the coverage error are zero because $\mathbf{V}_B$ is assumed to be zero and $\phi_{Ahik} = 1$ for every unit in the population.

The linearization variance estimator for the GREG, as shown in Section 2.3, is a function of the estimated assisting-model residuals ($e_{Ahik} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_A$). This variance estimator is discussed in standard sampling texts such as Särndal et al. (1992) and Lohr (1999). The linearization sample variance estimator (var) for $\hat{t}_{yR}$, under a stratified, multistage analytic survey design with PSUs selected with replacement and with the *naïve* assumptions that the estimated benchmark control totals are known without error and the sampling frame covers the population, is calculated as:

$$var_{Naïve}(\hat{t}_{yR}) = var(\hat{t}_{yGREG}) = \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{hi+} - \bar{\breve{u}}_{h++})^2 \tag{4.20}$$

where $\breve{u}_{hi+} = \sum_{k=1}^{n_{Ahi}} a_{hik} \pi_{hik}^{-1} e_{Ahik}$, the sum of (calibration) weighted model residuals within PSU hi; $\bar{\breve{u}}_{h++} = m_{Ah}^{-1} \sum_{i=1}^{m_{Ah}} \breve{u}_{hi+}$, the average weighted residual within stratum h; and a_{hik} is the calibration weight defined in (4.1).

One variance estimator for $\hat{t}_{yP}$ is obtained by substituting the estimated residuals associated with the group-mean model, $e_{Ahik} = y_{hik} - \hat{\bar{y}}_{Ag}$, and the EC-PSGR calibration weights, $a_{hik} = \hat{N}_{Bg}/\hat{N}_{Ag}$, into (4.20). Another asymptotically equivalent method-of-moments variance estimator for $\hat{t}_{yP}$ is calculated as follows by substituting the sample estimators for the population parameters:

$$var_{Naïve}\left(\hat{t}_{yP}\right) = var\left(\hat{t}_{yPSGR}\right) = \hat{\mathbf{N}}_B' \hat{\mathbf{V}}_A \hat{\mathbf{N}}_B. \tag{4.21}$$

Any variance formula developed for traditional calibration will underestimate the population sampling variance because the benchmark component in (4.16) is not accounted for in the calculation. However, highly precise benchmark estimates will likely contribute a negligible EC calibration variance component to the variance estimator. Thus the difference between the estimates for traditional and EC calibration for these situations also will be negligible assuming that the coverage error component is relatively small. In the next four sections, we present sample variance estimators that address a non-negligible EC calibration variance component.

4.4.2 Estimated-Control Taylor Linearization Variance Method

The EC linearization sample variance estimator for $AV(\hat{t}_{yR})$ (4.16) is derived by summing the approximately unbiased estimators for each of the three components V_1, V_2, and V_3. The first variance component is equivalent to the naive variance estimator represented in (4.20), i.e., $V_1 = var(\hat{t}_{yGREG}) \equiv var_{Naïve}(\hat{t}_{yR})$.

The second variance component is a function of the unknown unit-specific coverage propensities, ϕ_{Ahik}. Aggregate coverage estimates may be available from external sources or estimated using a combination of the analytic and benchmark survey data. For example, an estimated coverage probability is calculated as the ratio of the estimated population counts from the analytic and benchmark surveys either overall, i.e., $\hat{N}_A/\hat{N}_B$, or within certain key domains. This estimation technique relies on the assumption that the benchmark survey frame covers the population of interest. If coverage is associated with, for example, a demographic characteristic, then estimated coverage probabilities by those mutually exclusive domain categories may reduce the bias in the variance component. Using stratum-specific estimated coverage probabilities, $\hat{\bar{\phi}}_{Ah}$, we construct the following sample variance estimator with residuals $e_{Ahik} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_A$:

$$\hat{V}_2 = \sum_{h=1}^{H}\left(1 - \hat{\bar{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e^2_{Ahik}}{\pi_{hik}}$$

where $\hat{\bar{\phi}}_{Ah} = \hat{N}_{Ah}/\hat{N}_{Bh}$ with $\hat{N}_{Ah}$ and $\hat{N}_{Bh}$ defined as the estimated size of the h^{th} stratum defined by the analytic survey design estimated with the analytic and benchmark data, respectively. Because we are interested in *undercoverage* error variance, any $\hat{\bar{\phi}}_{Ah} > 1$ is truncated to one so that its contribution to the error variance is zero. Relying on the assumed sampling design for the analytic survey and the assumption that $E\left(\hat{\bar{\phi}}_{Ah}\right) = \bar{\phi}_{Ah}$, the corresponding population parameter, the expectation of $\hat{V}_2$ is determined as follows:

$$\begin{aligned} E\left(\hat{V}_2\right) &= E_{B,c_A}\left[E_A\left(\sum_{h=1}^{H}\left(1-\hat{\bar{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e_{Ahik}^2}{\pi_{hik}} \mid c_A, B\right)\right] \\ &\cong \sum_{h=1}^{H}\left(1-\bar{\phi}_{Ah}\right)\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} E_{c_A}\left(C_{Ahik}\right)E_{Ahik}^2 \\ &= \sum_{h=1}^{H}\left(1-\bar{\phi}_{Ah}\right)\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}E_{Ahik}^2. \end{aligned}$$

The design-based bias of this estimator is calculated as

$$\begin{aligned} Bias\left(\hat{V}_2\right) &= E\left(\hat{V}_2\right) - V_2 \\ &\cong \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}E_{Ahik}^2\left(\phi_{Ahik}-\bar{\phi}_{Ah}\right). \end{aligned} \tag{4.22}$$

If the coverage probabilities vary only by stratum, i.e., $\phi_{Ahik} \equiv \bar{\phi}_{Ah}$ for units within stratum h, then the bias of $\hat{V}_2$ is approximately zero. However, the bias is inflated if, for example, larger residuals are associated with coverage probabilities that differ from the stratum averages.

Combining an approximate method-of-moments estimator for the third variance component, $\hat{V}_3 = \hat{\mathbf{B}}_A'\hat{\mathbf{V}}_B\hat{\mathbf{B}}_A$, with the other sample components, we have the following sample variance estimator of $AV(\hat{t}_{yR})$ (4.16):

$$\begin{aligned} var_{ECTS}(\hat{t}_{yR}) &= \hat{V}_1 + \hat{V}_2 + \hat{V}_3 \\ &= var(\hat{t}_{yGREG}) \\ &\quad + \sum_{h=1}^{H}\left(1-\hat{\hat{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e_{Ahik}^2}{\pi_{hik}} \\ &\quad + \hat{\mathbf{B}}_A'\hat{\mathbf{V}}_B\hat{\mathbf{B}}_A \end{aligned} \tag{4.23}$$

The linearization sample variance estimator for the EC-PSGR is similarly derived by substituting the approximately unbiased sample estimators into the formula (4.19):

$$\begin{aligned} var_{ECTS}(\hat{t}_{yP}) &= \hat{\mathbf{N}}_B'\hat{\mathbf{V}}_A\hat{\mathbf{N}}_B \\ &\quad + \sum_{g=1}^{G}\left(1-\hat{\hat{\phi}}_{Ag}\right)\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{\delta_{ghik}e_{Ahik}^2}{\pi_{hik}} \\ &\quad + \hat{\hat{\mathbf{Y}}}_A'\hat{\mathbf{V}}_B\hat{\hat{\mathbf{Y}}}_A \end{aligned} \tag{4.24}$$

where $\hat{\mathbf{V}}_A \cong \hat{\mathbf{D}}\hat{\mathbf{\Sigma}}_{\hat{\boldsymbol{\theta}}}\hat{\mathbf{D}}'$, calculated using the analytic survey estimates corresponding to the terms defined for (4.19); $\hat{\hat{\phi}}_{Ag}$ is calculated as $\hat{N}_{Ag}/\hat{N}_{Bg}$ using the components from $\hat{t}_{yP}$ (4.3); $e_{Ahik} = y_{hik} - \hat{\hat{y}}_{Ag}$; and δ_{ghik}, as in (4.3), is an indicator of membership in poststratum g.

4.4.3 Fuller Two-Phase Jackknife Method

Isaki et al. (2004) applied a two-phase delete-one jackknife variance estimator developed by Fuller (1998) to account for estimated control totals. The premise behind Fuller's methodology (ECF2) is to take a spectral (eigenvalue) decomposition of the benchmark covariance matrix ($\hat{\mathbf{V}}_B$), develop benchmark adjustments that are a function of the resulting G eigenvalues and eigenvectors and to add the adjustments to the benchmark controls to create a set of replicate controls. A randomly chosen subset of the m_A replicates is calibrated to the G constructed replicate controls where the condition $m_A \geq G$ is required as shown below. Specifically, the benchmark control total for the r^{th} replicate of $\hat{t}_{yR}$ is defined as

$$\begin{aligned} \hat{\mathbf{t}}_{Bx(r)} &= \mathbf{t}_{Bx} + c_h \hat{\mathbf{z}}_{B(r)} \\ &= \mathbf{t}_{Bx} + c_h \delta_{(r)} \sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}_{Bg} \end{aligned} \tag{4.25}$$

where $\hat{\mathbf{t}}_{Bx} = \sum_{l \in s_B} w_l \mathbf{x}_l$, the vector of control totals estimated from the benchmark survey; c_h is a constant related to the chosen replication variance method ($c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$ for the delete-one jackknife); $\hat{\mathbf{z}}_{B(r)} = \delta_{(r)} \sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}_{Bg}$, the ECF2 replicate control total adjustment; $\delta_{(r)}$ is a zero/one indicator that identifies the G (out of m_A) randomly chosen replicates to receive an ECF2 adjustment; $\delta_{g|(r)} = 1$ if the g^{th} component of the benchmark covariance decomposition (out of G) is randomly chosen for the assignment given that replicate r is selected for an adjustment; and $\hat{\mathbf{z}}_{Bg} = \hat{\mathbf{q}}_g \sqrt{\hat{\lambda}_g}$, a function of an eigenvector $\hat{\mathbf{q}}_g$ and the associated eigenvalue $\hat{\lambda}_g$ such that $\hat{\mathbf{V}}_B = \sum_{g=1}^{G} \hat{\mathbf{z}}_{Bg} \hat{\mathbf{z}}'_{Bg}$, by definition. Given that $\delta_{(r)} = 1$ for a particular replicate, a single indicator $\delta_{g|(r)}$ $(1 \leq g \leq G)$ must also equal one; however, if $\delta_{(r)} = 0$, then *all* indicators $\delta_{g|(r)}$ equal zero.

A delete-one jackknife variance estimator can take multiple forms depending on the centering value. We chose to study the somewhat conservative variance estimator centered about the full-sample estimate for our research ($v4$ in Wolter, 2007, Section 4.5). The delete-one ECF2 jackknife variance estimator, $var_{ECF2}(\hat{t}_{yR})$, is calculated as follows for a stratified, multi-stage analytic survey design:

$$
\begin{aligned}
var_{ECF2}(\hat{t}_{yR}) &= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\ddot{t}_{yR(r)} - \hat{t}_{yGREG}\right)^2 \\
&= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{Ay(r)} + \left(\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)}\right)' \hat{\mathbf{B}}_{A(r)} - \hat{t}_{yGREG}\right)^2 \\
&= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{Ay(r)} + \left(\left\{\mathbf{t}_{Bx} + c_h \hat{\mathbf{z}}_{B(r)}\right\} - \hat{\mathbf{t}}_{Ax(r)}\right)' \hat{\mathbf{B}}_{A(r)} - \hat{t}_{yGREG}\right)^2 \\
&= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)}\right)^2 . \qquad (4.26)
\end{aligned}
$$

Note that the association of the r^{th} replicate to a particular (analytic survey) design stratum is defined through the stratum membership of the eliminated PSU. The replicate estimates in (4.26) are defined as:

- $\ddot{t}_{yR(r)} = \hat{t}_{Ay(r)} + (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{A(r)}$, the EC replicate estimator of the population total using the ECF2 method;

- $\hat{t}_{yGREG(r)} = \hat{t}_{Ay(r)} + (\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{A(r)}$, the corresponding fixed-control replicate estimator;

- $\hat{t}_{Ay(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \pi^{-1}_{hik} y_{hik}$, the replicate total of the y variable;

- $\hat{\mathbf{t}}_{Ax(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \pi^{-1}_{hik} \mathbf{x}_{hik}$, the replicate totals for the auxiliary variables estimated from the analytic survey; and

- $\hat{\mathbf{B}}_{A(r)} = \left[\sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik} \mathbf{x}'_{hik}\right]^{-1} \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik} y_{hik}$, the model coefficient vector calculated with analytic survey data for each replicate.

The hi^{th} PSU-subsampling weight for the r^{th} replicate, $\pi_{hi(r)}^{-1}$, is calculated under the following specification:

$$\pi_{hi(r)}^{-1} = \begin{cases} 0 & \text{if PSU } r \text{ and PSU } i \text{ are the same } (r = i) \\ 1 & \text{if } h \neq h' \text{ for } r \in s_{Ah} \text{ and } i \in s_{Ah'} \\ m_{Ah}/(m_{Ah} - 1) & \text{if } r \neq i \text{ but } h = h'. \end{cases} \tag{4.27}$$

Fuller (1998) approximates the squared term in (4.26) as

$$\begin{aligned} & \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} \\ &= \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \left(\mathbf{B}_A + O_P\left(m_A^{-1/2}\right)\right) \\ &= \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A + O_P\left(M/\sqrt{m_A m_B}\right) \\ &\cong \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \end{aligned} \tag{4.28}$$

by assuming $\hat{t}_{yGREG(r)} - \hat{t}_{yGREG} = O_P\left(M/\sqrt{m_A}\right)$, $\hat{\mathbf{B}}_{A(r)} = \mathbf{B}_A + O_P\left(m_A^{-1/2}\right)$ for the population parameter $\mathbf{B}_A = O(1)$ defined in (4.5), and $\hat{\mathbf{z}}'_{B(r)} = O_P\left(M/\sqrt{m_B}\right)$. Using (4.28) in $var_{ECF2}(\hat{t}_{yR})$ (4.26) and squaring the terms results in

$$\begin{aligned} var_{ECF2}(\hat{t}_{yR}) \cong & \sum_{h=1}^{H} \frac{(m_{Ah} - 1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)^2 \\ & + \sum_{h} \sqrt{\frac{(m_{Ah} - 1)}{m_{Ah}}} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \\ & + \mathbf{B}_A \hat{\mathbf{V}}_B \mathbf{B}_A. \end{aligned} \tag{4.29}$$

We apply M^{-2} to the variance above for convenience in comparing the orders of terms for the bounded quantities. The first component is associated with the variance of $\hat{t}_{yR}$ conditioned on the benchmark controls (i.e., a naïve variance estimator) and is $O_P\left(m_A^{-1}\right)$. Under standard conditions (see Rao & Wu, 1985), $\max\left\{M^{-1}\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)\right\}$ converges in probability to zero and $\hat{\mathbf{z}}_{B(r)}/M = O_P\left(m_B^{-1/2}\right)$ by assumption, so that the second component (divided by M^2) is $O_P\left(m_B^{-1/2}\right)$. The third component is $O_P\left(m_B^{-1}\right)$ and is related only to the variability of the benchmark controls because $\hat{\mathbf{V}}_B = \sum_{h=1}^{H}\sum_{r=1}^{m_{Ah}} \hat{\mathbf{z}}_{B(r)}\hat{\mathbf{z}}'_{B(r)}$, by definition. Fuller (1998) and Isaki et al. (2004) show that the first and third components are asymptotically equivalent to their respective components in $AV(\hat{t}_{yR})$ (4.16). Following the evaluation of the unconditional expectation given in (4.5),

$$
\begin{aligned}
& E_A\left[\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)\hat{\mathbf{z}}'_{B(r)} \mid B\right] \\
&= \left[E_A\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)\right]\ \hat{\mathbf{z}}'_{B(r)} \\
&= 0 \times \hat{\mathbf{z}}'_{B(r)} \equiv 0,
\end{aligned}
$$

thus demonstrating that the second term has expectation zero. However, the ECF2 does not incorporate the additional variation due to coverage error.

We propose the following modification to the ECF2 replicate estimators to account for the coverage error variance component. Let $\eta_{(r)}$ be a value randomly generated from a standard normal distribution for replicate r, i.e., $\eta_{(r)} \sim N(0,1)$.

For each replicate, we calculate the following values:

$$c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \tag{4.30}$$

where $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; $R_h = \sqrt{1/Hm_{Ah}}$; $\hat{\hat{\phi}}_{A(r)}$ is an estimate of the analytic survey coverage rate (error) using a combination of data from the complete benchmark survey and analytic survey replicate subsample (e.g., $\hat{N}_{A(r)}/\hat{N}_B$); $\hat{t}_{Ae2(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} e_{Ahik(r)}^2$ with $\hat{\mathbf{B}}_{A(r)}$ defined for expression (4.26), $\pi_{hi(r)}^{-1}$ defined in expression (4.27), and $e_{Ahik(r)} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{A(r)}$. This replicate value is similar to $\hat{V}_2$ used in $var_{ECTS}(\hat{t}_{yR})$ (4.23) in Section 4.4.2. As discussed in Section 4.4.2, if the value for $\hat{\hat{\phi}}_{A(r)}$ exceeds one, then the estimate is truncated to one to ensure the variance component is non-negative. The *modified ECF2* replicate estimates are then calculated using the following formula:

$$\begin{aligned}
\dddot{t}_{yR(r)} &= \ddot{t}_{yR(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&= \left[\hat{t}_{Ay(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}\right] + (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{A(r)} \\
&= \hat{t}_{yGREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \qquad (4.31)
\end{aligned}$$

with terms defined in expressions (4.26) and (4.30). The expectation of $\dddot{t}_{yR(r)} = \hat{t}_{Ay(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}$ is the same as $\ddot{t}_{yR(r)}$ in the original ECF2 method. This is shown by noting that the expectation of the coverage error term (4.30) is zero because of the inclusion of a standard normal random variable, i.e., $E_\eta\left(\eta_{(r)}\right) = 0$ by definition.

The modified ECF2 estimator, denoted by ECF2m in our research, is constructed as in (4.26) with $\dddot{t}_{yR(r)}$ substituted for $\ddot{t}_{yR(r)}$. Using the justification given for expression (4.28),

$$\begin{aligned}\dddot{t}_{yR(r)} - \hat{t}_{yGREG} \;\cong\;& \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \\ & + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}\end{aligned}$$

with $\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}$ assumed to be $O_P(M)$. The expectation of the ECF2m jackknife sample variance estimator is evaluated by examining the six resulting variance components. Note in the expression below that we abbreviate $(m_{Ah}-1)/m_{Ah}$ as c_h^{-2}:

$$\begin{aligned}
var_{ECF2m}(\hat{t}_{yR}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{t}_{yR(r)} - \hat{t}_{yGREG}\right)^2 \\
&\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\hat{t}_{yGREG(r)} - \hat{t}_{yGREG})^2 \\
&\quad + 2 \sum_{h} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \\
&\quad + \mathbf{B}'_A \hat{\mathbf{V}}_B \mathbf{B}_A \\
&\quad + 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&\quad + 2 \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \left(R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \right) \\
&\quad + \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)} \\
&\equiv \sum_{j=1}^{6} \hat{V}_j. \qquad (4.32)
\end{aligned}$$

Dividing both sides of the approximation in (4.32) by M^2 again facilitates the discussion of the relative convergence rates. The first three terms are the same as those discussed for $var_{ECF2}(\hat{t}_{yR})$ (4.26). The last three terms are specific to the estimated analytic survey coverage error. The fourth component is $O_P\left(M^{-3/2}\right)$ and converges in probability to zero under standard conditions (see Rao & Wu, 1985) by assuming $max\left\{M^{-1}\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)\right\}$. The fifth component is $O_P\left(1/\sqrt{Mm_B}\right)$ and has expectation zero because of the standard normal random variable $\eta_{(r)}$. The expectation of the sixth and final variance component in $var_{ECF2m}(\hat{t}_{yR})$, divided by M^2, is $O_P\left(M^{-1}\right)$ and is approximately equal to the following expression:

$$
\begin{aligned}
&E\left[\sum_{h=1}^{H}\sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1-\hat{\bar{\phi}}_{A(r)}\right)\hat{t}_{Ae2(r)}\right] \\
&\quad\cong \frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}} E_\eta\left(\eta_{(r)}^2\right)\left[1-E_{\phi_A}\left(\hat{\bar{\phi}}_{A(r)}\right)\right]E_A\left(\hat{t}_{Ae2(r)}\right)
\end{aligned}
$$

with $E_\eta\left(\eta_{(r)}^2\right) = 1$, the degrees of freedom for a chi-square random variable. Using the conditions of Rao & Wu (1985), we claim that the expectations of the replicate estimators are equal to their corresponding complete sample estimator, e.g., $E_{\bar{\phi}_A}\left(\hat{\bar{\phi}}_{A(r)}\right) = E_{\phi_A}\left(\hat{\bar{\phi}}_A\right)$. Therefore, $E_{\phi_A}\left(\hat{\bar{\phi}}_{A(r)}\right) = \bar{\phi}_A$ and $E_A\left(\hat{t}_{Ae2(r)}\right) \cong t_{Ae2}$ by assumption, with $t_{Ae2} = \sum_{hik\in U}\phi_{Ahik}E_{Ahik}^2$ and $E_{Ahik} = y_{hik} - \mathbf{x}_{hik}'\mathbf{B}_A$. In summary, sample variance components $\hat{V}_1$, $\hat{V}_3$, and $\hat{V}_6$ in (4.32) address variability in estimates from the analytic survey, the benchmark survey, and coverage error, respectively, and the remaining components are asymptotically equivalent to zero.

One additional finding from the ECF2 methodology presented in Fuller (1998) is important to our research. The author demonstrates that the jackknife variance of the replicate controls, $var_{ECF2}(\hat{\mathbf{t}}_{Bx})$, reproduces the estimated benchmark covariance matrix $\hat{\mathbf{t}}_{Bx}$ for *every* sample. This trait lends stability to the variance estimator as discussed in Section 4.4.4. We provide the derivation below using the notation adopted for our research. The definitions of the indicator variables $\delta_{(r)}$ and $\delta_{g|(r)}$ given for (4.25) are important to the work presented below. In particular, note that for a replicate r to which a $\hat{\mathbf{z}}_{Bg}$ is assigned, $\sum_{g=1}^{G} \delta_{g|(r)}\hat{\mathbf{z}}_{Bg} = \hat{\mathbf{z}}_{Bg(r)}$ where the $Bg(r)$ subscript denotes the particular g that is randomly selected.

$$\begin{aligned} var_{ECF2}(\hat{\mathbf{t}}_{Bx}) &= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Bx})(\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Bx})' \\ &= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} c_h^2 \delta_{(r)} \left(\sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}_{Bg} \right) \left(\sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}'_{Bg} \right) \\ &= \sum_{r=1}^{R} \delta_{(r)} \hat{\mathbf{z}}_{Bg(r)} \hat{\mathbf{z}}'_{Bg(r)} \\ &= \sum_{g=1}^{G} \hat{\mathbf{z}}_{Bg} \hat{\mathbf{z}}'_{Bg} \equiv \hat{\mathbf{V}}_B. \end{aligned} \quad (4.33)$$

In deriving (4.33), we use the fact that in summing $\delta_{(r)}\hat{\mathbf{z}}_{Bg(r)}\hat{\mathbf{z}}'_{Bg(r)}$ over the m_A replicates only G replicates receive an adjustment of a $\hat{\mathbf{z}}_{Bg}$ vector and each $\hat{\mathbf{z}}_{Bg}$ is assigned to one and only one replicate. Note that if $m_A < G$, the above evaluation would not hold because $\sum_{g=1}^{R<G} \hat{\mathbf{z}}_{Bg}\hat{\mathbf{z}}'_{Bg} \neq \hat{\mathbf{V}}_B$. The coverage adjustment in (4.30) was not included in the result above because we consider undercoverage only in the analytic survey sampling frame.

The modified ECF2 variance formula for estimated totals specializes to EC poststratification with G poststrata by first adapting the coverage error variance component in (4.30):

$$c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \tag{4.34}$$

where $\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' = \left[\left(1 - \hat{\phi}_{A1(r)}\right),\ ...,\ \left(1 - \hat{\phi}_{AG(r)}\right)\right]$, a G-length vector of estimated coverage rates within poststratum g; and $\hat{\mathbf{t}}_{Ae2(r)} = \left[\hat{t}_{Ae21(r)},\ ...,\ \hat{t}_{Ae2G(r)}\right]'$ with components $\hat{t}_{Ae2g(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \delta_{ghik} e_{Ahik(r)}^2$ and $e_{Ahik(r)} = y_{hik} - \hat{\hat{y}}_{Ag}$. The poststratum-specific coverage rates may be estimated as $\hat{\phi}_{Ag(r)} = \hat{N}_{Ag(r)} / \hat{N}_{Bg}$, where $\hat{N}_{Ag(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \pi_{hik}^{-1}$, if the benchmark survey frame is believed to correctly cover the population of interest. The ECF2m replicate estimates are functions of the coverage error components (4.34) and are calculated as follows:

$$\begin{aligned} \dddot{t}_{yP(r)} &= \hat{\mathbf{N}}_{B(r)}' \hat{\mathbf{N}}_{A(r)}^{-1} \hat{\mathbf{t}}_{Ay(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\ &= \left[\mathbf{N}_B + c_h \hat{\mathbf{z}}_{B(r)}\right]' \hat{\hat{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\ &= \mathbf{N}_B' \hat{\hat{\mathbf{Y}}}_{A(r)} + c_h \hat{\mathbf{z}}_{B(r)}' \hat{\hat{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \end{aligned} \tag{4.35}$$

where $\ddot{t}_{yP(r)} = \hat{\mathbf{N}}_{B(r)}' \hat{\mathbf{N}}_{A(r)}^{-1} \hat{\mathbf{t}}_{Ay(r)}$; $\hat{\hat{\mathbf{Y}}}_{A(r)} = \hat{\mathbf{N}}_{A(r)}^{-1} \hat{\mathbf{t}}_{Ay(r)}$; $\hat{\mathbf{t}}_{Ay(r)} = \left[\hat{t}_{Ay1(r)}, ..., \hat{t}_{AyG(r)}\right]'$ with elements that are functions of a zero/one indicator δ_{ghik} that signifies membership in the g^{th} poststratum, i.e., $\hat{t}_{Ayg(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \pi_{hik}^{-1} y_{hik}$; $\hat{\mathbf{N}}_{A(r)}$ is a diagonal matrix with elements $\left(\hat{N}_{A1(r)}, ..., \hat{N}_{AG(r)}\right)$ such that $\hat{N}_{Ag(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \times \pi_{hik}^{-1}$; and $\hat{\mathbf{N}}_{B(r)} = \mathbf{N}_B + c_h \hat{\mathbf{z}}_{B(r)}$. Substituting (4.34) and (4.35) in the expression

(4.32), we have

$$var_{ECF2m}(\hat{t}_{yP}) = \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{t}_{yP(r)} - \hat{t}_{yPSGR}\right)^2$$

$$\begin{aligned}
\cong\ & \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR})^2 \\
+\ & 2\sum_{h} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR}\right) \hat{\mathbf{z}}'_{B(r)} \bar{\mathbf{Y}}_A \\
+\ & \bar{\mathbf{Y}}'_A \hat{\mathbf{V}}_B \bar{\mathbf{Y}}_A \\
+\ & 2\sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR}\right) R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\
+\ & 2\sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} \hat{\mathbf{z}}'_{B(r)} \bar{\mathbf{Y}}_A \left(R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \right) \\
+\ & \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)} \qquad (4.36)
\end{aligned}$$

with the terms $c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$, $\hat{t}_{yPSGR} = \mathbf{N}'_B \hat{\mathbf{N}}_A^{-1} \hat{\mathbf{t}}_{Ay}$ (2.11), and others defined previously. The discussion of the asymptotics given for $var_{ECF2m}(\hat{t}_{yR})$ in (4.32) also applies to $var_{ECF2m}(\hat{t}_{yP})$.

The seven steps needed to calculate $var_{ECF2m}(\hat{t}_{yP})$ (4.36) are provided below where the total number of replicates (and analytic survey PSUs) is denoted as m_A. These steps are used in the simulation programs discussed in Section 4.5.

1. Calculate the full-sample estimate $\hat{t}_{yP}$ (4.3).

2. Determine the G eigenvalues $\hat{\lambda}_g$ and G-length eigenvectors $\hat{\mathbf{q}}_g$ from the spectral decomposition of $\hat{\mathbf{V}}_B$, and calculate the G replicate adjustments of the form $\hat{\mathbf{z}}_{Bg} = \hat{\mathbf{q}}_g \sqrt{\hat{\lambda}_g}$. Concatenate the $G \times G$ matrix of $\hat{\mathbf{z}}_{Bg}$'s, where $\hat{\mathbf{z}}_{Bg}$ represents the columns of this matrix, with a $G \times (m_A - G)$ matrix of zeroes. Randomly

sort the columns. Call this new $G \times m_A$ matrix $\mathbf{Z}$.

3. Create a $G \times m_A$ matrix, called $\mathbf{C}$, with column elements all equal to $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$. The m_A-length vector of jackknife stratum weights is calculated as $\mathbf{W}_{m_A} = (m_{Ah}-1)/m_{Ah}$.

4. Calculate the Hadamard (or element-wise) product of $\mathbf{Z}$ and $\mathbf{C}$ denoted as $\mathbf{Z} \bullet \mathbf{C}$ (Searle, 1982, pp. 49). Replicate the vector of poststratum counts estimated from the benchmark survey ($\hat{\mathbf{N}}_B$) into the columns of a $G \times m_A$ matrix and add to $\mathbf{Z} \bullet \mathbf{C}$. This new $G \times m_A$ matrix, called $\mathbf{N}_{Bm_A}$, contains the replicate benchmark controls for all m_A replicates. See the definition of $\hat{\mathbf{N}}_{B(r)}$ in expression (4.36).

5. Calculate the replicate estimates $\hat{\bar{\mathbf{Y}}}_{A(r)}$ with elements $\hat{\bar{y}}_{Ag(r)} = \hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}$ by removing in-turn one PSU from the analytic survey sample file, applying the PSU-subsampling weights (4.27), and summing the weighted values for the numerator and denominator within poststratum g. Call the resulting $G \times m_A$ matrix $\mathbf{B}_{m_A}$.

6. Create the following $G \times m_A$ matrices for the coverage error variance component (4.34): $\mathbf{R}_{m_A}$, with column elements all equal to $\sqrt{1/Hm_{Ah}}$; $\boldsymbol{\eta}_{m_A}$, with elements obtained from the standard normal distribution; $\boldsymbol{\phi}_{m_A}$, with column elements equal to $(1-\hat{N}_{Ag(r)}/\hat{N}_{Bg})$ for $(\hat{N}_{Ag(r)}/\hat{N}_{Bg}) \leq 1$ and zero otherwise; and $\mathbf{e}_{m_A}$ with column elements described above for $\hat{\mathbf{t}}_{Ae2(r)}$. Calculate the Hadamard product of these matrices and call it $\mathbf{E}$.

7. Calculate the m_A replicate estimates, $\ddot{t}_{yP(r)}$ in expression (4.35), by first multiplying the elements $\mathbf{N}_{Bm_A}$ by $\mathbf{B}_{m_A}$, adding $\mathbf{E}$ to the resulting matrix, and summing down the rows within a column. Next, subtract $\hat{t}_{yPSGR}$, $\hat{t}_{yP}$ in (4.3), from each of the m_A values and square the terms, multiply by $\mathbf{W}_{m_A}$, and sum across the m_A estimates. The resulting value is the estimated variance using the ECF2 method, $var_{ECF2m}(\hat{t}_{yP})$ in expression (4.36).

By excluding the sixth step given above, we are also able to calculate the variance of $\hat{t}_{yP}$ under the original ECF2 specification which does not inflate for the analytic survey coverage error. A comparison of the two variance estimators will suggest the level of underestimation associated with the exclusion of the error variance component.

4.4.4 Multivariate Normal Jackknife Method

The multivariate normal method (ECMV) involves a random perturbation of the controls totals for the *complete* set of replicates instead of adjusting only a subsample of replicates as with the original ECF2 method (Section 4.4.3). The ECMV relies on large sample theory so that the control total adjustments may be modeled as coming from a multivariate normal (MVN) distribution. The replicate controls for the ECMV have the form

$$\hat{\mathbf{t}}_{Bx(r)} = \hat{\mathbf{t}}_{Bx} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \tag{4.37}$$

where $\hat{\boldsymbol{\varepsilon}}_{B(r)}$ is a G-length vector of random variables from a multivariate normal distribution such that $\hat{\boldsymbol{\varepsilon}}_{B(r)} \overset{\text{iid}}{\sim} \text{MVN}_G(\mathbf{0}_G, \hat{\mathbf{V}}_B)$; $\mathbf{0}_G$ is a G-length vector of zeroes; $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; and $R_h = \sqrt{1/Hm_{Ah}}$.

The delete-one jackknife variance estimator for the ECMV is calculated with replicate estimates $\dddot{t}_{yR(r)}$ computed as described for the ECF2m in (4.31) but with $\hat{\mathbf{t}}_{Bx(r)}$ defined in (4.37). Note that we use the same technique as shown in expression (4.28) for the approximation below because $\hat{\boldsymbol{\varepsilon}}_{B(r)}$, like $\hat{\mathbf{z}}_{B(r)}$, is assumed to be $O_P\left(M/\sqrt{m_B}\right)$:

$$\begin{aligned}
&\hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&\quad = \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A + O_P\left(M/\sqrt{(m_A m_B)}\right) \\
&\qquad + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&\quad \cong \hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A \\
&\qquad + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}
\end{aligned} \tag{4.38}$$

Using this approximation, the ECMV jackknife variance formula is specified as:

$$
\begin{aligned}
var_{ECMV}(\hat{t}_{yR}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\dddot{t}_{yR(r)} - \hat{t}_{yGREG})^2 \\
\cong\ & \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)^2 \\
&+\ 2 \sum_{h=1}^{H} c_h^{-1} R_h \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A \\
&+\ \mathbf{B}'_A \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \right] \mathbf{B}_A \\
&+\ 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&+\ 2 \sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A\ \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&+\ \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)} \\
\equiv\ & \sum_{j=1}^{6} \hat{V}_j. \qquad (4.39)
\end{aligned}
$$

Components $\hat{V}_1$ and $\hat{V}_6$ are the same as shown for $var_{ECF2m}(\hat{t}_{yR})$ (4.32) and equal $O_P\left(m_A^{-1}\right)$ and $O_P\left(M^{-1}\right)$ after dividing by M^2 for convenience. These components account for the variation associated with the analytic survey estimates and the analytic survey coverage mechanism, respectively. Note that the expectation and bias of the sixth component, $\sum_h R_h^2 \sum_r \eta_{(r)}^2 \left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}$, is discussed in Section 4.4.3. The cross-product terms involving $\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)$ are asymptotically equal to zero by assuming $\max\left\{M^{-1}\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)\right\}$ converges in probability to zero under the conditions specified by Rao & Wu (1985). The rates of convergence for second and fourth variance components are $O_P\left(m_B^{-1/2}\right)$ and $O_P\left(M^{-3/2}\right)$, respectively. The fifth component is $O_P\left(M^{-1/2}m_B^{-1/2}\right)$ and has unconditional expectation zero because of the standard normal random variables, e.g., $\eta_{(r)}$. The term $\hat{V}_3$ addresses the variation associated with the benchmark control totals and is $O_P\left(m_B^{-1}\right)$. This variance component is shown to be unbiased as long as the estimated benchmark covariance matrix is also unbiased. The unconditional expectation is taken with respect to the MVN distribution (E_ε) as well as the benchmark (E_B) survey:

$$
\begin{aligned}
E&\left[\mathbf{B}_A'\left(\sum_{h=1}^{H}R_h^2\sum_{r=1}^{m_{Ah}}\hat{\boldsymbol{\varepsilon}}_{B(r)}\hat{\boldsymbol{\varepsilon}}_{B(r)}'\right)\mathbf{B}_A\right] \\
&= \mathbf{B}_A'\left(\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}}E_B\left[E_\varepsilon\left(\hat{\boldsymbol{\varepsilon}}_{B(r)}\hat{\boldsymbol{\varepsilon}}_{B(r)}' \mid B\right)\right]\right)\mathbf{B}_A \\
&= \mathbf{B}_A'\left(\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}}E_B\left(\hat{\mathbf{V}}_B\right)\right)\mathbf{B}_A \\
&= \mathbf{B}_A'\ \mathbf{V}_B\ \mathbf{B}_A, \qquad (4.40)
\end{aligned}
$$

if $E_B\left(\hat{\mathbf{V}}_B\right) = \mathbf{V}_B$. Therefore, we see that the $var_{ECMV}(\hat{t}_{yR})$ is an asymptotically unbiased estimator of the population sampling variance under the same conditions as noted for var_{ECF2m}.

The ECMV variance estimator under poststratification, $var_{ECMV}(\hat{t}_{yP})$, is calculated by substituting $\hat{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)}$ into the formula for $var_{ECF2m}(\hat{t}_{yP})$

given in (4.36):

$$\begin{aligned} var_{ECMV}(\hat{t}_{yP}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\ddot{t}_{yP(r)} - \hat{t}_{yPSGR}\right)^2 \\ &\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \Big(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \bar{\mathbf{Y}}_A \\ &\qquad + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}\Big)^2 \end{aligned} \qquad (4.41)$$

with terms defined in (4.37) and (4.35).

Unlike the Fuller method, however, $var_{ECMV}(\hat{\mathbf{t}}_{Bx}) \neq \hat{\mathbf{V}}_B$ as shown below:

$$\begin{aligned} var_{ECMV}(\hat{\mathbf{t}}_{Bx}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Bx})(\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Bx})' \\ &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} c_h^2 R_h^2 \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \\ &= \sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \neq \hat{\mathbf{V}}_B. \end{aligned} \qquad (4.42)$$

The ECMV method instead must rely on the design- and model-based properties of the estimator. The expectation of this estimator is evaluated with respect to the MVN distribution conditioned on the benchmark estimates (E_ε), and then with respect to the benchmark survey design (E_B) as shown in (4.40):

$$\begin{aligned} E[var_{ECMV}(\hat{\mathbf{t}}_{Bx})] &= E_B\left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} E_\varepsilon(\hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mid B)\right] \\ &= \frac{1}{H} \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} E_B(\hat{\mathbf{V}}_B) = E_B(\hat{\mathbf{V}}_B). \end{aligned} \qquad (4.43)$$

Only if $\hat{\mathbf{V}}_B$ is an unbiased estimator of $\mathbf{V}_B$, can we say that in expectation the population covariance matrix is reproduced with this method. This result naturally holds for the EC poststratified estimator where we substitute $\hat{\mathbf{N}}_B$ for $\hat{\mathbf{t}}_{Bx}$ in the expression above.

The stability of the variance estimators is directly related the variability of the sample estimates. The difference in the ECF2 and ECMV variance estimators is associated with the difference in the benchmark control total adjustments. The impact of widely varying replicate adjustments will have a direct effect on the stability of the variance estimates. Under the ECF2 method, $Var[var_{ECF2}(\hat{\mathbf{t}}_{Bx})] = Var_B(\hat{\mathbf{V}}_B)$. However, with the ECMV method,

$$\begin{aligned}
&Var\left[var_{ECMV}(\hat{\mathbf{t}}_{Bx})\right] \\
&= Var_B\left[E_\varepsilon\left(var_{ECMV}(\hat{\mathbf{t}}_{Bx})\right)\right] + E_B\left[Var_\varepsilon\left(var_{ECMV}(\hat{\mathbf{t}}_{Bx})\right)\right] \\
&= \frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r\in s_{Ah}}\left[Var_B\left(E_\varepsilon(\hat{\boldsymbol{\varepsilon}}_{B(r)}\hat{\boldsymbol{\varepsilon}}'_{B(r)})\right) + E_B\left(Var_\varepsilon(\hat{\boldsymbol{\varepsilon}}_{B(r)}\hat{\boldsymbol{\varepsilon}}'_{B(r)})\right)\right] \\
&= \frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}}Var_B\left[\hat{\mathbf{V}}_B\right] + \frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}}E_B\left[2tr(\hat{\mathbf{V}}_B^2)\right] \\
&= Var_B\left(\hat{\mathbf{V}}_B\right) + 2tr\left[E_B(\hat{\mathbf{V}}_B^2)\right] \qquad (4.44)
\end{aligned}$$

where $E_B(\hat{\mathbf{V}}_B^2) = Var_B\left(\hat{\mathbf{V}}_B\right) + \left[E_B(\hat{\mathbf{V}}_B)\right]^2$. The expression above suggests that var_{ECF2} and var_{ECMV} have similar asymptotic properties, i.e., $O\left(M^2/m_B\right)$. In practice, however, the ECMV is likely to be more variable than the ECF2 because of the second (positive) term above. We examine the variability in the variance estimates with our simulation study (Section 4.5).

4.4.5 Nadimpalli-Judkins-Chu Jackknife Method

Nadimpalli et al. (2004) describe a jackknife variance estimator similar to the ECMV in a conference proceedings paper. Though the primary focus of their research was an evaluation of regression models for smoothing monthly estimates from the Current Population Survey, the article contains a data analysis using their proposed EC calibration method (ECNJC). The article does not contain a theoretical evaluation of the ECNJC; we provide the theory below.

The ECNJC, like the ECMV, requires a random perturbation of the replicate control totals to account for the variability in the benchmark estimates. However, unlike either the ECF2 or the ECMV, this method accounts *only* for the benchmark variances instead of the complete benchmark covariance matrix, i.e., only the diagonal elements of $\hat{\mathbf{V}}_B$. The ECNJC replicate controls for $\hat{t}_{yR}$ are defined as

$$\hat{\mathbf{t}}_{Bx(r)} = \mathbf{t}_{Bx} + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \tag{4.45}$$

where $\hat{\mathbf{S}}_B$ is a diagonal matrix of estimated *standard errors* for the benchmark controls, i.e., $\hat{\mathbf{S}}_B = diag(\sqrt{\hat{\mathbf{V}}_B})$; $\boldsymbol{\eta}_{(r)}$ is a G-length vector of values randomly generated independently for each replicate from the standard normal distribution, $N(0,1)$; and, as with the other replication methods, $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$ and $R_h = \sqrt{1/Hm_{Ah}}$. The replicate controls ($\hat{\mathbf{N}}_{B(r)}$) for the $\hat{t}_{yP}$ are defined by substituting the benchmark poststratum counts ($\hat{\mathbf{N}}_B$) for the auxiliary variable totals ($\hat{\mathbf{t}}_{Bx}$) in (4.45) as noted for the other replicate methods.

The development of the formula for the ECNJC delete-one jackknife (sample) variance estimator of a population total begins with the definition of a replicate estimated total. Note that we approximate the following expression using $\mathbf{B}_A$ the vector of assisting-model coefficients specified by the analytic survey frame population:

$$\begin{aligned}
\ddot{t}_{yR(r)} &= \hat{t}_{Ay(r)} + \left(\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)}\right)' \hat{\mathbf{B}}_{A(r)} \\
&= \hat{t}_{Ay(r)} + \left(\left\{\mathbf{t}_{Bx} + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)}\right\} - \hat{\mathbf{t}}_{Ax(r)}\right)' \hat{\mathbf{B}}_{A(r)} \\
&= \hat{t}_{yGREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{A(r)} \\
&= \hat{t}_{yGREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{S}}_B \mathbf{B}_A \\
&\qquad + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{S}}_B O_P\left(m_A^{-1/2}\right) \\
&\cong \hat{t}_{yGREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \qquad (4.46)
\end{aligned}$$

where $\hat{\mathbf{t}}_{Bx(r)}$ is defined in (4.45); $\ddot{t}_{yR(r)} = \hat{t}_{Ay(r)} + \left(\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)}\right)' \hat{\mathbf{B}}_{A(r)}$; $\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) = O_P\left(M\right)$ as noted in the approximation for the other jackknife replicate estimates; and $\hat{\mathbf{S}}_B = O_P\left(M/\sqrt{m_B}\right)$, by assumption. Using the replicate estimate (4.46), the exact and approximate forms of the ECNJC jackknife variance estimator are:

$$\begin{aligned}
var_{ECNJC}(\hat{t}_{yR}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\ddot{t}_{yR(r)} - \hat{t}_{yGREG}\right)^2 \\
&\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)^2 \\
&\quad + 2\sum_{h=1}^{H} R_h c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \\
&\quad + \mathbf{B}'_A \left(\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B\right) \mathbf{B}_A. \qquad (4.47)
\end{aligned}$$

The first variance component accounts for the variation in the analytic survey and is $O_P\left(m_A^{-1}\right)$ by dividing $var_{ECNJC}(\hat{t}_{yR})$ by M^2. Note that this term is the same as shown for the other EC jackknife variance methods studied here. The second component is $O_P\left(m_B^{-1}\right)$ and has expectation zero, as with the other methods, because $E_A\left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)$ is assumed to be zero. Upon further examination, the third variance component is shown to address the variation within the benchmark control totals with order in probability m_B^{-1}. The expectation of this term is taken first with respect to standard normal distribution (E_η), and then the benchmark survey design (E_B):

$$
\begin{aligned}
&\mathbf{B}_A'\left(\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} E\left[\hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)}\boldsymbol{\eta}_{(r)}'\hat{\mathbf{S}}_B\right]\right)\mathbf{B}_A \\
&= \mathbf{B}_A'\left(\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}} E_B\left[\hat{\mathbf{S}}_B\ E_\eta\left(\boldsymbol{\eta}_{(r)}\boldsymbol{\eta}_{(r)}' \mid B\right)\ \hat{\mathbf{S}}_B\right]\right)\mathbf{B}_A \\
&= \mathbf{B}_A'\left(\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}} E_B\left(\hat{\mathbf{S}}_B\hat{\mathbf{S}}_B\right)\right)\mathbf{B}_A \\
&= \mathbf{B}_A' E_B\left(\hat{\mathbf{S}}_B^2\right)\mathbf{B}_A. \qquad (4.48)
\end{aligned}
$$

The term $\boldsymbol{\eta}_{(r)}\hat{\boldsymbol{\eta}}_{(r)}'$ evaluates to a diagonal matrix of dimension G because the components of $\boldsymbol{\eta}_{(r)}$ vector are independent standard normal variables. The expectation of each diagonal element is that of a chi-squared random variable with one degree of freedom ($\chi^2_{(1)}$) where $E\left(\chi^2_{(1)}\right) = 1$. Therefore, $E_\eta\left(\boldsymbol{\eta}_{(r)}\boldsymbol{\eta}_{(r)}' \mid B\right)$ in the calculation above. Note that the matrix $\hat{\mathbf{S}}_B^2$ is the square of the diagonal matrix $\hat{\mathbf{S}}_B$ (4.45) and is not necessarily equal to the benchmark covariance matrix $\hat{\mathbf{V}}_B$ used with the other methods. Therefore, the third variance component given in (4.47) will incorporate the variance in the benchmark estimates in expectation *only if* the population covariance terms are zero. However, if $\mathbf{V}_B$ is not diagonal, then var_{ECNJC} fails this test.

As with the original ECF2 discussed in Fuller (1998) and Isaki et al. (2004), we propose to modify the ECNJC replicate estimates to additionally account for the analytic survey coverage error variance by adding expression first given in (4.30):

$$\dddot{t}_{yR(r)} = \ddot{t}_{yR(r)} + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}, \tag{4.49}$$

with $\ddot{t}_{yR(r)}$ specified under the original ECNJC method (4.46). Using the modified replicate estimators $\dddot{t}_{yR(r)}$ and the approximation discussed for $var_{ECF2m}(\hat{t}_{yR})$ (4.32) and $var_{ECMV}(\hat{t}_{yR})$ (4.39), the modified ECNJC (ECNJCm) jackknife variance estimator is specified as:

$$\begin{aligned} var_{ECNJCm}(\hat{t}_{yR}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{t}_{yR(r)} - \hat{t}_{yGREG}\right)^2 \\ &\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \Bigg(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \\ &\qquad + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}}\Bigg)^2 . \end{aligned}$$

Continuing,

$$
\begin{aligned}
var_{ECNJCm}(\hat{t}_{yR}) \cong & \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right)^2 \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} R_h \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \\
&+ \mathbf{B}'_A \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B\right] \mathbf{B}_A \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yGREG(r)} - \hat{t}_{yGREG}\right) R_h \eta_{(r)} \sqrt{\left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&+ 2 \sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \; \eta_{(r)} \sqrt{\left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)}} \\
&+ \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\bar{\phi}}_{A(r)}\right) \hat{t}_{Ae2(r)} \qquad (4.50)
\end{aligned}
$$

The evaluation of the additional variance components follows the discussion given for $E\left[var_{ECF2m}(\hat{t}_{yR})\right]$ (4.32). Even though an additional positive variance component is added to the original ECNJC variance formulation, this term is of lower order than required to inflate for underestimation associated with the use of $\hat{\mathbf{S}}_B$.

The ECNJCm replicate estimator under EC poststratification is derived by specializing expression (4.49):

$$
\begin{aligned}
\dddot{t}_{yP(r)} &= \hat{\mathbf{N}}'_{B(r)} \hat{\mathbf{N}}^{-1}_{A(r)} \hat{\mathbf{t}}_{Ay(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\
&= \left[\mathbf{N}_B + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)}\right]' \hat{\bar{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\
&= \mathbf{N}'_B \hat{\bar{\mathbf{Y}}}_{A(r)} + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}'_{(r)} \hat{\bar{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}, \qquad (4.51)
\end{aligned}
$$

with terms defined for the ECF2m method in (4.35). The corresponding EC-PSGR jackknife variance estimator defined for the ECNJCm method is:

$$
\begin{aligned}
var_{ECNJCm}(\hat{t}_{yP}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\ddot{t}_{yP(r)} - \hat{t}_{yPSGR}\right)^2 \\
&\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR}\right)^2 \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} R_h \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR}\right) \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \\
&+ \mathbf{B}'_A \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \right] \mathbf{B}_A \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{yPSGR(r)} - \hat{t}_{yPSGR}\right) R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\
&+ 2 \sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \ \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}} \\
&+ \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)} \qquad (4.52)
\end{aligned}
$$

Use of the ECNJC would be plausible in two cases: (i) the complete benchmark covariance matrix for the controls is unavailable (e.g., estimates taken from a previous report), or (ii) the covariance terms are negative so that the replicate values defined in (4.45) would lead to a conservative variance estimate. The diagonal matrix $\hat{\mathbf{S}}_B$ would be correct if the auxiliary variables (or, in the case of poststratification, the estimated poststratum counts) were actually uncorrelated. However, this is unlikely especially for the EC poststratified estimator because of the multinomial structure of $\hat{\mathbf{N}}_B$. Given the setup for the ECNJC, the expectation of the variance estimator will *not* approximate (4.16); the bias is related to the difference between $\mathbf{V}_B$ and the expectation of $\hat{\mathbf{S}}_B$ squared, i.e., $E_B(\hat{\mathbf{S}}_B^2)$.

4.5 Simulation Study

We complement the theoretical evaluation of the variance estimators presented in the previous sections with an analysis of simulation results for the EC poststratified estimator of a total $\hat{t}_{yP}$ given in expression (4.3). The variance estimators include:

1. *Naïve*, the traditional calibration estimator defined in expression (4.21);
2. *ECTS*, the EC linearization estimator defined in (4.24);
3. *ECF2m*, the modified Fuller two-phase jackknife estimator (4.36) that includes an adjustment for analytic frame undercoverage (4.30);
4. *ECMV*, the Multivariate normal jackknife estimator (4.41); and,
5. *ECNJCm*, the modified Nadimpalli-Judkins-Chu jackknife estimator defined in (4.52).

We additionally compare the modified Fuller method (ECF2m) against *ECF2*, the original Fuller two-phase jackknife estimator (4.26) defined under EC poststratification, as well as the modified Nadimpalli-Judkins-Chu (ECNJCm) method against *ECNJC*, the original Nadimpalli-Judkins-Chu jackknife estimator defined in expression (4.47) for $\hat{t}_{yP}$. The former comparisons will suggest the use of one or more variance estimators for EC calibration, while the latter comparison will suggest the effectiveness of the undercoverage error variance component in properly inflating the overall variance estimates.

4.5.1 Simulation Parameters

The simulation population is a random subset of the 2003 National Health Interview Survey (NHIS) public-use file containing records for 21,664 U.S. residents. These records are categorized within 25 design strata, each containing six PSUs ($M_{Ah} = 6$). Samples for the analytic survey are selected from this "population" using a two-stage design. Two PSUs ($m_{Ah} = 2$) are selected *with replacement* using probabilities proportional to the total number of persons (PPS) within the PSU. From within each PSU, we selected a simple random sample (n_{Ahi}) of 20 and 40 persons *without replacement* resulting in a total sample size ($n_A = \sum_h \sum_{i=1}^{m_{Ah}} n_{Ahi}$) of 1,000 and 2,000, respectively. Two within-PSU sample sizes were considered for this study to evaluate the effects of smaller analytic survey variance components, calculated by increasing the level of n_A, on the variance of $\hat{t}_{yP}$. For each combination of PSU and size of the person-level samples (i.e., 50 PSUs and either 1,000 or 2,000 persons), we selected 4,000 simulation samples. We calculate the estimated population totals and associated variances for two binary NHIS variables in separate runs of the simulation program: NOTCOV=1 indicates that an adult *did not* have health insurance coverage in the 12 months prior to the NHIS interview ($t_y = 3,653$, approximately 17.1 percent of the population); and PDMED12M=1 indicates that an adult *delayed* medical care because of cost in the 12 months prior to the interview ($t_y = 1,522$, approximately 7.1 percent of the population). We exclude nonresponse from consideration in our current simulation study to minimize factors that could cloud our comparisons.

Table 4.1: Coverage Rates within the 16 Poststratification Cells by Outcome Variable.

Age	Not Covered by Health Insurance (NOTCOV)		Delayed Medical Care (PDMED12M)	
	Male	Female	Male	Female
< 5	0.9	0.9	0.9	0.9
5-17	0.8	0.8	0.8	0.8
18-24	0.5	0.5	0.6	0.5
25-44	0.5	0.5	0.6	0.5
45-64	0.8	0.8	0.6	0.5
65-69	0.9	0.9	0.9	0.5
70-74	0.9	0.9	0.9	0.7
75 +	0.9	0.9	0.9	0.8

Poststratification may reduce variances slightly. However, in household surveys, this technique is mainly used to correct for sampling frame undercoverage, as well as other problems inherent with surveys. Each of the 4,000 simulation samples is randomly selected from a sampling frame that suffers from differential undercoverage, such as those used for many telephone surveys. The 16 poststratification cells are defined by an eight-level *age* variable crossed with *gender*. The coverage rates for the 16 cells by outcome variable are provided in Table 4.1. These coverage rates were created based on the population means for each age by gender group. A coverage rate equal to 1.0 would indicate full coverage. Before each analytic survey sample is selected, a stratified random subsample is drawn from the full population using the coverage rates in Table 4.1 to create the analytic survey sampling frame. For example, 90 percent of the male population less than five years of age (age < 5, male) is randomly selected to be in the analytic survey sampling frame. This process

Table 4.2: Benchmark Control Total Correlations for Males by Age Groups Ranging from 18 to 69.

	18-24	25-44	45-64	65-69
18-24	1.00	0.37	0.29	0.01
25-44	0.37	1.00	0.31	0.10
45-64	0.29	0.31	1.00	0.19
65-69	0.01	0.10	0.19	1.00

of subsetting the population to the frame was independently implemented for each sample and for each outcome variable.

We suspect that the decision for researchers to use either a traditional or an EC calibration variance estimator will depend on the precision of the control totals. We calculated the population benchmark poststratum counts ($\mathbf{N}_B$) and covariance matrix ($\mathbf{V}_B$) from the complete NHIS public-use data file (92,148 records) and ratio adjusted the values to reflect a sample size comparable with our simulation population (N=21,664). A few example correlations for the covariance matrix $\mathbf{V}_B$ are provided in Table 4.2; the off-diagonal values range from -0.05 to 0.75 with a mean value of 0.22. From this matrix $\mathbf{V}_B$ we calculated four covariance matrices for the simulation $\left(\{\mathbf{V}_{Bl}\}_{l=1}^{4}\right)$ by dividing the original matrix by the adjustment factors 1.0, 3.6, 18, and 72. The adjustments reflect benchmark surveys with approximate effective sample sizes of 21,700, 6,000 ($\cong 21,700/3.6$), 1,200, and less than 500, respectively. The $\{\mathbf{V}_{Bl}\}_{l=1}^{4}$ are used directly in the calculation of the sample variance estimates in place of $\hat{\mathbf{V}}_B$. For example, the $\mathbf{V}_{Bl}$'s were (spectrally) decomposed for the 4,000 simulation samples to generate the ECF2 replicate control totals. From each of the four $\mathbf{V}_{Bl}$'s, we generated 4,000 estimated benchmark control total vectors

($\hat{\mathbf{N}}_B$) of length 16 using a multivariate normal (MVN) distribution such that $\hat{\mathbf{N}}_B \sim MVN_{16}(\mathbf{N}_B, \mathbf{V}_{Bl})$. These control total vectors were used to calculate the replicate controls $\hat{\mathbf{N}}_{B(r)}$ for all the jackknife methods. We chose not to randomly generate a $\hat{\mathbf{V}}_B$ for each $\hat{\mathbf{N}}_B$ using, for example, a Wishart distribution in order to simplify the simulation study. In short, the $\hat{\mathbf{N}}_B$'s varied from one simulated sample to another but the $\hat{\mathbf{V}}_B$'s did not.

In summary, the sources of variation accounted for in our simulation study can be classified into two groups — external and internal. External conditions vary across the set of simulation samples but do not vary within each set of 4,000 simulation samples. These include variation in the outcome variable (y=NOTCOV or PDMED12M), the size of the analytic survey sample ($n_A = 1,000$ or $2,000$), and the benchmark covariance matrix $\left(\{\mathbf{V}_{Bl}\}_{l=1}^{4}\right)$. Internal conditions vary within the set of simulations samples and include: creation of the analytic survey sampling frame, selection of the analytic survey sample units, generation of the benchmark control totals ($\hat{\mathbf{N}}_B$), selection of the G replicates to receive an ECF2 (spectral decomposition) adjustment factor, and generation of the multivariate and standard normal random variables for the ECMV and ECNJC methods.

The simulation was conducted in R® (Lumley, 2005; R Development Core Team, 2005) because of its extensive capabilities for analyzing survey data and efficiency in conducting simulation studies. We developed program code to calculate the linearization and replicate variance estimates for $\hat{t}_{yP}$ because the relevant code does not currently exist. The programs developed for the simulation studies are provided in Appendix A.

4.5.2 Evaluation Criteria

The empirical results for the variance estimators listed at the beginning of this section (Section 4.5) are compared using several measures across the j ($j = 1, ..., 4000$) simulation samples and two outcome variables (NOTCOV, PDMED12M). The measures include:

1. $100 \times \left[\left((1/4000)\sum_j var(\hat{t}_{yP_j}) - MSE\right)/MSE\right]$, the estimated percent bias of the variance estimator relative to the empirical $MSE = (1/4000)\sum_j(\hat{t}_{yP_j} - t_y)^2$;

2. $100 \times \left[\left((1/4000)\sum_j var(\hat{t}_{yP_j}) - VAR\right)/VAR\right]$, the estimated percent bias of the variance estimator relative to the empirical variance where $VAR = (1/4000)\sum_j(\hat{t}_{yP_j} - (1/4000)\sum_j \hat{t}_{yP_j})^2$;

3. $(1/4000)\sum_j I\left(|\hat{z}_j| \leq z_{1-\alpha/2}\right)$, the 95 percent confidence interval coverage rate where $\alpha = 0.05$, $\hat{z}_j = (\hat{t}_{yP_j} - t_y)/se(\hat{t}_{yP_j})$, and $se(\hat{t}_{yP_j}) = \sqrt{var(\hat{t}_{yP_j})}$;

4. $\sqrt{\frac{1}{(4000-1)}\sum_j\left[se(\hat{t}_{yP_j}) - (1/4000)\sum_j se(\hat{t}_{yP_j})\right]^2}$, the standard deviation of the estimated standard errors (se); and,

5. $100 \times \left[\left((1/4000)\ \sum_j se_*\left(\hat{t}_{yP_j}\right)\ -\ (1/4000)\ \sum_j se_{ECTS}\left(\hat{t}_{yP_j}\right)\right)\ /\ (1/4000)\ \sum_j se_{ECTS}\left(\hat{t}_{yP_j}\right)\right]$, the percent increase in the variation of the estimated standard errors for all studied estimators (se_*) relative to the ECTS variance estimator (se_{ECTS}).

Prior to comparing the variance estimators, we evaluate the relative bias of the estimated totals, $(1/4000)\sum_j\left(\hat{t}_{yP_j} - t_y\right)/t_y$ discussed in the next section.

4.5.3 Results for Point Estimators

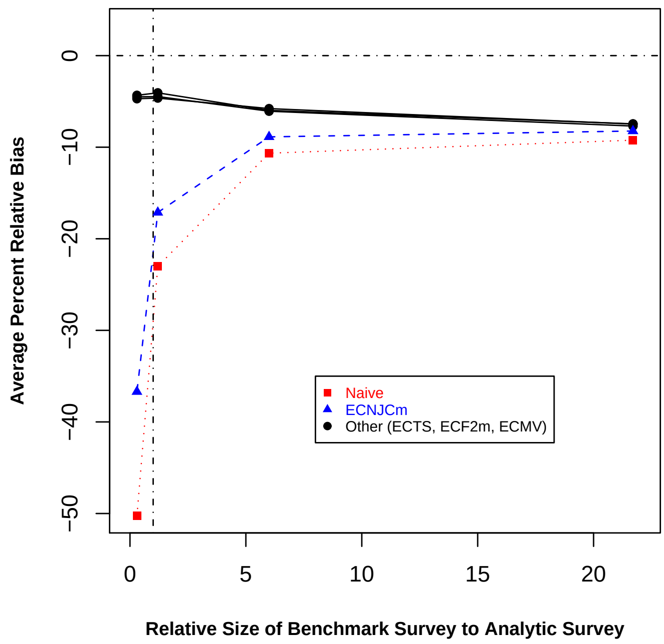
To justify the need for calibration, we initially evaluated the *pwr* estimated totals ($\hat{t}_{Ay} = \sum_{hik \in s_A} \pi_{hik}^{-1} y_{hik}$) for the two outcome variables. This estimator is known to be design-unbiased under pristine conditions (see Result 2.91 Särndal et al., 1992). The percent relative bias indicates that the point estimator is negatively biased, underestimating the population total by 38 percent for NOTCOV and 41 percent for PDMED12M. Also, the 95 percent confidence intervals for the empirical bias of NOTCOV and PDMED12M are (-1,852.2, -898.8) and (-854.2, -391.1), respectively, and do not cover a bias of zero. These large negative values show that some correction is needed to adjust for the non-negligible levels of undercoverage bias.

The percent relative bias for the poststratified estimator $\hat{t}_{yP}$ was much lower — the $\hat{t}_{yP}$ is positively biased by no more than 2 percent for both outcome variables. The EC 95 percent confidence intervals for the bias of NOTCOV and PDMED12M do contain zero as desired and are calculated as (-664.4, 819.2) and (-380.8, 422.3), respectively. Even though population values were not used for the calibration, the EC calibration using benchmark survey estimates greatly improved the MSE of our estimated totals. Estimated poststratum counts from the benchmark survey ($\hat{\mathbf{N}}_B$) were larger (at most 10 percent) than the corresponding values in the population ($\mathbf{N}$) for five out of the 16 poststrata. This is likely associated with the small positive percent relative bias seen for the EC poststratified estimator.

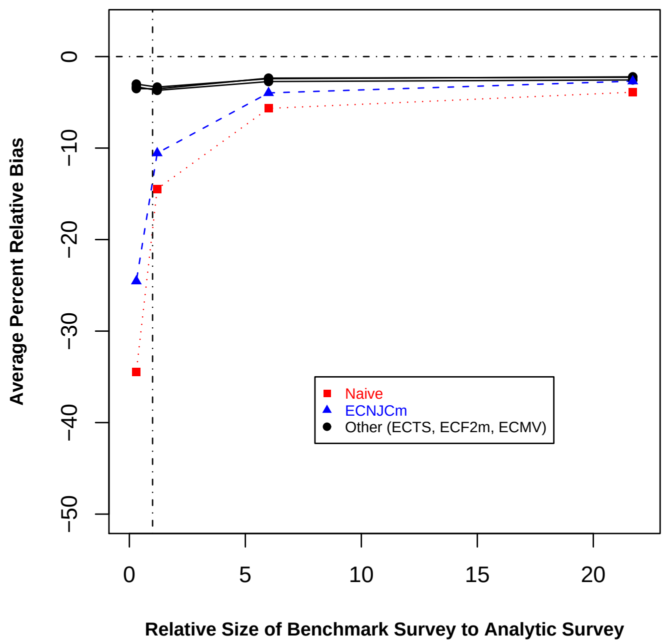
4.5.4 Comparison of Variance Estimators

Adding to the theoretical evaluation in Section 4.4, the empirical results for an effective variance estimator should possess a *percent bias relative to the empirical MSE* either near zero (unbiased) or somewhat positive (conservative measure). Figure 4.1 shows the general pattern of our results through the percent relative bias of five variance estimators (Naïve, ECTS, ECF2m, ECMV, and ECNJCm) by the increasing size (left to right on the x axis) of the benchmark survey relative to the 1,000 persons selected for the analytic survey (n_B/n_A) for NOTCOV (a) and PDMED12M (b). Note that in our study the increase in the benchmark survey size is directly related to an increase in the precision of the estimated control totals. The horizontal line represents zero bias, while the vertical line represents the effect for equal-sized analytic and benchmark surveys. Estimates for the Naïve and ECNJC estimators are represented by squares and triangles, respectively. The "Other EC" estimates (ECTS, ECF2m, and ECMV) are similar in value and are shown as circles.

The traditional poststratified estimator (Naïve) is most negatively biased among those compared as expected for both outcome variables. When the benchmark survey is smaller than the analytic survey (and therefore produces estimates less precise than the analytic survey), the Naïve estimator is negatively biased by as much as 50 percent for NOTCOV and 35 percent for PDMED12M. The level of bias improves as the relative size of the benchmark survey increases; however, the Naïve estimator still results in, at best, a 4 percent underestimate for the variables considered. The ECNJCm estimator fares slightly better than the Naïve estimator though the bias



(a) Total Number Not Covered by Health Insurance in Last 12 Months (NOTCOV)



(b) Total Number Delayed Medical Care Due to Cost in Last 12 Months (PDMED12M)

Figure 4.1: Percent Bias Relative to Empirical MSE of Five Variance Estimators by Relative Size of the Benchmark Survey to the Analytic Survey for 1,000 Analytic Survey Units.

is still larger than the other EC variance estimators — biases range from -8 to -37 percent for NOTCOV and from -3 to -25 percent for PDMED12M. The percent relative biases for the remaining estimators fall between -2 and -8 percent.

For a small benchmark survey relative to the size of the analytic survey (left of the vertical line), the levels of (absolute) bias dramatically increase for the Naïve and ECNJC estimators. either a negligible effect (NOTCOV) or an opposite (PDMED12M) effect is seen for the other EC variance estimators. The variance component associated with the benchmark survey, e.g., $\hat{\hat{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\hat{\mathbf{Y}}}_A$ in (4.24), becomes the dominate term within the EC variance estimators to the left of the (vertical) line of equality. Thus the benchmark variance component somewhat corrects for the negative bias associated with the analytic variance component. Additional research is needed to determine if a threshold exists for when such a counterbalance of bias can occur.

The percent biases relative to the empirical MSE generated from our simulation study are provided in Table 4.3. The 20 NOTCOV and PDMED12M estimates for $n_A = 1,000$ were used to generate Figure 4.1. Bias estimates for the Naïve and ECNJC estimators are larger than the other EC estimates for all our simulations. Differences are negligible for the remaining variance estimators under all conditions studied. Note that the relative sizes of 21.7 and 10.8 both imply benchmark survey sample sizes of about 21,700. Thus the variance components associated with the benchmark survey estimates are more prominent for the estimates in Table 4.3 based on $n_A = 2,000$. This leads to larger relative biases in these estimates, relative to those produced under $n_A = 1,000$, even though the analytic survey sample size is

Table 4.3: Percent Bias Estimates Relative to Empirical MSE for Five Variance Estimators by Outcome Variable and Relative Size of the Benchmark Survey to the Analytic Survey.

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-50.3	-23.0	-10.7	-9.2	-56.0	-31.0	-14.2	-12.2
	ECTS	-4.5	-4.5	-6.1	-7.7	-0.2	-8.4	-8.2	-10.1
	ECF2m	-4.7	-4.6	-5.8	-7.5	0.1	-8.2	-8.3	-10.1
	ECMV	-4.3	-4.1	-6.0	-7.5	-0.2	-8.1	-8.1	-10.0
	ECNJCm	-36.7	-17.1	-8.9	-8.2	-40.0	-24.2	-11.9	-11.1
PDMED12M	Naïve	-34.4	-14.5	-5.7	-3.9	-48.1	-23.4	-10.0	-10.1
	ECTS	-3.3	-3.7	-2.7	-2.6	-4.7	-6.4	-5.1	-7.8
	ECF2m	-3.5	-3.5	-2.4	-2.3	-4.6	-6.8	-5.2	-7.8
	ECMV	-3.0	-3.3	-2.4	-2.2	-4.3	-6.3	-5.0	-7.7
	ECNJCm	-24.5	-10.5	-4.0	-2.7	-35.1	-17.6	-7.6	-8.4

Table 4.4: Percent Bias Estimates Relative to Empirical Variance for Five Variance Estimators by Outcome Variable and Relative Size of the Benchmark Survey to the Analytic Survey.

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-49.5	-20.6	-6.9	-5.5	-54.9	-28.4	-8.8	-5.8
	ECTS	-3.1	-1.5	-2.1	-3.9	2.3	-5.0	-2.4	-3.6
	ECF2m	-3.3	-1.7	-1.8	-3.6	2.6	-4.8	-2.5	-3.5
	ECMV	-2.9	-1.1	-2.0	-3.6	2.3	-4.7	-2.3	-3.4
	ECNJCm	-35.8	-14.6	-5.1	-4.4	-38.5	-21.4	-6.4	-4.6
PDMED12M	Naïve	-33.9	-13.6	-4.7	-2.9	-47.3	-22.0	-8.2	-7.9
	ECTS	-2.4	-2.7	-1.7	-1.6	-3.2	-4.6	-3.1	-5.6
	ECF2m	-2.6	-2.5	-1.3	-1.3	-3.1	-5.0	-3.2	-5.5
	ECMV	-2.1	-2.3	-1.4	-1.2	-2.8	-4.6	-3.0	-5.4
	ECNJCm	-23.9	-9.6	-3.0	-1.7	-34.1	-16.1	-5.7	-6.1

larger.

The overall negative bias of our estimates is similar to the bias of linearization variance estimators shown for the combined ratio estimator ($\hat{t}_{Ayg}/\hat{N}_{Ag}$) in Section 4 of Rao & Wu (1985) and in Wu (1985). As noted in Section 4.5.3, the estimated totals are slightly larger than the corresponding population total. Therefore, we additionally examine the *percent bias relative to the empirical variance* to determine if the empirical bias is affecting our results. Table 4.4 shows a noticeable decrease in the negative biases in comparison to the values presented in Table 4.3.

The patterns exhibited for the percent relative bias are reflected in the *coverage rates for the 95 percent confidence interval* for the estimated totals (Table 4.5). The Naïve and ECNJC estimators are more likely to experience confidence intervals coverage rates below 95 percent. These rates approach the appropriate level as the precision of the benchmark survey estimates improves. However, the remaining EC variance estimators had coverage rates near acceptable levels regardless of the relative size of the surveys and therefore are more robust.

The discussion so far suggests that there are minimal theoretical, as well as empirical, differences between the ECTS, ECF2m, and ECMV methods. We look to the *standard deviation of the estimated standard errors* (SEs) in an attempt to distinguish the estimators. An examination of this variability can provide insight on the (empirical) stability of the variance estimators because an unstable variance estimator could generate a poor variance estimate based on the nuances of the particular sample selected. Table 4.6 contains the percent increase in the instability (i.e., variability) for all variance estimators against the ECTS. Minor differences

Table 4.5: Empirical 95 Percent Coverage Rates for Five Variance Estimators by Outcome Variable and Relative Size of the Benchmark Survey to the Analytic Survey.

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	83.5	91.7	93.7	93.6	81.2	89.5	92.8	93.4
	ECTS	95.6	94.4	94.2	94.0	95.7	94.3	93.7	93.7
	ECF2m	95.1	94.1	94.2	93.9	95.5	94.2	93.5	93.8
	ECMV	95.1	94.5	94.4	94.0	95.5	94.0	93.6	93.8
	ECNJCm	88.6	92.7	94.0	93.9	87.8	91.0	93.1	93.6
PDMED12M	Naïve	88.8	93.1	94.4	94.4	84.8	91.8	94.2	93.8
	ECTS	94.8	94.8	94.7	94.5	95.4	94.7	94.8	94.1
	ECF2m	94.8	94.8	94.8	94.5	95.2	94.7	94.7	94.1
	ECMV	95.0	94.8	94.7	94.4	94.7	94.8	94.8	94.0
	ECNJCm	91.1	93.7	94.5	94.4	89.0	92.8	94.4	93.9

in the stability of the estimates are seen for relatively large benchmark surveys. However, as the benchmark estimates themselves become less stable, the variation in the estimates also become less stable in comparison to the variation in the ECTS estimates for all simulation conditions studied. The largest increase is noted for the multivariate method (ECMV) and is attributed to the use of values from the multivariate normal distribution with the complete benchmark covariance matrix as discussed in Section 4.4.4.

We conclude this section with an examination of the effects of the undercoverage error variance component introduced into the original formulae for the Fuller and Nadimpalli-Judkins-Chu jackknife variance estimators. Table 4.7 shows the percentage point reduction in the bias of the variance estimates relative to the empirical

Table 4.6: Percent Increase in Instability of Variance Estimates Relative to EC Linearization Estimator (ECTS) by Outcome Variable and Relative Size of the Benchmark Survey.

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	17.3	7.5	2.1	0.7	23.5	11.2	3.0	1.0
	ECF2m	12.0	5.5	2.3	0.2	15.1	8.4	2.1	0.6
	ECMV	21.2	7.4	1.8	0.3	30.8	8.5	2.4	0.7
	ECNJCm	14.5	7.0	1.9	0.5	19.2	9.9	2.8	1.1
PDMED12M	Naïve	4.9	2.4	0.7	0.4	12.3	6.1	1.9	1.0
	ECF2m	7.7	3.8	1.1	0.4	12.0	6.3	2.1	0.7
	ECMV	11.5	4.0	0.9	0.5	22.6	7.6	2.2	1.1
	ECNJCm	5.3	2.6	0.7	0.3	13.4	6.2	2.2	0.9

Table 4.7: Percentage Point Reduction in Bias Relative to Empirical MSE Attributed to Coverage Error Variance by EC Variance Estimator, Outcome, and Relative Size of Benchmark Survey to the Analytic Survey.

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	ECF2	-0.2	-0.3	-0.4	-0.4	-0.2	-0.5	-0.5	-0.6
	ECJNC	-0.2	-0.3	-0.4	-0.4	-0.2	-0.5	-0.5	-0.6
PDMED12M	ECF2	-0.5	-0.7	-0.7	-0.7	-0.8	-1.0	-1.2	-1.3
	ECJNC	-0.5	-0.7	-0.7	-0.7	-0.7	-1.0	-1.2	-1.3

variance in using the modified variance estimators. Overall, the relative bias is reduced between 0.2 and 1.3 percentage points, with larger reductions occurring for the larger benchmark surveys. This is consistent with the coverage error variance component in (4.16) having order $O(M)$, i.e., not dependent on the sample size from either the analytic or benchmark surveys. The differences in the 95 percent coverage rates are not appreciable and are therefore not shown. This suggests that an undercoverage error adjustment is useful for the variance estimator; however, further research is needed to produce a more effective adjustment factor.

4.6 Summary of Research Findings

The theoretical and analytical work discussed in this chapter support the need for a new methodology to address calibration using estimated control totals, i.e., estimated-control (EC) calibration. Traditional variance estimators can severely underestimate the population sampling variance in estimated totals resulting in, for example, incorrect decisions for hypothesis tests and sub-optimal sample allocations when the design is optimized in the future. This is especially noticeable with relatively small benchmark surveys and has implications for studies such as the Web/RDD calibration example discussed at the end of Section 2.2.

The EC linearization variance estimators $var_{ECTS}(\hat{t}_{yR})$ (4.23) shows the most promise for EC calibration given the evaluation criteria used for our study. This estimator is effective at reducing the percent relative bias experienced with the Naïve variance estimator (4.20) when the benchmark survey is small relative to the analytic survey. The replication variance estimator var_{ECF2m} (4.32), the Fuller two-phase jackknife variance estimator augmented with an undercoverage error variance component, also reduces the relative bias and is recommended specifically for studies requiring replicate weights. These include, for example, public-use analysis files that are to be released without sampling design information to further protect data confidentiality and respondent privacy. The ECMV method is asymptotically equivalent to the recommended variance estimators; however, the instability of the estimates may make this variance estimator less attractive.

Implementation of the two recommended variance estimators requires specialized computer programs because the capabilities are currently not available in standard software. The linearization estimator may be more approachable because it involves a modification to available variance estimates (see Section 4.4.2 for further discussion). We provide a step-by-step guide to the procedures required for the var_{ECF2m} (Section 4.4.3) to facilitate the creation of the replicate weight program.

Chapter 5

Ratio of Two Estimated Population Totals

5.1 Introduction

The ratio estimator of a population mean, also known as a Hájek estimator (Hájek, 1971), is calculated as the estimated total for an outcome variable divided by the estimated population size ($\hat{N}$). The estimated population size is generally obtained by summing the final analysis weights for all sample cases. This estimator has for many sampling designs a smaller variance than the corresponding Horvitz-Thompson estimator (Horvitz & Thompson, 1952) or the *pwr* estimator (Särndal et al., 1992) divided by the *known* population size (N). We examine the ratio estimator for a population mean under estimated-control calibration (EC ratio-mean estimator) in the general regression setting, as well as under poststratification. Formulae for the EC general-regression (EC-GREG) and poststratified (EC-PSGR) ratio-mean estimators are provided in Section 5.2. We evaluate the bias in these estimators in Section 5.3, and compare the levels against those discussed for the estimator of the population total used in the numerator of the ratio (Section 4.3). The variance estimators included in the Chapter 4 evaluation are compared for the ratio-mean estimators in Section 5.4. We confirm our theoretical findings through a simulation study (Section 5.5) and summarize our theoretical and empirical results in the final section of the chapter (Section 5.6). The work in this chapter builds on the research presented in Chapter 4. Some of the Chapter 4 equations are repeated in Chapter 5 to complete the discussion, while others are merely referenced for the sake of brevity.

5.2 Point Estimators

The EC-GREG ratio-mean estimator is defined as

$$\hat{\bar{y}}_R = \frac{\hat{t}_{yR}}{\hat{N}_R} \tag{5.1}$$

where $\hat{t}_{yR} = \hat{t}_{Ay} + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_A$ as shown in expression (4.1). The denominator is an EC-GREG estimator of the population size and is calculated as

$$\begin{aligned}
\hat{N}_R &= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1} \\
&= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \left[1 + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})' \left(\sum_{s_A} \pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right)^{-1} \mathbf{x}_{hik}\right]\pi_{hik}^{-1} \\
&= \hat{N}_A + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_{AN} \qquad (5.2)
\end{aligned}$$

where π_{hik}^{-1} $(= 1/m_{Ah}\pi_{hi(1)}\pi_{k|hi})$ is the analytic survey design weight for the k^{th} sample unit in PSU i selected with-replacement within stratum h; a_{hik} is the EC calibration adjustment factor for the k^{th} unit; $\hat{N}_A = \sum_{hik\in s_A}\pi_{hik}^{-1}$, the estimated population size using only the analytic survey data (s_A); and

$$\hat{\mathbf{B}}_{AN} = \left(\sum_{hil \in s_A} \pi_{lik}^{-1} \mathbf{x}_{lik} \mathbf{x}'_{hil} \right)^{-1} \hat{\mathbf{t}}_{Ax}, \tag{5.3}$$

the model coefficient vector used in the numerator with $y_{hik} = 1$ for all sample units with $\hat{\mathbf{t}}_{Ax} = \sum_{hik \in s_A} \pi_{hik}^{-1} \mathbf{x}_{hik}$. Note that a_{hik} specified for $\hat{N}_R$ is the same as defined for $\hat{t}_{yR}$ in (4.1). Särndal et al. (1992, Section 7.13) refer to $\hat{\bar{y}}_R$ in (5.1) as a specific type of "ratio of population totals" estimator of the form $\hat{t}_{yR}/\hat{t}_{zR}$, where $z_{hik} = 1$ for our estimator. Therefore, our discussions of $\hat{\bar{y}}_R$ (and the poststratified ratio-mean estimator) can be extended to a ratio of any two population totals.

The EC-PSGR ratio-mean estimator is similarly defined as

$$\hat{\bar{y}}_P = \frac{\hat{t}_{yP}}{\hat{N}_P} \tag{5.4}$$

where $\hat{t}_{yP} = \sum_{g=1}^{G} \hat{N}_{Bg} \hat{N}_{Ag}^{-1} \hat{t}_{Ayg}$, the EC poststratified estimator of a total defined in expression (4.3) with estimates summed over the G poststrata. The population size, estimated through EC poststratification, simplifies to

$$\begin{aligned} \hat{N}_P &= \sum_{g=1}^{G} \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \left(\frac{\hat{N}_{Bg}}{\hat{N}_{Ag}} \delta_{ghik} \right) \pi_{hik}^{-1} \\ &= \sum_{g=1}^{G} \left(\frac{\hat{N}_{Bg}}{\hat{N}_{Ag}} \right) \hat{N}_{Ag} \equiv \hat{N}_B, \end{aligned} \tag{5.5}$$

the population size estimated from the benchmark survey where $\delta_{ghik} = 1$ if the k^{th} unit is a member of the poststratum g ($\delta_{ghik} = 0$ otherwise); $\hat{N}_{Ag} = \sum_{hik \in s_A} \delta_{ghik} \pi_{hik}^{-1}$;

and $\hat{N}_{Bg} = \sum_{l \in s_B} \delta_{gl} w_l$ ($\delta_{gl} = 1$ if $l \in s_{Bg}$, zero otherwise). Note that $\hat{N}_B$ may be written in matrix notation as $\hat{N}_B = \hat{\mathbf{N}}'_B \mathbf{1}_G$, where $\hat{\mathbf{N}}'_B = \left[\hat{N}_{B1}, \ ..., \ \hat{N}_{BG} \right]$, the G-length vector of estimated benchmark poststratum counts, and $\mathbf{1}_G$ is a G-length vector of ones. Because a poststratified estimator of a total can be expressed as $\hat{\mathbf{N}}'_B \hat{\bar{\mathbf{Y}}}_A$ for the appropriately chosen y, as discussed in (4.3), $\hat{\mathbf{B}}_{AN} = \hat{\bar{\mathbf{Y}}}_{AN} \equiv \mathbf{1}_G$ for an EC-PSGR estimator where $y_{hik} = 1$ for all analytic sample units.

Note that estimates calculated with the formulae defined for $\hat{\bar{y}}_R$ in (5.1) and $\hat{\bar{y}}_P$ in (5.4) are the same as those calculated for $\hat{\bar{y}}_{GREG}$ in (2.12) and $\hat{\bar{y}}_{PSGR}$ in (2.15). We use different notation primarily for variance estimation to identify situations when the benchmark controls are considered fixed, as with traditional calibration, in comparison to the EC calibration under study.

Using the formula of the ratio-mean estimators presented in this section, we evaluate the properties of the mean square error (MSE) by examining the bias and variance components separately. We begin in the next section by developing expressions for the bias under EC-GREG calibration and more specifically for the EC-PSGR ratio-mean estimator.

5.3 Bias of Point Estimators

Ratio estimators are asymptotically (but not exactly) unbiased because of the estimation required for the random denominator. Särndal et al. (1992, Section 7.3.1) and others, however, note that generally the bias of a ratio estimator is small. For example, the bias is stated as being $O_P(n_A^{-1})$ for a simple random sampling (SRS) design of size n_A. The bias of an EC ratio estimator differs slightly. The first-order bias approximation of $\hat{\bar{y}}_R$ (5.1) is

$$
\begin{aligned}
Bias(\hat{\bar{y}}_R) &\cong \frac{E(\hat{t}_{yR})}{E(\hat{N}_R)} - \bar{y} \\
&= \frac{1}{E(\hat{N}_R)}\left[Bias(\hat{t}_{yR}) + t_y - \bar{y}\ Bias(\hat{N}_R) - \bar{y}N\right] \\
&= \frac{1}{E(\hat{N}_R)}\left[Bias(\hat{t}_{yR}) - \bar{y}\ Bias(\hat{N}_R)\right] \qquad (5.6)
\end{aligned}
$$

where $\bar{y} = \sum_{k\in U} y_k/N$, the population mean of interest. As with the approximation $E(\hat{t}_{yR}) \cong t_{Ay} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B}_A$ given in (4.6), technical conditions, e.g., uniform integrability of certain terms, can be used to formally justify the approximation above (see Serfling, 1980, Thm. C, pg. 15). The unconditional expectation of $\hat{N}_R$ is approximated as

$$
\begin{aligned}
E(\hat{N}_R) &= E\left[\hat{N}_A + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_{AN}\right] \\
&\cong E_{c_A}\left[E_A\left(\hat{N}_A \mid c_A\right)\right] + \left(E_{c_B}\left[E_B\left(\hat{\mathbf{t}}_{Bx} \mid c_B\right)\right]\right. \\
&\qquad \left. - E_{c_A}\left[E_A\left(\hat{\mathbf{t}}_{Ax} \mid c_A\right)\right]\right)' E_{c_A}\left[E_A\left(\hat{\mathbf{B}}_{AN} \mid c_A\right)\right] \\
&\cong N_A + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B}_{AN} \equiv N_R \qquad (5.7)
\end{aligned}
$$

where E_A and E_B are the expectations taken with respect to the (independent) analytic and benchmark surveys, and E_{c_A} and E_{c_B} are the expectation under the coverage mechanisms for the respective sampling frames. As discussed in Section 4.3, $E(\hat{\mathbf{t}}_{Bx}) \cong \sum_{l\in U} \phi_{Bl}\mathbf{x}_l \equiv \mathbf{t}_{Bx}$, and $E(\hat{\mathbf{t}}_{Ax}) \cong \sum_{hik\in U} \phi_{Ahik}\mathbf{x}_{hik} \equiv \mathbf{t}_{Ax}$. Because $\hat{N}_A$ in (5.7) is a *pwr* estimator of a population total given the assumed analytic survey design, the expectation takes the following form:

$$
\begin{aligned}
E(\hat{N}_A) &= E_{c_A}\left[E_A\left(\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\pi_{hik}^{-1} \mid c_A\right)\right] \\
&\cong E_{c_A}\left(\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} C_{Ahik}\right) \\
&= \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik} \equiv N_A \qquad (5.8)
\end{aligned}
$$

where C_{Ahik} is a binary variable to indicate that the k^{th} unit is listed on the analytic survey frame such that $E_{c_A}(C_{Ahik}) = \phi_{Ahik}$, the mean of a Bernoulli distribution as detailed in Section 3.4. The expectation of the remaining term, $\hat{\mathbf{B}}_{AN}$, is similarly approximated as

$$
\begin{aligned}
E(\hat{\mathbf{B}}_{AN}) &= E_{c_A}\left[E_A\left(\hat{\mathbf{B}}_{AN} \mid c_A\right)\right] \\
&\cong \left[\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{l=1}^{N_{hi}} \phi_{Ahil}\mathbf{x}_{hil}\mathbf{x}'_{hil}\right]^{-1} \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\mathbf{x}_{hik} \\
&\equiv \mathbf{B}_{AN}. \qquad (5.9)
\end{aligned}
$$

The $Bias(\bar{y}_R)$ in (5.6) is approximately zero only if the estimators in the numerator and denominator are approximately unbiased, i.e., $E(\hat{N}_R) \cong N$ and $E(\hat{t}_{yR}) \cong t_y$. The conditions under which the $Bias(\hat{t}_{yR}) \cong 0$ are discussed in Section 4.3. These conditions also hold for $Bias(\hat{N}_R)$. The bias of the denominator is similarly defined as with $Bias(\hat{t}_{yR})$ in (4.7) and is specified as:

$$\begin{aligned} Bias(\hat{N}_R) &\cong NC_{ANE\phi} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_{AN} - \mathbf{B}_N) - N\bar{E}_N(1 - \bar{\phi}_A) \\ &\quad +(\mathbf{t}_{Bx} - \mathbf{t}_x)'\mathbf{B}_N \end{aligned}$$

where $E_{Nhik} = 1 - \mathbf{x}'_{hik}\mathbf{B}_N$, the residual calculated for the denominator under the population assisting model, with $\bar{E}_N = \sum_{hik\in U} E_{Nhik}/N$; $C_{ANE\phi} = \sum_{hik\in U} (E_{Nhik} - \bar{E}_N) \left(\phi_{Ahik} - \bar{\phi}_A\right)/N$, the population covariance between the coverage rates and the denominator residuals; and $\mathbf{B}_N = \left[\sum_{hil\in U} \mathbf{x}_{hil}\mathbf{x}'_{hil}\right]^{-1} \mathbf{t}_x$.

The bias of the EC ratio-mean estimator under poststratification evaluates to a form similar to $Bias(\hat{t}_{yP})$ given in (4.8):

$$\begin{aligned} Bias(\hat{\bar{y}}_P) &\cong \frac{E(\hat{t}_{yP})}{E(\hat{N}_P)} - \frac{t_y}{N} \\ &= \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} \left[t_{Ayg} - t_{yg}\frac{N_{Ag}}{N}\frac{N_B}{N_{Bg}} \right] \frac{1}{N_{Ag}} \\ &= \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} \left[\sum_{hik\in U_g} \phi_{Ahik}y_{hik} - N_g\bar{\phi}_{Ag}\bar{y}_g + N_g\bar{\phi}_{Ag}\bar{y}_g - t_{yg}\frac{N_{Ag}}{N}\frac{N_B}{N_{Bg}} \right] \frac{1}{N_{Ag}} \\ &= \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} \left[\sum_{hik\in U_g} \left(\phi_{Ahik} - \bar{\phi}_{Ag}\right)\left(y_{hik} - \bar{y}_g\right)/N_g \right] \frac{N_g}{N_{Ag}} \\ &\quad + \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} \left[N_g\bar{\phi}_{Ag}\bar{y}_g - t_{yg}\frac{N_{Ag}}{N}\frac{N_B}{N_{Bg}} \right] \frac{1}{N_{Ag}}. \\ &= \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} C_{Ay\phi g}\frac{1}{\bar{\phi}_{Ag}} + \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} \left[N_{Ag}\frac{t_{yg}}{N_g} - t_{yg}\frac{N_{Ag}}{N}\frac{N_B}{N_{Bg}} \right] \frac{1}{N_{Ag}}. \\ &= \sum_{g=1}^{G} \frac{N_{Bg}}{N_B} C_{Ay\phi g}\frac{1}{\bar{\phi}_{Ag}} + \frac{1}{N}\left[\sum_{g=1}^{G} t_{yg} \left\{ \frac{N_{Bg}}{N_B}\frac{N}{N_g} - 1 \right\} \right]. \end{aligned}$$

Therefore,

$$Bias(\hat{\bar{y}}_P) \cong \sum_{g=1}^{G} W_{Bg} C_{Ay\phi g} \frac{1}{\bar{\phi}_{Ag}} + \frac{1}{N}\left[\sum_{g=1}^{G} t_{yg}\left\{\frac{W_{Bg}}{W_g} - 1\right\}\right] \qquad (5.10)$$

where U_g is the complete set of population units in the g^{th} poststratum; $C_{Ay\phi g} = \sum_{hik \in U_g} (y_{hik} - \bar{y}_g)\left(\phi_{Ahik} - \bar{\phi}_{Ag}\right)/N_g$; $\bar{\phi}_{Ag} = N_g^{-1} \sum_{hik \in U} \delta_{ghik}\phi_{Ahik}$, the average analytic survey coverage rate within poststratum g; $\bar{y}_g = N_g^{-1}\sum_{hik \in U} \delta_{ghik} y_{hik} = t_{yg}/N_g$, the mean of y within poststratum g for the complete population; $W_{Bg} = N_{Bg}/N_B$ with $N_B = \sum_g N_{Bg}$; and $W_g = N_g/N$ with $N = \sum_g N_g$. The first bias component is zero if $C_{Ay\phi g}$, the covariance of y and the coverage propensities ϕ_A in poststratum g, is (approximately) zero. This can occur, for example, when poststratum variables are chosen so that the coverage probabilities are constant within each of the G cells (i.e., $\phi_{Ahik} = \phi_g$ for $k \in U_g$). If the benchmark proportions within the poststratum cells are the same as in the population (i.e., $W_{Bg} = W_g$), then the second bias component is zero. Only if both conditions are met will we have an approximately unbiased estimator.

The zero-bias criteria for the ratio-mean estimators discussed above appear to be more easily satisfied than those listed for the EC estimators of a total following expression (4.8). To examine this more fully, we rewrite the relative bias of $\hat{\bar{y}}_R$,

$RelBias\left(\hat{\bar{y}}_R\right)$, as a function of $\hat{t}_{yR}$ (4.1):

$$\begin{aligned} RelBias\left(\hat{\bar{y}}_R\right) &= \frac{Bias(\hat{\bar{y}}_R)}{\bar{y}} \\ &\cong \frac{1}{\bar{y}}\left[\frac{E(\hat{t}_{yR})}{E(\hat{N}_R)} - \bar{y}\right] \\ &= \frac{N}{t_y}\left[\frac{E(\hat{t}_{yR}) - t_y}{N_R} + \frac{t_y}{N_R} - \frac{t_y}{N}\right] \\ &= \frac{N}{N_R}\left[RelBias(\hat{t}_{yR}) + 1 - \frac{N_R}{N}\right], \end{aligned} \tag{5.11}$$

where $E(\hat{N}_R) \equiv N_R$ as shown in (5.7). Figure 5.1 displays the level of $RelBias(\hat{\bar{y}}_R)$ for $RelBias(\hat{t}_{yR}) = 0.1$, 0.2, and 0.3, and the coverage ratios N_R/N ranging from 0.4 to 1.6. The vertical line represents a coverage ratio of 1.0, i.e., $N_R = N$, where $RelBias(\hat{\bar{y}}_R) \cong RelBias(\hat{t}_{yR})$. Coverage ratios to the left of this line denote $RelBias(\hat{\bar{y}}_R)$ for a negatively biased $\hat{N}_R$. For example, $N_R/N = 0.6$ indicates that the EC-GREG estimates undercover the population size by 40 percent, i.e., $100 \times (1 - N_R/N)$. At this level, $RelBias(\hat{\bar{y}}_R) \cong 0.83$ is more than eight times larger than the corresponding $RelBias(\hat{t}_{yR}) = 0.1$. By contrast, an N_R that is too large ($N_R/N > 1$) results in $RelBias(\hat{t}_{yR}) > RelBias(\hat{\bar{y}}_R)$. For example, with $RelBias(\hat{t}_{yR}) = 0.3$, $RelBias(\hat{\bar{y}}_R) \cong 0$ for a 30 percent overestimate ($N_R/N = 1.3$), and increases to $RelBias(\hat{\bar{y}}_R) \cong 0.2$ for a 60 percent overestimate. By noting the slope of the line, we see that the difference between $RelBias(\hat{\bar{y}}_R)$ and $RelBias(\hat{t}_{yR})$ is larger when the population size is underestimated (left of the vertical line) by the calibration system in comparison with an overestimate. Therefore, the figure suggests that the zero-bias criterion for $\hat{\bar{y}}_R$ is not necessarily easier to satisfy than

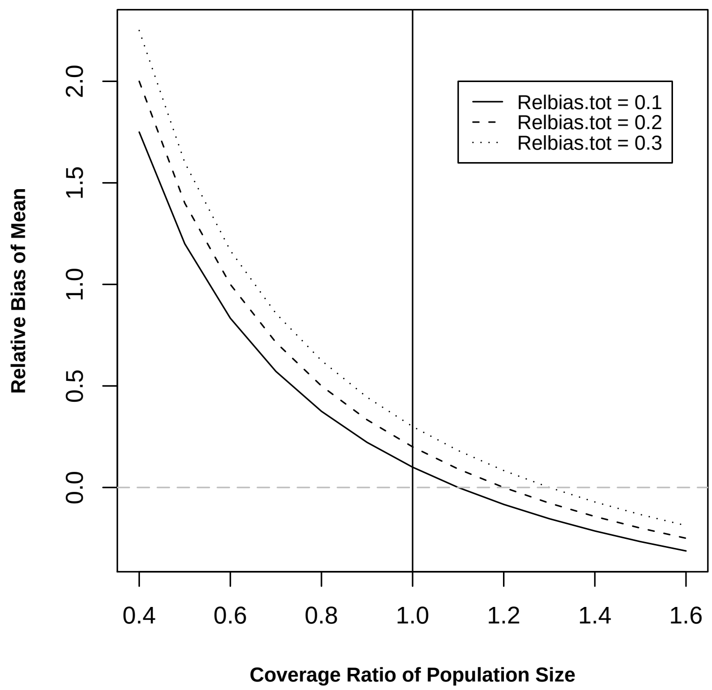


Figure 5.1: Relative Bias of Mean by Coverage Ratio ($\hat{N}_R/N$) of Population Size for Relative Bias of Total equal to 0.1, 0.2, and 0.3.

$\hat{t}_{yR}$ because we have conditions where one but not both are unbiased. The major influence is related to the denominator of the ratio-mean estimator.

5.4 Variance Estimation

Having addressed the bias of $\hat{\bar{y}}_R$ and $\hat{\bar{y}}_P$, we move on to an evaluation of the variance estimators presented in Chapter 4. We begin this section by examining the approximate population sampling variance — the parameter that the sample variance estimators should equal in expectation — and then turn our attention to the sample variance estimators of interest for our research.

The population sampling variance (AV) for the EC-GREG ratio-mean estimator $\hat{\bar{y}}_R$ (5.1) is approximated through a first-order Taylor series expansion about the expected values $\theta = (t_{yR}, N_R)$ with $\bar{y}_R = t_{yR}/N_R$:

$$AV(\hat{\bar{y}}_R) = \left(\frac{1}{N_R}\right)^2 \left[AV(\hat{t}_{yR}) + \bar{y}_R^2 \; AV(\hat{N}_R) - \; 2\bar{y}_R \; ACov(\hat{t}_{yR}, \; \hat{N}_R)\right]. \tag{5.12}$$

The approximate population sampling variance for the EC-GREG estimated total of y is defined in expression (4.16) as:

$$\begin{aligned} AV(\hat{t}_{yR}) &= AV(\hat{t}_{yGREG}) \\ &\quad + \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\,(1-\phi_{Ahik})\,E_{Ahik}^2 \\ &\quad + \mathbf{B}_A'\mathbf{V}_B\mathbf{B}_A. \end{aligned}$$

The corresponding variance components for the estimated population size, $\hat{N}_R$, is calculated using the following expansion:

$$\begin{aligned} AV(\hat{N}_R) &= E_B\left[AV_{c_A,A}(\hat{N}_R \mid B)\right] + AV_B\left[E_{c_A,A}(\hat{N}_R \mid B)\right] \\ &= E_B\left[E_{c_A}\left\{AV_A(\hat{N}_R \mid c_A, B) \mid B\right\}\right] \\ &\quad + E_B\left[AV_{c_A}\left\{E_A(\hat{N}_R \mid c_A, B) \mid B\right\}\right] \\ &\quad + AV_B\left[E_{c_A}\left\{E_A(\hat{N}_R \mid c_A, B) \mid B\right\}\right] \\ &\equiv V_1 + V_2 + V_3. \end{aligned} \tag{5.13}$$

Note that we assume complete coverage for the benchmark survey in determining the approximate population sampling variance. The asymptotic variance relies on the approximation $\hat{N}_R \cong \left\{\sum_{hik \in s_A} \pi_{hik}^{-1} E_{ANk}\right\} + \hat{\mathbf{t}}_{Bx}'\mathbf{B}_{AN}$ where $E_{ANhik} = 1 - \mathbf{x}_{hik}'\mathbf{B}_{AN}$, the assisting model residuals, and $\mathbf{B}_{AN} \cong E(\hat{\mathbf{B}}_{AN})$ from (5.9). The population sampling variance of the estimated population size under traditional calibration, $V_1 \equiv AV(\hat{N}_{GREG})$, is calculated as shown for $AV(\hat{t}_{yGREG})$ in expression (4.14) by substituting E_{ANhik} in place of $E_{Ahik} = y_{hik} - \mathbf{x}'_{hik}\mathbf{B}_A$. Therefore, the first variance component in (5.13) is

$$V_1 \cong \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \pi_{hi(1)} \left(\frac{t_{ANEhi}}{\pi_{hi(1)}} - t_{ANEh} \right)^2 + \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{i=1}^{M_h} \frac{E_{c_A}\left(V_{ANhi}\right)}{\pi_{hi(1)}}$$

where $t_{ANEhi} = \sum_{k=1}^{N_{hi}} \phi_{Ahik} E_{ANhik}$; $t_{ANEh} = \sum_{i=1}^{M_h} t_{ANEhi}$; and V_{ANhi} is the within-PSU population sampling variance associated with the estimated population size. The second component in (5.13) addresses the variation due to the analytic survey sampling frame coverage error,

$$V_2 \cong \sum_{h=1}^{H} \sum_{i=1}^{M_h} \sum_{k=1}^{N_{hi}} \phi_{Ahik} \left(1 - \phi_{Ahik}\right) E_{ANhik}^2$$

and, again, is similar to the V_2 defined in expression (4.15) for the estimator in the numerator of $\hat{\bar{y}}_R$. The V_3 variance component in (5.13) is approximated by

$$V_3 \equiv \mathbf{B}'_{AN} \mathbf{V}_B \mathbf{B}_{AN}$$

where $\mathbf{V}_B = Var_B\left(\hat{\mathbf{t}}_{Bx}\right)$, the population sampling covariance matrix associated with the vector of estimated control totals, and $\mathbf{B}_{AN}$ (5.9) is the G-length vector of assisting model coefficients specified for the analytic survey frame population. By combining the variance component approximations, the asymptotic population sampling variance of $\hat{N}_R$ is:

$$
\begin{aligned}
AV(\hat{N}_R) &= AV(\hat{N}_{GREG}) \\
&\quad + \sum_{h=1}^{H} \sum_{i=1}^{M_h} \sum_{k=1}^{N_{hi}} \phi_{Ahik} \left(1 - \phi_{Ahik}\right) E_{ANhik}^2 \\
&\quad + \mathbf{B}'_{AN} \mathbf{V}_B \mathbf{B}_{AN}. \qquad (5.14)
\end{aligned}
$$

The relative influences of the variance components for $M^{-2}AV(\hat{N}_R)$ are the same as specified for $M^{-2}AV(\hat{t}_{yR})$: $V_1 = O\left(m_A^{-1}\right)$, $V_2 = O\left(M^{-1}\right)$, and $V_3 = O\left(m_B^{-1}\right)$ where m_A and m_B are the number of PSUs selected for the analytic and benchmark surveys, and M is the total number of PSUs in the population.

The covariance term in (5.12), $ACov(\hat{t}_{yR},\ \hat{N}_R)$, is defined as follows using the residual approximations to $\hat{t}_{yR}$ and $\hat{N}_R$ discussed previously:

$$
\begin{aligned}
& ACov\left(\hat{t}_{yR}, \hat{N}_R\right) \\
&= ACov\left(\sum_{hik \in s_A} \pi_{hik}^{-1} E_{Ahik} + \hat{\mathbf{t}}'_{Bx} \mathbf{B}_A,\ \sum_{hik \in s_A} \pi_{hik}^{-1} E_{ANhik} + \hat{\mathbf{t}}'_{Bx} \mathbf{B}_{AN}\right) \\
&= ACov\left(\sum_{hik \in s_A} \pi_{hik}^{-1} E_{Ahik},\ \sum_{hik \in s_A} \pi_{hik}^{-1} E_{ANhik}\right) \\
&\quad + ACov\left(\hat{\mathbf{t}}'_{Bx} \mathbf{B}_A,\ \hat{\mathbf{t}}'_{Bx} \mathbf{B}_{AN}\right), \qquad (5.15)
\end{aligned}
$$

because the analytic and benchmark surveys are assumed to be independent, i.e., $ACov\left(\sum_{hik \in s_A} \pi_{hik}^{-1} E_{Ahik},\ \hat{\mathbf{t}}'_{Bx} \mathbf{B}_{AN}\right) \equiv 0$. To evaluate the covariance of the weighted residuals in (5.15), we use the unconditional variance formula given in, e.g., Casella

& Berger (2002, Theorem 4.4.7):

$$\begin{aligned}
&ACov\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{Ahik},\ \sum_{hik\in s_A}\pi_{hik}^{-1}E_{ANhik}\right)\\
&\quad= ACov_{c_A}\left[E_A\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{Ahik}\mid c_A\right),\ E_A\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{ANhik}\mid c_A\right)\right]\\
&\qquad+ E_{c_A}\left[ACov_A\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{Ahik},\sum_{hik\in s_A}\pi_{hik}^{-1}E_{ANhik}\mid c_A\right)\right]. \qquad (5.16)
\end{aligned}$$

The first term in (5.16) equals

$$\begin{aligned}
&ACov_{c_A}\left[E_A\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{Ahik}\mid c_A\right),\ E_A\left(\sum_{hik\in s_A}\pi_{hik}^{-1}E_{ANhik}\mid c_A\right)\right]\\
&\quad\cong ACov_{c_A}\left[\sum_{hik\in U}C_{Ahik}E_{Ahik},\ \sum_{hik\in U}C_{Ahik}E_{ANhik}\right]\\
&\quad= \sum_{hik\in U}ACov_{c_A}\left(C_{Ahik}E_{Ahik},\ C_{Ahik}E_{ANhik}\right)\\
&\qquad+\sum_{hik\neq(hik)'\in U}ACov_{c_A}\left(C_{Ahik}E_{Ahik},\ C_{A(hik)'}E_{AN(hik)'}\right)\\
&\quad= \sum_{hik\in U}ACov_{c_A}\left(C_{Ahik}E_{Ahik},\ C_{Ahik}E_{ANhik}\right)\\
&\quad= \sum_{hik\in U}AVar_{c_A}\left(C_{Ahik}\right)E_{Ahik}E_{ANhik}\\
&\quad= \sum_{hik\in U}\phi_{Ahik}\left(1-\phi_{Ahik}\right)E_{Ahik}E_{ANhik}
\end{aligned}$$

because Bernoulli random variables are independent by definition with a convergence rate of $O\left(M^{-1}\right)$. The second term in (5.16) equals $ACov\left(\hat{t}_{yGREG},\hat{N}_{GREG}\right)$ and converges at a rate of m_A^{-1}. The remaining term in $ACov\left(\hat{t}_{yR},\hat{N}_R\right)$, $ACov\left(\hat{\mathbf{t}}'_{Bx}\mathbf{B}_A,\right.$

$\hat{\mathbf{t}}'_{Bx}\mathbf{B}_{AN})$ given in expression (5.15), evaluates to $\mathbf{B}'_A\mathbf{V}_B\mathbf{B}_{AN}$. Therefore,

$$
\begin{aligned}
ACov\left(\hat{t}_{yR}, \hat{N}_R\right) &= ACov\left(\hat{t}_{yGREG}, \hat{N}_{GREG}\right) \\
&+ \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\left(1-\phi_{Ahik}\right) E_{Ahik}E_{ANhik} \\
&+ \mathbf{B}'_A\mathbf{V}_B\mathbf{B}_{AN}. \qquad (5.17)
\end{aligned}
$$

As shown for $AV(\hat{t}_{yR})$ in (4.16), $AV(\hat{\bar{y}}_R)$ can be expressed as a function of terms associated with traditional calibration plus terms related to coverage and EC calibration by rearranging the components within (4.16), (5.14), and (5.17):

$$
\begin{aligned}
AV\left(\hat{\bar{y}}_R\right) &= \left(\frac{1}{N_R}\right)^2 \left[AV(\hat{t}_{yR}) + \bar{y}_R^2\ AV(\hat{N}_R) -\ 2\bar{y}_R\ ACov(\hat{t}_{yR}, \hat{N}_R)\right] \\
&= \left(\frac{1}{N_R}\right)^2 \left[AV\left(\hat{t}_{yGREG}\right) + \bar{y}_R^2 AV(\hat{N}_{GREG}) - 2\bar{y}_R ACov\left(\hat{t}_{yGREG}, \hat{N}_{GREG}\right)\right] \\
&+ \left(\frac{1}{N_R}\right)^2 \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right)\left[E_{Ahik}^2 + \bar{y}_R^2 E_{ANhik}^2 - 2\bar{y}_R E_{Ahik}E_{ANhik}\right] \\
&+ \left(\frac{1}{N_R}\right)^2 \left[\mathbf{B}'_A\mathbf{V}_B\mathbf{B}_A + \bar{y}_R^2\mathbf{B}'_{AN}\mathbf{V}_B\mathbf{B}_{AN} - 2\bar{y}_R\mathbf{B}'_A\mathbf{V}_B\mathbf{B}_{AN}\right] \\
&= AV\left(\hat{\bar{y}}_{GREG}\right) \\
&+ \left(\frac{1}{N_R}\right)^2 \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right)\left(E_{Ahik} - \bar{y}_R E_{ANhik}\right)^2 \\
&+ \left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A - \bar{y}_R\mathbf{B}_{AN}\right)' \mathbf{V}_B \left(\mathbf{B}_A - \bar{y}_R\mathbf{B}_{AN}\right) \qquad (5.18)
\end{aligned}
$$

The asymptotic population sampling variance for the EC poststratified ratio-mean estimator $\hat{\bar{y}}_P$ is defined as follows through a specialization of expression (5.18):

$$\begin{aligned} AV\left(\hat{\bar{y}}_P\right) &= \left(\frac{1}{N_P}\right)^2 \left[AV(\hat{t}_{yP}) + \bar{y}_P^2 AV(\hat{N}_P) - 2\bar{y}_P \; ACov(\hat{t}_{yP}, \hat{N}_P)\right] \\ &= \left(\frac{1}{N_B}\right)^2 AV\left(\hat{t}_{yPSGR}\right) \\ &\quad + \left(\frac{1}{N_B}\right)^2 \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \delta_{ghik}\phi_{Ahik}\left(1-\phi_{Ahik}\right) E_{Ahik}^2 \\ &\quad + \left(\frac{1}{N_B}\right)^2 \left(\bar{\mathbf{Y}}_A - \bar{y}_P \mathbf{1}_G\right)' \mathbf{V}_B \left(\bar{\mathbf{Y}}_A - \bar{y}_P \mathbf{1}_G\right) \end{aligned} \qquad (5.19)$$

where $AV(\hat{t}_{yP})$ is defined in (4.19); $AV\left(\hat{t}_{yPSGR}\right) = \mathbf{N}_B' \; E_{c_A}\left(\mathbf{V}_A\right) \; \mathbf{N}_B$; $\mathbf{B}_A \equiv \bar{\mathbf{Y}}_A$, $\mathbf{B}_{AN} \equiv \mathbf{1}_G$, and $\hat{N}_{PSGR} \equiv N_B$ under EC poststratification; and $ACov\left(\hat{t}_{yPSGR}, N_B\right)$ $\equiv 0$ because N_B is a population parameter. Note also that $E_{ANhik} = 1 - \mathbf{x}_{hik}'\mathbf{B}_{AN}$ evaluates to zero under poststratification because $\mathbf{x}_{hik}$ is a G-length vector with one in the g^{th} position to indicate poststratum membership and zeros elsewhere, and $\mathbf{B}_{AN}$ is a G-length vector of ones, i.e., $\mathbf{1}_G$. Finally, ϕ_{Ahik} is the probability that unit hik is listed on the analytic survey sampling frame.

The approximate population sampling variance for $\hat{\bar{y}}_R$ is a function of the approximate population sampling variance for its numerator, $AV(\hat{t}_{yR})$, as shown in (5.18). To better understand the relative effects of EC calibration on $\hat{t}_R$ and $\hat{\bar{y}}_R$, we examine the difference between EC variance components in $AV(\hat{t}_{yR})$ and $AV(\hat{\bar{y}}_R)$. We multiply $AV(\hat{t}_{yR})$ by N_R^{-2} to reduce its size to a level comparable with $AV(\hat{\bar{y}}_R)$

in the following expression:

$$
\begin{aligned}
&\left(\frac{1}{N_R}\right)^2 \left[AV\left(\hat{t}_{yR}\right) - AV\left(\hat{t}_{yGREG}\right)\right] - \left[AV\left(\hat{\bar{y}}_R\right) - AV\left(\hat{\bar{y}}_{GREG}\right)\right] \\
&\quad = \left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_A + \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right) E_{Ahik}^2\right) \\
&\qquad - \left(\frac{1}{N_R}\right)^2 \Big(\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_A + \bar{y}_R^2\mathbf{B}_{AN}'\mathbf{V}_B\mathbf{B}_{AN} - 2\bar{y}_R\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_{AN} \\
&\qquad\qquad + \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right)\left(E_{Ahik} - \bar{y}_R E_{ANhik}\right)^2\Big) \\
&\quad \cong \bar{y}_R \left(\frac{1}{N_R}\right)^2 \left[2\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_{AN} - \bar{y}_R\mathbf{B}_{AN}'\mathbf{V}_B\mathbf{B}_{AN}\right], \qquad (5.20)
\end{aligned}
$$

by dropping the lower-order $O(M)$ coverage error terms. We assume y to be strictly non-negative, as in our simulation study, so that we can claim $\bar{y}_R \geq 0$. If $\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_{AN} > \frac{1}{2}\bar{y}_R\mathbf{B}_{AN}'\mathbf{V}_B\mathbf{B}_{AN}$, i.e., the covariance between the numerator and denominator variables in $\hat{\bar{y}}_R$ is larger than $\frac{1}{2}\bar{y}_R$ times the denominator variance, then the difference is positive. This implies that the variance inflation due to EC calibration will be less noticeable in the ratio of two estimated totals in comparison with the estimator of a single total. Conversely, if $\mathbf{B}_{AN}'\mathbf{V}_B\mathbf{B}_{AN} > 2\mathbf{B}_A'\mathbf{V}_B\mathbf{B}_{AN}/\bar{y}_R$, then the difference is negative and the variance inflation in $\hat{\bar{y}}_R$ estimates will exceed levels for $\hat{t}_R$.

5.4.1 Linearization Variance Estimation for Traditional Calibration

A linearization variance estimator developed for $\hat{\bar{y}}_R$ (5.1) under traditional calibration accounts only for the variation within the analytic survey sample. This variance estimator, which excludes positive variance components for the analytic survey coverage error and benchmark estimates, will underestimate the population sampling variance, $AV(\hat{\bar{y}}_R)$, given in expression (5.18). The formula for this naïve variance estimator given below is derived by substituting the sample estimates for the appropriate population parameters:

$$\begin{aligned}
var_{Naïve}(\hat{\bar{y}}_R) &\equiv var(\hat{\bar{y}}_{GREG}) \\
&= \left(\frac{1}{\hat{N}_{GREG}}\right)^2 \left[var\left(\hat{t}_{yGREG}\right) + \hat{\bar{y}}_R^2 var\left(\hat{N}_{GREG}\right)\right. \\
&\qquad \left. - \; 2\hat{\bar{y}}_R \, cov\left(\hat{t}_{yGREG}, \; \hat{N}_{GREG}\right)\right] \\
&= \left(\frac{1}{\hat{N}_{GREG}}\right)^2 \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{hi+} - \bar{\breve{u}}_{h++})^2 \qquad (5.21)
\end{aligned}$$

for $\hat{\bar{y}}_{GREG}$ defined in expression (2.12). Särndal et al. (1992) refer to $var(\hat{\bar{y}}_{GREG})$ as a g-weighted (sample) variance estimator developed for traditional weight calibration. The values

$$\breve{u}_{hi+} = \sum_{k \in s_{Ahi}} a_{hik} \pi_{hik}^{-1} \left(e_{Ahik} - \hat{\bar{y}}_R e_{ANhik}\right)$$

are linear substitutes derived from the first-order linear approximation to $\hat{\bar{y}}_{GREG}$. The linear substitutes are functions of the GREG model residuals for the numerator, $e_{Ahik} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_A$ with estimated model coefficient vector $\hat{\mathbf{B}}_A$ defined in expression (4.2), as well as the denominator, $e_{ANhik} = 1 - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{AN}$ with $\hat{\mathbf{B}}_{AN}$ defined in (5.3). The sample variance is centered around the stratum-specific means of the linear substitute estimates, $\bar{\breve{u}}_{h++} = m_{Ah}^{-1} \sum_{i \in s_{Ah}} \breve{u}_{hi+}$. We can reduce the underestimation in (5.21) by including a sample estimate of the undercoverage error:

$$\begin{aligned} var_{Naïve,c_A}(\hat{\bar{y}}_R) &= var_{Naïve}(\hat{\bar{y}}_{GREG}) \\ &+ \left(\frac{1}{\hat{N}_{GREG}}\right)^2 \sum_{h=1}^{H}\left(1-\hat{\bar{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e_{Ahik}^2}{\pi_{hik}}. \end{aligned} \tag{5.22}$$

Even with this addition, however, (5.22) will still underestimate its population parameter. The magnitude, as discussed before, is related to the precision of the benchmark survey estimates, i.e., the size of $\hat{V}_B$.

The corresponding naïve sample variance estimator for the EC poststratified ratio mean, $\hat{\bar{y}}_P$ given in (5.4), takes a similar form and also underestimates the true population sampling variance:

$$\begin{aligned} var_{Naïve}(\hat{\bar{y}}_P) &\equiv var(\hat{\bar{y}}_{PSGR}) \\ &= \left(\frac{1}{N_B}\right)^2 \left[var\left(\hat{t}_{yPSGR}\right) - 2\hat{\bar{y}}_P\ cov\left(\hat{t}_{yPSGR},\ \hat{N}_B\right)\right] \\ &= \left(\frac{1}{N_B}\right)^2 var\left(\hat{t}_{yPSGR}\right) \\ &= \left(\frac{1}{N_B}\right)^2 \hat{\mathbf{N}}_B'\hat{\mathbf{V}}_A\hat{\mathbf{N}}_B \end{aligned} \tag{5.23}$$

where $\hat{\mathbf{V}}_A$ is the sample covariance matrix for $\hat{\bar{\mathbf{Y}}}_A = \left[\hat{\bar{y}}_{A1}, ..., \hat{\bar{y}}_{AG}\right]'$ discussed in (4.24). Note that $cov\left(\hat{t}_{yPSGR},\ \hat{N}_B\right)$ in (5.23) is zero because we assume independence between the analytic and benchmark surveys. The coverage error-adjusted sample variance estimator takes the same form as above:

$$
\begin{aligned}
var_{Naïve,c_A}(\hat{\bar{y}}_P) &= var_{Naïve}(\hat{\bar{y}}_P) \\
&+ \left(\frac{1}{N_B}\right)^2 \sum_{g=1}^{G} \left(1 - \hat{\bar{\phi}}_{Ag}\right) \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{\delta_{ghik} e_{Ahik}^2}{\pi_{hik}} \qquad (5.24)
\end{aligned}
$$

where $e_{Ahik} = y_{hik} - \hat{\bar{y}}_{Ag}$ and $\hat{\bar{y}}_{Ag} = \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik} y_{hik} / \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik}$, a function of the g^{th} poststratum membership indicator δ_{ghik}. This sample variance estimator will also underestimate $AV(\hat{\bar{y}}_P)$ (5.19) due to the missing benchmark variance component. The next section contains a discussion of sample variance estimators that address all three sources of variation in the EC calibration system.

5.4.2 Estimated-Control Taylor Linearization Variance Method

Linearization variance estimators for ratios are widely used in survey research with traditional weight calibration. However, as discussed throughout Chapter 4, the application of these variance estimators to data calibrated to survey estimates can result in non-negligible levels of bias and erroneous conclusions for hypothesis tests. The same holds for EC ratio-mean estimators discussed in this chapter. Sample variance formulae for the EC calibrated estimators can take multiple forms; we present the linearization variance estimators in this section followed by the Jackknife methods first discussed in Chapter 4.

The EC Taylor series linearization sample variance of an EC-GREG ratio estimator of a population mean can be decomposed into components associated with traditional calibration, coverage error in the analytic survey sampling frame, and variation in the benchmark controls as shown below:

$$\begin{aligned}
var_{ECTS}(\hat{\bar{y}}_R) &= \left(\frac{1}{\hat{N}_R}\right)^2 \left[var_{ECTS}(\hat{t}_{yR}) + \hat{\bar{y}}_R^2\ var_{ECTS}(\hat{N}_R) - 2\ \hat{\bar{y}}_R\ cov(\hat{t}_{yR}, \hat{N}_R)\right] \\
&= \left(\frac{1}{\hat{N}_R}\right)^2 \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{hi+} - \bar{\breve{u}}_{h++})^2 \\
&\quad + \left(\frac{1}{\hat{N}_R}\right)^2 \sum_{h=1}^{H} \left(1 - \hat{\bar{\phi}}_{Ah}\right) \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} \left(e_{Ahik} - \hat{\bar{y}}_R e_{ANhik}\right)^2 \\
&\quad + \left(\frac{1}{\hat{N}_R}\right)^2 \left(\hat{\mathbf{B}}_A - \hat{\bar{y}}_R \hat{\mathbf{B}}_{AN}\right)' \hat{\mathbf{V}}_B \left(\hat{\mathbf{B}}_A - \hat{\bar{y}}_R \hat{\mathbf{B}}_{AN}\right) \qquad (5.25)
\end{aligned}$$

where the first term equals the naïve sample variance estimator, $var(\hat{\bar{y}}_{GREG})$ given in expression (5.21), under the assumed stratified, multistage design for the analytic survey with PSUs selected with replacement and is $O_P\left(m_A^{-1}\right)$. The second term is a function of the average coverage propensity within stratum h, $\hat{\bar{\phi}}_{Ah}$, and the assisting model residuals for the numerator and denominator of $\hat{\bar{y}}_R$ and is $O_P\left(M^{-1}\right)$. The third set of variance components in (5.25) is $O_P\left(m_B^{-1}\right)$ and increases the sample variance to account for the precision of the estimated control totals. This is captured in $\hat{\mathbf{V}}_B$, the benchmark control total covariance matrix. Therefore, $var_{ECTS}(\hat{\bar{y}}_R) = \max\left[O_P\left(m_A^{-1}\right), O_P\left(m_B^{-1}\right)\right]$.

The EC linearization sample variance of an EC poststratified ratio estimator is similarly defined as follows:

$$\begin{aligned}
var_{ECTS}(\hat{\bar{y}}_P) &= \left(\frac{1}{\hat{N}_P}\right)^2 \Big[var_{ECTS}(\hat{t}_{yP}) + \hat{\bar{y}}_P^2 var_{ECTS}(\hat{N}_B) \\
&\qquad - 2\,\hat{\bar{y}}_P\ cov(\hat{t}_{yP}, \hat{N}_B)\Big] \\
&= \left(\frac{1}{\hat{N}_B}\right)^2 \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{hi+} - \bar{\breve{u}}_{h++})^2 \\
&\quad + \left(\frac{1}{\hat{N}_B}\right)^2 \sum_{g=1}^{G} \left(1 - \hat{\bar{\phi}}_{Ag}\right) \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \frac{\delta_{ghik} e_{Ahik}^2}{\pi_{hik}} \\
&\quad + \left(\frac{1}{\hat{N}_B}\right)^2 \left(\hat{\bar{\mathbf{Y}}}_A - \hat{\bar{y}}_P \mathbf{1}_G\right)' \hat{\mathbf{V}}_B \left(\hat{\bar{\mathbf{Y}}}_A - \hat{\bar{y}}_P \mathbf{1}_G\right), \qquad (5.26)
\end{aligned}$$

where the first term in the final expression is equal to $var_{Naïve}(\hat{\bar{y}}_P)$ with $\breve{u}_{hi+} = \sum_{k=1}^{n_{Ahi}} (\hat{N}_{Bg}/\hat{N}_{Ag}) \pi_{hik}^{-1} \left(y_{hik} - \hat{\bar{y}}_{Ag}\right)$ and $\bar{\breve{u}}_{h++} = m_{Ah}^{-1} \sum_{i \in s_{Ah}} \breve{u}_{hi+}$. Note that $\hat{N}_P \equiv \hat{N}_B$ under EC poststratification as shown in (5.5). The sample coverage error estimator multiplied by N_B^2 is the same as specified for $var_{ECTS}(\hat{t}_{yP})$ in expression (4.24) because the residuals under the model specified by the denominator (i.e., $\hat{N}_B$) are zero. The term $var_{ECTS}(\hat{N}_B)$ is a scalar variance estimate of the estimated population size $\hat{N}_B$ calculated from the benchmark survey data.

As discussed in Chapter 4, the sample variance estimators are asymptotically unbiased only if the components are calculated using consistent estimators of their corresponding population parameter components. Having addressed linearization variance estimation for the EC ratio-mean estimator, we next examine the set of jackknife replication estimators identified for our research.

5.4.3 Fuller Two-Phase Jackknife Method

Isaki et al. (2004) describe the Fuller (1998) two-phase jackknife method in general terms for any type of smooth estimator applied to weighted sample data. This excludes, for example, quantile estimation because jackknife variance estimators are consistent only for smooth functions (Rao & Shao, 1999). In the following section, we briefly describe the Fuller two-phase jackknife method for EC calibration (ECF2) as it relates to a ratio estimator of a population mean. A more complete discussion of the method is left to Section 4.4.3 because much of the mechanics used to implement the ECF2 is the same for EC estimated totals and ratios of estimated totals. Some material is repeated here for completeness of the discussion.

The delete-one ECF2 jackknife variance estimator for the EC-GREG ratio-mean estimator is expressed as

$$var_{ECF2}(\hat{\bar{y}}_R) = \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{\bar{y}}_{R(r)} - \hat{\bar{y}}_{GREG})^2 \tag{5.27}$$

where m_{Ah} is the number of analytic survey PSUs, and $\hat{\bar{y}}_{GREG}$ (2.12) is the EC-GREG ratio-mean estimator. The r^{th} ECF2 replicate estimator of the population mean, $\ddot{\bar{y}}_{R(r)}$, is calculated using the following formula:

$$\begin{aligned} \ddot{\bar{y}}_{R(r)} &= \frac{\ddot{t}_{yR(r)}}{\ddot{N}_{R(r)}} \\ &= \frac{\hat{t}_{Ay(r)} + (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{A(r)}}{\hat{N}_{A(r)} + (\hat{\mathbf{t}}_{Bx(r)} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{AN(r)}}. \end{aligned} \tag{5.28}$$

The EC replicate estimator of the population total using the ECF2 method, $\ddot{t}_{yR(r)}$, is detailed in expression (4.26). The replicate components within the estimator for the population size, $\ddot{N}_{R(r)}$, are defined as follows: $\hat{N}_{A(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1}$, the population size estimated from the replicate-adjusted analytic survey weights; $\hat{\mathbf{B}}_{AN(r)} = \left[\sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik} \mathbf{x}'_{hik}\right]^{-1} \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik}$, the assisting model coefficient vector with $y_{hik} = 1$; and $\pi_{hi(r)}^{-1}$, the PSU-subsampling weight defined in expression (4.27). The replicate-specific vector of estimated benchmark controls, $\hat{\mathbf{t}}_{Bx(r)}$, is defined in (4.25) as

$$\begin{aligned} \hat{\mathbf{t}}_{Bx(r)} &= \hat{\mathbf{t}}_{Bx} + c_h \hat{\mathbf{z}}_{B(r)} \\ &= \hat{\mathbf{t}}_{Bx} + c_h \delta_{(r)} \sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}_{Bg} \end{aligned} \qquad (5.29)$$

for $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; $\hat{\mathbf{z}}_{Bg}$, a component of the spectral decomposition of $\hat{\mathbf{V}}_B$ such that $\hat{\mathbf{V}}_B = \sum_{g=1}^{G} \hat{\mathbf{z}}_{Bg} \hat{\mathbf{z}}'_{Bg}$; and the indicator functions $\delta_{(r)}$ and $\delta_{g|(r)}$ to identify the G replicates chosen from a total of m_A ($\sum_h m_{Ah}$) for an adjustment, and the $\hat{\mathbf{z}}_{Bg}$ used in the adjustment, respectively. Therefore, the replicate estimates may be specified as follows by substituting (5.29) into (5.28):

$$\ddot{y}_{R(r)} = \frac{\hat{t}_{yGREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)}}{\hat{N}_{GREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{AN(r)}}. \qquad (5.30)$$

The ECF2 variance estimator for the EC poststratified ratio-mean estimator is defined by specializing the EC-GREG terms in (5.27) to the EC-PSGR setting:

$$var_{ECF2}(\hat{\ddot{y}}_P) = \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{y}_{P(r)} - \hat{\ddot{y}}_P)^2 \qquad (5.31)$$

for $\hat{\bar{y}}_P$ expressed in (5.4). The ECF2 poststratified replicate estimator is defined as

$$\ddot{\bar{y}}_{P(r)} = \frac{\ddot{t}_{yP(r)}}{\ddot{N}_{P(r)}} \tag{5.32}$$

The numerator term is defined in matrix form as:

$$\begin{aligned} \ddot{t}_{yP(r)} &= \hat{\mathbf{N}}'_{B(r)}\hat{\mathbf{N}}^{-1}_{A(r)}\hat{\mathbf{t}}_{Ay(r)} \\ &= \mathbf{N}'_B\hat{\mathbf{N}}^{-1}_{A(r)}\hat{\mathbf{t}}_{Ay(r)} + c_h\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{N}}^{-1}_{A(r)}\hat{\mathbf{t}}_{Ay(r)} \\ &= \hat{t}_{yPSGR(r)} + c_h\hat{\mathbf{z}}'_{B(r)}\hat{\bar{\mathbf{Y}}}_{A(r)} \end{aligned} \tag{5.33}$$

where $\hat{\mathbf{N}}_{B(r)} = \mathbf{N}_B + c_h\hat{\mathbf{z}}'_{B(r)}$; $\hat{\mathbf{N}}_{A(r)}$ is a diagonal matrix of dimension G with components $\hat{N}_{Ag(r)} = \sum_{hik\in s_A}\delta_{ghik}\pi^{-1}_{hi(r)}\pi^{-1}_{hik}$; $\hat{\mathbf{t}}_{Ay(r)} = \left[\hat{t}_{Ay1(r)},\ ...,\ \hat{t}_{AyG(r)}\right]'$ with components $\hat{t}_{Ayg(r)} = \sum_{hik\in s_A}\delta_{ghik}\pi^{-1}_{hi(r)}\pi^{-1}_{hik}y_{hik}$; and $\hat{\bar{\mathbf{Y}}}_{A(r)} = \hat{\mathbf{N}}^{-1}_{A(r)}\hat{\mathbf{t}}_{Ay(r)}$. The denominator is similarly expressed as

$$\begin{aligned} \ddot{N}_{P(r)} &= \hat{\mathbf{N}}'_{B(r)}\mathbf{1}_G \\ &= \mathbf{N}'_B\mathbf{1}_G + c_h\hat{\mathbf{z}}'_{B(r)}\mathbf{1}_G \\ &= N_B + c_h\hat{\mathbf{z}}'_{B(r)}\mathbf{1}_G. \end{aligned} \tag{5.34}$$

where $\hat{\mathbf{B}}_{AN(r)}$ is specified for $\ddot{\bar{y}}_{R(r)}$ (5.30), the general form under the EC-GREG, is equivalent to $\mathbf{1}_G$ with EC-PSGR as discussed for $\hat{N}_P$ (5.5).

The components in the ECF2 variance estimator (5.27) reproduces in expectation the corresponding population sampling variance components listed in $AV(\hat{\bar{y}}_R)$

(5.18) using the rationale discussed for $\hat{t}_{yR}$ in Section 4.4.3. The sample variance estimator in its current form fails to capture the undercoverage error variance. The amount of negative bias is related to the residuals from the assisting model and the coverage rates are close to one. As in Section 4.4.3, we suggest the following modification to the original ECF2 method to account for this missing variance component. Define the modified ECF2 (ECF2m) replicate estimates as follows:

$$\dddot{y}_{R(r)} = \frac{\dddot{t}_{yP(r)}}{\dddot{N}_{P(r)}}$$
$$= \frac{\ddot{t}_{yP(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}}}{\ddot{N}_{P(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}}$$
$$= \frac{\hat{N}^{-1}_{GREG(r)} \left(\hat{t}_{yGREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}} \right)}{1 + \hat{N}^{-1}_{GREG(r)} \left(c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{AN(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}} \right)} \quad (5.35)$$

where $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; $R_h = \sqrt{1/Hm_{Ah}}$; $\hat{\phi}_{A(r)}$ is an estimate of the analytic survey coverage rate (error) using a combination of data from the complete benchmark survey and analytic survey replicate subsample; $\hat{t}_{Ae2(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \pi^{-1}_{hik} \times e^2_{Ahik(r)}$ with $\hat{\mathbf{B}}_{A(r)}$ defined in expression (4.26) and $e_{Ahik(r)} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{A(r)}$; $\hat{t}_{ANe2(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \pi^{-1}_{hik} e^2_{ANhik(r)}$ with $e_{ANhik(r)} = 1 - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{AN(r)}$; and $\hat{\mathbf{B}}_{AN(r)}$ defined for $\ddot{y}_{R(r)}$ (5.28). Note that $\ddot{t}_{yP(r)}$ is also used in the ECF2 modification for $\hat{t}_{yR}$ (4.31). The modified delete-one ECF2 jackknife sample variance estimator is then specified as

$$var_{ECF2m}(\hat{\bar{y}}_R) = \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\dddot{y}_{R(r)} - \hat{\bar{y}}_{GREG})^2. \quad (5.36)$$

The expectation of (5.36) is evaluated by first approximating the denominator of $\dddot{y}_{R(r)}$ (5.35) with a geometric series:

$$
\begin{aligned}
&\left[1+\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right)\right]^{-1}\\
&\quad=1-\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right)\\
&\qquad+\left[\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right)\right]^2+\ldots\\
&\quad\cong 1-\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right) \qquad (5.37)
\end{aligned}
$$

where the term $\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right)$ has order in probability $m_B^{-1/2}$. The approximation is justified by Lehmann (1999, Thm 2.1.3) for convergence of a function of two random variables; $\hat{N}_{GREG(r)}^{-1}$ is small given our assumption that the estimated population size is large so that the claim $\hat{N}_{GREG(r)}^{-1}c_h\left(\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right)\to 0$ as $M\to\infty$ is reasonable; and by using the conditions of Rao & Wu (1985) for convergence of replicate estimates to the population parameter. Therefore, we approximate the modified ECF2 replicate estimates minus the full-sample estimate as follows:

$$
\begin{aligned}
&\dddot{y}_{R(r)}-\hat{\bar{y}}_{GREG}\\
&\quad\cong\left(\hat{\bar{y}}_{GREG(r)}+\frac{1}{\hat{N}_{GREG(r)}}c_h\left\{\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{A(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}}\right\}\right)\\
&\qquad\times\left(1-\frac{1}{\hat{N}_{GREG(r)}}c_h\left\{\hat{\mathbf{z}}'_{B(r)}\hat{\mathbf{B}}_{AN(r)}+R_h\eta_{(r)}\sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right\}\right).\\
&\qquad-\ \hat{\bar{y}}_{GREG}.
\end{aligned}
$$

Continuing, we have

$$\begin{aligned}
&\dddot{y}_{R(r)} - \hat{\hat{y}}_{GREG} \\
&\quad\cong \hat{\hat{y}}_{GREG(r)} - \hat{\hat{y}}_{GREG} \\
&\qquad + \frac{1}{\hat{N}_{GREG(r)}} c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + \frac{1}{\hat{N}_{GREG(r)}} c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}} \\
&\qquad - \frac{\hat{\hat{y}}_{GREG(r)}}{\hat{N}_{GREG(r)}} c_h \left[\hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{AN(r)} + R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right] \\
&\quad\cong \hat{\hat{y}}_{GREG(r)} - \hat{\hat{y}}_{GREG} \\
&\qquad + \frac{1}{N_R} c_h \hat{\mathbf{z}}'_{B(r)} \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right) \\
&\qquad + \frac{1}{N_R} c_h R_h \eta_{(r)} \left[\sqrt{(1-\hat{\phi}_{A(r)})} \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right)\right] \qquad (5.38)
\end{aligned}$$

where (5.37) is used in the first approximation; by assuming $\hat{\hat{y}}_{GREG(r)} - \hat{\hat{y}}_{GREG} = O_P\left(m_A^{-1/2}\right)$, $\hat{N}_{GREG(r)} - \hat{N}_{GREG} = O_P\left(M/\sqrt{m_A}\right)$, $\left(\hat{t}_{Ae2(r)},\ \hat{t}_{ANe2(r)}\right) = O_P(M)$, and $\hat{\mathbf{z}}_{B(r)} = O_P\left(M/\sqrt{m_B}\right)$ for the second approximation; and finally, by assuming $\left(\hat{\hat{y}}_{GREG(r)}, \hat{\mathbf{B}}_{A(r)}, \hat{\mathbf{B}}_{AN(r)}\right) = (\bar{y}_R, \mathbf{B}_A, \mathbf{B}_{AN}) + O_P\left(m_A^{-1/2}\right)$ and $\hat{N}_{GREG(r)} = N_R + O_P\left(M/\sqrt{m_A}\right)$ to allow the substitution of the population parameters in the third approximation as shown for the estimated total in expression (4.28). Using the approximation (5.38), the ECF2m jackknife variance estimator is calculated as

$$
\begin{aligned}
var_{ECF2m}(\hat{\bar{y}}_R) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{y}_{R(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
&\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
&+ \left(\frac{1}{N_R}\right)^2 \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\phi}_{A(r)}\right) \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right)^2 \\
&+ \left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)' \hat{\mathbf{V}}_B \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right) \\
&+ 2 \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right) \left(\text{other components in } \dddot{y}_{R(r)}\right) \\
&+ 2 \frac{1}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h \eta_{(r)} \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_A \sqrt{1 - \hat{\phi}_{A(r)}} \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right) \\
&- 2 \frac{\bar{y}_R}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h \eta_{(r)} \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_{AN} \sqrt{1 - \hat{\phi}_{A(r)}} \\
&\qquad \times \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right). \qquad (5.39)
\end{aligned}
$$

The first term is the traditional calibration variance components with order in probability m_A^{-1}, while the second term addresses the variation associated with the coverage error in the analytic survey sampling frame and is $O_P\left(M^{-1}\right)$. The third component, a function of $\hat{\mathbf{V}}_B$, accounts for the variation in the benchmark controls totals and is $O_P\left(m_B^{-1}\right)$. The fourth set of terms has expectation zero under the conditions of Rao & Wu (1985) to say that $\max\left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right)$ converges in probability to zero. The remaining components are $O_P\left(M^{-1/2} m_B^{-1/2}\right)$ under the assumption that $M^{-1}\hat{\mathbf{z}}_{B(r)} = O_P\left(m_B^{-1/2}\right)$ and M^{-1} times the residuals sums (e.g., $\hat{t}_{Ae2(r)}$) is $O_P\left(M^{-1/2}\right)$. The terms all have expectation zero because of the inclusion of the standard normal random variables $\eta_{(r)}$. The first three terms approximate their associated population sampling variance components only if the sample estimates are at least approximately unbiased — see, for example, the discussion of the coverage error bias for expression (4.22).

By substituting the EC poststratified terms into the expression (5.39), we have

$$var_{ECF2m}(\hat{\bar{y}}_P) = \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{y}_{P(r)} - \hat{\bar{y}}_{PSGR}\right)^2 \tag{5.40}$$

where $c_h^{-2} = (m_{Ah} - 1)/m_{Ah}$, and $\hat{\bar{y}}_{PSGR} = \hat{t}_{yPSGR}/N_B$ defined in (2.15) with $\hat{t}_{yPSGR} = \sum_{g=1}^{G} N_{Bg}\left(\hat{t}_{Ayg}/\hat{N}_{Ag}\right)$. The modified ECF2 poststratified replicate estimates are defined as follows:

$$\begin{aligned} \dddot{y}_{P(r)} &= \frac{\dddot{t}_{yP(r)}}{\dddot{N}_{P(r)}} \\ &= \frac{\hat{t}_{yPSGR(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\bar{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}}{N_B + c_h \hat{\mathbf{z}}'_{B(r)} \mathbf{1}_G} \end{aligned} \tag{5.41}$$

where $\hat{\bar{\mathbf{Y}}}_{A(r)} = \hat{\mathbf{N}}_{A(r)}^{-1} \hat{\mathbf{t}}_{Ay(r)}$; $\hat{\mathbf{t}}_{Ay(r)} = \left[\hat{t}_{Ay1(r)}, ..., \hat{t}_{AyG(r)}\right]'$ with elements that are a function of a zero/one indicator δ_{ghik} that signifies membership in the g^{th} poststratum, i.e., $\hat{t}_{Ayg(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \pi_{hik}^{-1} y_{hik}$; $\hat{\mathbf{N}}_{A(r)}$ is a diagonal matrix with elements $\left(\hat{N}_{A1(r)}, ..., \hat{N}_{AG(r)}\right)$ such that $\hat{N}_{Ag(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \pi_{hik}^{-1}$; $\mathbf{1}_G$ is a G-length vector of ones and is equivalent to the assisting model coefficient vector for $y_{hik} = 1$ in the denominator; $\hat{\mathbf{t}}_{Ae2(r)} = \left[\hat{t}_{A1e2(r)}, \ ..., \hat{t}_{AGe2(r)}\right]'$ where $\hat{t}_{Age2(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \delta_{ghik} \pi_{hik}^{-1} e_{Ahik(r)}^2$ with $e_{Ahik(r)} = y_{hik} - \bar{y}_{Ag(r)}$, and $\bar{y}_{Ag(r)} = \hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}$;

and $\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' = \left[\left(1 - \hat{\hat{\phi}}_{A1(r)}\right), \ ..., \ \left(1 - \hat{\hat{\phi}}_{AG(r)}\right)\right]$, a G-length vector of estimated coverage rate by poststratum for replicate r also shown in expression (4.34). Note that under EC poststratification, the assisting model residual defined by the denominator variable $e_{ANhik(r)} = y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{AN} = 0$ because $y_{hik} = 1$ for all units in the sample, $\hat{\mathbf{B}}_{AN} = \mathbf{1}_G$ as shown in (5.5), and $\mathbf{x}'_{hik}$ is a G-length vector of ones and zeroes to indicate membership in a particular poststratum. This is an intuitive finding given that we have shown $\hat{N}_{PSGR}$ reduces to N_B (2.15), a value independent of the analytic survey. The asymptotic property of $var_{ECF2m}(\hat{\hat{y}}_P)$ is the same as discussed for $var_{ECF2m}(\hat{\hat{y}}_R)$ (5.39) and is not repeated.

The eight steps used to calculate $var_{ECF2m}(\hat{\hat{y}}_P)$ in our simulation study detailed in Section 5.5 are provided below. The total number of replicates generated for a simulation sample is equal to the number of sample PSUs, i.e., $m_A = \sum_h m_{Ah}$.

1. Calculate the full-sample estimate $\hat{\hat{y}}_P$ using expression (5.4).

2. Determine the G eigenvalues $\hat{\lambda}_g$ and G-length eigenvectors $\hat{\mathbf{q}}_g$ from the spectral decomposition of $\hat{\mathbf{V}}_B$, and calculate the G replicate adjustments of the form $\hat{\mathbf{z}}_{Bg} = \hat{\mathbf{q}}_g\sqrt{\hat{\lambda}_g}$. Concatenate the $G \times G$ matrix of $\hat{\mathbf{z}}_{Bg}$'s, where $\hat{\mathbf{z}}_{Bg}$ represents the columns of this matrix, with a $G \times (m_A - G)$ matrix of zeros. Randomly sort the columns. Call this new $G \times m_A$ matrix $\mathbf{Z}$.

3. Create a $G \times m_A$ matrix, called $\mathbf{C}$, with column elements all equal to $c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$. The m_A-length vector of jackknife stratum weights is calculated as $\mathbf{W}_{m_A} = (m_{Ah} - 1)/m_{Ah}$.

4. Calculate the Hadamard (or element-wise) product of $\mathbf{Z}$ and $\mathbf{C}$ denoted as $\mathbf{Z} \bullet \mathbf{C}$ (Searle, 1982, pp. 49). Replicate the vector of poststratum counts estimated from the benchmark survey ($\hat{\mathbf{N}}_B$) into the columns of a $G \times m_A$ matrix and add to $\mathbf{Z} \bullet \mathbf{C}$. This new $G \times m_A$ matrix, called $\mathbf{N}_{Bm_A}$, contains the replicate benchmark controls for all m_A replicates — see the definition of $\hat{\mathbf{N}}_{B(r)}$ defined for expression (5.33).

5. Calculate the replicate estimates $\hat{\hat{\mathbf{Y}}}_{A(r)}$ with elements $\hat{\hat{y}}_{Ag(r)} = \hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}$ by removing in-turn one PSU from the analytic survey sample file, applying the PSU-subsampling weights (4.27), and summing the weighted values for the numerator and denominator within poststratum g. Call the resulting $G \times m_A$ matrix $\mathbf{B}_{m_A}$.

6. Create the following $G \times m_A$ matrices for the coverage error variance component: $\mathbf{R}_{m_A}$, with column elements all equal to $\sqrt{1/Hm_{Ah}}$; $\boldsymbol{\eta}_{m_A}$, with elements obtained from the standard normal distribution; $\boldsymbol{\phi}_{m_A}$, with column elements equal to $(1 - \hat{N}_{Ag(r)}/\hat{N}_{Bg})$ for $(\hat{N}_{Ag(r)}/\hat{N}_{Bg}) \leq 1$ and zero otherwise; and $\mathbf{e}_{m_A}$ with column elements described above for $\hat{\mathbf{t}}_{Ae2(r)}$. Calculate the Hadamard product of these matrices and call it $\mathbf{E}$.

7. Calculate the m_A replicate estimates, $\dddot{t}_{yP(r)}$ (5.41), by first multiplying the elements $\mathbf{N}_{Bm_A}$ by $\mathbf{B}_{m_A}$, adding $\mathbf{E}$ to the resulting matrix, and summing within the columns of the resulting matrix. Calculate the corresponding denominator estimates, $\dddot{N}_{P(r)}$ (5.41), by summing within the columns of the $G \times m_A$ matrix $\mathbf{N}_{Bm_A}$. Divide the m_A numerator estimates by the m_A denominator estimates

to create the replicate estimates $\dddot{y}_{P(r)}$ (5.41).

8. Finally, subtract $\hat{\ddot{y}}_P$ from $\dddot{y}_{P(r)}$ and square the m_A terms, multiply by $\mathbf{W}_{m_A}$, and sum across the m_A estimates. The resulting value is the estimated variance using the ECF2 method, $var_{ECF2m}(\hat{\ddot{y}}_P)$ given in expression (5.40).

By excluding the sixth step given above, we are also able to calculate the variance under the original ECF2 specification which does not inflate for the analytic survey coverage error. A comparison of the two variance estimators will suggest the level of underestimation associated with the exclusion of the error variance component.

5.4.4 Multivariate Normal Jackknife Method

The multivariate normal approach (ECMV) to EC calibration perturbs all of the m_A $(= \sum_h m_{Ah})$ replicate estimates using an adjustment to the benchmark controls detailed initially in expression (4.37):

$$\hat{\mathbf{t}}_{Bx(r)} = \hat{\mathbf{t}}_{Bx} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \tag{5.42}$$

where $\hat{\boldsymbol{\varepsilon}}_{B(r)}$ is a G-length vector of random variables from a multivariate normal distribution such that $\hat{\boldsymbol{\varepsilon}}_{B(r)} \overset{\text{iid}}{\sim} \text{MVN}_G(\mathbf{0}_G, \hat{\mathbf{V}}_B)$; $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$, as with the ECF2; and $R_h = \sqrt{1/Hm_{Ah}}$, a function of the number of analytic survey strata (H) and the number of samples PSUs within stratum h (m_{Ah}). We additionally incorporate an adjustment to the replicate estimates to account for the analytic survey frame coverage error. This adjustment is the same as specified for the EC-

GREG of a population total in expression (4.30). Following the convention used for the modified ECF2 method (Section 5.4.3), the adjustments are applied to the numerator and denominator of the ratio-mean replicate estimator:

$$\dddot{y}_{R(r)} = \frac{\hat{t}_{yGREG(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}}}{\hat{N}_{GREG(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{AN(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}}. \tag{5.43}$$

The ECMV replicate estimates (5.43) are used in the jackknife sample variance estimator given in expression (5.27). To evaluate the expectation of this variance estimator, we first approximate the replicate ratio-mean estimator using a geometric series as shown in detail for the modified ECF2 (5.38):

$$\begin{aligned}
&\dddot{y}_{R(r)} - \hat{\bar{y}}_{GREG} \\
&\quad \cong \left(\hat{\bar{y}}_{GREG(r)} + \frac{1}{\hat{N}_{GREG(r)}} c_h R_h \left\{\hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{A(r)} + R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}}\right\}\right) \\
&\qquad \times \left(1 - \frac{1}{\hat{N}_{GREG(r)}} c_h R_h \left\{\hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{AN(r)} + R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}\right\}\right) \\
&\qquad - \hat{\bar{y}}_{GREG} \\
&\quad \cong \hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG} \\
&\qquad + \frac{1}{N_R} c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A + \frac{1}{N_R} c_h R_h^2 \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Ae2(r)}} \\
&\qquad - \frac{\bar{y}_R}{N_R} c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_{AN} - \frac{\bar{y}_R}{N_R} c_h R_h^2 \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANe2(r)}}.
\end{aligned} \tag{5.44}$$

Note that the rate of convergence is the same as with the ECF2 method because the $\hat{\mathbf{z}}_{B(r)}$ and $\hat{\boldsymbol{\varepsilon}}_{B(r)}$ adjustments have the same orders in probability by construction. The lower-order terms involving $\hat{N}_{R(r)}^{-2}$ are again eliminated from the approximation.

The approximation for the EC poststratified ratio-mean estimator again follows the pattern established in Section 5.4.3:

$$
\begin{aligned}
\dddot{y}_{P(r)} &= \frac{\hat{t}_{yPSGR(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\hat{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}}{N_B + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{1}_G} \\
&\cong \left(\hat{\hat{y}}_{PSGR(r)} + \frac{c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\hat{\mathbf{Y}}}_{A(r)}}{N_B} + \frac{c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}}{N_B} \right) \\
&\quad \times \left(1 - \frac{c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{1}_G}{N_B}\right). \qquad (5.45)
\end{aligned}
$$

As discussed in Section 4.4.4, the ECMV jackknife variance estimator has the same asymptotic properties as the ECF2 for estimated totals. This asymptotic equivalence holds for the ratio of two calibrated totals discussed in this chapter. The expectation of $var_{ECMV}\left(\hat{\bar{y}}_R\right)$ is evaluated with respect to the analytic and benchmark survey sample designs, the sampling frame coverage mechanisms, and the multivariate normal random variable resulting in

$$
\begin{aligned}
var_{ECMV}(\hat{\bar{y}}_R) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{y}_{R(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
\cong & \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
& + \left(\frac{1}{N_R}\right)^2 \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\phi}_{A(r)}\right) \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right)^2 \\
& + \left(\frac{1}{N_R}\right)^2 (\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN})' \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)}\right] (\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}) \\
& + \ 2 \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right) \left(\text{other components in } \dddot{y}_{R(r)}\right) \\
& + \ 2 \frac{1}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_A \sqrt{1 - \hat{\phi}_{A(r)}} \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right) \\
& - \ 2 \frac{\bar{y}_R}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_{AN} \sqrt{1 - \hat{\phi}_{A(r)}} \\
& \qquad\qquad \times \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right). \qquad (5.46)
\end{aligned}
$$

The first and second terms address the variation within the analytic survey sample and coverage error for the associated sampling frame, and are $O_P\left(m_A^{-1}\right)$ and $O_P\left(M^{-1}\right)$, respectively. As shown in expression (4.40), the expectation of the third component is taken with respect to the specified MVN distribution, $\text{MVN}_G(\mathbf{0}, \hat{\mathbf{V}}_B)$, and evaluates to $\left(\frac{1}{N_R}\right)^2 (\mathbf{B}_A - \bar{y}_R\mathbf{B}_{AN})' \mathbf{V}_B (\mathbf{B}_A - \bar{y}_R\mathbf{B}_{AN})$ with $O_P\left(m_B^{-1}\right)$ provided that $E\left(\hat{\mathbf{V}}_B\right) = \mathbf{V}_B$. The fourth term has expectation zero under the conditions of Rao & Wu (1985) to say that $\max\left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right)$ converges in probability to zero with order $(Mm)_B^{-1/2}$. The last two components are $O_P\left(m_B^{-1/2}\right)$ under the assumption that $M^{-1}\hat{\boldsymbol{\varepsilon}}_{B(r)} = O_P\left(m_B^{-1/2}\right)$ and M^{-1} times the residuals sums (e.g.,

$\hat{t}_{Ae2(r)}$) is $O_P(1)$. The terms all have expectation zero because of the inclusion of the standard normal random variables $\eta_{(r)}$. Therefore, the ECMV variance estimator is asymptotically equivalent to the modified ECF2 variance estimator for a ratio-mean estimator, as well as with the estimated totals as discussed in Section 4.4.4. However, the use of the MVN distribution should again produce variance estimates with more variability than those calculated for the ECF2m. This is examined in the simulation study (Section 5.5).

5.4.5 Nadimpalli-Judkins-Chu Jackknife Method

The approach discussed in Nadimpalli et al. (2004) assumes that only the diagonal of the complete benchmark covariance matrix is available (or necessary) for EC calibration unlike the other jackknife methods discussed previously. As with the other jackknife methods, the ECNJC method also adjusts the replicate ratio estimates for variation in the benchmark estimates. The following is the ECNJC adjustment repeated from Section 4.4.5 and is used for all smooth point estimators examined in our research:

$$\hat{\mathbf{t}}_{Bx(r)} = \hat{\mathbf{t}}_{Bx} + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)}$$

where $\hat{\mathbf{S}}_B = diag(\sqrt{\hat{\mathbf{V}}_B})$, the diagonal matrix of estimated benchmark control standard errors; and $\boldsymbol{\eta}_{(r)} \sim N(0,1)$, a G-length vector of values generated from the standard normal distribution. The remaining terms are the same as defined for $\hat{\mathbf{t}}_{Bx}$ under the ECMV method in expression (5.42). The original ECNJC method does not account for the coverage error variance in the analytic sampling frame. We augment their formulation with a term that accounts for the additional variance component and include a modified ECNJC (ECNJCm) replicate ratio-mean estimator in the jackknife variance formula — see (4.30) for an EC-GREG estimator and (4.34) for an EC-PSGR estimator. The following are the EC-GREG and EC-PSGR replicate estimators, respectively, for the modified ECNJC jackknife:

$$\dddot{y}_{R(r)} = \frac{\hat{t}_{yGREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{(1 - \hat{\phi}_{A(r)}) \hat{t}_{Ae2(r)}}}{\hat{N}_{GREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{AN(r)} + c_h R_h \eta_{(r)} \sqrt{(1 - \hat{\phi}_{A(r)}) \hat{t}_{ANe2(r)}}} \tag{5.47}$$

and

$$\dddot{y}_{P(r)} = \frac{\hat{t}_{yPSGR(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\bar{\mathbf{Y}}}_{A(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Ae2(r)}}}{N_B + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{1}_G}. \tag{5.48}$$

The expectation of the jackknife variance estimator for the ECNJCm ratio-mean estimator is evaluated as with the other methods after making a geometric (series) approximation to the denominator of $\dddot{y}_{R(r)}$. The approximation is calculated by substituting $\boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B$ for $\hat{\boldsymbol{\varepsilon}}'_{B(r)}$ in expression (5.45) and is used in the following expansion of the sample variance estimator:

$$
\begin{aligned}
var_{ECNJCm}(\hat{\bar{y}}_R) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{\bar{y}}_{R(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
\cong & \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right)^2 \\
& + \left(\frac{1}{N_R}\right)^2 \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\phi}_{A(r)}\right) \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right)^2 \\
& + \left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)' \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B\right] \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right) \\
& + 2 \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{\bar{y}}_{GREG(r)} - \hat{\bar{y}}_{GREG}\right) \left(\text{other components in } \dddot{\bar{y}}_{R(r)}\right) \\
& + 2 \frac{1}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_A \sqrt{1 - \hat{\phi}_{A(r)}} \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right) \\
& - 2 \frac{\bar{y}_R}{N_R^2} \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \mathbf{B}_{AN} \sqrt{1 - \hat{\phi}_{A(r)}} \\
& \qquad \times \left(\sqrt{\hat{t}_{Ae2(r)}} - \bar{y}_R \sqrt{\hat{t}_{ANe2(r)}}\right). \qquad (5.49)
\end{aligned}
$$

All of the variance components in (5.49), except for the third term, follow the same arguments given for $var_{ECF2m}(\hat{\bar{y}}_R)$ in (5.39) and $var_{ECMV}(\hat{\bar{y}}_R)$ in (5.46). Using the work shown in expression (4.48), the expectation of the third variance component equals $\left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)' E_B\left(\hat{\mathbf{S}}_B^2\right) \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)$. If the true population covariance matrix for the benchmark control totals, $\mathbf{V}_B$, is not diagonal, then this component is not asymptotically equivalent to $\left(\frac{1}{N_R}\right)^2 \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)' \mathbf{V}_B \left(\mathbf{B}_A - \bar{y}_R \mathbf{B}_{AN}\right)$. The magnitude of the under or overestimation is related to the sign of the off-diagonal values in $\mathbf{V}_B$.

5.5 Simulation Study

We use the simulation study described in detail in Section 4.5 to compare the empirical properties of the variance estimators for the ratio of two EC-PSGR totals $\hat{\bar{y}}_P$ given in expression (5.4). The following abbreviations are used as labels for the variance estimators:

- *Naïve*, the traditional calibration estimator defined in (5.23);
- *ECTS*, the EC linearization estimator defined in (5.26);
- *ECF2*, the traditional Fuller two-phase jackknife estimator defined in (5.31);
- *ECF2m*, the modified Fuller two-phase jackknife estimator (5.40) that includes an adjustment for analytic frame undercoverage;
- *ECMV*, the Multivariate normal jackknife estimator defined in by substituting $\dddot{\bar{y}}_{P(r)}$ defined in expression (5.45) in the jackknife variance formula (5.40);
- *ECNJC*, the traditional Nadimpalli-Judkins-Chu jackknife estimator which uses the replicate estimator $\dddot{\bar{y}}_{P(r)}$ (5.48) without a coverage error term; and,
- *ECNJCm*, the modified Nadimpalli-Judkins-Chu jackknife estimator defined in (5.48).

We additionally compare these results with those presented for the estimated total in the previous chapter.

5.5.1 Simulation Parameters

We summarize the necessary aspects of the simulation study here for clarity and leave the details to Section 4.5. Samples are selected from the 2003 National Health Interview Survey (NHIS), the analytic survey sampling frame, in two stages: (*i*) $m_{Ah} = 2$ PSUs are selected with replacement from each of 25 design strata with probabilities proportional to the number of U.S. residents within each PSU; and (*ii*) either 20 or 40 residents are randomly selected without replacement (SRS) from each sampled PSU resulting in a total analytic survey sample size (n_A) of 1,000 and 2,000, respectively. We again select 4,000 simulation samples from a randomly generated frame after introducing the undercoverage rates shown in Table 4.1. We calculate the estimated population means and associated variances for two binary NHIS variables in separate runs of the simulation program: NOTCOV=1 indicates that an adult *did not* have health insurance coverage in the 12 months prior to the NHIS interview ($\bar{y} \cong 0.17$); and PDMED12M=1 indicates that an adult *delayed* medical care because of cost in the 12 months prior to the interview ($\bar{y} \cong 0.07$). Inclusion of nonresponse in the simulation study is reserved for future work. Four benchmark covariance matrices are used to produce separate EC calibration estimates to reflect varying levels of precision in the control totals. The estimated matrices reflect benchmark surveys with approximate effective sample sizes of 21,700, 6,000, 1,200, and less than 500, respectively. Key R® (Lumley, 2005; R Development Core Team, 2005) programs developed specifically for this simulation study are provided in Appendix A.

5.5.2 Evaluation Criteria

The empirical results for the variance estimators listed at the beginning of this section (Section 5.5) are compared using five measures across the j $(j = 1, ..., 4000)$ simulation samples and two outcome variables (*NOTCOV* and *PDMED12M*). The measures include:

1. $100 \times \left[\left(\frac{1}{4000}\sum_j var(\hat{\bar{y}}_{P_j}) - MSE\right)/MSE\right]$, the estimated percent bias of the variance estimator relative to the empirical $MSE = \frac{1}{4000}\sum_j(\hat{\bar{y}}_{P_j} - \bar{y})^2$;

2. $100 \times \left[\left(\frac{1}{4000}\sum_j var(\hat{\bar{y}}_{P_j}) - VAR\right)/VAR\right]$, the estimated percent bias of the variance estimator relative to the empirical variance where $VAR = \frac{1}{4000}\sum_j \left(\hat{\bar{y}}_{P_j} - \frac{1}{4000}\sum_j \hat{\bar{y}}_{P_j}\right)^2$;

3. $\frac{1}{4000}\sum_j I\left(|\hat{z}_j| \leq z_{1-\alpha/2}\right)$, the 95 percent confidence interval coverage rate where $\alpha = 0.05$, $\hat{z}_j = (\hat{\bar{y}}_{P_j} - \bar{y})/se(\hat{\bar{y}}_{P_j})$, and $se(\hat{\bar{y}}_{P_j}) = \sqrt{var(\hat{\bar{y}}_{P_j})}$;

4. $\sqrt{\frac{1}{(4000-1)}\sum_j \left[se(\hat{\bar{y}}_{P_j}) - \frac{1}{4000}\sum_j se(\hat{\bar{y}}_{P_j})\right]^2}$, the standard deviation of the estimated standard errors (se); and,

5. $100 \times \left[\left(\frac{1}{4000}\sum_j se_*\left(\hat{\bar{y}}_{P_j}\right) - \frac{1}{4000}\sum_j se_{ECTS}\left(\hat{\bar{y}}_{P_j}\right)\right)/\frac{1}{4000}\sum_j se_{ECTS}\left(\hat{\bar{y}}_{P_j}\right)\right]$, the percent increase in the variation of the estimated standard errors for all studied estimators (se_*) relative to the ECTS variance estimator (se_{ECTS}).

These criteria are also used to compare the results for estimated totals and ratio means. Prior to comparing the variance estimators, we evaluate the relative bias of the estimated totals, $\frac{1}{4000}\sum_j\left(\hat{\bar{y}}_{P_j} - \bar{y}\right)/\bar{y}$ discussed in the next section.

Table 5.1: Percent Relative Bias Averaged Across Samples and Benchmark Covariance Matrices for Percents and Totals of Outcome Variables by Point Estimator

Estimator	Not Covered by Health Insurance (NOTCOV) $n_{Ahi} = 20$	$n_{Ahi} = 40$	Delayed Medical Care (PDMED12M) $n_{Ahi} = 20$	$n_{Ahi} = 40$
$\hat{\bar{y}}_{HJ}$	-11.7	-11.6	-9.2	-9.1
$\hat{\bar{y}}_{P}$	0.7	0.8	0.9	1.1
$\hat{t}_{yPWR}$	-37.7	-37.6	-40.8	-40.7
$\hat{t}_{yP}$	2.0	2.2	1.5	1.6

HJ = Hájek estimator; PWR = p-expanded with-replacement estimator.

5.5.3 Results for Point Estimators

Calibration inflates the variance of point estimates when the variability in the analysis weights is increased. Without a greater decrease in the squared bias, the MSE of the estimates increases — an undesirable occurrence. We begin in Table 5.1 with an examination of the MSE by comparing the relative bias (*RelBias*) of the mean estimates using only the design weights, i.e., Hájek estimators, against those estimators that incorporate an EC-PSGR adjustment. Negative values in Table 5.1 indicate underestimation, while positive values suggest estimates in excess of the true values. Relative biases of zero are ideal; however, values near zero are also acceptable and more realistic with simulation studies. The relative bias for the Hájek estimator of the population mean ($\hat{\bar{y}}_{HJ}$) calculated from the 4,000 simulation samples identifies underestimates in excess of nine percent. The outcome variable NOTCOV has higher levels of underestimation in comparison with PDMED12M even though the latter condition is rarer in the population (17 versus 7 percent). EC poststratification corrects for undercoverage resulting in a slight overestimate of the population mean by approximately one percent, thereby providing a justification for the weight adjustment procedure. The same conclusion is obtained for the estimated totals originally discussed in Chapter 4 and reproduced in the last two rows of Table 5.1.

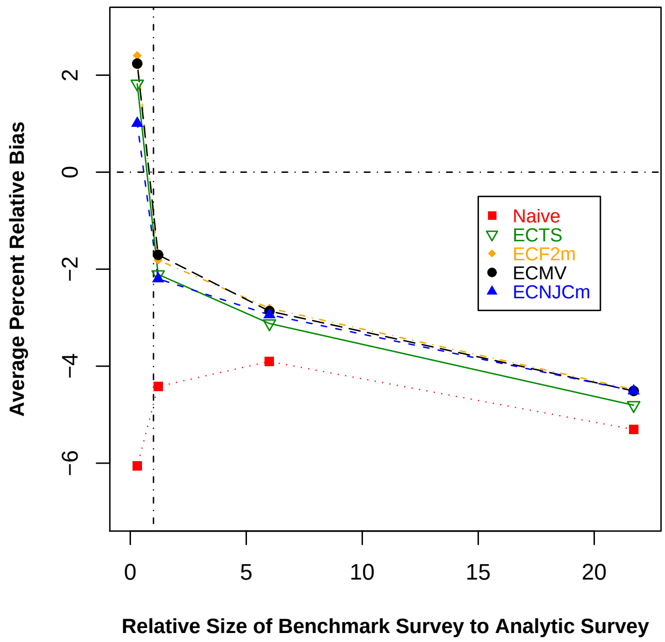
Comparing the relative biases for the two EC poststratified estimators, we see that $RelBias(\hat{\bar{y}}_P) < RelBias(\hat{t}_{yP})$ for both within-PSU samples sizes and outcome variables. The difference is less pronounced for PDMED12M in comparison with NOTCOV. The benchmark controls for the simulation $\hat{\mathbf{N}}_{B(r)}$ are generated under a multivariate normal distribution such that $\hat{\mathbf{N}}_{B(r)} \sim MVN(\hat{\mathbf{N}}_B,\ \hat{\mathbf{V}}_B)$ as detailed in Section 4.5.1. The average of the poststratum benchmark totals was verified to be very close (though not exact) to the values in $\hat{\mathbf{N}}_B$. However, the average of $\sum_g \hat{N}_{Bg(r)}$ exceeded N for our study. Therefore, by our discussion of Figure 5.1, we expect and see in Table 5.1 that $RelBias(\hat{\bar{y}}_P) < RelBias(\hat{t}_{yP})$ due to the overestimation of N by $\hat{N}_B$.

5.5.4 Comparison of Variance Estimators

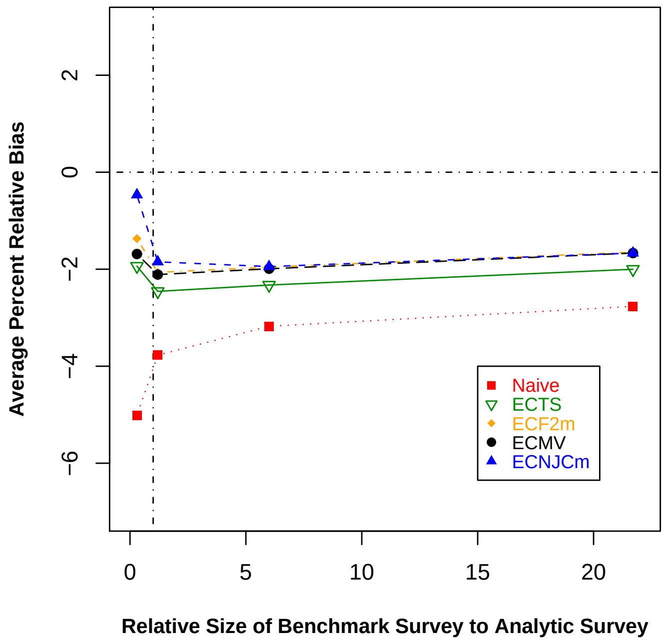
Having addressed the relative bias of the point estimators, we next compare the relative biases for our variance estimators. The *percent biases relative to the empirical MSE* for the variance of the estimated means range between -8 percent and just over 3 percent across the simulation parameters with most values falling below the desired zero percent level. This range is much smaller than the range calculated for the estimated totals, i.e., -56 percent to roughly zero percent, and is associated with the contribution of the benchmark controls to the estimated variances. As discussed for expression (5.20), a large, positive covariance between the numerator and denominator of the estimated mean (i.e., $\bar{\mathbf{Y}}'_A \mathbf{V}_B \mathbf{1}_G$) relative to a function of the numerator variance (i.e., $\bar{y}_P \mathbf{1}'_G \mathbf{V}_B \mathbf{1}_G / 2$) will reduce the influence of the benchmark covariance matrix on the overall variance. In our simulation population, the covariance is 1.10 and 1.00 times as large as the variance term for NOTCOV and PDMED12M, respectively. Averaged across the simulation samples, the relative increase in size of the covariance is 1.08 for NOTCOV and 0.98 for PDMED12M.

The pattern of bias across the sizes of the benchmark and analytic surveys for the estimated means also differs from the total estimates shown in, for example, Figure 4.1. Figure 5.2 displays the estimated percent relative bias of the five variance estimators (y axis) in estimating the MSE of our two outcome variables (NOTCOV and PDMED12M) by the relative size (n_B/n_A) of the benchmark survey to the analytic survey of size $n_A = 1,000$ (x axis). The horizontal line represents zero bias. The vertical line represents studies for which the analytic and benchmark surveys are equal in size as well as a relatively equal-sized contribution to the overall variance. The relative biases for the ECTS, ECF2m, and ECMV variance estimators for estimated totals were similar — see Figure (4.1). In this chapter, however, there is a slight visual distinction between their values due to the smaller scale of the y axis.

For both outcome variables, the traditional poststratified variance estimator (Naïve) is most negatively biased as noted in our theoretical examination. This



(a) Average Number Not Covered by Health Insurance in Last 12 Months (NOTCOV)



(b) Average Number Delayed Medical Care Due to Cost in Last 12 Months (PDMED12M)

Figure 5.2: Percent Bias Relative to Empirical MSE of Five Variance Estimators by Relative Size of the Benchmark Survey to the Analytic Survey for 1,000 Analytic Survey Units

holds for the four relative survey sizes included in our study. The relative bias for NOTCOV shown in Figure 5.2(a) is smallest when the benchmark survey is approximately six times larger than the analytic survey. As the relative benchmark size decreases, the negative bias falls below 6 percent. This is a stark contrast to the EC variance estimators presented here, as well as, the relative bias of -50 percent calculated for the estimated total. A similar interpretation can be used for PDMED12M in Figure 5.2(b).

A comparison of the biases for the EC variance estimators shows similar patterns within the relative sizes of the surveys for both outcome variables. When the relative size of the benchmark survey is greater than the analytic survey (right of the vertical line), the empirical EC variance estimates are all too small but only by levels as much as 5 percent for NOTCOV and 3 percent for PDMED12M. Once the benchmark size drops below 1,000, the EC variance estimators become conservative and overestimate the NOTCOV population parameter by less than 2.5 percent. Underestimation by as much as 2 percent is seen with the PDMED12M variance estimates (Figure 5.2(b)). We believe that these levels of negative bias would likely disappear with a larger analytic survey sample size. The dramatic change in the biases from $n_B/n_A = 1.2$ to $n_B/n_A = 0.3$ suggests that additional research is needed to determine a threshold for when a benchmark adjustment will result in overly conservative variance estimates. We also note that the relative biases for the ECTS are slightly lower than the other EC variance estimators. This is attributed to linearization variance estimators producing, in general, more stable estimates than replication variance estimators (Krewski & Rao, 1981).

Table 5.2: Percent Bias Estimates Relative to Empirical MSE for Five Variance Estimators by Mean Outcome Variable and Relative Size of the Benchmark Survey to the Analytic Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-6.1	-4.4	-3.9	-5.3	-7.9	-5.1	-4.5	-5.6
	ECTS	1.8	-2.1	-3.1	-4.8	2.3	-2.1	-3.4	-4.9
	ECF2m	2.4	-1.8	-2.8	-4.5	3.1	-1.9	-3.3	-4.7
	ECMV	2.2	-1.7	-2.9	-4.5	2.5	-2.0	-3.4	-4.7
	ECNJCm	1.0	-2.2	-2.9	-4.5	0.3	-2.4	-3.5	-4.8
PDMED12M	Naïve	-5.0	-3.8	-3.2	-2.8	-7.6	-7.1	-6.0	-7.8
	ECTS	-1.9	-2.5	-2.3	-2.0	-3.4	-5.1	-4.5	-6.5
	ECF2m	-1.4	-2.1	-2.0	-1.6	-3.0	-5.0	-4.4	-6.3
	ECMV	-1.7	-2.1	-2.0	-1.7	-3.1	-5.1	-4.4	-6.3
	ECNJCm	-0.5	-1.8	-1.9	-1.7	-1.9	-4.6	-4.4	-6.3

The summary measures used to produce Figure 5.2 are contained in the first set of four columns of Table 5.2, i.e., columns associated with $n_A = 1,000$. The second set of columns within this table contains the percent relative biases of the outcome variables for an analytic survey of size $n_A = 2,000$. Many of the same conclusions derived for the $n_A = 1,000$ estimates can be repeated for the estimates derived under $n_A = 2,000$.

Overall we can see that there are no striking differences in the EC relative biases for all conditions unlike the comparisons made for the estimated totals in Chapter 4. The contrast between the percent relative bias for estimated totals and means within each method is most noticeable with the ECNJCm. The ECNJCm values follow closely with the Naïve estimator in Figure 4.1, though levels of bias are much less. The ECNJCm values in Figure 5.2 and Table 5.2, however, are closer in value to the other EC variance estimators. This suggests that when a complete benchmark covariance matrix is not available, estimated (ratio) means may be less biased than the corresponding totals used in the numerator of the ratio. Additional theory is needed, however, to generalize this finding.

We additionally examine the *percent bias relative to the empirical variance* to determine if the empirical bias is affecting our results. Overall, the percent relative biases were improved by no more than 1.4 percentage points. We have chosen to suppress this tabular information because of the similarities with estimates provided in Table 5.2.

The next criterion used to compare the variance estimators is the empirical *coverage rates for the 95 percent confidence interval* (CI) associated with the two outcome variables. Coverage rates for the estimated means under all simulation conditions were fairly stable and near the desired level of 95 percent. We additionally do not detect a linear trend with the increasing size of the benchmark survey. Hence, we show only the minimum, maximum, and range of the coverage rates in Table 5.3 by outcome variable, variance estimator, and relative size of the benchmark and analytic surveys. The minimum coverage rate across the values in the table rests with the Naïve variance estimator though the differences are not excessive.

Because of the visual uniformity of the results in Table 5.3, we ran a linear regression to determine the correlates of CI coverage rates. The covariates included simulation results

Table 5.3: Minimum, Maximum, and Range of Empirical 95 Percent Coverage Rates for Five Variance Estimators Across Relative Size of the Benchmark Survey to the Analytic Survey by Mean Outcome Variable

Outcome Variable	Variance Estimator	$n_A = 1,000$ 95 Pct Coverage Min	Max	Range	$n_A = 2,000$ 95 Pct Coverage Min	Max	Range
NOTCOV	Naïve	93.6	94.2	0.6	92.8	94.0	1.2
	ECTS	93.8	95.0	1.2	93.8	94.3	0.6
	ECF2m	93.8	94.9	1.1	93.7	94.2	0.5
	ECMV	93.8	94.8	0.9	93.7	94.2	0.5
	ECNJCm	93.9	94.9	1.0	93.5	94.0	0.5
PDMED12M	Naïve	94.0	94.3	0.3	93.8	94.8	1.1
	ECTS	94.3	94.6	0.2	94.2	95.0	0.8
	ECF2m	94.4	94.9	0.5	94.2	95.0	0.7
	ECMV	94.3	94.7	0.3	94.3	94.8	0.5
	ECNJCm	94.3	94.8	0.5	94.2	94.8	0.6

- the relative bias of the point estimates (Table 5.1),
- the relative bias of the variance estimators (Table 5.2), and
- the calculated bias ratio,

and simulation conditions

- outcome variable (NOTCOV and PDMED12M),
- size of the analytic survey (1,000 and 2,000),
- the relative size of the benchmark survey (four sizes), and
- variance estimator (Naïve, ECTS, ECF2m, ECMV, and ECNJCm).

Särndal et al. (1992, Section 5.2) define the bias ratio of an estimator $\hat{\theta}$, $BR(\hat{\theta})$, as the bias, $E(\hat{\theta}) - \theta)$, divided by the root population sampling variance, $\sqrt{Var(\hat{\theta})}$.

Table 5.4: Minimum, Maximum, and Range of Empirical 95 Percent Coverage Rates for Five Variance Estimators Across Relative Size of the Benchmark Survey to the Analytic Survey by Total of Outcome Variable

Outcome Variable	Variance Estimator	$n_A = 1,000$ 95 Pct Coverage Min	Max	Range	$n_A = 2,000$ 95 Pct Coverage Min	Max	Range
NOTCOV	Naïve	83.5	93.7	10.1	81.2	93.4	12.2
	ECTS	94.0	95.6	1.6	93.7	95.7	2.0
	ECF2m	93.9	95.1	1.2	93.5	95.5	2.0
	ECMV	94.0	95.1	1.1	93.6	95.5	2.0
	ECNJCm	88.6	94.0	5.4	87.8	93.6	5.8
PDMED12M	Naïve	88.8	94.4	5.5	84.8	94.2	9.4
	ECTS	94.5	94.8	0.3	94.1	95.4	1.2
	ECF2m	94.5	94.8	0.4	94.1	95.2	1.2
	ECMV	94.4	95.0	0.6	94.0	94.8	0.9
	ECNJCm	91.1	94.5	3.5	89.0	94.4	5.4

This bias ratio affects the desired CI coverage rates through the formula $P(|Z + BR(\hat{\theta})| \leq z_{1-\alpha/2})$ for $Z = [\hat{\theta} - E(\hat{\theta})]/\sqrt{Var(\hat{\theta})}$. Bias ratios larger than one can either reduce or increase the coverage rates, depending on the positive or negative bias term, while small bias ratios have minimal effects on the rates. Among the model covariates included in the linear model ($R^2 = 0.78$), only the relative size of the benchmark survey and the variance estimator were not significantly associated with the confidence interval coverage rates. The remaining covariates were highly significant at levels less than 0.001.

In comparison to these fairly stable rates, the range of the 95 percent confidence coverage rates is wider in general for the estimated totals (Table 5.4). The increased range in the coverage rates is especially noticeable for the Naïve and ECNJCm

Table 5.5: Percent Increase in Instability of Variance Estimates Relative to EC Linearization Estimator (ECTS) by Outcome Variable and Relative Size of the Benchmark Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$				Relative Size $n_B/(n_A = 2,000)$			
		0.3	1.2	6.0	21.7	0.2	0.6	3.0	10.8
NOTCOV	Naïve	3.9	1.0	0.4	0.3	4.6	1.5	0.5	0.4
	ECF2m	2.9	0.9	0.5	0.0	3.8	1.3	0.2	0.2
	ECMV	3.5	1.2	0.6	0.0	3.6	0.9	0.3	0.4
	ECNJCm	3.2	0.7	0.4	0.1	3.2	1.1	0.3	0.3
PDMED12M	Naïve	0.8	0.5	0.3	0.3	1.6	0.9	0.7	0.6
	ECF2m	1.5	0.7	0.3	0.2	1.7	0.6	0.6	0.5
	ECMV	1.4	0.5	0.2	0.2	1.7	0.6	0.6	0.5
	ECNJCm	1.3	0.5	0.4	0.2	1.9	0.7	0.6	0.5

variance estimators. We ran the same linear regression specified above to determine the correlates of the coverage rates for estimated total. For this linear model ($R^2 = 0.96$), size of the analytic survey, outcome variable, and type of variance estimator was not significantly associated with the coverage rates. It is interesting to note that unlike the regression model for the estimated ratio-means, the relative size of the benchmark survey was highly significant.

The discussion so far suggests that there are minimal theoretical, as well as empirical, differences between the ECTS, ECF2m, and ECMV methods. A comparison of the *variation in the variance estimates* suggests that the ECTS variance estimator is most stable among those examined though the relative increase for the other estimators was less than five percent (Table 5.5). This corresponds with the theoretical discussion given in Krewski & Rao (1981). The difference in the stability of the ECF2 and ECMV methods is less noticeable with estimated means than with

Table 5.6: Percentage Point Reduction in Bias Relative to Empirical MSE Attributed to Coverage Error Variance Averaged over ECF2 and ECNJC Variance Estimators and Size of Benchmark Survey, by Type of Point Estimator and Size of the Analytic Survey

Point Estimator	Outcome Variable	Analytic Survey Size $n_A = 1{,}000$	$n_A = 2{,}000$
$\hat{\bar{y}}_P$	NOTCOV	-0.41	-0.55
	PDMED12M	-0.73	-1.28
$\hat{t}_{yP}$	NOTCOV	-0.34	-0.45
	PDMED12M	-0.64	-1.08

the estimated totals displayed in Table 4.6.

Our final analysis involves an examination of the *undercoverage error variance component* introduced into the original formulae for the Fuller and Nadimpalli-Judkins-Chu jackknife variance estimators. Table 5.6 shows the percentage point reduction in the bias of the variance estimates relative to the empirical variance by including an undercoverage error component. The values are averaged across benchmark survey size and EC variance estimator due to the similarities in the results. On average, the relative percent bias is reduced between 0.4 and 1.3 percentage points with the larger reductions occurring as the analytic survey sample size increases. This pattern is also seen for the estimated totals ($\hat{t}_{yP}$) shown in the second half of Table 5.6; however, the percentage point decrease in bias is slightly higher for the ratio mean ($\hat{\bar{y}}_P$). Additionally, the increase in the 95 percent coverage rates associated with the coverage error component is less then 0.4 percentage points for both methods. This suggests that an undercoverage error adjustment is useful for the variance estimator. However, as discussed in Chapter 4, further research is needed in an attempt to develop a more effective coverage error variance component.

5.6 Summary of Research Findings

Many of the same conclusions noted for EC-calibrated totals in Section 4.6 are echoed for the ratio of two EC-calibrated totals. The traditional GREG variance estimators can underestimate the population sampling variance though our empirical results suggest that the severity is less with ratio-mean estimators. The level of underestimation is related to the precision of the benchmark control totals. The original and modified ECNJC methods can also produce estimates that are too small if the missing population covariance values are negative. Our simulation study suggests that the bias in the ECNJCm variance estimates is less pronounced with the ratio means than with totals though additional theory is needed to support this claim.

Our recommendation therefore points to the remaining EC calibration variance estimators; a specific recommendation is less clear cut in this chapter in contrast with Chapter 4. Theoretically, the newly developed linearization variance estimator (ECTS), the modified Fuller two-phase jackknife estimator (ECF2m), and the multivariate normal jackknife estimator (ECMV) are asymptotically equivalent. The empirical results suggest that the differences among the three methods in practice are negligible. Choosing between the ECTS and one of the jackknife replication methods must be based on the type of analysis or public-use file desired for the study.

Relevant steps and computer code are provided for the ECF2m and ECMV methods to facilitate their implementation. As mentioned previously, additional work is required to improve the variance component associated with any (non-random) undercoverage in the analytic survey sampling frame.

Chapter 6

Domain Estimation

6.1 Introduction

Domain (or subpopulation) estimation is an integral part of the design and analysis phases of the survey. As discussed in Chapter 2, calibration domain point and variance estimators have been studied but the literature currently does not extend to estimated-control (EC) calibration. Research on EC calibration for domain totals and ratio-means begins with our work presented in this chapter. Here we assume that the domain of interest is large enough to allow direct estimation instead of additionally addressing situations when small area estimation techniques are required.

The research presented in the next sections relies heavily on the theoretical work presented in Chapters 4 and 5. We reference certain formulae from these chapters and discuss the modifications required for domain estimation, instead of presenting similar results. When appropriate, we detail issues with EC calibration that are specific to domain estimation. However, explicit formulae for domain point estimators are described to maximize clarity.

Our research on EC calibration for totals and ratio-means within sizeable domains is presented in the next sections. We detail the formulae and design bias for the new generalized regression (EC-GREG) and poststratified (EC-PSGR) estimators of totals within a domain in Section 6.2. The ratio of two EC-GREG totals and of two EC-PSGR totals within a domain (domain ratio-mean estimators) is similarly defined and evaluated in Section 6.3. This section additionally contains a comparison of bias levels for overall and domain-specific total and ratio-mean estimators. As in Chapter 5, the mean of an outcome within a domain is the ratio of particular interest. Our findings, however, generalize to the ratio of any two calibrated domain-specific totals. We evaluate the set of variance estimators identified for our research in two sections — variance estimation for domain totals in Section 6.4, and for ratio-means in Section 6.5. Comparisons are made between the variance estimators for domain and overall units to suggest under what conditions EC calibration may have a stronger influence. We present empirical domain-estimator results from a simulation study in Section 6.6. The findings, both theoretical and empirical, are summarized in the final section (Section 6.7).

6.2 Estimation of Domain Totals

The general label *estimators for domain totals* includes the population domain size, as well as the total number of population units within a domain with a characteristic (outcome) of interest. The formulae for the EC-GREG and EC-PSGR estimators are expressed in terms of an outcome variable in Section 6.2.1. The design-based bias of these estimators follows in Section 6.2.2.

6.2.1 Calibration Estimators

The EC-GREG estimated population total for domain d is calculated as follows:

$$\begin{aligned}\hat{t}_{ydR} &= \hat{t}_{Ayd} + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_{Ad} \\ &= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \left[1 + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{l=1}^{n_{Ahi}} \pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right)^{-1} \mathbf{x}_{hik}\right] \pi_{hik}^{-1} y_{dhik} \\ &= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1}\delta_{dhik}y_{hik} \qquad (6.1)\end{aligned}$$

where $\hat{t}_{Ayd} = \sum_{hik \in s_A} \pi_{hik}^{-1} y_{dhik}$, the *pwr* total estimator of y for domain d using the analytic survey design with $y_{dhik} = \delta_{dhik} y_{hik}$; $a_{hik} = 1 + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \left(\sum_{hil \in s_A} \pi_{hil}^{-1} \times \mathbf{x}_{hil}\mathbf{x}_{hil}'\right)^{-1} \mathbf{x}_{hik}$, the calibration adjustment factor also used in the overall estimated total $\hat{t}_{yR}$ (4.1); and, $\delta_{dhik} = 1$ if unit k in PSU i within stratum h is a member of domain d ($\delta_{dhik} = 0$ otherwise). Under the *regression model approach*, this calibration estimator is generated through an assisting model specified by $E_\epsilon(y_{dhik}) = \mathbf{x}_k'\mathbf{B}_d$ and $Var_\epsilon(y_k) = \sigma^2$, where E_ϵ and Var_ϵ represent the expectation and variance evaluated with respect to the model; and

$$\mathbf{B}_d = \left[\sum_{hil \in U} \mathbf{x}_{hil}\mathbf{x}_{hil}'\right]^{-1} \sum_{hik \in U} \mathbf{x}_{hik}\delta_{dhik}y_{hik}. \qquad (6.2)$$

Särndal et al. (1992, Section 10.6) refer to this assisting model as a *separate ratio model* because the slope coefficients are defined *within* and not across the domains. The vector of sample coefficients for the working domain-specific model in (6.1) is defined as:

$$\hat{\mathbf{B}}_{Ad} = \left[\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{l=1}^{n_{Ahi}} \pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right]^{-1} \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1}\mathbf{x}_{hik}\delta_{dhik}y_{hik}. \qquad (6.3)$$

The G-length vectors $\hat{\mathbf{t}}_{Bx} = \sum_{l \in s_B} w_l \mathbf{x}_l$ and $\hat{\mathbf{t}}_{Ax} = \sum_{hik \in s_A} \pi_{hik}^{-1}\mathbf{x}_{hik}$ in (6.1) contain estimates of the auxiliary variable totals from the complete benchmark and analytic samples, respectively. These estimators are also used in the EC-GREG estimator of the overall population total of y, $\hat{t}_{yR}$ defined in expression (4.1). We could have used $\hat{\mathbf{t}}_{Bxd} = \sum_{l \in s_B} w_l \delta_{dl} \mathbf{x}_l$ and $\hat{\mathbf{t}}_{Axd} = \sum_{hik \in s_A} \pi_{hik}^{-1}\delta_{dhik}\mathbf{x}_{hik}$ to form a domain-specific calibration adjustment factor. However, this would violate our requirement of one set of analysis weights because such an adjustment would need to be produced for *each* analysis domain — see Section 2.4 for a more detailed discussion. Note that the only difference between $\hat{t}_{yR}$ (4.1) and $\hat{t}_{ydR}$ (6.1) is that $\delta_{dhik}y_{hik}$ is used in place of y_{hik} for domain estimation. This modification to previously presented formulae is seen throughout this chapter.

The estimated domain total under EC poststratification, a specialized EC-GREG estimator, is a function of two indicator variables: $\delta_{ghik} = 1$ if the $(hik)^{th}$ unit is in poststratum g (zero otherwise), and δ_{dhik} described for $\hat{t}_{ydR}$ (6.1). The estimated total is calculated as follows:

$$\begin{aligned}
\hat{t}_{ydP} &= \hat{\mathbf{N}}'_B \hat{\mathbf{N}}_A^{-1} \hat{\mathbf{t}}_{Ayd} = \hat{\mathbf{N}}_B \hat{\bar{\mathbf{Y}}}_{Ad} \\
&= \sum_{g=1}^{G} \hat{N}_{Bg} \hat{N}_{Ag}^{-1} \hat{t}_{Aydg} \\
&= \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \left[\sum_{g=1}^{G} \hat{N}_{Bg} \hat{N}_{Ag}^{-1} \, \delta_{ghik} \right] \pi_{hik}^{-1} \delta_{dhik} y_{hik} \\
&= \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \sum_{g=1}^{G} a_{hik} \pi_{hik}^{-1} \delta_{dhik} y_{hik} \qquad (6.4)
\end{aligned}$$

where $\hat{\mathbf{N}}'_B = \left[\hat{N}_{B1}, \; ..., \; \hat{N}_{BG} \right]$, the vector of EC-PSGR benchmark controls with $\hat{N}_{Bg} = \sum_{l \in s_B} w_l \delta_{gl}$ and δ_{gl}, the poststratum-indicator variable for the benchmark survey; $\hat{\mathbf{N}}_A$ is a diagonal matrix of G poststratum counts estimated from the analytic survey data with elements $\hat{N}_{Ag} = \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik}$; $\hat{\mathbf{t}}_{Ayd} = \left[\hat{t}_{Ayd1}, \; ..., \; \hat{t}_{AydG} \right]'$, the vector of analytic survey estimated population totals for variable y within domain d and poststratum g such that $\hat{t}_{Aydg} = \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik} \delta_{dhik} y_{hik}$; $\hat{\bar{\mathbf{Y}}}_{Ad} = \hat{\mathbf{N}}_A^{-1} \hat{\mathbf{t}}_{Ayd} = \left[\hat{\bar{y}}_{Ad1}, \; ..., \; \hat{\bar{y}}_{AdG} \right]'$, the vector of estimated coefficients under the group-mean assisting model specified by $E_\epsilon(y_{dhik}) = \bar{y}_{Adg}$ and $Var_\epsilon(y_k) = \sigma^2$ with $\hat{\bar{y}}_{Adg} = \hat{t}_{Aydg} / \hat{N}_{Ag}$; $\bar{y}_{Adg} = t_{Aydg} / N_{Ag}$ with $t_{Aydg} = \sum_{hik \in U} \delta_{ghik} \delta_{dhik} y_{hik}$ and $N_{Ag} = \sum_{hik \in U} \delta_{ghik}$; and $a_{hik} = \sum_{g=1}^{G} \hat{N}_{Bg} \hat{N}_{Ag}^{-1} \, \delta_{ghik} = \hat{N}_{Bg} \hat{N}_{Ag}^{-1}$, the calibration adjustment factor. Note that a_{hik} specified in (6.4) is the same as defined for the overall EC-PSGR estimator of a total, $\hat{t}_{yP}$ given in expression (4.3), and is neither a function of the outcome variable y nor the domain indicator δ_{dhik}.

6.2.2 Bias of the Estimators

The design-based bias is a function of the expected value of an estimator and the population parameter being estimated. The population domain total is denoted as

$$t_{yd} = \sum_{k \in U} \delta_{dk} y_k.$$

Because $\hat{t}_{ydR}$ (6.1) is a nonlinear function of sample estimators, we evaluate the expectation of the linearized expression through a first-order Taylor series approximation:

$$\begin{aligned}
\hat{t}_{ydR} &= \hat{t}_{Ayd} + (\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax})'\hat{\mathbf{B}}_{Ad} \\
&\cong t_{ydR} + \left(\hat{t}_{Ayd} - t_{Ayd}\right) + \mathbf{B}'_{Ad}\left(\hat{\mathbf{t}}_{Bx} - \mathbf{t}_{Bx}\right) \\
&\quad - \mathbf{B}'_{Ad}\left(\hat{\mathbf{t}}_{Ax} - \mathbf{t}_{Ax}\right) + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\left(\hat{\mathbf{B}}_{Ad} - \mathbf{B}_{Ad}\right) \\
&= t_{ydR} + \max\left\{O_P\left(M/\sqrt{m_{Ad}}\right),\ O_P\left(M/\sqrt{m_B}\right)\right\}. \qquad (6.5)
\end{aligned}$$

where m_{Ad} denotes the number of analytic survey PSUs containing at least one element of the domain from a total of M_{Ad} domain PSUs on the analytic survey sampling frame. The complete population contains M_d domain PSUs where $M_{Ad} \leq M_d$ by definition. For this first-order approximation, we assume the population parameters $t_{ydR} = t_{Ayd} + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'\mathbf{B}_{Ad}$, $\mathbf{t}_{Bx}$, and $\mathbf{t}_{Ax}$ are all $O(M)$ where M is the total number of PSUs in the complete population; $\mathbf{B}_{Ad} = O(1)$; $\left(\hat{t}_{Ayd} - t_{Ayd}\right) = O_P\left(M_d/\sqrt{m_{Ad}}\right)$; $\left(\hat{\mathbf{t}}_{Bx} - \mathbf{t}_{Bx}\right) = O_P\left(M/\sqrt{m_B}\right)$; $\left(\hat{\mathbf{t}}_{Ax} - \mathbf{t}_{Ax}\right) = O_P\left(M/\sqrt{m_A}\right)$; and $\left(\hat{\mathbf{B}}_{Ad} - \mathbf{B}_{Ad}\right) = O_P\left(m_{Ad}^{-1/2}\right)$. Note that the $O_P\left(M/\sqrt{m_{Ad}}\right)$ term dominates both the $O_P\left(M/\sqrt{m_A}\right)$ and $O_P\left(M_d/\sqrt{m_{Ad}}\right)$ terms in (6.5) because we assume that the domain is a subset of the analytic survey sample and complete population, i.e., $m_A \geq m_{Ad}$ and $M \geq M_d$. Therefore,

$$\begin{aligned} E\left(\hat{t}_{ydR}\right) &= t_{ydR} + \max\left\{O\left(M/\sqrt{m_{Ad}}\right),\ O\left(M/\sqrt{m_B}\right)\right\} \\ &\cong t_{ydR}. \end{aligned} \tag{6.6}$$

Following the approach used for $E\left(\hat{\mathbf{B}}_A\right)$ in (4.5), the expectation of the model coefficient vector evaluates to

$$\begin{aligned} E(\hat{\mathbf{B}}_{Ad}) &= E_{c_A}\left[E_A\left(\hat{\mathbf{B}}_{Ad}|c_A\right)\right] \\ &\cong \left[\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{l=1}^{N_{hi}} E_{c_A}\left(C_{Ahil}\right)\mathbf{x}_{hil}\mathbf{x}'_{hil}\right]^{-1} \\ &\quad \times \sum_{h}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} E_{c_A}\left(C_{Ahik}\right)\mathbf{x}_{hik}\delta_{dhik}y_{hik} \\ &= \left[\sum_{hil\in U}\phi_{Ahil}\mathbf{x}_{hil}\mathbf{x}'_{hil}\right]^{-1}\sum_{hik\in U}\phi_{Ahik}\mathbf{x}_{hik}\delta_{dhik}y_{hik} \equiv \mathbf{B}_{Ad} \end{aligned} \tag{6.7}$$

where $C_{Ahik} = 1$ indicates that the k^{th} population unit ($k \in U$) is listed on the analytic sampling frame (zero otherwise) with $E_{c_A}(C_{Ahik}) = \phi_{Ahik}$. Note that the subscript Ad above identifies the population model-coefficient vector associated with the domain-specific subset of the population covered by the analytic survey sampling frame, i.e., U_{Ad}. As discussed in Section 3.5, U_{Ad} and m_{Ad} are assumed to be of sufficient size for direct estimation. This implies that the coverage mechanism is not systematic and therefore, does not exclude all units within the domain of interest. Using the method shown in (6.7), $E\left(\hat{t}_{Ayd}\right) = E_{c_A}\left[E_A\left(\hat{t}_{Ayd}|c_A\right)\right] \cong \sum_{hik\in U}\phi_{Ahik}y_{hik} \equiv t_{Ayd}$ and $E(\hat{\mathbf{t}}_{Ax}) = E_{c_A}\left[E_A\left(\hat{\mathbf{t}}_{Ax}|c_A\right)\right] \cong \sum_{hik\in U}\phi_{Ahik}\mathbf{x}_{hik} \equiv \mathbf{t}_{Ax}$. The expectation of the benchmark control total vector equates to $\mathbf{t}_{Bx} = \sum_{l\in U}\phi_{Bl}\mathbf{x}_l$ where $C_{Bl} = 1$ identifies the population units listed on the benchmark survey frame such that $E(C_{Bl}) = \phi_{Bl}$ (zero otherwise).

Using $E\left(\hat{t}_{ydR}\right)$ in expression (6.6) and following the steps shown for $Bias(\hat{t}_{yR})$ in (4.7), the design-based bias for $\hat{t}_{ydR}$ is defined as follows:

$$\begin{aligned} Bias\left(\hat{t}_{ydR}\right) &\cong NC_{AE\phi d} - N(1-\bar{\phi}_A)\bar{E}_d + (\mathbf{t}_{Bx} - \mathbf{t}_{Ax})'(\mathbf{B}_{Ad} - \mathbf{B}_d) \\ &+ (\mathbf{t}_{Bx} - \mathbf{t}_x)'\mathbf{B}_d \end{aligned} \tag{6.8}$$

where $E_{dhik} = \delta_{dhik}y_{hik} - \mathbf{x}'_{hik}\mathbf{B}_d$, the population-level assisting model residual for domain d; $\bar{\phi}_A$ is the coverage rate for the analytic survey sampling frame; and $C_{AE\phi d} = \sum_{hik\in U}\left(E_{dhik} - \bar{E}_d\right)\left(\phi_{Ahik} - \bar{\phi}_A\right)/N$, the covariance between the coverage rates and the domain assisting-model residuals.

The four bias components in (6.8) each can be eliminated under the following conditions. (i) If the auxiliary variables ($\mathbf{x}_{hik}$) are correlated with y in domain d and with the coverage mechanism, and the working model is sufficiently close to the domain-specific population assisting model, then the random variation left unexplained by the model (in theory) should be uncorrelated with the coverage propensities, i.e., $C_{AE\phi d} \cong 0$. Under this scenario, the first bias component $NC_{AE\phi d}$ is approximately zero. (ii) If the design matrix contains a column of ones (intercept) so that the overall estimated population size is included as an auxiliary variable, then by definition $\bar{E}_d = 0$ and the second bias component is eliminated. (*iii*) If the coverage mechanism is such that it does not negatively affect the population model-coefficient vector within the domain, then $\mathbf{B}_{Ad} \cong \mathbf{B}_d$ and the third term is at least approximately zero. (*iv*) Finally, if $\mathbf{t}_{Bx} = \mathbf{t}_x$, as with traditional calibration, the last component is zero. Therefore, the estimator $\hat{t}_{ydR}$ will be asymptotically design unbiased only if *all* these conditions are satisfied. This occurrence is unlikely especially when examining multiple domains.

The bias for the corresponding EC-PSGR estimator of a domain total, $\hat{t}_{ydP}$ defined in (6.4), follows the development of $Bias\left(\hat{t}_{yP}\right)$ discussed in (4.8) and is specified as follows:

$$Bias\left(\hat{t}_{ydP}\right) \cong \sum_{g=1}^{G} \left\{ t_{ydg} \left(\frac{N_{Bg}}{N_g} - 1 \right) + N_{Bg} C_{Ay\phi dg} \frac{1}{\bar{\phi}_{Ag}} \right\} \tag{6.9}$$

where N_g is the complete population size within poststratum g; N_{Ag} and N_{Bg} are the poststratum sizes for the populations defined by the analytic and benchmark sampling frames; $C_{Ay\phi dg} = N_g^{-1} \sum_{hik \in U_g} \left(\delta_{dhik} y_{hik} - \bar{y}_{dg}\right) \left(\phi_{Ahik} - \bar{\phi}_{Ag}\right)$, the population covariance between the outcome variable within domain d and the coverage rates within poststratum g; $\bar{y}_{dg} = t_{ydg}/N_g$, the g^{th} poststratum mean of y in domain d; and $\bar{\phi}_{Ag} = N_{Ag}/N_g$, the average coverage rate within the poststratum under the analytic survey design. If the benchmark survey does not cover the target population correctly, so that $N_{Bg} \neq N_g$, then the first bias component, $t_{ydg}(N_{Bg}/N_g - 1)$, will be either positive (overestimate) or negative (underestimate) depending on the magnitude of the bias. This component will be strictly negative if the benchmark survey suffers undercoverage, and can accumulate across the poststrata to a sizeable negative bias depending on the magnitude of the outcome variable. Otherwise, this component is zero because the benchmark survey covers the population of interest. The second component may be negative if large y values within the domain are more likely to be excluded from a sampling frame. If, however, the coverage rates are the same within poststratum (i.e., $\phi_{Ahik} = \bar{\phi}_{Ag}$ for all units in poststratum g), then the second bias component is zero. As discussed for $Bias\left(\hat{t}_{yP}\right)$ in (4.8), the conditions under which both components are zero are unlikely to occur.

6.3 Estimation of Domain Means

Functions of domain totals are also important to survey data analysis. In this section, we provide an equation to estimate the ratio of two EC-GREG domain totals, focusing specifically on the mean of an outcome variable within a domain of interest (Section 6.3.1). This general formula is also expressed in terms of EC post-stratification. The design-based bias, as with the formulae for the point estimators, is a function of the domain total biases and is shown in Section 6.3.2.

6.3.1 Calibration Estimators

The estimated totals presented earlier in the chapter are used in this section to generate estimates for the population mean of y within domain d. We again focus on the Hájek estimator of the population mean within the domain instead of assuming that the population domain size, needed for the denominator, is known.

The Hájek EC-GREG ratio-mean of y in domain d is calculated as follows:

$$\hat{\bar{y}}_{dR} = \frac{\hat{t}_{ydR}}{\hat{N}_{dR}} \tag{6.10}$$

for $\hat{t}_{ydR}$ defined in expression (6.1), and

$$\begin{aligned} \hat{N}_{dR} &= \sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1}\delta_{dhik} \\ &= \hat{N}_{Ad} + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \hat{\mathbf{B}}_{ANd} \end{aligned} \tag{6.11}$$

with $\hat{N}_{Ad} = \sum_{hik\in s_A} \pi_{hik}^{-1}\delta_{dhik}$, the *pwr* domain population size estimated from the analytic survey; and a_{hik} is defined for $\hat{t}_{ydR}$. The vector of model coefficients, defined for the denominator estimator of the population domain size, is specified as

$$\hat{\mathbf{B}}_{ANd} = \left[\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{l=1}^{n_{Ahi}} \pi_{hil}^{-1}\mathbf{x}_{hil}\mathbf{x}_{hil}'\right]^{-1} \hat{\mathbf{t}}_{Axd} \tag{6.12}$$

with $\hat{\mathbf{t}}_{Axd} = \sum_{hik\in s_A} \pi_{hik}^{-1}\mathbf{x}_{hik}\delta_{dhik}$. Note that the formula associated with $\hat{N}_{dR}$ is based on the expression specified for $\hat{t}_{ydR}$ with $y_{hik} = 1$.

The Hájek EC-PSGR estimated population mean is expressed as

$$\hat{\bar{y}}_{dP} = \frac{\hat{t}_{ydP}}{\hat{N}_{dP}} \tag{6.13}$$

where the formula for $\hat{t}_{ydP} = \hat{\mathbf{N}}_B\hat{\mathbf{N}}_A^{-1}\hat{\mathbf{y}}_{Ayd} \equiv \hat{\mathbf{N}}_B\hat{\bar{\mathbf{Y}}}_{Ad}$ derived in (6.4). We note in expression (5.5) that the estimated population count used in the denominator of $\hat{\bar{y}}_P$ reduces to the sum of the estimated benchmark control totals, i.e., $\hat{N}_P \equiv$

$\hat{N}_B$. However, the simplification does not occur with domain estimation. The denominator in (6.13) is defined as

$$\begin{aligned}
\hat{N}_{dP} &= \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}} \left(\frac{\hat{N}_{Bg}}{\hat{N}_{Ag}}\right) \delta_{ghik}\pi_{hik}^{-1}\delta_{dhik} \\
&= \sum_{g=1}^{G} \left(\frac{\hat{N}_{Bg}}{\hat{N}_{Ag}}\right) \hat{N}_{Adg} \\
&\equiv \hat{\mathbf{N}}_B \hat{\mathbf{N}}_A^{-1} \hat{\mathbf{N}}_{Ad} \\
&\equiv \hat{\mathbf{N}}_B \hat{\bar{\mathbf{Y}}}_{ANd} \qquad (6.14)
\end{aligned}$$

where $\hat{\mathbf{N}}_{Ad} = \left[\hat{N}_{Ad1}, ..., \hat{N}_{AdG}\right]'$, a G-length vector of domain population totals by poststratum estimated from the analytic survey such that $\hat{N}_{Adg} = \sum_{hik \in s_A} \delta_{ghik}\pi_{hik}^{-1} \times \delta_{dhik}$; and $\hat{\bar{\mathbf{Y}}}_{ANd} = \hat{\mathbf{N}}_A^{-1}\hat{\mathbf{N}}_{Ad}$, the estimated proportion of domain units within each of the poststrata. The remaining terms are defined following the expression for $\hat{t}_{ydP}$ (6.4).

6.3.2 Bias of the Estimators

Ratio estimators are approximately unbiased only if all components are approximately unbiased. We note in Section 5.3 that the bias of the overall ratio estimator is small in general. The same holds true for a domain ratio estimator with a sufficient number of domain PSUs (i.e., the PSU contains at least one member of the domain). However, convergence to the population domain parameter is slower because the number of degrees of freedom is reduced. The domain population ratio-mean, the parameter of interest for the Hájek estimators given in the previous section, is defined as:

$$\bar{y}_d = \frac{t_{yd}}{N_d} = \frac{\sum_{k \in U} \delta_{dk} y_k}{\sum_{k \in U} \delta_{dk}}.$$

The bias of the non-linear EC-GREG domain ratio-mean $\hat{\bar{y}}_{dR}$ (6.10) is approximated using a first-order Taylor linearization:

$$\begin{aligned} Bias\left(\bar{y}_{dR}\right) &= E\left(\bar{y}_{dR}\right) - \bar{y}_d \\ &\cong E\left[\frac{1}{N_d}\left(\hat{t}_{ydR} - t_{yd}\right) - \frac{\bar{y}_d}{N_d}\left(\hat{N}_{dR} - N_d\right)\right] \\ &= \frac{1}{N_d}\left[Bias\left(\hat{t}_{ydR}\right) - \bar{y}_d Bias\left(\hat{N}_{dR}\right)\right]. \end{aligned} \tag{6.15}$$

The estimator $\bar{y}_{dR}$ is approximately unbiased only if both bias components are approximately zero. The numerator bias, $Bias\left(\hat{t}_{ydR}\right)$, is specified in expression (6.8) followed by the conditions under which this bias is negligible. The denominator bias, $Bias\left(\hat{N}_{dR}\right)$, is expressed in the same form as

$$\begin{aligned} Bias\left(\hat{N}_{dR}\right) &\cong NC_{ANE\phi d} - N(1-\bar{\phi}_A)\bar{E}_{Nd} + \left(\mathbf{t}_{Bx} - \mathbf{t}_{Ax}\right)'\left(\mathbf{B}_{ANd} - \mathbf{B}_{Nd}\right) \\ &\quad + \left(\mathbf{t}_{Bx} - \mathbf{t}_x\right)'\mathbf{B}_{Nd} \end{aligned} \tag{6.16}$$

where $E_{Ndhik} = \delta_{dhik} - \mathbf{x}'_{hik}\mathbf{B}_{Nd}$, the population residual for domain d under the assisting model specified for the domain estimator with $\mathbf{B}_{Nd} = \left[\sum_{hil \in U} \mathbf{x}_{hil}\mathbf{x}'_{hil}\right]^{-1} \times \sum_{hik \in U} \mathbf{x}_{hik}\delta_{dhik}$; $\bar{\phi}_A$ is the coverage rate for the analytic survey sampling frame; $C_{AEN\phi d} = \sum_{hik \in U}\left(E_{Ndhik} - \bar{E}_{Nd}\right)\left(\phi_{Ahik} - \bar{\phi}_A\right)/N$, the covariance between the coverage rates and the assisting model residuals for the domain estimator in the denominator of $\bar{y}_{dR}$; $\mathbf{B}_{ANd} = \left[\sum_{hil \in U} \phi_{Ahil} \mathbf{x}_{hil} \mathbf{x}_{hil}'\right]^{-1} \sum_{hik \in U} \phi_{Ahik} \mathbf{x}_{hik} \delta_{dhik} \cong E\left(\hat{\mathbf{B}}_{ANd}\right)$; and $E\left(\hat{N}_{dR}\right) \cong N_{dR} + \max\left\{O_P\left(M/\sqrt{m_{Ad}}\right),\ O_P\left(M/\sqrt{m_B}\right)\right\}$ with $N_{dR} = N_{Ad} + \left(\hat{\mathbf{t}}_{Bx} - \hat{\mathbf{t}}_{Ax}\right)' \mathbf{B}_{ANd}$. Similar conditions noted for $Bias\left(\hat{t}_{ydR}\right)$ will result in low levels of $Bias\left(\hat{N}_{dR}\right)$, such as no association between the auxiliary variables and the coverage probabilities within domain d.

The bias for the EC poststratified domain ratio-mean, $\hat{\bar{y}}_{dP}$ (6.13), follows this same pattern:

$$Bias\left(\bar{y}_{dP}\right) \cong \frac{1}{N_d}\left[Bias\left(\hat{t}_{ydP}\right) - \bar{y}_d Bias\left(\hat{N}_{dP}\right)\right] \tag{6.17}$$

where $Bias\left(\hat{t}_{ydP}\right)$ is given in expression (6.9). This formula is also used for $\hat{N}_{dP}$ with $y_{hik} = 1$ resulting in

$$Bias\left(\hat{N}_{dP}\right) \cong \sum_{g=1}^{G}\left\{N_{dg}\left(\frac{N_{Bg}}{N_g} - 1\right) + N_{Bg} C_{A\phi dg} \frac{1}{\bar{\phi}_{Ag}}\right\} \tag{6.18}$$

where $C_{A\phi dg} = \sum_{hik \in U_g}\left(\delta_{dhik} - \bar{d}_g\right)\left(\phi_{Ahik} - \bar{\phi}_{Ag}\right)/N_g$, the covariance between the domain indicators and the coverage propensities in poststratum g, and $\bar{d}_g = N_{dg}/N_g$, the proportion of population domain members in poststratum g. The remaining terms are the same as specified for $Bias\left(\hat{t}_{ydP}\right)$. Substituting the formulae for $Bias\left(\hat{t}_{ydP}\right)$ and $Bias\left(\hat{N}_{dP}\right)$ into (6.17) gives the complete expression for the bias of $\bar{y}_{dP}$:

$$Bias\left(\bar{y}_{dP}\right) \cong \frac{1}{N_d}\sum_{g=1}^{G}\left[N_{dg}\left(\bar{y}_{dg}-\bar{y}_d\right)\left(\frac{N_{Bg}}{N_g}-1\right)\right]$$
$$+\ \frac{1}{N_d}\sum_{g=1}^{G}\left[N_{Bg}\sum_{hik\in U_g}\left(\delta_{dhik}\left(y_{hik}-\bar{y}_d\right)-\frac{1}{N_g}\sum_{hik\in U_g}\delta_{dhik}\left(y_{hik}-\bar{y}_d\right)\right)\right.$$
$$\left.\times\ \left(\phi_{Ahik}-\bar{\phi}_{Ag}\right)/N_g\frac{1}{\bar{\phi}_{Ag}}\right] \qquad (6.19)$$

where $\bar{y}_{dg} = t_{ydg}/N_{dg}$ with $t_{ydg} = \sum_{hik\in U}\delta_{ghik}\delta_{dhik}y_{hik}$ and $N_{dg} = \sum_{hik\in U}\delta_{ghik}\delta_{dhik}$. The first bias component is zero if the benchmark survey sampling frame covers the population within poststratum g and the poststratum total is calculated using an unbiased estimator, i.e., $N_{Bg} \equiv N_g$. If the benchmark frame does not cover the poststratum population, then the bias component is positive or negative depending on the deviation between the poststratum and overall domain means. The second bias component is zero if the poststrata are formed so that the coverage propensities are the same for the domain members, i.e., $\phi_{Ahik}\delta_{dhik} = \bar{\phi}_{Ag}$. This is a stronger condition than specified for the bias of the overall ratio-mean in (5.10).

6.4 Variance Estimation for Domain Totals

Having addressed bias in the EC-GREG estimators of a population domain total in Section 6.2, we next examine the properties of the associated variance estimators. We begin by specifying the approximate population sampling variance and compare this expression against the expectation of the sample variance estimators identified for our research.

6.4.1 Population Sampling Variance

The unconditional population sampling variance for the EC-GREG estimator of a population total within domain d is evaluated with respect to the analytic and benchmark survey designs (A and B subscripts) and the analytic survey frame coverage mechanism (subscript c_A) through the following derivation:

$$\begin{aligned} AV\left(\hat{t}_{ydR}\right) &= E_B\left[E_{c_A}\left\{AV_A\left(\hat{t}_{ydR}|c_A,B\right)|B\right\}\right] \\ &\quad + E_B\left[AV_{c_A}\left\{E_A\left(\hat{t}_{ydR}|c_A,B\right)|B\right\}\right] \\ &\quad + AV_B\left[E_{c_A}\left\{E_A\left(\hat{t}_{ydR}|c_A,B\right)|B\right\}\right] \\ &\equiv AV\left(\hat{t}_{ydGREG}\right) \\ &\quad + \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}}\phi_{Ahik}\left(1-\phi_{Ahik}\right)E_{Adhik}^2 \\ &\quad + \mathbf{B}_{Ad}'\,\mathbf{V}_B\,\mathbf{B}_{Ad}. \end{aligned} \qquad (6.20)$$

This expression is obtained by applying the methods used for $AV\left(\hat{t}_{yR}\right)$ in Section 4.4 and substituting y_{hik} with $\delta_{dhik}y_{hik}$, $\mathbf{B}_A$ with $\mathbf{B}_{Ad}$ specified in (6.7), and E_{Ahik} with $E_{Adhik} = \delta_{dhik}y_{hik} - \mathbf{x}_{hik}'\mathbf{B}_{Ad}$. The first component in (6.20) is the approximate population sampling variance for the domain total under the traditional calibration assumptions with order $O(M^2/m_{Ad})$; the explicit formula is derived by substituting E_{Adhik} for E_{Ahik} in (4.14). The second component addresses the coverage error in the analytic survey specific to the set of domain population units and is $O(M)$. The final component inflates $AV(\hat{t}_{ydR})$ for the estimated benchmark control totals and is $O(M^2/m_B)$. The orders of magnitude differ from those presented in Chapter 4 in that the analytic survey variance component is now associated with the number of sample domain PSUs (m_{Ad}) instead of the total number of sample PSUs (m_A). Note that all of the variance components in (6.20) are by definition positive contributors to the overall variance.

By expressing the model-coefficient vector and residuals from (6.20) in terms of the group-mean model for domain d, the population sampling variance for the EC-PSGR domain estimator is specified as

$$\begin{aligned} AV\left(\hat{t}_{ydP}\right) &\equiv AV\left(\hat{t}_{ydPSGR}\right) \\ &+ \sum_{g=1}^{G}\sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \delta_{ghik}\phi_{Aghik}\left(1-\phi_{Aghik}\right)E_{Adhik}^{2} \\ &+ \bar{\mathbf{Y}}_{Ad}'\,\mathbf{V}_B\,\bar{\mathbf{Y}}_{Ad} \end{aligned} \qquad (6.21)$$

where $\bar{\mathbf{Y}}_{Ad} = \mathbf{N}_A^{-1}\mathbf{t}_{Ayd} = [\bar{y}_{Ad1},\ ...,\ \bar{y}_{AdG}]'$, the vector of population assisting-model coefficients with $\bar{y}_{Adg} = t_{Aydg}/N_{Ag}$; and $E_{Adhik} = \delta_{dhik}y_{hik} - \bar{y}_{Adg}$. Following the development of $AV\left(\hat{t}_{yPSGR}\right)$ given below (4.18),

$$AV\left(\hat{t}_{ydPSGR}\right) \equiv \mathbf{N}_B' E_{c_A}\left(\mathbf{V}_{Ad}\right)\mathbf{N}_B \qquad (6.22)$$

where $\mathbf{N}_B = [N_{B1},\ ...,\ N_{BG}]'$, the vector of totals for the G poststrata within the population associated with the benchmark sampling frame; and $\mathbf{V}_{Ad} = Var_A(\hat{\bar{\mathbf{Y}}}_{\mathbf{Ad}})$ $\cong \mathbf{D}_d\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}d}\mathbf{D}_d'$ with

$$\mathbf{D}_d = \left[diag\left(\left\{\frac{1}{N_{Ag}}\right\}_{g=1}^{G}\right), diag\left(\left\{\frac{-\bar{y}_{Adg}}{N_{Ag}}\right\}_{g=1}^{G}\right)\right]$$

and

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \sigma_{(\hat{t}_{Ayd1},\hat{t}_{Ayd1})} & \cdots & \sigma_{(\hat{t}_{Ayd1},\hat{N}_{AG})} \\ \vdots & \ddots & \vdots \\ \sigma_{(\hat{N}_{AG},\hat{t}_{Ayd1})} & \cdots & \sigma_{(\hat{N}_{AG},\hat{N}_{AG})} \end{bmatrix}.$$

6.4.2 Traditional Calibration Variance

Variance estimation for traditional calibration only recognizes the variation within the analytic survey. As discussed in, for example, Särndal et al. (1992, Section 10.6), the traditional linearization sample variance estimator for a GREG domain total is a function of the estimated residuals for the chosen assisting model. In the case of a stratified, multistage analytic survey design with PSUs selected with replacement and the separate ratio model, the linearization sample variance estimator for $\hat{t}_{ydR}$ is calculated as:

$$var_{Naïve}(\hat{t}_{ydR}) = var(\hat{t}_{ydGREG}) = \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{dhi+} - \bar{\breve{u}}_{dh++})^2 \qquad (6.23)$$

where $\breve{u}_{dhi+} = \sum_{k=1}^{n_{Ahi}} a_{hik}\pi_{hik}^{-1}e_{Adhik}$, the sum of (calibration) weighted model residuals for units within domain d within PSU hi; $e_{Adhik} = \delta_{dhik}y_{hik} - \mathbf{x}'_{hik}\hat{\mathbf{B}}_{Ad}$; a_{hik} is the calibration weight defined for $\hat{t}_{ydR}$ in (6.1); and $\bar{\breve{u}}_{dh++} = m_{Ah}^{-1}\sum_{i=1}^{m_{Ah}} \breve{u}_{dhi+}$, the average weighted residual within stratum h. Note that this sample variance estimator is a function of residuals calculated from all sample PSUs (m_{Ah}) and does not exclude

PSUs without at least one domain member. Because domain membership within a PSU is a random event (by assumption), the non-domain PSUs could contain domain members given a different sample. Therefore, the zero estimate is included as a contribution to the overall variance estimate.

The Naïve sample variance estimator for the EC-PSGR domain total generated through poststratification is defined either as a method-of-moments estimator,

$$var_{Naïve}\left(\hat{t}_{ydP}\right) = var\left(\hat{t}_{ydPSGR}\right) = \hat{\mathbf{N}}_B' \hat{\mathbf{V}}_{Ad} \hat{\mathbf{N}}_B \tag{6.24}$$

where $\hat{\mathbf{V}}_{Ad} \cong \hat{\mathbf{D}}_d \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}d} \hat{\mathbf{D}}_d'$, calculated using the analytic survey estimates corresponding to the terms defined for (6.22), and $\hat{\mathbf{N}}_B$ defined for $\hat{t}_{ydP}$ in (6.4); or by substituting $e_{Adhik} = \delta_{dhik} y_{hik} - \hat{\bar{y}}_{Adg}$ in the formula for $var_{Naïve}(\hat{t}_{ydR})$ (6.23) where $\hat{\bar{y}}_{Adg} = \hat{t}_{Aydg}/\hat{N}_{Ag} = \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik} \delta_{dhik} y_{hik} / \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{ghik}$. The term $\hat{t}_{ydPSGR} = \sum_g N_{Bg} \left(\hat{t}_{Aydg}/N_{Ag}\right)$ is the traditional poststratified domain total discussed in Section 2.4.

As discussed in Section 6.4.1, the population sampling variance is the sum of three positive variance components. Consequently, a variance estimate for an EC-calibrated total calculated with a traditional variance estimator will be too small. Hence, the use of the *Naïve variance estimator* label. The magnitude of the underestimation is suggested in the next section where we discuss linearization variance estimators that account for the three components.

6.4.3 Estimated-Control Linearization Variance

The formula for the EC linearization sample variance estimator of $\hat{t}_{ydR}$ (6.1), denoted as ECTS, is obtained by substituting sample estimators for the components in $AV\left(\hat{t}_{ydR}\right)$ (6.20), thereby accounting for all sources of variation. The ECTS sample variance estimator is expressed as:

$$\begin{aligned} var_{ECTS}\left(\hat{t}_{ydR}\right) &= var\left(\hat{t}_{ydGREG}\right) \\ &+ \sum_{h=1}^{H}\left(1-\hat{\bar{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e_{Adhik}^{2}}{\pi_{hik}} \\ &+ \hat{\mathbf{B}}_{Ad}'\hat{\mathbf{V}}_{B}\hat{\mathbf{B}}_{Ad}. \end{aligned} \tag{6.25}$$

The first sample variance component, $var\left(\hat{t}_{ydGREG}\right)$, is the traditional calibration variance estimator given in expression (6.23) and accounts for the variation within the analytic survey. The second component estimates the coverage error variance in the analytic survey sampling frame with $e_{Adhik} = \delta_{dhik}y_{hik} - \mathbf{x}_{hik}'\hat{\mathbf{B}}_{Ad}$ and $\hat{\bar{\phi}}_{Ah}$ is an estimate of the sampling frame coverage rate in stratum h. The estimates may be calculated as $\hat{N}_{Ah}/\hat{N}_{Bh}$, the ratio of the stratum sizes estimated from the analytic and benchmark survey data, if the benchmark survey is believed to adequately cover the complete population of interest. Using the formula for the bias shown in expression (4.22),

$$\begin{aligned} &Bias\left[\sum_{h=1}^{H}\left(1-\hat{\bar{\phi}}_{Ah}\right)\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{e_{Adhik}^{2}}{\pi_{hik}}\right] \\ &\quad\cong \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}}\phi_{Ahik}E_{Adhik}^{2}\left(\phi_{Ahik}-\bar{\phi}_{Ah}\right). \end{aligned} \tag{6.26}$$

If the coverage probabilities vary only by stratum, i.e., $\phi_{Ahik} \equiv \bar{\phi}_{Ah}$ for units within stratum h, then the associated bias is approximately zero. However, the bias is inflated if larger residuals are associated with coverage probabilities that differ from the stratum averages. The third component in (6.25) estimates the variation in the benchmark control totals where $\hat{\mathbf{B}}_{Ad}$ is the estimated coefficient vector specified for $\hat{t}_{ydR}$ (6.1), and $\hat{\mathbf{V}}_B = var\left(\hat{\mathbf{t}}_{Bx}\right)$, the estimated covariance matrix for the benchmark controls.

The order of convergence for the first domain variance component in (6.25) is $O_P(M^2/m_{Ad})$ and is of lower order than the corresponding component for an overall total, $O_P(M^2/m_A)$. The coverage error and benchmark variance components for $\hat{t}_{ydR}$ and $\hat{t}_{yR}$ are the same and equal $O_P(M)$ and $O_P(M^2/m_B)$, respectively. This suggests that the benchmark controls will have less influence on the variance of the domain estimators than with the overall estimators if $m_A \cong m_B$.

An expression for the ECTS sample variance estimator of an EC-PSGR domain total is defined as:

$$\begin{aligned} var_{ECTS}(\hat{t}_{yP}) &= var\left(\hat{t}_{ydPSGR}\right) \\ &+ \sum_{g=1}^{G}\left(1-\hat{\hat{\phi}}_{Ag}\right)\sum_{h=1}^{H}\sum_{i=1}^{m_{Ah}}\sum_{k=1}^{n_{Ahi}}\frac{\delta_{ghik}e_{Adhik}^2}{\pi_{hik}} \\ &+ \hat{\hat{\mathbf{Y}}}_{Ad}'\hat{\mathbf{V}}_B\hat{\hat{\mathbf{Y}}}_{Ad} \end{aligned} \qquad (6.27)$$

where $var\left(\hat{t}_{ydPSGR}\right)$ is defined in expression (6.24); $\hat{\hat{\phi}}_{Ag} = \hat{N}_{Ag}/\hat{N}_{Bg}$, for example; and $e_{Adhik} = \delta_{dhik}y_{hik} - \hat{\bar{y}}_{Adg}$ with $\hat{\bar{y}}_{Adg} = \hat{t}_{Aydg}/\hat{N}_{Ag}$. The remaining terms are defined for $\hat{t}_{ydP}$ in expression (6.4).

6.4.4 Fuller Two-Phase Jackknife Method

Isaki et al. (2004) apply the Fuller jackknife variance estimator (Fuller, 1998), labeled here as the ECF2 method, to account for variations in the benchmark controls. We demonstrate that the modified ECF2 (ECF2m) in Chapter 4, augmented to additionally account for the coverage error variance in the analytic survey sampling frame, has a lower relative bias than the original ECF2. In this section, we translate the ECF2m formulae for estimation of domain totals. As discussed previously, our domain estimators are functions of overall and not domain-specific analytic survey auxiliary variables and benchmark controls. Because the change to the ECF2m and the other jackknife methods only affects the analytic survey components, references to Chapter 4 text allow us to abbreviate this discussion without loss of clarity.

The delete-one ECF2m jackknife variance estimator for the EC-GREG domain total $\hat{t}_{ydR}$ defined in (6.1) requires the calculation of replicate estimates using the following formula:

$$\begin{aligned} \dddot{t}_{ydR(r)} &= \hat{t}_{ydGREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{Ad(r)} \\ &\quad + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \end{aligned} \qquad (6.28)$$

where $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; $R_h = \sqrt{1/Hm_{Ah}}$;

- $\hat{t}_{ydGREG(r)} = \hat{t}_{Ady(r)} + (\mathbf{t}_{Bx} - \hat{\mathbf{t}}_{Ax(r)})'\hat{\mathbf{B}}_{Ad(r)}$;
- $\hat{t}_{Ady(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \pi^{-1}_{hik} \delta_{dhik} y_{hik}$, the replicate total of the y in domain d;

- $\hat{\mathbf{t}}_{Ax(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik}$, the replicate totals for the auxiliary variables estimated from the analytic survey;

- $\hat{\mathbf{z}}_{B(r)} = \delta_{(r)} \sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}_{Bg}$, the ECF2 replicate control-total adjustment such that $\hat{\mathbf{V}}_B = \sum_{g=1}^{G} \hat{\mathbf{z}}_{Bg} \hat{\mathbf{z}}'_{Bg}$;

- $\hat{\mathbf{B}}_{Ad(r)} = \left[\sum_{hil \in s_A} \pi_{hi(r)}^{-1} \pi_{hil}^{-1} \mathbf{x}_{hil} \mathbf{x}'_{hil} \right]^{-1} \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik} \delta_{dhik} y_{hik}$, the model coefficient vector for domain d calculated for each analytic survey replicate;

- $\eta_{(r)}$ is the randomly generated value from a standard normal distribution for replicate r;

- $\hat{t}_{Aed2(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} e_{Adhik(r)}^2$ with $e_{Adhik(r)} = \delta_{dhik} y_{hik} - \mathbf{x}'_{hik} \hat{\mathbf{B}}_{Ad(r)}$; and,

- $\pi_{hi(r)}^{-1}$ is the PSU-subsampling weight for the r^{th} replicate defined in (4.27).

As shown in (4.28) for the ECF2, the second term in (6.28) can be approximated as

$$c_h \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_{Ad} \tag{6.29}$$

by assuming $\mathbf{t}_{Bx} = O(M)$ resulting in $\hat{t}_{yGREG(r)} = O_P(M)$, $\hat{\mathbf{B}}_{Ad(r)} = \mathbf{B}_{Ad} + O_P\left(m_{Ad}^{-1/2}\right)$ for the population domain parameter $\mathbf{B}_{Ad} = O(1)$ defined in (6.7), and $\hat{\mathbf{z}}'_{B(r)} = O_P\left(M/\sqrt{m_B}\right)$. Using the replicate estimator defined in (6.28) and the approximation in (6.29), the delete-one ECF2m jackknife variance estimator is expressed as:

$$
\begin{aligned}
var_{ECF2m}(\hat{t}_{ydR}) = \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{t}_{ydR(r)} - \hat{t}_{ydGREG}\right)^2 \\
\cong \;& \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG})^2 \\
&+ \; 2 \sum_{h} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG}\right) \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_{Ad} \\
&+ \; \mathbf{B}'_{Ad} \hat{\mathbf{V}}_B \mathbf{B}_{Ad} \\
&+ \; 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG}\right) R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \\
&+ \; 2 \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} \hat{\mathbf{z}}'_{B(r)} \mathbf{B}_{Ad} \left(R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \right) \\
&+ \; \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}. \qquad (6.30)
\end{aligned}
$$

To facilitate the evaluation of $E[var_{ECF2m}(\hat{t}_{ydR})]$, we divide (6.30) by M^2 and discuss each component in turn. The first variance component estimates the variation associated only with the analytic survey design and is $O_P(m_{Ad}^{-1})$. The second component is $O_P(m_B^{-1/2})$ under the assumption that max $\left\{M^{-1}\left(\hat{t}_{yGREG(r)} - \hat{t}_{ydGREG}\right)\right\}$ converges in probability to zero (Rao & Wu, 1985, see standard conditions in). The third component estimates the variation within the benchmark control totals and is $O_P(m_B^{-1})$ by assumption. The fourth component divided by M^2, as with the second component, converges in probability to zero and is $O_P(1/\sqrt{Mm_A})$ with $\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)} = O_P(M)$. The fifth component is $O_P(1/\sqrt{Mm_B})$ and has expectation zero by the inclusion of the standard normal random variable, $\eta_{(r)}$. The sixth and final term is $O_P(M^{-1})$ and estimates the coverage error variance component associated with the analytic survey sampling frame by noting $E\left(\eta^2_{(r)}\right) = 1$. Thus, $var_{ECF2m}\left(\hat{t}_{ydR}\right)$ is an approximately unbiased estimator of $AV\left(\hat{t}_{ydR}\right)$ provided that the sample estimators used in the replicate estimates are unbiased and the data are without (non-random) error. Note that by removing the coverage error term from $\dddot{t}_{ydR(r)}$ (6.28) that we are able to produce a sample variance estimator for domain totals under the original specification for the Fuller method denoted as ECF2 in our research.

The delete-one ECF2m jackknife variance estimate for EC-PSGR domain totals, a specific type of EC-GREG estimate, is calculated as

$$var_{ECF2m}(\hat{t}_{ydP}) = \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\dddot{t}_{ydP(r)} - \hat{t}_{ydPSGR}\right)^2 \tag{6.31}$$

where

$$\begin{aligned}\dddot{t}_{ydP(r)} &= \hat{t}_{ydPSGR(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\bar{\mathbf{Y}}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}} \\ &\cong \hat{t}_{ydPSGR(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \bar{\mathbf{Y}}_{Ad} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}. \end{aligned} \tag{6.32}$$

The approximation is justified by using the same assumptions as given for (6.29) with $\hat{t}_{ydPSGR(r)} = \sum_{g=1}^{G} N_{Bg} \hat{N}^{-1}_{Ag(r)} \hat{t}_{Aydg(r)}$; $\mathbf{N}'_B$ defined for expression (6.22); $\hat{N}_{Ag(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \delta_{ghik} \pi^{-1}_{hik}$; and $\hat{t}_{Aydg(r)} = \sum_{hik \in s_A} \pi^{-1}_{hi(r)} \delta_{ghik} \pi^{-1}_{hik} \delta_{dhik} y_{hik}$. The last component in (6.32),

$$c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\bar{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}, \tag{6.33}$$

produces replicate estimates of the analytic survey frame coverage error by poststratum where $\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' = \left[\left(1 - \hat{\phi}_{A1(r)}\right), \ ..., \ \left(1 - \hat{\phi}_{AG(r)}\right)\right]$, a G-length vector of estimated coverage rates within poststratum g; and $\hat{\mathbf{t}}_{Aed2(r)} = \left[\hat{t}_{Aed21(r)}, ..., \hat{t}_{Aed2G(r)}\right]'$ with components $\hat{t}_{Aed2g(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \delta_{ghik} e_{Adhik(r)}^2$ and $e_{Adhik(r)} = (\delta_{dhik} y_{hik} - \hat{\bar{y}}_{Adg(r)})$ with $\hat{\bar{y}}_{Adg} = \hat{t}_{Aydg(r)} / \hat{N}_{Ag(r)}$. Provided that the benchmark survey covers the population under study, the coverage rates $\hat{\phi}_{Ag(r)}$ can be estimated as $\hat{\phi}_{Ag(r)} = \hat{N}_{Ag(r)} / \hat{N}_{Bg}$. The remaining terms are defined for $\ddot{t}_{ydR(r)}$ below expression (6.28). The evaluation of the variance components follows the discussion given for $var_{ECF2m}(\hat{t}_{ydR})$ in (6.30).

The seven-step process used to calculate $var_{ECF2m}(\hat{t}_{yP})$ given in expression (4.36) is given at the end of Section 4.4.3. By replacing the outcome variable y_{hik} with a domain-specific outcome variable $y_{dhik} = \delta_{dhik} y_{hik}$, we are able to use these same steps to create estimates for $var_{ECF2m}(\hat{t}_{ydP})$ in (6.31).

6.4.5 Multivariate Normal Jackknife Method

The ECMV method (ECMV) incorporates the random value from a multivariate normal distribution with mean equal to a G-length vector of zeroes ($\mathbf{0}_G$) and covariance equal to the estimated covariance matrix for the benchmark control totals ($\hat{\mathbf{V}}_B$). The ECMV jackknife sample variance estimator for domain totals is derived in a similar manner as shown for the ECF2m in (6.30):

$$
\begin{aligned}
var_{ECMV}(\hat{t}_{ydR}) &= \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} (\dddot{t}_{ydR(r)} - \hat{t}_{ydGREG})^2 \\
&\cong \sum_{h=1}^{H} c_h^{-2} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG}\right)^2 \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} R_h \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG}\right) \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_{Ad} \\
&+ \mathbf{B}'_{Ad} \left[\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \right] \mathbf{B}_{Ad} \\
&+ 2 \sum_{h=1}^{H} c_h^{-1} \sum_{r=1}^{m_{Ah}} \left(\hat{t}_{ydGREG(r)} - \hat{t}_{ydGREG}\right) R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \\
&+ 2 \sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} \hat{\boldsymbol{\varepsilon}}'_{B(r)} \mathbf{B}_{Ad} \; \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \\
&+ \sum_{h=1}^{H} \sum_{r=1}^{m_{Ah}} R_h^2 \eta_{(r)}^2 \left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)} \qquad (6.34)
\end{aligned}
$$

where $\hat{\boldsymbol{\varepsilon}}_{B(r)}$ is a G-length vector of random variables from the specified multivariate normal distribution, i.e., $\hat{\boldsymbol{\varepsilon}}_{B(r)} \overset{\text{iid}}{\sim} \text{MVN}_G(\mathbf{0}_G, \hat{\mathbf{V}}_B)$; and

$$
\begin{aligned}
\dddot{t}_{ydR(r)} &= \hat{t}_{ydGREG(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\mathbf{B}}_{Ad(r)} \\
&\quad + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}, \qquad (6.35)
\end{aligned}
$$

the ECMV replicate estimator. The approximations to the replicate estimator used in (6.34) are obtained as with the ECF2m in (6.29), by assuming $\hat{\mathbf{B}}_{Ad(r)} = \mathbf{B}_{Ad} + O_P\left(m_{Ad}^{-1/2}\right)$ and eliminating the lower-order term. The expectation of $var_{ECMV}(\hat{t}_{ydR})$ mirrors the discussion given for $var_{ECF2m}(\hat{t}_{ydR})$ following expression (6.30). Note that the third variance component has expectation $\mathbf{B}'_{Ad}\mathbf{V}_B\mathbf{B}_{Ad}$ using the work demonstrated in expression (4.40). Thus, in expectation, $var_{ECMV}\left(\hat{t}_{ydR}\right) \cong AV\left(\hat{t}_{ydR}\right)$ given that the component estimators are approximately unbiased. Additionally, as with the overall estimated total, the ECF2m and ECMV methods are asymptotically equivalent.

The replicate estimator in (6.35) is specialized for EC poststratification to provide details for our simulation study presented in Section 6.6. The ECMV replicates estimates for an EC-PSGR domain total are calculated with the following formula:

$$\begin{aligned} \dddot{t}_{ydP(r)} &= \hat{t}_{ydPSGR(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{B(r)} \hat{\hat{\mathbf{Y}}}_{Ad(r)} \\ &\quad + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}} \end{aligned} \tag{6.36}$$

with $\hat{\boldsymbol{\varepsilon}}_{B(r)}$ defined for $var_{ECMV}(\hat{t}_{ydR})$ in expression (6.34) and the remaining terms are the same as defined for the ECF2m replicate estimates in (6.33). The $\dddot{t}_{ydP(r)}$ estimates are substituted in the jackknife variance formula given in (6.31) to calculate $var_{ECMV}(\hat{t}_{ydP})$.

6.4.6 Nadimpalli-Judkins-Chu Jackknife Method

The jackknife variance estimator developed by Nadimpalli et al. (2004) is similar to the ECMV method developed for our research. However, their method assumes that only the variance estimates for the benchmark controls are available, i.e., $diag\left(\hat{\mathbf{V}}_B\right)$. As discussed in Section 4.5.4, the lack of information on the benchmark controls can result in variance estimates that are too small. The same holds for domain estimation as well.

The replicate estimates for the modified ECNJC (ECNJCm) method account for the undercoverage error in the analytic survey and are defined as

$$\begin{aligned}\dddot{t}_{ydR(r)} &= \hat{t}_{ydGREG(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{Ad(r)} \\ &\quad + c_h R_h \eta_{(r)} \sqrt{\left(1 - \hat{\hat{\phi}}_{A(r)}\right) \hat{t}_{Aed2(r)}} \end{aligned} \tag{6.37}$$

for an EC-GREG estimator of a domain total, and as

$$\begin{aligned}\dddot{t}_{ydP(r)} &= \hat{t}_{ydPSGR(r)} + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\hat{\mathbf{Y}}}_{Ad(r)} \\ &\quad + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}} \end{aligned} \tag{6.38}$$

for the corresponding EC-PSGR estimator. The term $\hat{\mathbf{S}}_B = diag\left(\sqrt{\hat{\mathbf{V}}_B}\right)$, and $\boldsymbol{\eta}_{(r)}$ is a G-length vector of standard normal random values independently generated for each replicate.

The expression for $var_{ECNJCm}(\hat{t}_{ydR})$ is obtained by substituting the replicate estimates (6.37) into the EC-GREG jackknife variance formula shown for the ECMV in expression (6.34). The expectation of the components of the ECNJCm variance estimator follow the discussion given for the overall total subsequent to expression (4.48). Namely, the component associated with the variability only in the benchmark estimates, $\mathbf{B}'_{Ad}\left(\sum_{h=1}^{H} R_h^2 \sum_{r=1}^{m_{Ah}} E\left[\hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B\right]\right)\mathbf{B}_{Ad}$, is not in general an unbiased estimator of $\mathbf{B}'_{Ad}\mathbf{V}_B\mathbf{B}_{Ad}$. This estimator is either negatively or positively biased depending on the sign of the covariance terms within $\mathbf{V}_B$, the population sampling covariance matrix for the benchmark estimates.

6.5 Variance Estimation for Domain Means

The bias of the EC-GREG ratio-mean within a domain is a function of the bias in both the numerator and denominator. The same holds true for the other component within the MSE, i.e., the variance. We begin this section by defining the approximate population sampling variance for the domain ratio-mean. A theoretical evaluation of the five sample variance estimator under study is provided in the subsequent sections.

6.5.1 Population Sampling Variance

We approximate the population sampling variance of $\hat{\bar{y}}_{dR} = \hat{t}_{ydR}/\hat{N}_{dR}$ (6.10), the ratio-mean estimator within domain d, through a first-order Taylor linearization about the components of $\bar{y}_{dR} = t_{ydR}/N_{dR}$. The population parameter t_{ydR} is defined for $\hat{t}_{ydR}$ in (6.5), and N_{dR} is defined for $Bias\left(\hat{N}_{dR}\right)$ in (6.16). The approximate population sampling variance is expressed as

$$AV(\hat{\bar{y}}_{dR}) = \left(\frac{1}{N_{dR}}\right)^2 \left[AV(\hat{t}_{ydR}) + \bar{y}_{dR}^2 AV(\hat{N}_{dR}) - 2\bar{y}_{dR} ACov(\hat{t}_{ydR}, \hat{N}_{dR})\right] \quad (6.39)$$

with $AV(\hat{t}_{ydR})$ specified in expression (6.20). The approximate population sampling variance for the denominator of $\hat{\bar{y}}_{dR}$ is similarly defined as:

$$\begin{aligned} AV(\hat{N}_{dR}) &= AV(\hat{N}_{dGREG}) \\ &+ \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\left(1-\phi_{Ahik}\right) E_{ANdhik}^{2} \\ &+ \mathbf{B}_{ANd}'\mathbf{V}_B\mathbf{B}_{ANd} \end{aligned} \tag{6.40}$$

where $\mathbf{B}_{ANd}$ is given for (6.16) and $E_{ANdhik} = \delta_{dhik} - \mathbf{x}_{hik}'\mathbf{B}_{ANd}$. The first variance component $AV(\hat{N}_{dGREG})$, a traditional calibration variance estimator, is calculated as shown for $AV(\hat{t}_{yGREG})$ in expression (4.14) by substituting E_{Ahik} with E_{ANdhik}. The remaining term in (6.39) follows the development of $ACov(\hat{t}_{yR},\ \hat{N}_R)$ given in expression (5.17) and equals

$$\begin{aligned} ACov\left(\hat{t}_{ydR}, \hat{N}_{dR}\right) &= ACov\left(\hat{t}_{ydGREG}, \hat{N}_{dGREG}\right) \\ &+ \sum_{h=1}^{H}\sum_{i=1}^{M_h}\sum_{k=1}^{N_{hi}} \phi_{Ahik}\left(1-\phi_{Ahik}\right) E_{Adhik}E_{ANdhik} \\ &+ \mathbf{B}_{Ad}'\mathbf{V}_B\mathbf{B}_{ANd}. \end{aligned} \tag{6.41}$$

Substituting the expressions into (6.39), we have

$$\begin{aligned} AV\left(\hat{\bar{y}}_{dR}\right) &= \left(\frac{1}{N_{dR}}\right)^2 \Big(AV(\hat{t}_{ydGREG}) + \bar{y}_{dR}^2 AV(\hat{N}_{dGREG}) \\ &\quad - 2\bar{y}_{dR} ACov(\hat{t}_{ydGREG}, \hat{N}_{dGREG})\Big) \\ &+ \left(\frac{1}{N_{dR}}\right)^2 \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right)\left(E_{Adhik} - \bar{y}_{dR} E_{ANdhik}\right)^2 \\ &+ \left(\frac{1}{N_{dR}}\right)^2 \left(\mathbf{B}_{Ad} - \bar{y}_{dR}\mathbf{B}_{ANd}\right)' \mathbf{V}_B \left(\mathbf{B}_{Ad} - \bar{y}_{dR}\mathbf{B}_{ANd}\right) \end{aligned} \tag{6.42}$$

where the first term is equivalent to $AV\left(\hat{\bar{y}}_{dGREG}\right)$.

The approximate population sampling variance of $\hat{\bar{y}}_{dP}$ is defined as follows by substituting the appropriate values within expression (6.42):

$$\begin{aligned} AV\left(\hat{\bar{y}}_{dP}\right) &= \left(\frac{1}{N_{dP}}\right)^2 \Big(AV\left(\hat{t}_{ydPSGR}\right) + \bar{y}_{dP}^2 AV(\hat{N}_{dPSGR}) \\ &\quad - 2\bar{y}_{dP} ACov(\hat{t}_{ydPSGR}, \hat{N}_{dPSGR})\Big) \\ &+ \left(\frac{1}{N_{dP}}\right)^2 \sum_{hik\in U} \phi_{Ahik}\left(1-\phi_{Ahik}\right)\left(E_{Adhik} - \bar{y}_{dP} E_{ANdhik}\right)^2 \\ &+ \left(\frac{1}{N_{dP}}\right)^2 \left(\bar{\mathbf{Y}}_{Ad} - \bar{y}_{dP}\bar{\mathbf{Y}}_{ANd}\right)' \mathbf{V}_B \left(\bar{\mathbf{Y}}_{Ad} - \bar{y}_{dP}\bar{\mathbf{Y}}_{ANd}\right) \end{aligned} \tag{6.43}$$

where $E_{Adhik} = \delta_{dhik} - \mathbf{x}'_{hik}\mathbf{B}_{ANd}$ and $E_{ANdhik} = \delta_{dhik} - \bar{d}_{Ag}$ with $\bar{d}_{Ag} = N_{Adg}/N_{Ag}$, the proportion of domain members in poststratum g within the population defined by the analytic survey sampling frame. The first term in (6.43) equals $AV\left(\hat{\bar{y}}_{dPSGR}\right)$ with the following components: (i) $AV\left(\hat{t}_{ydPSGR}\right)$ given in expression (6.22); (ii)

$$AV\left(\hat{N}_{dPSGR}\right) \equiv \mathbf{N}'_B E_{c_A}\left(\mathbf{V}_{ANd}\right)\mathbf{N}_B \tag{6.44}$$

where $\mathbf{V}_{ANd} = Var_A(\hat{\bar{\mathbf{Y}}}_{\mathbf{ANd}}) \cong \mathbf{D_d}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}\mathbf{d}}\mathbf{D'_d}$ with

$$\mathbf{D}_d = \left[diag\left(\left\{ \frac{1}{N_{Ag}} \right\}_{g=1}^{G} \right), diag\left(\left\{ \frac{-N_{Adg}}{N_{Ag}^2} \right\}_{g=1}^{G} \right) \right]$$

and

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \sigma_{(\hat{N}_{Ad1},\hat{N}_{Ad1})} & \cdots & \sigma_{(\hat{N}_{Ad1},\hat{N}_{AG})} \\ \vdots & \ddots & \vdots \\ \sigma_{(\hat{N}_{AG},\hat{N}_{Ad1})} & \cdots & \sigma_{(\hat{N}_{AG},\hat{N}_{AG})} \end{bmatrix};$$

and, (iii) $ACov\left(\hat{t}_{ydPSGR}, \hat{N}_{dPSGR}\right) = \mathbf{N}'_B ACov\left(\bar{\mathbf{Y}}_{Ad}, \bar{\mathbf{Y}}_{ANd}\right)\mathbf{N}'_B \cong \mathbf{N}'_B\mathbf{D}_d\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}d}\mathbf{D}'_d\mathbf{N}_B$ where

$$\begin{aligned} \mathbf{D}_d &= \left[diag\left(\left\{ \frac{1}{N_{Ag}} \right\}_{g=1}^{G} \right),\ diag\left(\left\{ \frac{-t_{Aydg}}{N_{Ag}^2} \right\}_{g=1}^{G} \right), \right. \\ &\quad \left. diag\left(\left\{ \frac{1}{N_{Ag}} \right\}_{g=1}^{G} \right),\ diag\left(\left\{ \frac{-N_{Adg}}{N_{Ag}^2} \right\}_{g=1}^{G} \right) \right], \end{aligned}$$

a $G \times 4G$ matrix of first-order derivatives, and $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}d}$ is a $4G \times 4G$ matrix of population sampling covariances for each pair of matrices within $\left(\bar{\mathbf{Y}}_{Ad}, \bar{\mathbf{Y}}_{ANd}\right)$, i.e., $\hat{\mathbf{t}}_{Ayd}$, $\hat{\mathbf{N}}_{Ad}$, and $\hat{\mathbf{N}}_A$.

6.5.2 Traditional Calibration Variance

The linearization sample variance estimator for $\hat{\bar{y}}_{dR}$ (6.10) is developed under the assumption that the benchmark controls are fixed population values. This naïve variance estimator is also assumed in expectation to be a reasonable approximation to $AV\left(\hat{\bar{y}}_{dR}\right)$ and is calculated as:

$$
\begin{aligned}
var_{Na\ddot{\imath}ve}(\hat{\bar{y}}_{dR}) &\equiv var(\hat{\bar{y}}_{dGREG}) \\
&= \left(\frac{1}{\hat{N}_{dR}}\right)^2 \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{dhi+} - \bar{\breve{u}}_{dh++})^2 \qquad (6.45)
\end{aligned}
$$

for $\hat{\bar{y}}_{dGREG}$ defined in expression (2.29); $\breve{u}_{dhi+} = \sum_{k \in s_{Ahi}} a_{hik} \pi_{hik}^{-1} \left(e_{Adhik} - \hat{\bar{y}}_{dR} e_{ANdhik}\right)$ with $e_{Adhik} = \delta_{dhik} y_{hik} - \mathbf{x}'_{hik} \hat{\mathbf{B}}_{Ad}$ and $e_{ANdhik} = \delta_{dhik} - \mathbf{x}'_{hik} \hat{\mathbf{B}}_{ANd}$; and $\bar{\breve{u}}_{dh++} = m_{Ah}^{-1} \sum_{i \in s_{Ah}} \breve{u}_{dhi+}$. The EC-PSGR version of (6.45) is defined as

$$
\begin{aligned}
var_{Na\ddot{\imath}ve}(\hat{\bar{y}}_{dP}) &\equiv var(\hat{\bar{y}}_{dPSGR}) \\
&= \left(\frac{1}{\hat{N}_{dP}}\right)^2 \sum_{h=1}^{H} \frac{m_{Ah}}{m_{Ah}-1} \sum_{i=1}^{m_{Ah}} (\breve{u}_{dhi+} - \bar{\breve{u}}_{dh++})^2 \qquad (6.46)
\end{aligned}
$$

where $\breve{u}_{dhi+} = \sum_{k \in s_{Ahi}} a_{hik} \pi_{hik}^{-1} \left(e_{Adhik} - \hat{\bar{y}}_{dP} e_{ANdhik}\right)$ with $e_{Adhik} = \delta_{dhik} y_{hik} - \hat{\bar{y}}_{Adg}$ and $e_{ANdhik} = \delta_{dhik} - \hat{\bar{d}}_{Ag}$ with $\hat{\bar{d}}_{Ag} = \hat{N}_{Adg} / \hat{N}_{Ag}$. The discussion given in previous sections about the traditional variance estimator also applies here. Namely, this estimator is negatively biased for $AV\left(\hat{\bar{y}}_{dR}\right)$ due to the missing benchmark and analytic survey frame coverage error components.

6.5.3 Estimated-Control Linearization Variance

The EC sample linearization variance estimator is developed by adding components to the naïve estimator, $var(\hat{\bar{y}}_{dGREG})$, given in expression (6.45). The sample variance estimator for the EC-GREG estimator of a ratio mean within domain d is expressed as:

$$\begin{aligned} var_{ECTS}(\hat{\hat{y}}_{dR}) &= var(\hat{\hat{y}}_{dGREG}) \\ &+ \left(\frac{1}{\hat{N}_{dR}}\right)^2 \sum_{h=1}^{H} \left(1-\hat{\hat{\phi}}_{Ah}\right) \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} \left(e_{Adhik} - \hat{\hat{y}}_{dR} e_{ANdhik}\right)^2 \\ &+ \left(\frac{1}{\hat{N}_{dR}}\right)^2 \left(\hat{\mathbf{B}}_{Ad} - \hat{\hat{y}}_{dR}\hat{\mathbf{B}}_{ANd}\right)' \hat{\mathbf{V}}_B \left(\hat{\mathbf{B}}_{Ad} - \hat{\hat{y}}_{dR}\hat{\mathbf{B}}_{ANd}\right) \end{aligned} \tag{6.47}$$

where the terms are defined for $\hat{\hat{y}}_{dR}$ (6.10) and following (6.45). The corresponding EC-PSGR sample variance estimator, used in the simulation study (Section 6.6), is defined as:

$$\begin{aligned} var_{ECTS}(\hat{\hat{y}}_{dP}) &= var(\hat{\hat{y}}_{dPSGR}) \\ &+ \left(\frac{1}{\hat{N}_{dP}}\right)^2 \sum_{h=1}^{H} \left(1-\hat{\hat{\phi}}_{Ah}\right) \sum_{i=1}^{m_{Ah}} \sum_{k=1}^{n_{Ahi}} \pi_{hik}^{-1} \left(e_{Adhik} - \hat{\hat{y}}_{dP} e_{ANdhik}\right)^2 \\ &+ \left(\frac{1}{\hat{N}_{dP}}\right)^2 \left(\hat{\hat{\mathbf{Y}}}_{Ad} - \hat{\hat{y}}_{dP}\hat{\hat{\mathbf{Y}}}_{ANd}\right)' \hat{\mathbf{V}}_B \left(\hat{\hat{\mathbf{Y}}}_{Ad} - \hat{\hat{y}}_{dP}\hat{\hat{\mathbf{Y}}}_{ANd}\right) \end{aligned} \tag{6.48}$$

with residuals defined for expression (6.46).

6.5.4 Fuller Two-Phase Jackknife Method

The modified delete-one Fuller (ECF2m) jackknife variance estimator, as well as the other jackknife methods discussed in the subsequent sections, use the following general formula to calculate the sample estimates for an EC-GREG domain ratio-mean estimator:

$$var(\hat{\bar{y}}_{dR}) = \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{\bar{y}}_{dR(r)} - \hat{\bar{y}}_{dGREG})^2. \tag{6.49}$$

The EC-PSGR version is similarly defined as

$$var(\hat{\bar{y}}_{dP}) = \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{\bar{y}}_{dP(r)} - \hat{\bar{y}}_{dPSGR})^2. \tag{6.50}$$

Each method requires the calculation of replicate estimates using a different approach. The ECF2m replicate estimates for the EC-GREG domain ratio-mean are calculated as

$$\begin{aligned}\dddot{\bar{y}}_{dR(r)} &= \frac{\dddot{t}_{ydR(r)}}{\dddot{N}_{dR(r)}} \\ &= \frac{\hat{t}_{ydGREG(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)}) \hat{t}_{Aed2(r)}}}{\hat{N}_{dGREG(r)} + \left(c_h \hat{\mathbf{z}}'_{B(r)} \hat{\mathbf{B}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)}) \hat{t}_{ANed2(r)}} \right)}\end{aligned} \tag{6.51}$$

where $\hat{\mathbf{B}}_{ANd(r)} = \left[\sum_{hil \in s_A} \pi_{hi(r)}^{-1} \pi_{hil}^{-1} \mathbf{x}_{hil} \mathbf{x}'_{hil}\right]^{-1} \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} \mathbf{x}_{hik} \ \delta_{dhik}$, and $\hat{t}_{ANed2(r)} = \sum_{hik \in s_A} \pi_{hi(r)}^{-1} \pi_{hik}^{-1} e_{ANdhik(r)}^2$ with $e_{ANdhik(r)} = \delta_{dhik} - \mathbf{x}'_{hik} \hat{\mathbf{B}}_{ANd(r)}$. The remaining terms are defined for expression (6.28). By substituting (6.51) into (6.49), we obtain an explicit expression for $var_{ECF2m}(\hat{\bar{y}}_{dR})$. The approximation techniques shown for $var_{ECF2m}(\hat{\bar{y}}_R)$, beginning with a geometric approximation of the ratio-mean in expression (5.37), and is used here to demonstrate that $var_{ECF2m}(\hat{\bar{y}}_{dR})$ is asymptotically equivalent to $AV\left(\hat{\bar{y}}_{dR}\right)$ given in (6.42) provided that the sample estimates used in (6.51) are (approximately) unbiased. The rates of convergence for the variance components in $var_{ECF2m}(\hat{\hat{y}}_{dR})$ mirror the discussion given for $var_{ECF2m}(\hat{\hat{y}}_R)$ following (4.32) after replacing the number of sample PSUs (m_A) with the number of domain sample PSUs (m_{Ad}).

By substituting the following ECF2m replicate estimates into the EC-PSGR jackknife sample variance estimator in (6.50), we are able to calculate $var_{ECF2m}(\hat{\hat{y}}_{dP})$:

$$\begin{aligned}
\dddot{y}_{dP(r)} &= \frac{\dddot{t}_{ydP(r)}}{\dddot{N}_{dP(r)}} \qquad (6.52)\\
&= \frac{\hat{t}_{ydPSGR(r)} + c_h \hat{\mathbf{z}}'_{B(r)} \hat{\hat{\mathbf{Y}}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}}{\hat{N}_{dPSGR(r)} + \left(c_h \hat{\mathbf{z}}'_{B(r)} \hat{\hat{\mathbf{Y}}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{ANed2(r)}}\right)}.
\end{aligned}$$

6.5.5 Multivariate Normal Jackknife Method

The multivariate normal method (ECMV) introduces a multivariate normal random variable into the numerator and denominator of $\hat{\hat{y}}_{dR}$ generated for each jackknife replicate. The ECMV replicate estimates for an EC-GREG and EC-PSGR domain ratio-mean estimator are calculated using the following formulae, respectively:

$$\begin{aligned}
\dddot{y}_{dR(r)} &= \frac{\dddot{t}_{ydR(r)}}{\dddot{N}_{dR(r)}} \qquad (6.53)\\
&= \frac{\hat{t}_{ydGREG(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\mathbf{B}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{(1 - \hat{\phi}_{A(r)}) \hat{t}_{Aed2(r)}}}{\hat{N}_{dGREG(r)} + \left(c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\mathbf{B}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{(1 - \hat{\phi}_{A(r)}) \hat{t}_{ANed2(r)}}\right)},
\end{aligned}$$

and

$$\dddot{y}_{dP(r)} = \frac{\dddot{t}_{ydP(r)}}{\dddot{N}_{dP(r)}} \tag{6.54}$$

$$= \frac{\hat{t}_{ydPSGR(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\hat{\mathbf{Y}}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}}{\hat{N}_{dPSGR(r)} + \left(c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\hat{\mathbf{Y}}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{ANed2(r)}}\right)}.$$

The replicate estimates are substituted into (6.49) and (6.50), respectively, to derive $var_{ECMV}(\hat{\hat{y}}_{dR})$ and $var_{ECMV}(\hat{\hat{y}}_{dP})$. The asymptotic evaluation of $var_{ECMV}(\hat{\hat{y}}_{dR})$ provided in Section 5.4.4 also holds for domain estimation after substituting m_{Ad} with m_A and is not repeated here. Therefore, the ECF2m and ECMV jackknife variance estimators for the ratio-mean estimators are asymptotically equivalent and both are approximately unbiased for $AV(\hat{\hat{y}}_{dR})$.

6.5.6 Nadimpalli-Judkins-Chu Jackknife Method

Based on the results from Section 5.4.5, we know that the ECNJCm method, a simplification of the ECMV, can underestimate the variance of estimated totals and, to a lesser degree, the variance of estimated ratio-means. A theoretical evaluation for the domain ratio-mean also suggest a biased variance estimator. The ECNJCm replicate estimates for an EC-GREG and EC-PSGR domain ratio-mean estimator are calculated using the following formulae, respectively:

$$\dddot{y}_{dR(r)} = \frac{\dddot{t}_{ydR(r)}}{\dddot{N}_{dR(r)}} \tag{6.55}$$

$$= \frac{\hat{t}_{ydGREG(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\mathbf{B}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{Aed2(r)}}}{\hat{N}_{dGREG(r)} + \left(c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{(1-\hat{\phi}_{A(r)})\hat{t}_{ANed2(r)}} \right)},$$

and

$$\dddot{y}_{dP(r)} = \frac{\dddot{t}_{ydP(r)}}{\dddot{N}_{dP(r)}} \tag{6.56}$$

$$= \frac{\hat{t}_{ydPSGR(r)} + c_h R_h \hat{\boldsymbol{\varepsilon}}_{B(r)} \hat{\hat{\mathbf{Y}}}_{Ad(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{Aed2(r)}}}{\hat{N}_{dPSGR(r)} + \left(c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\hat{\mathbf{Y}}}_{ANd(r)} + c_h R_h \eta_{(r)} \sqrt{\left(\mathbf{1}_G - \hat{\hat{\boldsymbol{\phi}}}_{A(r)}\right)' \hat{\mathbf{t}}_{ANed2(r)}} \right)}.$$

The replicate estimates are substituted into (6.49) and (6.50), respectively, to derive $var_{ECNJCm}(\hat{\bar{y}}_{dR})$ and $var_{ECNJCm}(\hat{\bar{y}}_{dP})$. The asymptotic evaluation provided in Section 5.4.5 also holds for domain estimation indicating that this variance estimator will have higher levels of relative bias than the other jackknife methods studied in our research. Whether the ECNJCm variance estimator over- or underestimates the true population sampling variance depends on the sign of the off-diagonal terms in $\mathbf{V}_B$.

6.6 Simulation Study

The simulation study described in detail in Section 4.5 is used to confirm the theoretical evaluation presented in the previous sections. We compare the empirical properties of five variance estimators for (i) an EC-PSGR estimator of a total within a domain, $\hat{t}_{ydP}$ (6.4), and (ii) the ratio of two EC-PSGR totals within a domain $\hat{\bar{y}}_{dP}$ (6.13). The following abbreviations are used as labels for the variance estimators:

- *Naïve*, the traditional calibration estimator defined in (6.24) for totals, and in (6.46) for ratio-means;
- *ECTS*, the EC linearization estimator defined in (6.27) for totals, and in (6.48) for ratio-means;
- *ECF2m*, the modified Fuller two-phase jackknife estimator defined in (6.31) for totals, and in (6.50) with replicate estimates (6.52) for ratio-means;
- *ECMV*, the Multivariate normal jackknife estimator defined for totals with replicate estimates (6.36) substituted in the variance formula (6.31), and for ratio-means with the replicate estimates (6.54) substituted in (6.50);
- *ECNJCm*, the modified Nadimpalli-Judkins-Chu jackknife estimator defined in (6.31) with replicate estimates (6.38) for totals, and for ECF2m in (6.50) with replicate estimates (6.56) for ratio-means.

We additionally compare these results with those presented in Sections 4.5 and 5.5. Based on the positive results for the modified ECF2 and ECNJC methods from the previous chapters, we forgo a discussion of the original methods for domain estimation.

6.6.1 Simulation Parameters

Results from a simulation study are used to examine the empirical properties for the domain estimators discussed in this chapter. We select 4,000 (analytic survey) simulation samples using a stratified, multi-stage design from an incomplete frame generated from the 2003 NHIS. The analytic survey sample size and the effective size of the benchmark survey are varied to examine the affects of differential influences on the overall variance. Additional details on the basic set-up of the simulation study are provided in Section 4.5.1 and are not repeated here.

We calculate the estimated population totals and means within a domain, as well as the variance estimates, for two NHIS binary variables: NOTCOV=1 indicates that an adult *did not* have health insurance coverage in the 12 months prior to the NHIS interview ($\bar{y} \cong 0.17$); and PDMED12M=1 indicates that an adult *delayed* medical care because of cost in the 12 prior to the interview ($\bar{y} \cong 0.07$). Total and mean estimates are calculated for records with NHIS variable HISCODI2=1 to create a *Hispanic ethnicity* domain for this study. Approximately 23 percent of the U.S. residents in our target population (records on the NHIS data file) are self-classified as Hispanic. Within this domain, 35.4 percent did not have health insurance (NOTCOV=1) and 7.0 percent delayed medical care (PDMED12M=1) in the 12 months prior to the interview. Simulation programs were developed and run in R® (Lumley, 2005; R Development Core Team, 2005) for this empirical study. The primary programs are included as Appendix A.

6.6.2 Evaluation Criteria

The empirical results for the variance estimators listed in Section 6.6 are compared using four measures across the 4,000 simulation samples and two outcome variables (*NOTCOV* and *PDMED12M*) within the Hispanic (d) domain. In the following list of empirical measures, $\hat{\theta}_{dP} = \left(\hat{t}_{dP}, \hat{\bar{y}}_{dP}\right)$ generate expressions for the estimated totals and ratio-means, respectively. The corresponding population parameters are denoted as $\theta_d = (t_d, \bar{y}_d)$ and are calculated from the 2003 NHIS population of size $N = 21,664$. The measures include:

1. $100 \times \left[\left(\frac{1}{4000}\sum_j var(\hat{\theta}_{dP_j}) - MSE_d\right)/MSE_d\right]$, the estimated percent bias of the variance estimator relative to the empirical $MSE_d = \frac{1}{4000}\sum_j(\hat{\theta}_{dP_j} - \theta_d)^2$;

2. $\frac{1}{4000}\sum_j I\left(|\hat{z}_j| \leq z_{1-\alpha/2}\right)$, the 95 percent confidence interval coverage rate where $\alpha = 0.05$, $\hat{z}_j = (\hat{\theta}_{dP_j} - \theta_d)/se(\hat{\theta}_{dP_j})$, and $se(\hat{\theta}_{dP_j}) = \sqrt{var(\hat{\theta}_{dP_j})}$;

3. $\sqrt{\frac{1}{(4000-1)}\sum_j\left[se(\hat{\theta}_{dP_j}) - \frac{1}{4000}\sum_j se(\hat{\theta}_{dP_j})\right]^2}$, the standard deviation of the estimated standard errors (se); and,

4. $100\times\left[\left(\frac{1}{4000}\sum_j se_*\left(\hat{\theta}_{dP_j}\right) - \frac{1}{4000}\sum_j se_{ECTS}\left(\hat{\theta}_{dP_j}\right)\right)/\frac{1}{4000}\sum_j se_{ECTS}\left(\hat{\theta}_{dP_j}\right)\right]$, the percent increase in the variation of the estimated standard errors for all studied estimators (se_*) relative to the ECTS variance estimator (se_{ECTS}).

We initially evaluate the relative bias of the point estimators, $\frac{1}{4000}\sum_j\left(\hat{\theta}_{dP_j} - \theta_d\right)/\theta_d$, to justify the use of estimated-control weight calibration. These criteria are also used to compare the results for the overall estimates given in Sections 4.5.4, 5.5.3, and 5.5.4.

Table 6.1: Percent Relative Bias Averaged Across Samples and Benchmark Covariance Matrices for Totals and Percents of Total Outcome within the Hispanic Domain by Point Estimator

Estimator	Not Covered by Health Insurance (NOTCOV) $n_{Ahi}=20$	Not Covered by Health Insurance (NOTCOV) $n_{Ahi}=40$	Delayed Medical Care (PDMED12M) $n_{Ahi}=20$	Delayed Medical Care (PDMED12M) $n_{Ahi}=40$
$\hat{t}_{ydPWR}$	-37.5	-37.5	-38.7	-38.4
$\hat{t}_{ydP}$	1.1	1.2	-0.2	0.3
$\hat{\bar{y}}_{dHJ}$	-9.1	-9.2	-7.7	-7.3
$\hat{\bar{y}}_{dP}$	1.3	1.3	1.1	1.6

PWR = p-expanded with-replacement estimator, HJ = Hájek estimator.

Note that the estimated percent bias of the variance estimators relative to the empirical variance (see measure 2 in Section 4.5.2) were also examined. However, these results are not presented in this chapter due to the similarities with the discussions given previously.

6.6.3 Results for Point Estimators

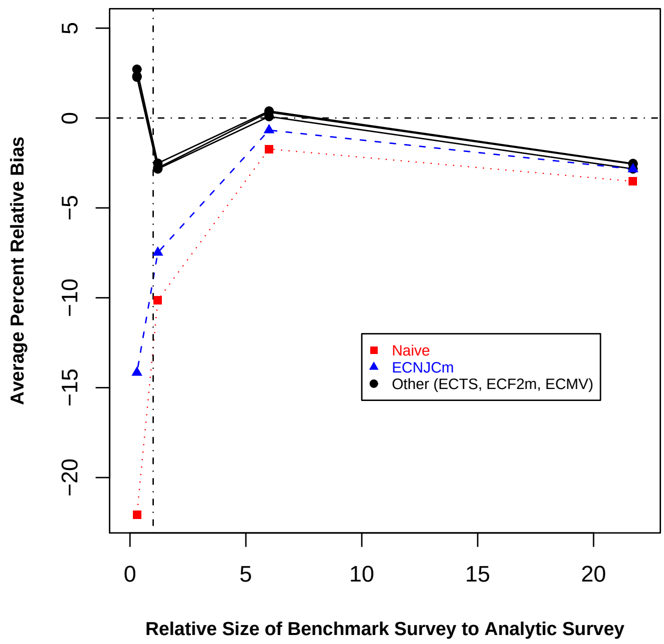
Data from a particular sample survey may have errors that negatively affect the estimates using an otherwise unbiased estimator. The estimators included in Table 6.1 are all (approximately) unbiased and should produce percent relative biases for the domain estimates near zero. Because we introduce undercoverage error in the analytic survey sampling frame, the uncalibrated point estimators $\hat{t}_{ydPWR} = \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{dhik} y_{hik}$ and $\hat{\bar{y}}_{dHJ} = \hat{t}_{ydPWR} / \sum_{hik \in s_A} \pi_{hik}^{-1} \delta_{dhik}$ are all negatively biased. The NOTCOV and PDMED12M estimates within the Hispanic domain show negative biases in excess of 37 percent. The corresponding biases for the domain ratio-means are much lower but still underestimate the population means by as much as 7 percent. Calibrating the design weights to the set of estimated benchmark control totals improves the negative biases dramatically. The percent relative biases for the estimated totals within our domain, $\hat{t}_{ydP}$, are either close to zero or no more than a 1.2 percent overestimate. The levels for the domain ratio-mean, $\hat{\bar{y}}_{dP}$, are comparable and exceed the population means by less than 2 percent. Therefore, with the levels of undercoverage introduced in our simulation study, the EC calibration procedure was a benefit. Note that the percent relative biases presented here correspond with the overall estimates given in Table 5.1.

6.6.4 Comparison of Variance Estimators for Estimated Totals

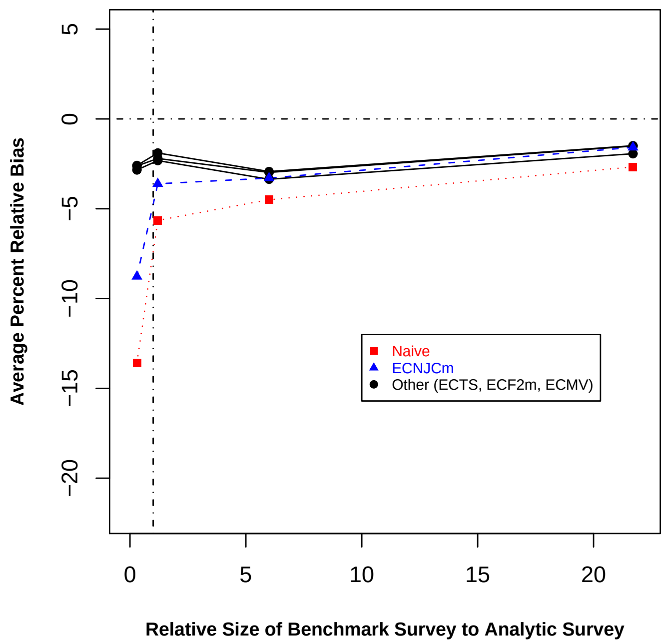
Empirical analyses of values from unbiased variance estimators should result in percent *biases relative to the empirical MSE* at or near zero. However, as seen in the previous chapters with overall point estimators, the levels of bias can vary with the relative size of the benchmark survey as well as the choice of variance estimator. Figure 6.1 contains the pattern of bias for the five variance estimators by the increasing size (left to right on the x axis) of the benchmark survey relative to the 1,000 persons selected for the analytic survey (n_B/n_A) for NOTCOV (a) and PDMED12M (b). The horizontal line represents zero bias, while the vertical line represents the effect for equal-sized analytic and benchmark surveys. Estimates for the Naïve and ECNJC estimators are represented by squares and triangles, respectively. The "Other EC" estimates (ECTS, ECF2m, and ECMV) are close in value and are shown as circles. This pattern is similar for analytic survey samples of size 2,000.

Simulation study results for the relative biases are least favorable for the traditional poststratified (Naïve) variance estimator as expected from our theoretical evaluation. The Naïve variance estimator underestimates the empirical MSE by as much as 22 percent for NOTCOV and 14 percent for PDMED12M. This naturally occurs when the benchmark variance component is the largest and not accounted for with this estimator. The slight improvement in the bias noted for n_B/n_A=6.0 is related only to a decrease in the empirical MSE. We suspect that additional simulation results will remove this anomaly by producing a more stable set of MSE values.

The EC jackknife variance estimators all contain a component associated with traditional poststratification. Therefore, the relative biases should mimic the bias levels exhibited for the Naïve variance estimator until the relative influence of the benchmark variance component becomes sizeable, i.e., the relative size of the benchmark is small. This pattern is seen in Figure 6.1 with a benchmark survey at least six times larger than the analytic survey. The bias is improved for the EC variance estimators because of the variance increase due to the coverage error and benchmark components. However, when the size of the benchmark survey is equal to or smaller than the analytic survey, changes occur in the picture. The biases in the figure for the ECNJCm variance estimator are smaller than the Naïve variance estimator but still fall below levels for the other estimators especially for small benchmark



(a) Total Number of Hispanics Not Covered by Health Insurance in Last 12 Months (NOTCOV)



(b) Total Number of Hispanics Who Delayed Medical Care Due to Cost in Last 12 Months (PDMED12M)

Figure 6.1: Percent Bias Relative to Empirical MSE of Five Variance Estimators by Relative Size of the Benchmark Survey to the Analytic Survey for 1,000 Analytic Survey Units

Table 6.2: Percent Bias Estimates Relative to Empirical MSE for Five Variance Estimators by Total Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey to the Analytic Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-22.1	-10.1	-1.7	-3.5	-26.2	-13.5	-0.1	-3.3
	ECTS	2.3	-2.8	0.1	-2.8	1.5	-5.0	2.1	-2.4
	ECF2m	2.3	-2.8	0.3	-2.6	1.6	-5.0	2.1	-2.4
	ECMV	2.7	-2.5	0.4	-2.6	1.6	-5.1	2.2	-2.3
	ECNJCm	-14.2	-7.5	-0.7	-2.9	-18.0	-10.8	0.9	-2.7
PDMED12M	Naïve	-13.6	-5.6	-4.5	-2.7	-20.1	-6.7	-5.4	2.4
	ECTS	-2.8	-2.3	-3.3	-1.9	-3.9	-1.1	-3.4	3.8
	ECF2m	-2.6	-2.2	-3.0	-1.5	-4.0	-1.3	-3.3	3.7
	ECMV	-2.6	-1.9	-2.9	-1.5	-3.8	-1.0	-3.4	3.8
	ECNJCm	-8.8	-3.6	-3.3	-1.6	-14.1	-3.9	-3.9	3.6

surveys. The negative bias of the ECNJCm variance estimator decreases to levels of 14 percent for NOTCOV and 9 percent for PDMED12M because of the missing off-diagonal terms in the benchmark covariance matrix. By contrast, the benchmark components in the "other" EC jackknife variance estimators (ECTS, ECF2m, and ECMV) assist in reducing the negative bias associated with the Naïve variance estimator. A positive relative bias of no more than 3 percent for NOTCOV suggests that this set of EC variance estimators can be slightly conservative when benchmark control totals are taken from relatively small benchmark surveys (n_B/n_A=0.3). Instability in the empirical MSEs, as discussed for the Naïve variance estimator above, also explains the slight increase in the negative bias for PDMED12M.

The relative biases used to produce Figure 6.1 are displayed in the $n_B/(n_A =$

Table 6.3: Empirical 95 Percent Coverage Rates for Five Variance Estimators by Total Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey to the Analytic Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	89.5	91.8	92.6	92.8	88.5	91.1	92.4	92.1
	ECTS	93.8	93.0	92.9	92.8	94.0	92.7	92.8	92.3
	ECF2m	93.8	93.0	92.9	92.8	93.7	92.5	92.8	92.3
	ECMV	93.7	92.7	92.9	92.9	94.0	92.5	92.8	92.2
	ECNJCm	90.9	92.2	92.8	92.8	90.3	91.5	92.5	92.2
PDMED12M	Naïve	88.9	90.2	90.8	90.7	88.5	91.4	91.6	92.1
	ECTS	91.0	90.6	90.9	90.9	92.0	92.5	91.9	92.3
	ECF2m	91.0	90.6	90.8	91.0	92.0	92.5	92.0	92.2
	ECMV	90.8	90.7	90.9	91.0	91.6	92.4	91.8	92.2
	ECNJCm	90.0	90.5	90.9	91.0	90.1	91.9	91.8	92.3

1,000) column of Table 6.2. The second column contains results for larger analytic survey sample sizes ($n_A = 2,000$). An interpretation similar to the one given for the figure also holds for this set of results.

The second comparative measure is the *empirical coverage rates for the 95 percent confidence intervals*. The values from our simulation study are provided in Table 6.3. Overall, we see a general pattern of stability in the coverage rates for the ECTS, ECF2m, and ECMV variance estimators across the eight relative sizes within each outcome variable. Differences in the rates across this set of variance estimators are minimal, and all have higher rates than either the Naïve or the ECNJCm variance estimators. Coverage rates for the Naïve estimator are largest when the benchmark variance components are inconsequential and fall well below 95 percent as the size

Table 6.4: Percent Increase in Instability of Variance Estimates Relative to EC Linearization Estimator (ECTS) by Total Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	5.2	2.0	0.5	0.2	8.5	2.9	0.8	0.4
	ECF2m	3.2	1.3	0.5	0.3	3.7	1.4	0.7	0.2
	ECMV	5.1	2.0	0.4	0.2	6.7	1.3	0.5	0.2
	ECNJCm	4.3	1.9	0.6	0.3	7.1	2.6	0.7	0.3
PDMED12M	Naïve	-0.1	0.1	0.1	0.1	2.2	1.1	0.5	0.4
	ECF2m	1.5	0.6	0.3	0.4	1.7	0.5	0.5	0.1
	ECMV	1.1	0.7	0.4	0.3	3.1	1.1	0.5	0.1
	ECNJCm	0.5	0.4	0.3	0.3	2.1	1.2	0.5	0.2

of this variance component increases — the same pattern as shown for the relative biases. A similar interpretation is given for the ECNJCm coverage rates with rates slightly higher than those for the Naïve estimator. Coverage rates for the estimated total number of Hispanics who delayed medical care (PDMED12M) are lower than those rates exhibited for NOTCOV. This also holds for the overall estimates given in Table 4.5 and is associated with the prevalence of the outcome variables in the population.

As with estimated totals examined in Chapter 4, our research suggests that there are minimal theoretical and empirical differences between the ECTS, ECF2m, and ECMV methods for domain estimation. The *variation in the estimated standard errors* for the methods, an indication of stability of the estimator, is presented in Table 6.4. We primarily see that the stability of the ECF2m and ECMV estimates are similar; however, the ECF2m is slightly more stable than the ECMV when the relative size of the benchmark survey is small. Both methods are more variable than the ECTS as expected (Krewski & Rao, 1981).

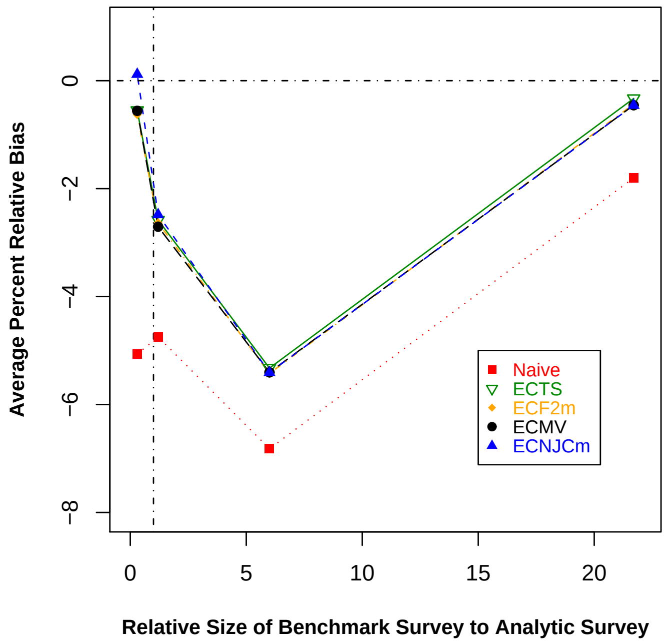
6.6.5 Comparison of Variance Estimators for Estimated Means

Differences in the set of EC variance estimators were less noticeable for ratio-means than totals as noted in Chapter 5. The same statement applies to estimation for domain ratio-means discussed in this section.

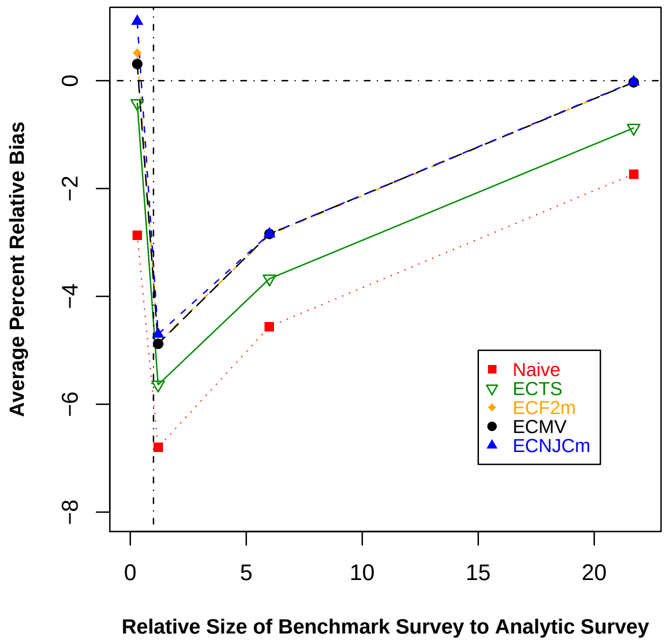
The *percent biases relative to the empirical MSE* for the variance of the estimated domain means range between -9 and 1 percent across the simulation parameters with almost all values falling below the desired zero percent level. Note that a slight positive bias suggests a conservative estimator; this trait is desired over negative biases. This range of values is comparable with the range for overall ratio-means (-8 to 3 percent shown in Table 5.2) and less than the range for the domain totals (-27 to 3 percent shown in Table 6.2).

Figure 6.2 contains a visual display of the estimated percent relative biases (y axis) in estimating the MSE of our two outcome variables within the Hispanic population, by the relative size (n_B/n_A) of the benchmark survey to the analytic survey of size $n_A = 1,000$ (x axis). The horizontal line represents zero bias. The vertical line effectively represents equal-sized analytic and benchmark surveys.

As with Figure 6.1, we note the similarities in the patterns for the relative biases for the variance estimators until the line of equality. The Naïve variance esti-



(a) Average Number of Hispanics Not Covered by Health Insurance in Last 12 Months (NOTCOV)



(b) Average Number of Hispanics Who Delayed Medical Care Due to Cost in Last 12 Months (PDMED12M)

Figure 6.2: Percent Bias Relative to Empirical MSE of Five Variance Estimators by Relative Size of the Benchmark Survey to the Analytic Survey for 1,000 Analytic Survey Units

mator, as with the other analyses presented in this body of work, is most negatively biased among those estimators examined for all simulation conditions included in our study. As expected, the least amount of bias can be seen when the benchmark survey is more than 21 times as large as the analytic survey. The ECTS variance estimator improves upon the bias of the Naïve, and actually produces comparable levels between -1 and -2 percent at the two ends of the relative size scale ($n_B/n_A = 0.3$ and $n_B/n_A = 21.7$). Hence, it appears that the reduction in the analytic survey variance component is counterbalanced by the increase in the benchmark variance components. The point at which the counterbalance occurs is a potential research topic. The biases for the remaining EC jackknife variance estimators are numerically close to the linearization estimators; however, the ECNJCm is positively biased for $n_B/n_A \cong 0.3$.

Values used to generate Figure 6.2 are provided in Table 6.5 for analytic survey sample sizes of $n_A = 1,000$. The pattern in the relative biases for the domain ratio-mean with $n_A = 1,000$ is closer to the pattern given for the domain totals shown in Figure 6.1. This suggests that domain ratio-means may be more sensitive to the variability in the EC benchmark controls in comparison with the other point estimators studied here, and also sensitive to the number of simulation samples.

The empirical *coverage rates for the 95 percent confidence intervals* shown in Table 6.6 range from 90.6 to 94.0 percent with many values (especially for PDMED12M) falling below 93 percent. Minor fluctuations occur across the relative sizes of the benchmark survey for the domain ratio-means, the same trait noted for the overall ratio-means. However, the coverage rates presented here, and

Table 6.5: Percent Bias Estimates Relative to Empirical MSE for Five Variance Estimators by Mean Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey to the Analytic Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-5.1	-4.7	-6.8	-1.8	-8.4	-9.0	-3.8	-8.1
	ECTS	-0.6	-2.6	-5.3	-0.3	-2.6	-6.1	-1.5	-6.1
	ECF2m	-0.6	-2.6	-5.4	-0.4	-3.6	-7.1	-2.6	-7.1
	ECMV	-0.6	-2.7	-5.4	-0.5	-3.5	-7.1	-2.6	-7.1
	ECNJCm	0.1	-2.5	-5.4	-0.5	-2.9	-7.0	-2.6	-7.1
PDMED12M	Naïve	-2.9	-6.8	-4.6	-1.7	-7.2	-4.3	-6.4	-2.4
	ECTS	-0.4	-5.6	-3.7	-0.9	-4.1	-2.4	-4.8	-0.9
	ECF2m	0.5	-4.9	-2.9	0.0	-3.7	-2.1	-4.6	-0.7
	ECMV	0.3	-4.9	-2.8	0.0	-4.0	-2.2	-4.7	-0.7
	ECNJCm	1.1	-4.7	-2.8	0.0	-3.2	-2.0	-4.6	-0.7

Table 6.6: Empirical 95 Percent Coverage Rates for Five Variance Estimators by Mean Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey to the Analytic Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	93.2	92.9	92.5	93.0	92.6	91.8	93.2	92.7
	ECTS	94.0	93.2	92.7	93.4	93.8	92.5	93.6	93.0
	ECF2m	94.0	93.0	92.6	93.2	93.2	92.3	93.5	92.8
	ECMV	93.7	93.0	92.6	93.3	93.5	92.2	93.4	92.7
	ECNJCm	94.0	93.1	92.6	93.3	93.6	92.4	93.4	92.8
PDMED12M	Naïve	91.6	90.6	91.4	91.1	90.9	92.3	92.0	91.7
	ECTS	92.1	90.8	91.6	91.2	91.7	92.7	92.1	92.1
	ECF2m	92.1	90.8	91.7	91.2	91.7	92.5	92.1	92.0
	ECMV	92.1	90.8	91.7	91.2	91.5	92.5	92.1	91.9
	ECNJCm	92.3	90.8	91.6	91.1	91.9	92.5	92.1	92.0

Table 6.7: Percent Increase in Instability of Variance Estimates Relative to EC Linearization Estimator (ECTS) by Mean Outcome within the Hispanic Domain and Relative Size of the Benchmark Survey

Outcome Variable	Variance Estimator	Relative Size $n_B/(n_A = 1,000)$ 0.3	1.2	6.0	21.7	Relative Size $n_B/(n_A = 2,000)$ 0.2	0.6	3.0	10.8
NOTCOV	Naïve	2.1	0.9	0.6	0.5	2.8	1.3	0.9	0.8
	ECF2m	1.7	1.8	1.5	1.5	2.2	1.5	1.5	1.4
	ECMV	1.8	1.7	1.5	1.5	2.2	1.6	1.4	1.4
	ECNJCm	2.2	1.5	1.6	1.5	2.4	1.5	1.4	1.4
PDMED12M	Naïve	0.3	0.2	0.1	0.1	0.9	0.5	0.4	0.4
	ECF2m	1.3	0.9	0.8	0.8	1.0	1.0	0.7	0.7
	ECMV	1.0	0.8	0.9	0.7	0.8	1.0	0.7	0.7
	ECNJCm	1.1	0.8	0.8	0.7	0.6	1.0	0.6	0.7

also in Table 6.3, are lower than the desired level of 95 percent. Further research is needed in an attempt to improve the coverage rates for the EC domain estimators.

A comparison of the stability in the estimates (Table 6.7) again shows that the ECTS variance estimator produces more stable estimates than any of the variance estimators studied. Note that the decrease in stability for the EC variance estimators is more consistent across the relative survey sizes in comparison with our other analyses — see, for example, Table 5.5.

6.7 Summary of Research Findings

To summarize, the empirical results for estimated domain totals and ratio-means mirror comments given for the corresponding overall estimates. The empirical results for the EC calibration estimators are not as strong as in Chapters 4 and 5 but the comparative differences still exist. We recommend against the use of traditional calibration variance estimators for domain estimation. A theoretical and empirical evaluation suggests that the underestimation can be sizeable. Use of the ECNJCm method, when a complete benchmark covariance matrix is not accessible, is more applicable to ratio-means than with estimated domain totals. The choice between the EC linearization method (ECTS) and one of the EC jackknife methods (ECF2m and ECMV) may be more related to preference of the analysis file structure. If design variables are to be suppressed for disclosure avoidance, then either the ECF2m or the ECMV will suffice.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Traditional methods are generally applied to calibration estimators even when the assumptions, such as population benchmark totals and perfect sampling frames, are violated. Our research presented in this dissertation examines the use of calibration control totals estimated from an independent (benchmark) survey on a different (analytic) survey with units selected from an incomplete sampling frame. We label this methodology as estimated-control (EC) calibration. As shown in the three research chapters (Chapters 4, 5, and 6), traditional calibration variance estimators under certain conditions fail to capture all of the variation associated with the survey estimates. Underestimation is most dramatic when the benchmark survey is smaller than the analytic survey as demonstrated for estimated totals within and across domains. Underestimation is also present for controls estimated from relatively large benchmark surveys, though the level of bias is less pronounced than with small benchmark surveys. Ratios of two estimated totals by domain and overall are less affected by the size of the benchmark survey than population total estimators, but some negative bias is still present. In addition to variance estimation, we define a formula for the bias of the point estimators as a function of the benchmark control bias and the element-wise probabilities of being included on the analytic survey sampling frame.

Taylor linearization and jackknife variance estimators are developed to address the benchmark-control estimation and the sampling frame undercoverage error, as well as the variation within the analytic survey data. The analytic sample is obtained from a general design with primary sampling units selected with replacement from within first-stage strata. Both types of EC calibration variance estimators are adapted from prior research and are shown, both theoretically and empirically, to be superior to formulae developed under the traditional weight calibration assumptions discussed in Chapter 2. Based on a comparison of the EC calibration variance estimators, we recommend either the EC Taylor linearization variance estimator (ECTS) or the modified Fuller jackknife variance estimator (ECF2m) for use with EC calibration total and ratio-of-totals estimators when the complete control total covariance matrix is available. The choice between the linearization and the replication variance estimators is related to the type of analysis data file to be produced. When only the diagonal elements of the covariance matrix are available, the modified Nadimpalli-Judkins-Chu variance estimator (ECNJCm), a simplification of the multivariate normal variance estimator (ECMV), may be used. However, unlike levels seen for the ratio of two totals in our simulation studies, negative biases can be substantial with the ECNJCm for the variance of estimated totals. The accompanying computer code written in R® translates our research into practical tools for the survey statistician.

Weight calibration continues to be an important instrument for survey researchers, especially given the increased use of data collection modes not accessible by all members of a population (e.g., Web surveys). EC calibration is a mechanism that allows benchmarking to specialized control totals that are not available in the large-scale surveys. The attempt to reduce bias through weight calibration must be counterbalanced with the increase in variance properly captured with our methodology.

7.2 Future Work

A basic framework for EC calibration is presented in the pages of this dissertation. However, EC calibration remains a rich source of research. The following is a list of important questions generated by our current work:

1. What modification to the current coverage error component will make this adjustment more robust?

2. Is there a threshold that exists to suggest when traditional variance estimators are acceptable with EC calibrated estimators?

3. Is there a measure that will determine when a benchmark estimate is too imprecise for use in EC calibration?

Extensions to our current work may address the following questions:

1. What are the degrees of freedom associated with statistical tests that use the EC calibrated estimates?

2. What are the effects of nonresponse in one or both surveys on EC calibration?

3. How might *non-sampling errors* in both surveys change the properties of EC calibrated estimators?

4. How might EC calibration for cross-sectional surveys and independent benchmark surveys be adapted for *two-phase designs* which may include dependent benchmark controls and *panel surveys*?

5. Are the properties of balanced repeated replication (BRR) variance estimators more favorable than the jackknife for EC calibration estimators?

6. What are the effects of EC calibration on point estimators other than totals and ratios of two totals?

7. What are the theoretical and empirical properties of *non-linear* EC calibration such as the logistic GREG (LGREG) estimators discussed in Duchesne (2003)?

8. What are the theoretical and empirical properties of *constrained* EC calibration?

Chapter A

Simulation Programs

The simulation study programs are provided in the following sections. With the exception of the SAS-callable SUDAAN® program (Research Triangle Institute, 2004) included in the first section, the programs were written in R®.

A.1 Calculate Benchmark Estimates

```
/******************************************************************/
/* Program: NHIS Covar.sas                                        */
/* Name:    J.Dever                                               */
/* Date:    06/07/07                                              */
/* Purpose: Produce covariance matrix from NHIS data.             */
/******************************************************************/
options nocenter pageno=1 errors=1 orientation=portrait nofmterr;

LIBNAME in  "...\NHIS\Data2003\";
LIBNAME out "...\Dissertation\Programs\Data\";
LIBNAME outxp xport "...\Dissertation\Programs\Data\COVMATRX.xpt";

TITLE1 "Dissertation/JSM07 - NHIS Covariance Matrix";

******************************************************************;
** Process NHIS Data Using SUDAAN.                              **;
******************************************************************;
PROC CONTENTS DATA=in.PERSONSX;  RUN cancel;

PROC SORT DATA=in.PERSONSX OUT=PERSONSX;  BY STRATUM PSU;  RUN;

PROC CROSSTAB DATA=PERSONSX DESIGN=WR DEFT2;
 SETENV COLWIDTH=30 DECWIDTH=10;
 WEIGHT WTFA;
 NEST STRATUM PSU;
 SUBGROUP R_AGE1 SEX;
```

```
 LEVELS 8 2;
 TABLES SEX * R_AGE1;
 PRINT /*NSUM WSUM COVWGT*/ / STYLE=NCHS;
 OUTPUT / WGTCOV=ALL FILENAME=out.COVMATRX REPLACE;
RUN;

PROC PRINT DATA=out.COVMAT01;  RUN;

*********************************************************************;
** Process Covariance Matrix.                                      **;
*********************************************************************;
PROC CONTENTS DATA=out.COVMAT01;  RUN cancel;

PROC PRINT DATA=out.COVMAT01 UNIFORM NOOBS;
  VAR B011-B018 B020-B027 EST_ID IDNUM NCELL PROCNUM ROWNUM TABLENO;
RUN cancel;

DATA outxp.COVMATRX(KEEP=B011-B018 B020-B027);
  SET out.COVMAT01;
                   ** Subset to covar matrix, exclude "total" rows **;
  IF IDNUM=2 &
     ROWNUM in (11, 12, 13, 14, 15, 16, 17, 18,
                20, 21, 22, 23, 24, 25, 26, 27);
RUN;
```

A.2 Generate Benchmark Covariance Matrices

```
#---------------------------------------------------------------------
# Program: Estimated Controls.R
# Name:    J.Dever
# Date:    09/26/07
# Project: Dissertation / JSM07
# Purpose: Create object containing estimated controls from full
#          2003 NHIS public-use file randomly generated based on
#          specified (adjusted) covariance matrix under a
#          multivariate normal assumption.  Original program
#          entitled Random Controls2.R from NCHS project with
#          R.Valliant, J.Kim updated for SURV699G - Weighting
#          and Imputation final.  Additionally revised to add
#          variance of estimated overall total.
#---------------------------------------------------------------------
                                              #Set working directory
rm(list=ls(all=TRUE))
setwd(".../Dissertation/Programs/Data/")
```

```
                                                     #Load R libraries
require(MASS)
require(foreign)
                                     #Random seed for MVnormal function
set.seed(82841)
                                         #Maximum number of simulations
n.sims <- 5000

#----------------------------------------------------------------------
# Load SAS XPORT File Containing Covariance Matrix and Relabel
#----------------------------------------------------------------------

NHIS03.cov <- read.xport(".../Programs/Data/COVMATRX.xpt")
dim(NHIS03.cov)
NHIS03.cov

names(NHIS03.cov) <- c("R_AGE11.SEX1","R_AGE12.SEX1","R_AGE13.SEX1",
                       "R_AGE14.SEX1","R_AGE15.SEX1","R_AGE16.SEX1",
                       "R_AGE17.SEX1","R_AGE18.SEX1","R_AGE11.SEX2",
                       "R_AGE12.SEX2","R_AGE13.SEX2","R_AGE14.SEX2",
                       "R_AGE15.SEX2","R_AGE16.SEX2","R_AGE17.SEX2",
                       "R_AGE18.SEX2")
rownames(NHIS03.cov) <- c("R_AGE11.SEX1","R_AGE12.SEX1",
                       "R_AGE13.SEX1","R_AGE14.SEX1","R_AGE15.SEX1",
                       "R_AGE16.SEX1","R_AGE17.SEX1","R_AGE18.SEX1",
                       "R_AGE11.SEX2","R_AGE12.SEX2","R_AGE13.SEX2",
                       "R_AGE14.SEX2","R_AGE15.SEX2","R_AGE16.SEX2",
                       "R_AGE17.SEX2","R_AGE18.SEX2")

#----------------------------------------------------------------------
# Control Variables
#----------------------------------------------------------------------

control.vars <- as.data.frame(cbind(R_AGE1 = sort(rep(1:8,2)),
                                    SEX = rep(1:2,8)))
control.vars <- control.vars[order(control.vars$SEX),]
control.vars

rownames(control.vars) <- c("R_AGE11.SEX1","R_AGE12.SEX1",
                       "R_AGE13.SEX1","R_AGE14.SEX1","R_AGE15.SEX1",
                       "R_AGE16.SEX1","R_AGE17.SEX1","R_AGE18.SEX1",
                       "R_AGE11.SEX2","R_AGE12.SEX2","R_AGE13.SEX2",
                       "R_AGE14.SEX2","R_AGE15.SEX2","R_AGE16.SEX2",
                       "R_AGE17.SEX2","R_AGE18.SEX2")
control.vars
```

```
#-------------------------------------------------------------------
# Vector of Pop Totals from Edited NHIS Frame used in Simulations
#-------------------------------------------------------------------

NHIS03.popcts <- c(10148500, 27153554, 13901790, 40810498, 33082381,
                   4575653,  3714185,  6211982,
                   9707257, 25961781, 13867450, 41956867, 35164332,
                   5184382,  4701061,  9868567)

names(NHIS03.popcts) <- c("R_AGE11.SEX1","R_AGE12.SEX1",
                      "R_AGE13.SEX1","R_AGE14.SEX1","R_AGE15.SEX1",
                      "R_AGE16.SEX1","R_AGE17.SEX1","R_AGE18.SEX1",
                      "R_AGE11.SEX2","R_AGE12.SEX2","R_AGE13.SEX2",
                      "R_AGE14.SEX2","R_AGE15.SEX2","R_AGE16.SEX2",
                      "R_AGE17.SEX2","R_AGE18.SEX2")
NHIS03.popcts

#-------------------------------------------------------------------
# Overall Estimated Population Count and Variance
#-------------------------------------------------------------------

NHIS03.pop <- sum(NHIS03.popcts)
NHIS03.pop

NHIS03.popVar <- (2919389.5935)**2
NHIS03.popVar

#-------------------------------------------------------------------
# Adjustment Factor to Reduce Size of Random Control
#-------------------------------------------------------------------

rc.adj <- (21664 / sum(NHIS03.popcts))

NHIS03.adj.popcts <- round(NHIS03.popcts * rc.adj)
NHIS03.adj.cov    <- as.matrix(NHIS03.cov) * (rc.adj**2)
NHIS03.adj.popVar <- NHIS03.popVar * (rc.adj**2)

NHIS03.pop.adj0 <- as.data.frame(rbind(NHIS03.adj.popcts,
                                       t(control.vars)))
NHIS03.pop.adj0
sum(NHIS03.pop.adj0[1,])

cbind(NHIS03.adj.popcts, sqrt(diag(NHIS03.adj.cov)),
      sqrt(NHIS03.adj.popVar))

#-------------------------------------------------------------------
# Generate Random Control Totals (Covariance Adjustment = 1.0)
```

```
#-------------------------------------------------------------------

cm.adj1 <- 92000 / 92000
cm.adj1

NHIS03.cov.adj1 <- as.matrix(NHIS03.adj.cov) * cm.adj1
NHIS03.cov.adj1

MV.Norm <- round(mvrnorm(n=n.sims, mu=NHIS03.adj.popcts,
                         Sigma=NHIS03.cov.adj1))

NHIS03.pop.adj1 <- as.data.frame(rbind(as.data.frame(MV.Norm),
                                       t(control.vars)))
NHIS03.pop.adj1[c(1:5, n.sims:nrow(NHIS03.pop.adj1)),]
rbind(NHIS03.adj.popcts, mean=apply(NHIS03.pop.adj1,2,mean),
      min=apply(NHIS03.pop.adj1,2,min),
      max=apply(NHIS03.pop.adj1,2,max),
      se =sqrt(apply(NHIS03.pop.adj1,2,var)))

NHIS03.popVar.adj1 <- NHIS03.adj.popVar * cm.adj1
NHIS03.popVar.adj1

#-------------------------------------------------------------------
# Generate Random Control Totals (Covariance Adjustment = 3.6)
#-------------------------------------------------------------------

cm.adj2 <- 92000 / 25000
cm.adj2

NHIS03.cov.adj2 <- as.matrix(NHIS03.adj.cov) * cm.adj2
NHIS03.cov.adj2

MV.Norm <- round(mvrnorm(n=n.sims, mu=NHIS03.adj.popcts,
                         Sigma=NHIS03.cov.adj2))

NHIS03.pop.adj2 <- as.data.frame(rbind(as.data.frame(MV.Norm),
                                       t(control.vars)))
NHIS03.pop.adj2[c(1:5,n.sims:nrow(NHIS03.pop.adj2)),]
rbind(NHIS03.adj.popcts, mean=apply(NHIS03.pop.adj2,2,mean),
      min=apply(NHIS03.pop.adj2,2,min),
      max=apply(NHIS03.pop.adj2,2,max),
      se =sqrt(apply(NHIS03.pop.adj2,2,var)))

NHIS03.popVar.adj2 <- NHIS03.adj.popVar * cm.adj2
NHIS03.popVar.adj2

#-------------------------------------------------------------------
```

```
# Generate Random Control Totals (Covariance Adjustment = 18)
#---------------------------------------------------------------------

cm.adj3 <- 92000 /  5000
cm.adj3

NHIS03.cov.adj3 <- as.matrix(NHIS03.adj.cov) * cm.adj3
NHIS03.cov.adj3

MV.Norm <- round(mvrnorm(n=n.sims, mu=NHIS03.adj.popcts,
                         Sigma=NHIS03.cov.adj3))

NHIS03.pop.adj3 <- as.data.frame(rbind(as.data.frame(MV.Norm),
                                       t(control.vars)))
NHIS03.pop.adj3[c(1:5,n.sims:nrow(NHIS03.pop.adj3)),]
rbind(NHIS03.adj.popcts, mean=apply(NHIS03.pop.adj3,2,mean),
      min=apply(NHIS03.pop.adj3,2,min),
      max=apply(NHIS03.pop.adj3,2,max),
      se =sqrt(apply(NHIS03.pop.adj3,2,var)))

NHIS03.popVar.adj3 <- NHIS03.adj.popVar * cm.adj3
NHIS03.popVar.adj3

#---------------------------------------------------------------------
# Generate Random Control Totals (Covariance Adjustment = 72)
#---------------------------------------------------------------------

cm.adj4 <- 92000 /  1250
cm.adj4

NHIS03.cov.adj4 <- as.matrix(NHIS03.adj.cov) * cm.adj4
NHIS03.cov.adj4

MV.Norm <- round(mvrnorm(n=n.sims, mu=NHIS03.adj.popcts,
                         Sigma=NHIS03.cov.adj4))

NHIS03.pop.adj4 <- as.data.frame(rbind(as.data.frame(MV.Norm),
                                       t(control.vars)))
NHIS03.pop.adj4[c(1:5,n.sims:nrow(NHIS03.pop.adj4)),]
rbind(NHIS03.adj.popcts, mean=apply(NHIS03.pop.adj4,2,mean),
      min=apply(NHIS03.pop.adj4,2,min),
      max=apply(NHIS03.pop.adj4,2,max),
      se =sqrt(apply(NHIS03.pop.adj4,2,var)))

NHIS03.popVar.adj4 <- NHIS03.adj.popVar * cm.adj4
NHIS03.popVar.adj4
```

```
save.image(".../Programs/Data/Estimated Controls.RData")
```

A.3 Simulation Call Program

```
#----------------------------------------------------------------------
# Program: NOTCOV Hsp cadj1 n20.R
# Name:    J.Dever
# Project: Dissertation / Sim Domains / Appendix Code
# Date:    11/03/08
# Purpose: Reduce code for dissertation appendix and text.
#          Use random controls with variable=NOTCOV, covariance
#          adjustment=1.0, and n_hi=20 for 2,000 simulation runs.
#          Domain = Hispanic Race (HISCODI2=1)
#----------------------------------------------------------------------
                                               #Set working directory
rm(list=ls(all=TRUE))
setwd(".../Dissertation/Programs/")

require(MASS)                                       #Load R libraries
require(survey)
require(nlme)
memory.size()                                   #Increase memory size
round(memory.limit()/1048576.0, 2)
memory.limit(size=2000)
                                 #Sampling Frame, External Controls
source("nhis25.new.dmp")
attach("Estimated Controls.RData")
                                                       #Sim functions
source("Sim.ECPS.fcn")
source("Rep.VarEst.fcn")
source("chk.PS.fcn")
source("clus.sam.fcn")
source("cov.rate.fcn")

#----------------------------------------------------------------------
# Simulation program
#----------------------------------------------------------------------

NOTCOV.hsp.cadj1.n20 <- Sim.ECPS(pop           =nhis25.new,
                                 y.col         ="NOTCOV",
                                 y.val         =1,
                                 d.col         ="HISCODI2",
                                 d.val         =1,
                                 unit.id       ="ID",
```

```
                                 str.col       ="new.str",
                                 clus.id       ="new.psu",
                                 PS.col        =c("R_AGE1", "SEX"),
                                 nh            =rep(2,25),
                                 nh.sub        =20,
                                 substrat      ="substrat",
                                 sub.vals      =1,
                                 sel.meth      ="ppswr",
                                 no.sams       =2000,
                                 cov.prob      =c(0.9,0.8,0.5,0.5,
                                                  0.8,0.9,0.9,0.9,
                                                  0.9,0.8,0.5,0.5,
                                                  0.8,0.9,0.9,0.9),
                                 seed          =81311,
                                 m.cell        =2,
                                 ex.cntrls     =T,
                                 ex.cntrls.pop =NHIS03.pop.adj1,
                                 ex.cntrls.cov =NHIS03.cov.adj1,
                                 ex.cntrls.var =NHIS03.popVar.adj1,
                                 cert.PSUs     =T,
                                 sam.prt       =100)

rm("nhis25.new")                    #Eliminate pop file to save space
save.image("NOTCOV Hsp cadj1 n20.RData")
```

A.4 Primary Simulation Program

```
Sim.ECPS <- function(pop, y.col, y.val=1., d.col="ones", d.val=1.,
                     str.col, PS.col, clus.id, unit.id, nh, nh.sub,
                     substrat, sub.vals, sel.meth, no.sams, cov.prob,
                     seed, m.cell, ex.cntrls, ex.cntrls.pop,
                     ex.cntrls.start=0., ex.cntrls.cov, ex.cntrls.var,
                     PS.chk=F, cert.PSUs, sam.prt) {

  # Simulation for poststratified estimates using estimated
  # controls (ECPS). Original code taken from NCHS PS-cell collapse
  # project with R.Valliant.
  #
  # pop               = population
  # y.col             = variable for estimating total
  # y.val             = variable value for estimating total (convert
  #                     to 0/1 variable)
  # d.col             = variable for conducting domain analysis
  # d.val             = variable value used in domain analyses
```

```
  # str.col          = stratum column name or no.
  # PS.col           = poststratification column names or nos. PS
  #                    can be defined as cross of several variables
  # clus.id          = cluster ID column name or no.
  # unit.id          = unique ID for unit of observation (person)
  # nh               = num of clusters sampled within each stratum
  #                    (vector) If no substrata are used for
  #                    sampling within clusters, nh.sub is single
  #                    value, substrata var (sub.vals) should
  #                    have same value for every unit
  # nh.sub           = sample size for each substratum (vector)
  # substrat         = substrata, column name or no.
  # sub.vals         = values taken by substrat variable
  # sel.meth         = method fo selecting clusters (ppswr or srs)
  # no.sams          = no. of samples
  # cov.prob         = response probability vector
  # seed             = seed for random no. generator
  # m.cell           = minimum cell size for no. of covered units
  # ex.cntrls        = T/F external controls used for wt adju
  # ex.cntrls.pop    = data file name with external control counts
  # ex.cntrls.start = (start + sim no) = line in ex.cntrls.pop list
  #                    used as controls,allows diff controls per prg
  # ex.cntrls.cov    = name of data file containing external control
  #                    var-covar matrix
  # ex.cntrls.var    = var(Nhat.B) from benchmark survey
  # PS.chk           = T/F to run check on replicate algorithsm
  # cert.PSUs        = T/F if size>nh.sub, select all units w/in PSU
  # sam.prt          = how often to print current sample no,
  #                    e.g., every 10, 25, 100, etc.
  #
  # Last Update: 11/03/2008 Old code removed from prog for appendix

set.seed(seed)
cat("begin ", date(), "\n")

#__________Initialization section____________________________________

                      # Variable containing all ones (default domain)
pop$ones <- rep(1, nrow(pop))
                                                  # Domain indicator
pop$delta.d <- as.numeric(pop[,d.col] == d.val)
                              # Analysis variable by domain indicator
pop$yd.col <- pop[,y.col] * pop$delta.d

                # select units with nonmissing y
                # Note: this will result in ID's being nonconsecutive
                #   in the reduced pop after missing y's eliminated
```

```
pop <- pop[!is.na(pop[, "yd.col"]), ]
pop <- pop[, -1]                          # Put the ID's back in order
ID <- 1:nrow(pop)
pop <- cbind(ID, pop)

N.PS <- prod(dim(table(pop[, PS.col])))            # No. poststrata

str.id <- unique(pop[, str.col])                       # Stratum IDs

H <- length(str.id)                              # No. of design strata

                                          # Error check on input specs
if (length(cov.prob) != N.PS) {
   stop("length(cov.prob) != no. of PS\n")}
if (length(nh) != H) stop("length(nh) != H\n")
if (any(nh != 2)) stop("nh not 2 for all strata {
   (chk.psu.dups only works for nh = 2)\n")}
if (any( table(pop[, str.col], pop[, clus.id]) < min(nh.sub)))
   cat("At least one psu has fewer than nh.sub units\n")

G <- nrow(ex.cntrls.cov)
R <- sum(nh)
if(R < G) stop("Insufficient number of replicates\n")
if(N.PS != G) stop("Poststrata in survey and external controls are
    not compatible\n")

                                           # Analyses objects (totals)
out.tot <- matrix(0., nrow = no.sams, ncol = 12)
dimnames(out.tot) <- list(NULL, c(
                     "T.pop",          #(Pseudo-)Pop total
                     "ECPS.tot",       #PS estimated total
                     "PWR.tot",        #Unadjusted estimated total
                     "Naive.tot",      #SE - Traditional PS
                     "ECTSr.tot",      #SE - Linear w/o trace
                     "ECF2.tot",       #SE - Fuller method
                     "ECMV.tot",       #SE - MV method
                     "ECNJC.tot",      #SE - NJC method
                     "ECTSr.totcov",   #SE (cov adj) - Linear
                     "ECF2.totcov",    #SE (cov adj) - Fuller method
                     "ECMV.totcov",    #SE (cov adj) - MV method
                     "ECNJC.totcov"    #SE (cov adj) - NJC method
))
                                      # Analyses objects (ratio means)
out.mu <- matrix(0., nrow = no.sams, ncol = 12)
dimnames(out.mu)  <- list(NULL, c(
                     "P.pop",        #(Pseudo-)Pop mual
                     "ECPS.mu",      #PS estimated mual
```

```
                     "Hajek.mu",    #Unadjusted estimated mual
                     "Naive.mu",    #SE - Traditional PS
                     "ECTSr.mu",    #SE - Linear w/o trace
                     "ECF2.mu",     #SE - Fuller method
                     "ECMV.mu",     #SE - MV method
                     "ECNJC.mu",    #SE - NJC method
                     "ECTSr.mucov", #SE (cov adj) - Linear
                     "ECF2.mucov",  #SE (cov adj) - Fuller method
                     "ECMV.mucov",  #SE (cov adj) - MV method
                     "ECNJC.mucov"  #SE (cov adj) - NJC method
))

num.skip.sam <- 0.
n.clus <- sum(nh.sub)                      # units sampled w/ cluster
nh.cl = rep(n.clus, sum(nh))                        # vector of units
n.tot <- sum(nh * nh.sub)                         # total sample size

sam.id   <- vector("numeric", length = n.tot)
base.wts <- vector("numeric", length = n.tot)
y        <- vector("numeric", length = n.tot)
no.PS    <- vector("numeric", length = no.sams)

A <- c(0,cumsum(nh*sum(n.clus)))

#__________Pop tabs__________________________________________________

                                  # Total y (level) by domain (level)
T.pop <- sum(pop[(pop[, y.col] == y.val) & (pop[, d.col] == d.val),
             "ones"])

                      # Prop of Total y (level) within domain (level)
P.pop <- T.pop / sum(pop[(pop[, d.col] == d.val),"ones"])
                               # String of poststratum variable names
ps.for <- NULL
for (i in 1:length(PS.col)){
  ps.for <- paste(ps.for, "+", PS.col[i])
}
                                                 # Code for ECF2 test
if(!ex.cntrls) {
  PS.pop <- xtabs(as.formula(paste("~", ps.for)), data = pop)
  PS.index <- array(1:length(PS.pop), dim = dim(PS.pop),
                    dimnames = dimnames(PS.pop) )
}

  #____Frame needed to feed into postStratify function____
  #   1st column is PS index, 2nd is pop counts in each PS
```

```
if(!ex.cntrls) {
  PS.pop.frm <- data.frame(PS.new1 = as.vector(PS.index),
                           Tot = as.vector(PS.pop) )
}

  #____PS.new1 calc only works for 2-dimensional poststrata____

                            # Max level for 1st of 2 poststratum vars
maxI <- max(unique(pop[, PS.col[1] ]))
                             # Calculate poststratum IDs for pop file
PS.new1 <- maxI * (pop[, PS.col[2]]-1) +  pop[, PS.col[1]]
pop <- cbind(pop, PS.new1)
                                     # Sorted list of poststratum IDs
PS.all <- sort(unique(pop[, "PS.new1"]))
cov.mat <- matrix(0., nrow = no.sams, ncol = length(PS.all) )

#__________Simulation loop__________________________________________

for(i in 1.:no.sams) {

  if((i %% sam.prt) == 0.) {
    cat("i =", i, date(), "\n")
  }
                # Set switches for whether units are covered by frame
  c.sw <- cov.rate(pop=pop, c.prob=cov.prob, cells="PS.new1")
  keep.sw <- skip.sw <- empty.PS.sw <- FALSE

  #__________Pop tabs for random controls______________________

  if(ex.cntrls) {
                     # Identify external controls from generated list

    pop.ext <- as.data.frame(t(ex.cntrls.pop[c((i + ex.cntrls.start),
               (nrow(ex.cntrls.pop) - 1), nrow(ex.cntrls.pop)),]))
    names(pop.ext)[1] <- "Tot"

    PS.pop <- xtabs(as.formula(paste("~", ps.for)), data = pop.ext)
    PS.pop <- PS.pop *
              matrix(pop.ext[order(pop.ext[,PS.col[2]]),"Tot"],
              nrow=nrow(PS.pop), ncol=ncol(PS.pop))
    PS.index <- array(1:length(PS.pop), dim = dim(PS.pop),
    dimnames = dimnames(PS.pop) )

                                           # Process external controls
    PS.pop.mrg <- as.data.frame(PS.pop)[,-3]
    PS.pop.mrg$PS.new1 <- as.vector(PS.index)
```

```
 external.CTot <- pop.ext
 external.CTot <- merge(PS.pop.mrg, external.CTot,
                        by.x=c(PS.col), by.y=c(PS.col))
 if(nrow(external.CTot) == 0) {
   stop("Frame / External Controls are not compatible.\n")}

             # File of ext controls and associated poststratum IDs
   PS.pop.frm <- external.CTot[order(external.CTot$PS.new1),]
   PS.pop.frm <- PS.pop.frm[, c(3,4)]
 }

#__________Draw simulation sample____________________________

 while(!keep.sw) {
   drop.sw <- NULL
   for (h in str.id){
                         # Select sample from "covered" units only
     poph <- pop[pop[,str.col]==h & c.sw,]
     h.id<-(1:length(str.id))[str.id==h]
     clus.dat.h <- clus.sam(pop       =poph,     clus.id =clus.id,
                            unit.id   ="ID",     n.cl    =nh[h.id],
                            sel.meth  =sel.meth, substrat=substrat,
                            n.substrat=nh.sub,   sub.vals=sub.vals,
                            cert.PSUs =cert.PSUs)

     sam.id[(A[h.id]+1):A[h.id+1]] <- clus.dat.h[[1]][,1]
     base.wts[(A[h.id]+1):A[h.id+1]] <- clus.dat.h[[1]][,2]
     drop.sw <- c(drop.sw, clus.dat.h[[3]])
     drop.sw <- any(drop.sw)

     if (any(drop.sw == TRUE)) {
       cat("bad sample", "h=",h, "\n")
     }
   }  #for h loop

          #___Check for missing poststrata, cells with all missing
          #   units, or cells where nonmissing count is < nh.sub___
   t1 <- pop[sam.id,  ]
   skip.sw <- chk.PS(sdat=t1, cl.all=PS.all, cl.col="PS.new1",
                    r.sw=c.sw[sam.id], min.size = m.cell)

   if (skip.sw | drop.sw){
     num.skip.sam <- num.skip.sam + 1
   }

   if(!(skip.sw | drop.sw)) {
                                                # Create sample file
```

```
      sam.dat <- cbind(pop[sam.id, c(unit.id, str.col, PS.col,
                       "PS.new1", clus.id, y.col, "ones", "delta.d")])

                      # Binary version of categorical y with domain
      sam.dat$bin.yvar <- as.numeric((sam.dat[,y.col] == y.val) *
                                      sam.dat$delta.d)

                  # Assign cluster ID for 2 PSUs per stratum design
      sam.dat[, clus.id] <- rep(c(rep(1,nh.sub), rep(2,nh.sub)), H)

                             # Max categories for analysis variable
      y.level <- max(unique(sam.dat[,y.col]))

#__________Point estimates___________________________________

##### PWR/Hajek estimates #####

            # NOTE: Will get warning message if zero occurrences of
            #   characteristic of interest with poststratum

      sam.dsgn <- svydesign(id = as.formula(paste("~", clus.id)),
                       strata = as.formula(paste("~", str.col)),
                      weights = base.wts,
                         data = sam.dat, nest = TRUE)

      PWR.tot <- svyby(as.formula(paste("~interaction(",y.col,")")),
                   as.formula(paste("~interaction(","delta.d",")")),
                   sam.dsgn, svytotal)
      PWR.tot <- as.numeric(PWR.tot[PWR.tot[, 1] == 1, -1][y.val])

      Hajek.mu  <- svyby(as.formula(paste("~interaction(",y.col,")")),
                   as.formula(paste("~interaction(","delta.d",")")),
                   sam.dsgn, svymean)
      Hajek.mu <- as.numeric(Hajek.mu[Hajek.mu[,1] == 1, -1][y.val])

##### Poststratified estimates - mean and total #####

      PS.dsgn <- postStratify(sam.dsgn, strata = ~PS.new1,
                         population = PS.pop.frm, partial = T)

      if(PS.chk == T) {
        ECPS.tot.est <- as.matrix(PS.pop.frm$Tot)
        ECPS.tot.SE  <- 0.
        ECPS.mu.est  <- 0.
        ECPS.mu.SE   <- 0.
      }
      else {
```

```
      ECPS.tot <- svyby(as.formula(paste("~interaction(",y.col,")")),
                  as.formula(paste("~interaction(","delta.d",")")),
                  PS.dsgn, svytotal)
      ECPS.tot.est <- as.numeric(ECPS.tot[ECPS.tot[,1]==1,
                                 -1][y.val])
      ECPS.tot.SE  <- as.numeric(ECPS.tot[ECPS.tot[,1]==1,
                                 -1][(y.val + y.level)])

      ECPS.mu  <- svyby(as.formula(paste("~interaction(",y.col,")")),
                  as.formula(paste("~interaction(","delta.d",")")),
                  PS.dsgn, svymean)
      ECPS.mu.est  <- as.numeric(ECPS.mu[ECPS.mu[,1] == 1,-1][y.val])
      ECPS.mu.SE   <- as.numeric(ECPS.mu[ECPS.mu[,1] == 1,-1][(y.val
                                 + y.level)])
    }

    PS.That <- svytable(as.formula(paste("~",PS.col[1],"+",
                        PS.col[2])), design = sam.dsgn)
    PS.cnt <- xtabs(as.formula(paste("~",PS.col[1],"+",PS.col[2])),
                    data = sam.dat)

    cov.mat[i, ] <- as.vector(PS.That/PS.pop)

                                              # Extract g-weights
    g.wts <- (1/PS.dsgn$prob) / base.wts


#__________Estimated Variance________________________________

##### Estimated-control Taylor Series variance #####

    if(ex.cntrls) {
      if(PS.chk == F) {
                       # Total est'd y within domain by poststratum
        t.Aydg <- as.data.frame(svytable(as.formula(paste("~",
                  y.col, "+", "PS.new1", "+", "delta.d")),
                  sam.dsgn))
        t.Aydg <- as.matrix(t.Aydg[(t.Aydg[,y.col] == y.val &
                  t.Aydg$delta.d == 1),"Freq"])

                          # Total est'd domain total by poststratum
        t.ANdg <- as.data.frame(svytable(as.formula(paste("~",
                  "ones", "+", "PS.new1", "+", "delta.d")),
                   sam.dsgn))
        t.ANdg <- as.matrix(t.ANdg[(t.ANdg$delta.d == 1),"Freq"])

                               # Est'd number in pop by poststratum
```

```
        N.Ag  <- as.matrix(svyby(~ones, ~PS.new1, sam.dsgn,
                           svytotal)[,2])

                                  # Control totals by poststratum
        N.Bg  <- as.matrix(pop.ext[,1])

                            # Est.s under traditional calibration
        N.dgPSGR  <- N.Bg * (1/N.Ag) * t.ANdg
                                          # PS model coefficients
        B.hat.A  <- t.Aydg / N.Ag
        B.hat.AN <- t.ANdg / N.Ag


##### Estimated-control TS variance w/coverage error #####

                     # Merge model coeff vectors onto sample file
        betas <- as.data.frame(cbind(1:16, B.hat.A, B.hat.AN))
        names(betas) <- c("PS.new1","B.hat.A","B.hat.AN")

        resids <- merge(sam.dat[,c("PS.new1","bin.yvar",
                                   "delta.d","ones")],
        betas, by.x="PS.new1", by.y="PS.new1")

  #_____ Sum squared residuals by poststratum (wt=base.wt) ____

                                             # Numerator component
        resids$sqrd.resid.num <- base.wts * (resids$bin.yvar -
                                     resids$B.hat.A)^2
        tot.resids.g.num <- as.matrix(by(resids$sqrd.resid.num,
                                resids$PS.new1, sum))

                                           # Denominator component
        resids$sqrd.resid.den <- base.wts * (resids$ones -
                                     resids$B.hat.AN)^2
        tot.resids.g.den <- as.matrix(by(resids$sqrd.resid.den,
                                resids$PS.new1, sum))

                                            # Covariance component
        resids$sqrd.resid.cov <- base.wts * (resids$bin.yvar -
                                     resids$B.hat.A) *
                                     (resids$ones - resids$B.hat.AN)
        tot.resids.g.cov <- as.matrix(by(resids$sqrd.resid.cov,
                                resids$PS.new1, sum))

  #____ Var component for coverage error ____

        one.minus.phi <- (1 - N.Ag/N.Bg)
```

```
          one.minus.phi[one.minus.phi < 0] <- 0.

          coverr.adj1.tot <- t(one.minus.phi) %*% tot.resids.g.num
          coverr.adj1.mu  <- t(one.minus.phi) %*%
                             (tot.resids.g.num + ECPS.mu.est^2 *
                              tot.resids.g.den - 2 * ECPS.mu.est *
                              tot.resids.g.cov)

##### Residualized ECTS variance #####

          ECTSr.tot <- sqrt(ECPS.tot.SE**2 +
                       (t(B.hat.A) %*% ex.cntrls.cov %*% B.hat.A))

          ECTSr.mu  <- sqrt(ECPS.mu.SE**2 + (1/sum(N.dgPSGR))**2 *
                       ((t(B.hat.A - ECPS.mu.est * B.hat.AN) %*%
                         ex.cntrls.cov %*% (B.hat.A - ECPS.mu.est *
                         B.hat.AN))))

  #_____ Linear SEs for EC estimates with coverage component _____

          ECTSr.totcov <- sqrt(ECTSr.tot**2 + coverr.adj1.tot)
          ECTSr.mucov  <- sqrt(ECTSr.mu**2  + (1/sum(N.dgPSGR))**2 *
                          coverr.adj1.mu)
        }


##### Fuller (1998) Jackknife Method (Not Balanced) #####

  #_____Eigenvalue decomposition_____

        spec.decmp <- eigen(ex.cntrls.cov, symmetric=T)

            #Calculate random components for calibration
            #(columns of the z.matrix corresponds to z(r) in notes)
        lambda <- matrix(spec.decmp$values, byrow=T,
                         nrow=nrow(spec.decmp$vectors),
                         ncol=ncol(spec.decmp$vectors))

        z.matrix <- sqrt(lambda) * spec.decmp$vectors

  #_____JK Adjustments_____

        for(k in 1.:length(nh)) {
          if(k == 1) { PSUs.rep <- c(rep(nh[k],nh[k])) }
          else   { PSUs.rep <- c(PSUs.rep, rep(nh[k],nh[k])) }
        }
        c.h <- sqrt(PSUs.rep / (PSUs.rep - 1))
```

```
      c.h <- matrix(rep(c.h, G), nrow=G, byrow=T)

      R.h <- 1 / sqrt(H * PSUs.rep)
      R.h <- matrix(rep(R.h, G), nrow=G, byrow=T)

#_____Replicate PS controls_____

      col.order <- sample(1:R, R)

      zero.matrix <- matrix(rep(as.matrix(rep(0,G)), R - G),
                           nrow=G)

      z.adj <- cbind(zero.matrix, z.matrix)
      z.adj <- z.adj[, col.order]

      N.hats.B.FUL <- matrix(rep(PS.pop.frm$Tot, R), nrow=G) +
                            (z.adj * c.h)

#_____Matrix of PS group indicators_____

      PS.matrix  <- PSd.matrix <-
          matrix(0., nrow=nrow(sam.dat), ncol=G)

      for(k in 1.:G) {
        PS.matrix[,k]   <- as.numeric(sam.dat$PS.new1 == k)
        PSd.matrix[, k] <- as.numeric((sam.dat$PS.new1 == k) &
                                 (sam.dat$delta.d == 1))
      }

#_____Matrix of PS group indicators x analysis var_____

      if(PS.chk == T) { PS.yvar.mat <- PS.matrix }
      else  { PS.yvar.mat <- matrix(rep(sam.dat$bin.yvar, G), ncol=G)
                                  * PS.matrix }

#_____Replicate weight adjustments_____

      rep.dsgn <- as.svrepdesign(sam.dsgn, type="JKn")
      JKn.adj.wts <- weights(rep.dsgn)

      base.wts.R <- matrix(rep(base.wts, R), byrow=F, ncol=R)
      g.wts.R  <- matrix(rep(g.wts, R), byrow=F, ncol=R)

                                  # Design wt * PSU subsmp wt
      rep.wts <- base.wts.R * JKn.adj.wts

                         # (Design wt * g wt) * PSU subsmp wt
```

```
      rep.calib.wts <- (base.wts.R * g.wts.R) * JKn.adj.wts

#_____PS Slope estimates_____

      for(k in 1.:R) {
                           # Numerator of beta.hat, per PS group
        t.Aydg.rep <- t(PS.yvar.mat) %*% as.matrix(rep.wts[,k])

                         # Denominator of beta.hat, per PS group
        N.Ag.rep   <- t(PS.matrix) %*% as.matrix(rep.wts[,k])

                 # Denominator beta for ratio mean, per PS group
        t.ANdg.rep <- t(PSd.matrix) %*% as.matrix(rep.wts[,k])

                          # Estimated domain totals per PS group
        N.hat_Adg <- t(PSd.matrix) %*% as.matrix(rep.wts[,k])

        if(k == 1) {
          B.hat.Arep  <- as.matrix(t.Aydg.rep / N.Ag.rep)
          B.hat.ANrep <- as.matrix(t.ANdg.rep / N.Ag.rep)
          N.hats.A  <- as.matrix(N.Ag.rep)
        }
        else {
          B.hat.Arep  <- cbind(B.hat.Arep,
                         as.matrix(t.Aydg.rep / N.Ag.rep))
          B.hat.ANrep <- cbind(B.hat.ANrep,
                         as.matrix(t.ANdg.rep / N.Ag.rep))
          N.hats.A  <- cbind(N.hats.A,  as.matrix(N.Ag.rep))
        }
      }
#_____Coverage error variance component_____

              #merge Hajek avg.s per poststratum onto sample file
      betas <- as.data.frame(cbind(B.hat.Arep, B.hat.ANrep,
               PS.new1 = 1:G))
      resids <- merge(sam.dat[,c("PS.new1","bin.yvar","ones",
                "delta.d")], betas, by.x="PS.new1",
                by.y="PS.new1")

         #sum of squared base-wtd residuals by PS (num, den, cov)
      wtd.resids.R.num  <- as.data.frame(cbind(PS.new1 =
                 resids$PS.new1, rep.wts * (resids$bin.yvar -
                 resids[,c(5:(4 + R))])^2))
      tot.resids.gR.num <- gsummary(wtd.resids.R.num, sum,
                 groups=wtd.resids.R.num$PS.new1)

      wtd.resids.R.den <- as.data.frame(cbind(PS.new1 =
```

```
                resids$PS.new1, rep.wts * (resids$ones -
                resids[,c((5 + R):ncol(resids))])^2))
        tot.resids.gR.den <- gsummary(wtd.resids.R.den, sum,
                groups=wtd.resids.R.den$PS.new1)

        wtd.resids.R.cov  <- as.data.frame(cbind(PS.new1 =
                resids$PS.new1, rep.wts * (resids$bin.yvar -
                resids[,c(5:(4 + R))]) * (resids$ones -
                resids[,c((5 + R):ncol(resids))])))
        tot.resids.gR.cov <- gsummary(wtd.resids.R.cov, sum,
                groups=wtd.resids.R.cov$PS.new1)

              #sum of squared g-wtd residuals by PS (num, den, cov)
        g.wtd.resids.R.num <- as.data.frame(cbind(PS.new1 =
                resids$PS.new1, rep.calib.wts *
                (resids$bin.yvar - resids[,c(5:(4 + R))])^2))
        tot.resids.gR.g.num <- gsummary(g.wtd.resids.R.num, sum,
                groups=g.wtd.resids.R.num$PS.new1)

        g.wtd.resids.R.den  <- as.data.frame(cbind(PS.new1 =
                resids$PS.new1, rep.calib.wts * (resids$ones -
                resids[,c((5 + R):ncol(resids))])^2))
        tot.resids.gR.g.den <- gsummary(g.wtd.resids.R.den, sum,
                groups=g.wtd.resids.R.den$PS.new1)

        g.wtd.resids.R.cov  <- as.data.frame(cbind(PS.new1 =
                resids$PS.new1, rep.calib.wts *
                (resids$bin.yvar - resids[,c(5:(4 + R))]) *
                (resids$ones - resids[,c((5 + R):ncol(resids))])))
        tot.resids.gR.g.cov <- gsummary(g.wtd.resids.R.cov, sum,
                groups=g.wtd.resids.R.cov$PS.new1)

          #varcomp for coverage error - fixed (per sample) N.hats.B
        one.minus.phi.R <- (1 - N.hats.A /
                        matrix(rep(PS.pop.frm$Tot, R), nrow=G))
        one.minus.phi.R[one.minus.phi.R < 0] <- 0.

#_____JK variance estimates_____

        stdnorm.gR <- matrix(rnorm(G * R), nrow=G)

        ECF2.method <- Rep.VarEst(
                    PS.chk              = PS.chk,
                    N.hats.B            = N.hats.B.FUL,
                    B.hat.Arep          = B.hat.Arep,
                    B.hat.ANrep         = B.hat.ANrep,
                    PSUs.rep            = PSUs.rep,
```

```
                  one.minus.phi.R      = one.minus.phi.R,
                  tot.resids.gR.num    = tot.resids.gR.num,
                  tot.resids.gR.den    = tot.resids.gR.den,
                  tot.resids.gR.cov    = tot.resids.gR.cov,
                  tot.resids.gR.g.num = tot.resids.gR.g.num,
                  tot.resids.gR.g.den = tot.resids.gR.g.den,
                  tot.resids.gR.g.cov = tot.resids.gR.g.cov,
                  c.h                  = c.h,
                  R.h                  = R.h,
                  stdnorm.gR           = stdnorm.gR,
                  ECPS.tot             = ECPS.tot.est,
                  ECPS.mu              = ECPS.mu.est)

   ECF2.tot    <- ECF2.method$Rep.VarEst.tot
   ECF2.totcov <- ECF2.method$Rep.VarEst.totcov
   ECF2.mu     <- ECF2.method$Rep.VarEst.mu
   ECF2.mucov  <- ECF2.method$Rep.VarEst.mucov
   Vhat.B.ECF2 <- ECF2.method$Vhat.B.est


##### MV Normal Jackknife Method #####

   MV.Norm <- t(mvrnorm(n=R, mu=rep(0, nrow(ex.cntrls.cov)),
                Sigma=ex.cntrls.cov))

   N.hats.B.MVN <- matrix(rep(PS.pop.frm$Tot, R), nrow=G, byrow=F)
                   + c.h * R.h * MV.Norm

   ECMV.method <- Rep.VarEst(
                  PS.chk               = PS.chk,
                  N.hats.B             = N.hats.B.MVN,
                  B.hat.Arep           = B.hat.Arep,
                  B.hat.ANrep          = B.hat.ANrep,
                  PSUs.rep             = PSUs.rep,
                  one.minus.phi.R      = one.minus.phi.R,
                  tot.resids.gR.num    = tot.resids.gR.num,
                  tot.resids.gR.den    = tot.resids.gR.den,
                  tot.resids.gR.cov    = tot.resids.gR.cov,
                  tot.resids.gR.g.num = tot.resids.gR.g.num,
                  tot.resids.gR.g.den = tot.resids.gR.g.den,
                  tot.resids.gR.g.cov = tot.resids.gR.g.cov,
                  c.h                  = c.h,
                  R.h                  = R.h,
                  stdnorm.gR           = stdnorm.gR,
                  ECPS.tot             = ECPS.tot.est,
                  ECPS.mu              = ECPS.mu.est)
```

```
        ECMV.tot    <- ECMV.method$Rep.VarEst.tot
        ECMV.totcov <- ECMV.method$Rep.VarEst.totcov
        ECMV.mu     <- ECMV.method$Rep.VarEst.mu
        ECMV.mucov  <- ECMV.method$Rep.VarEst.mucov
        Vhat.B.ECMV <- ECMV.method$Vhat.B.est


##### Nadimpalli-Judkins-Chu (2004) Jackknife Method #####

        SN <- matrix(rnorm(G * R), nrow=G)

        N.hats.B.NJC <- matrix(rep(PS.pop.frm$Tot, R), nrow=G,
                        byrow=F) + c.h * R.h * SN *
                        matrix(rep(sqrt(diag(ex.cntrls.cov)), R),
                        nrow=G, byrow=F)

       ECNJC.method <- Rep.VarEst(
                       PS.chk                = PS.chk,
                       N.hats.B              = N.hats.B.NJC,
                       B.hat.Arep            = B.hat.Arep,
                       B.hat.ANrep           = B.hat.ANrep,
                       PSUs.rep              = PSUs.rep,
                       one.minus.phi.R       = one.minus.phi.R,
                       tot.resids.gR.num     = tot.resids.gR.num,
                       tot.resids.gR.den     = tot.resids.gR.den,
                       tot.resids.gR.cov     = tot.resids.gR.cov,
                       tot.resids.gR.g.num   = tot.resids.gR.g.num,
                       tot.resids.gR.g.den   = tot.resids.gR.g.den,
                       tot.resids.gR.g.cov   = tot.resids.gR.g.cov,
                       c.h                   = c.h,
                       R.h                   = R.h,
                       stdnorm.gR            = stdnorm.gR,
                       ECPS.tot              = ECPS.tot.est,
                       ECPS.mu               = ECPS.mu.est)

        ECNJC.tot    <- ECNJC.method$Rep.VarEst.tot
        ECNJC.totcov <- ECNJC.method$Rep.VarEst.totcov
        ECNJC.mu     <- ECNJC.method$Rep.VarEst.mu
        ECNJC.mucov  <- ECNJC.method$Rep.VarEst.mucov
        Vhat.B.ECNJC <- ECNJC.method$Vhat.B.est
      }

#__________Save Results___________________________________

      if(PS.chk == T) {
        out.tot[i, ] <- c(as.vector(T.pop)[1], sum(ECPS.tot),
                          PWR.tot[1],  rep(0,15))
```

```
          out.mu[i, ]  <- c(as.vector(P.pop)[1], sum(ECPS.mu),
                            Hajek.mu[1],   rep(0,15))
        }
        else {
          out.tot[i, ] <- c(as.vector(T.pop)[1],
                            ECPS.tot.est,  PWR.tot[1],
                            ECPS.tot.SE,
                            ECTSr.tot,     ECF2.tot,
                            ECMV.tot,      ECNJC.tot,
                            ECTSr.totcov,  ECF2.totcov,
                            ECMV.totcov,   ECNJC.totcov)
          out.mu[i, ]  <- c(as.vector(P.pop)[1],
                            ECPS.mu.est,   Hajek.mu[1],
                            ECPS.mu.SE,
                            ECTSr.mu,      ECF2.mu,
                            ECMV.mu,       ECNJC.mu,
                            ECTSr.mucov,   ECF2.mucov,
                            ECMV.mucov,    ECNJC.mucov)
        }

        if(!skip.sw) {
          keep.sw <- TRUE
        }

      }   # skip.sw
    }   # keep.sw
  }   # no.sams

  cat("end ", date(), "\n")
  c.rate <- apply(cov.mat, 2, mean)
  c.rate <- matrix(c.rate, nrow = dim(PS.pop)[1],
                   ncol = dim(PS.pop)[2], byrow = FALSE)

  list(seed         = seed,
       num.skip.sam = num.skip.sam,
       c.rate       = round(c.rate,2),
       Vhat.B.ECF2  = Vhat.B.ECF2,
       Vhat.B.ECNJC = Vhat.B.ECNJC,
       Vhat.B.ECMV  = Vhat.B.ECMV,
       out.tot      = out.tot,
       out.mu       = out.mu)
}
```

A.5 Replicate Variance Estimates

```
Rep.VarEst <- function(PS.chk, N.hats.B, B.hat.Arep, B.hat.ANrep,
                       PSUs.rep, one.minus.phi.R, tot.resids.gR.num,
                       tot.resids.gR.den, tot.resids.gR.cov,
                       tot.resids.gR.g.num, tot.resids.gR.g.den,
                       tot.resids.gR.g.cov, c.h, R.h, stdnorm.gR,
                       ECPS.tot, ECPS.mu) {

  # General code to calculate replicate variance estimates
  #
  # PS.chk              = T/F if code invoked to check reproduction
  #                       of benchmark covar matrix
  # N.hats.B            = adjusted benchmark est's specific to
  #                       EC method
  # B.hat.Arep          = sample model coefficients (est'd total)
  # B.hat.ANrep         = sample model coefficients for
  #                       denominator of (ratio) mean
  # PSUs.rep            = number of replicates per stratum
  # one.minus.phi.R     = coverage error adjustment
  # tot.resids.gR.num   = base-wtd sqrd residuals for numerator of
  #                       ratio mean
  # tot.resids.gR.den   = base-wtd sqrd residuals for denominator of
  #                       ratio mean
  # tot.resids.gR.cov   = base-wtd sqrd residuals for covariance of
  #                       ratio mean
  # tot.resids.gR.g.num = g-wtd sqrd residuals for numerator of
  #                       ratio mean
  # tot.resids.gR.g.den = g-wtd sqrd residuals for denominator of
  #                       ratio mean
  # tot.resids.gR.g.cov = g-wtd sqrd residuals for covariance of
  #                       ratio mean
  # c.h                 = sqrt(m_Ah / (m_Ah - 1))
  # R.h                 = sqrt(1 / (H * m_Ah))
  # stdnorm.gR          = standard normal random values
  # ECPS.tot            = poststratified estimate of total
  #                       (centering value)
  # ECPS.mu             = poststratified estimate of ration mean
  #                       (centering value)

  if(PS.chk == T) {
    tot.reps <- N.hats.B * B.hat.Arep

    for(k in 1.:R) {
      if(k == 1) {
        Vhat.B.cmp <- ((PSUs.rep[k] - 1) / PSUs.rep[k]) *
                      ((tot.reps[,k] - ECPS.tot) %*%
                      t(tot.reps[,k] - ECPS.tot)) }
```

```
      else {
        Vhat.B.cmp <- Vhat.B.cmp + ((PSUs.rep[k] - 1) / PSUs.rep[k]) *
                      ((tot.reps[,k] - ECPS.tot) %*%
                      t(tot.reps[,k] - ECPS.tot)) }
    }
    Rep.VarEst.tot <- 0.
    Vhat.B.cmp <- Vhat.B.cmp / no.sams
    if(i == 1.) {Vhat.B.est <- Vhat.B.cmp}
    else        {Vhat.B.est <- Vhat.B.est + Vhat.B.cmp}
  }
  else {
                             #_____No coverage error component_____

                                                     #Estimated totals
    tot.reps <- apply(N.hats.B * B.hat.Arep, 2, sum)

    diff.vec <- as.matrix(tot.reps - ECPS.tot)
    Rep.VarEst.tot <- sqrt(t(diff.vec) %*%
                    (diff.vec * as.matrix((PSUs.rep - 1) / PSUs.rep)))
    Vhat.B.est <- 0.

#---------------------------------------------------------------------
                                                #Estimated ratio means
    Nhat.reps <- apply(N.hats.B * B.hat.ANrep, 2, sum)
    mu.reps   <- tot.reps / Nhat.reps
    diff.vec  <- as.matrix(mu.reps - ECPS.mu)
    Rep.VarEst.mu  <- sqrt(t(diff.vec) %*%
                   (diff.vec * as.matrix((PSUs.rep - 1) / PSUs.rep)))

                                #_____Coverage error component_____

                                                     #Estimated totals
    coverr.adj1.tot <- one.minus.phi.R * tot.resids.gR.num[,-1]
    coverr.adj2.tot <- one.minus.phi.R * tot.resids.gR.g.num[,-1]

                                                #Estimated ratio means
    coverr.adj1.mu  <- one.minus.phi.R *
                       (tot.resids.gR.num[,-1] + ECPS.mu^2 *
                        tot.resids.gR.den[,-1] -
                        2 * ECPS.mu * tot.resids.gR.cov[,-1])
    coverr.adj2.mu  <- one.minus.phi.R *
                       (tot.resids.gR.g.num[,-1] + ECPS.mu^2 *
                        tot.resids.gR.g.den[,-1] - 2 * ECPS.mu *
                        tot.resids.gR.g.cov[,-1])

                                                     #Estimated totals
    tot.reps.c1        <- apply(N.hats.B * B.hat.Arep +
```

```
                          c.h * R.h * stdnorm.gR *
                          sqrt(coverr.adj1.tot), 2, sum)
    diff.vec           <- as.matrix(tot.reps.c1 - ECPS.tot)
    Rep.VarEst.totcov  <- sqrt(t(diff.vec) %*%
                          (diff.vec * as.matrix((PSUs.rep - 1)
                           / PSUs.rep)))

                                                     #Estimated ratio means
    Nhat.reps.c1  <- apply(N.hats.B * B.hat.ANrep +
                     c.h * R.h * stdnorm.gR *
                     sqrt(coverr.adj1.mu), 2, sum)
    mu.reps   <- tot.reps.c1 / Nhat.reps.c1
    diff.vec  <- as.matrix(mu.reps - ECPS.mu)
    Rep.VarEst.mucov  <- sqrt(t(diff.vec) %*%
                         (diff.vec * as.matrix((PSUs.rep - 1)
                         / PSUs.rep)))
  }

  list(Rep.VarEst.tot    = Rep.VarEst.tot,
       Rep.VarEst.totcov = Rep.VarEst.totcov,
       Rep.VarEst.mu     = Rep.VarEst.mu,
       Rep.VarEst.mucov  = Rep.VarEst.mucov,
       Vhat.B.est        = Vhat.B.est)
}
```

A.6 Generate Analytic Survey Sampling Frames

```
cov.rate <- function(pop, c.prob, cells) {

  # Assign coverage indicators at pop level.  Written by R.Valliant.
  #
  # pop    = population
  # c.prob = vector of coverage probs - must be in the numeric
  #          order of coverage cells
  # cells  = name of col in pop that gives coverage cells

  N <- nrow(pop)
  Nc <- table(pop[, cells])
  H <- length(unique(pop[, cells]))
  cell.id <- sort(unique(pop[, cells]))
  cell.list <- pop[, cells]

  p.cov <- rep(0., N)
  for(h in cell.id) {
```

```
    p.cov[cell.list == h] <- c.prob[(1:length(cell.id))[cell.id==h]]
  }
  c.sw <- (runif(N) <= p.cov)
  c.sw
}
```

A.7 Select Analytic Survey Samples

```
clus.sam <- function(pop, clus.id, unit.id, n.cl, sel.meth, substrat,
                     n.substrat, sub.vals, cert.PSUs) {

  # Select a two-stage cluster sample after randomizing order
  # of the clusters.  Code written by R.Valliant
  #
  # pop        = population matrix
  # clus.id    = name / number of column for cluster identification
  # unit.id    = variable to indicate unique units of observation
  # n.cl       = no. of sample clusters
  # sel.meth   = "ppswr" for pps cluster sample
  #            = "srs" for simple random sample of clusters
  # substrat   = the substratum variable (HISP)
  # n.substrat = a vector of sample sizes for substrat
  # sub.vals   = the values of substrat (for HISP = (0,1))
  # cert.PSUs  = T/F if size > nh.sub, select all units within PSU

  Mi.vec <- table(pop[, clus.id])
  M <- sum(Mi.vec)
  N <- length(Mi.vec)#

  #_____ Select sample of clusters _____

  if(sel.meth == "ppswr") {
    cl.sam <- sample(1:N, n.cl, replace = TRUE, prob = Mi.vec/M) }
  if(sel.meth == "srs") {
    cl.sam <- sort(sample(1:N, n.cl, replace = TRUE, prob = Mi.vec)) }

  cl.sam.id <- names(Mi.vec)[cl.sam]
  Mi.sam <- Mi.vec[cl.sam]

  #_____ Calculate Cluster selection probabilities _____

  if (sel.meth == "ppswr"){
    phi  <- n.cl*Mi.sam/M }
```

```
#_____ Select subsamples from sample clusters _____

cl.sam.data <- matrix(0,nrow=n.cl*sum(n.substrat),ncol=2)
cc <- cumsum(c(0,n.substrat))
n.subtot <- sum(n.substrat)
N.hat <- NULL
drop.sw <- FALSE

for(i in 1:n.cl) {
  sam <- match(as.numeric(pop[, clus.id]), cl.sam.id[i],
               nomatch = 0)
  sam[sam > 0] <- 1
  sam.rows <- (1:M)[sam == 1]
  phi.c <- rep(phi[i],n.subtot)
  Mi.sub <- table(pop[sam.rows,substrat])
  sam.id <- vector("numeric", sum(n.substrat))

  for(ss in 1:length(n.substrat)) {
    S1 <- sam.rows[pop[sam.rows, substrat]==sub.vals[ss]]

                       # Check that PSU pop count >= subsample size
    if (n.substrat[ss] > length(S1)){
      if(!cert.PSUs) {
        cat("i=", i, "n.substrat[ss]=", n.substrat[ss],
            "length(S1)=", length(S1),"\n")
        cat("cl.sam", cl.sam, "\n")
        drop.sw <- TRUE
        phij <- 0.
      }
      else{
        phij <- 1.
        subsam.vec <- S1
        sam.id[(cc[ss]+1):cc[ss+1]] <- pop[subsam.vec, unit.id]
      }
    }
    else{
      phij <- n.substrat/Mi.sub
      subsam.vec <- sample(S1, n.substrat[ss])
      sam.id[(cc[ss]+1):cc[ss+1]] <- pop[subsam.vec, unit.id]
    }
  }
  phij.vec <- rep(phij,n.substrat)
  wij.vec <- 1/(phij.vec*phi.c)
  N.hat <- rbind(N.hat,by(wij.vec,names(wij.vec),sum))
  cl.sam.data[((i-1)*n.subtot+1):(i*n.subtot),] <-
    cbind(sam.id, wij.vec)
}
```

```
  list(cl.sam.data=cl.sam.data, N.hat=apply(N.hat,2,sum),
       drop.sw=drop.sw)
}
```

A.8 Check Poststratum Sizes

```
chk.PS <- function(sdat, cl.all, cl.col, r.sw, min.size = 0) {

  # Check to see whether all poststrata are in sample and have
  # minimum sample sizes.  Code written by R.Valliant.
  #
  # sdat     = matrix of sample data
  # cl.all   = vector of all PS in pop
  # cl.col   = column of sdat for PS
  # r.sw     = vector of coverage indicators for sdat sample units
  # min.size = minimum sample size allowed per poststratum

  skip.sw <- FALSE

  cl.sam <- unique(sdat[, cl.col])
  if(!all(is.element(cl.all, cl.sam))) {
    skip.sw <- TRUE
  }

  cnt <- table(sdat[, cl.col], as.numeric(r.sw))
  if(any(cnt < min.size)) {
    skip.sw <- TRUE
  }
  skip.sw
}
```

A.9 Simulation Analysis Program

```
ECPS.SimStats <- function(adj, pop.val, ds.name, estr="tot") {

  # Calculate summary stats for simulation runs
  #
  # adj      = covariance adjustment factor number
  # pop.val  = population estimate used in comparisons
  # ds.name  = name of data file containing sim results
```

```
# estr     = type of estimator

#_____Program parameters_____

  max.col <- ncol(ds.name)

  varnm.lst <- c("CovAdj", colnames(ds.name)[-c(1:3)])

#_____Absolute Biases of Point Estimates_____

  AbsBias.Pt <- c(adj, apply(abs(ds.name[,c(2,3)] -
                  ds.name[,pop.val]), 2, mean, na.rm=T))
  names(AbsBias.Pt)[1] <- "CovAdj"

#_____RelBiases of Point Estimates_____

  RelBias.Pt <- c(adj, apply((ds.name[,c(2,3)] - ds.name[,pop.val]) /
                  ds.name[,pop.val], 2, mean, na.rm=T))
  names(RelBias.Pt)[1] <- "CovAdj"

  RelBias2.Pt <- c(adj, (apply(ds.name[,c(2,3)], 2, mean, na.rm=T) -
                   ds.name[1,pop.val]) / ds.name[1,pop.val])
  names(RelBias2.Pt)[1] <- "CovAdj"

#_____Estimated Bias Ratios_____

  biasratio.R <- (ds.name[, 2] - ds.name[, pop.val]) /
                  ds.name[, 4:max.col]
  BiasRatio <- c(adj, apply(biasratio.R, 2, mean, na.rm=T))
  names(BiasRatio) <- varnm.lst

  biasratio.R <- abs(ds.name[1:10, 2] - ds.name[1:10, pop.val]) /
                     ds.name[1:10, 4:max.col]
  absBiasRatio <- c(adj, apply(biasratio.R, 2, mean, na.rm=T))
  names(absBiasRatio) <- varnm.lst

#_____Square Root of MSE_____

  RtMSE <- c(adj, sqrt(apply((ds.name[,c(2,3)] - ds.name[,pop.val])^2,
            2, mean, na.rm=T)))
  names(RtMSE)[1] <- "CovAdj"

#_____Average Estimated SE_____

  AvgSE <- c(adj, apply(ds.name[,4:max.col], 2, mean, na.rm=T))
  names(AvgSE)[1] <- "CovAdj"
```

```
#_____Stability of Estimated SE_____

  StdErr.SE   <- c(adj, sqrt(apply(ds.name[,4:max.col], 2,
                  var, na.rm=T)))
  names(StdErr.SE)[1] <- "CovAdj"

#_____RelBias of Avg Estimated Variances to MSE_____

  RelBias.Var.mse <- c(adj, (apply(ds.name[,4:max.col]^2, 2,
                      mean, na.rm=T) /
                      apply((ds.name[,rep(2, (max.col - 3))] -
                      ds.name[,pop.val])^2, 2, mean, na.rm=T) - 1))
  names(RelBias.Var.mse)[1] <- "CovAdj"

#_____RelBias of Avg Estimated Variances to Var(t.hat)_____

  var.t.hat <- var(ds.name[,2])
  RelBias.Var.vrt <- c(adj, ((apply(ds.name[,4:max.col]^2, 2,
                      mean, na.rm=T) - rep(var.t.hat, (max.col - 3))) /
                      rep(var.t.hat, (max.col - 3))))
  names(RelBias.Var.vrt)[1] <- "CovAdj"

#_____Estimated MSE_____

  mse.R <- ds.name[,4:max.col]^2 + (ds.name[,rep(2, (max.col - 3))]
                  - ds.name[,pop.val])^2
  RelBias.mse <- c(adj, (apply(mse.R, 2, mean, na.rm=T) /
                        apply((ds.name[,rep(2, (max.col - 3))] -
                        ds.name[,pop.val])^2, 2, mean, na.rm=T) - 1))
  names(RelBias.mse)[1] <- "CovAdj"

#_____Simulation SE of Point Estimates_____

  EmpSE <- c(adj, sqrt(apply(ds.name[,c(2,3)], 2, var, na.rm=T)))
  names(EmpSE)[1] <- "CovAdj"

#_____Empirical 95% CI Coverage_____

  t.stat    <- (ds.name[,rep(2, (max.col - 3))] - ds.name[,pop.val])
               / ds.name[,4:max.col]

  CI.Cov    <- c(adj, apply(abs(t.stat) <= qt(0.975, df = 25), 2,
                mean, na.rm=T))
  names(CI.Cov) <- varnm.lst

#_____Coefficient of Variation_____
```

```
    CV.var <- c(adj, sqrt(apply(ds.name[,4:max.col], 2, var, na.rm=T))
                / apply(ds.name[,4:max.col], 2, mean, na.rm=T))
    names(CV.var)  <- varnm.lst


    list(AbsBias.Pt      = AbsBias.Pt,
         RelBias.Pt      = RelBias.Pt,
         RelBias2.Pt     = RelBias2.Pt,
         BiasRatio       = BiasRatio,
         absBiasRatio    = absBiasRatio,
         RtMSE           = RtMSE,
         AvgSE           = AvgSE,
         StdErr.SE       = StdErr.SE,
         RelBias.Var.mse = RelBias.Var.mse,
         RelBias.Var.vrt = RelBias.Var.vrt,
         RelBias.mse     = RelBias.mse,
         EmpSE           = EmpSE,
         CI.Cov          = CI.Cov,
         CV.var          = CV.var)
}
```

Bibliography

Ash, S. (2003). Simultaneous calibration estimators for two-phase samples. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 395–400).

Axelson, M. (2000). Variance estimation for the two-phase regression estimator - a calibration approach. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 799–804).

Binder, D. A. (1995). Linearization methods for single phase and two-phase samples: A cookbook approach. *Survey Methodology*, *22*(1), 17–22.

Blumberg, S. J., & Luke, J. V. (2008). Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2007. White paper published by the Division of Health Interview Statistics, National Center for Health Statistics.
URL www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200805.pdf

Bolling, K., Grant, C., & Sinclair, P. (2006). 2006-07 British Cime Survey (England and Wales), Technical Report Volume I. Tech. rep., Home Office, UK - Research, Development and Statistics area of the Crime Reduction and Community Safety Group.
URL http://rds.homeoffice.gov.uk/rds/pdfs07/bcs0607tech1.pdf

Bray, R., Hourani, L., Rae, K., Dever, J., Brown, J., Vincus, A., Pemberton, M., Marsden, M., Faulkner, D., & Vandermaas-Peeler, R. (2003). 2002 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Tech. Rep. RTI/7841/006-FR, U.S. Department of Defense prepared by RTI International.
URL http://dodwws.rti.org/2002WWFinalReportComplete05-04.pdf.

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. California: Duxbury Press, 2nd ed.

Centers for Disease Control and Prevention (2006). *Technical Information and Data for the Behavioral Risk Factor Surveillance System (BRFSS) - BRFSS Weighting Formula*. U.S. Department of Health and Human Services, Atlanta, GA.

Chambers, R. L., & Skinner, C. J. (2003). *Analysis of Survey Data*. England: John Wiley and Sons, Inc.

Chao, A., & Tsay, P. (1998). A sample coverage approach to multiple-system estimation with application to census undercount. *Journal of the American Statistical Association*, *93*(441), 283–293.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376–382.

Duchesne, P. (2003). Estimation of a proportion with survey data. *Journal of Statistics Education*, *11*(3), 1–24.
URL `http://www.amstat.org/publications/jse/v11n3/duchesne.pdf`

Estevao, V. M., & Särndal, C.-E. (2000). A functional form approach to calibration. *Journal of Official Statistics*, *16*(4), 379–399.

Estevao, V. M., & Särndal, C.-E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *Journal of Official Statistics*, *18*(2), 233–255.

Estevao, V. M., & Särndal, C.-E. (2004). Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *Journal of Official Statistics*, *20*(4), 645–669.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sakmpling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 598–603).

Fuller, W. A. (1998). Replication variance estimation for two-phase samples. *Statistica Sinica*, *8*, 1153–1164.

Fuller, W. A. (2000). Two-phase sampling. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 23–30).

Fuller, W. A. (2004). Replication variance estimation for the two-phase regression estimator. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 3525–3531).

Fuller, W. A., & Rao, J. (2001). A regression composite estimator with application to the canadian labour force survey. *Survey Methodology*, *27*(1), 45–51.

Gonzalez, J. M., & Eltinge, J. L. (2007). Multiple matrix sampling: A review. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 3069–3075).

Griffiths, R. (2007). Stratification for radio listening estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 1–8).

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley and Sons, Inc.

Hájek, J. (1971). *Discussion of "An essay on the logical foundations of survey sampling, Part 1" by D. Basu in Foundations of Statistical Inference*. Toronto: Holt, Rinehart, and Winston.

Hansen, M. H., & Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, *14*, 333–362.

Hedlin, D., Dalvey, H., Chambers, R., & Kokic, P. (2001). Does the model matter for greg estimation? a business survey example. *Journal of Official Statistics*, *17*(4), 527–544.

Hidiroglou, M. A., & Patak, Z. (2004). Domain estimation using linear regression. *Survey Methodology*, *30*(1), 67–78.

Hidiroglou, M. A., & Patak, Z. (2006). Raking ratio estimation: An application to the canadian retail trade survey. *Journal of Official Statistics*, *22*(1), 71–80.

Horvitz, D., & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

Huang, E., & Fuller, W. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section, American Statistical Association*, (pp. 300–305).

Isaki, C. T., Tsay, J. H., & Fuller, W. A. (2004). Weighting sample data subject to independent controls. *Survey Methodology*, *30*(1), 35–44.

Jayasuriya, B. R., & Valliant, R. (1996). An application of regression and calibration estimation to post-stratification in a household survey. *Survey Methodology*, *22*(2), 127–137.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society, Series A (General)*, *142*(2), 210–222.

Kim, J. J., Li, J., & Valliant, R. (2007). Cell collapsing in poststratification. *Survey Methodology*, *33*(2), 139–150.

Kim, J. K., & Sitter, R. R. (2003). Efficient replication variance estimation for two-phase sampling. *Statistica Sinica*, *13*, 641–653.

Korn, E. L., & Graubard, B. I. (1999). *Analysis of Health Surveys*. John Wiley and Sons, Inc.

Kott, P. S. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, *17*, 521–52617.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, *32*(2), 133–142.

Krewski, D., & Rao, J. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, *9*(5), 1010–1019.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag, Inc.

Lehtonen, R., & Veijanen, A. (1998). Logistic generalized regression estimators. *Survey Methodology*, *24*(1), 51–56.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling Error in Surveys*. John Wiley and Sons, Inc.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. New York: Duxbury Press.

Lohr, S. L., & Prasad, N. (2003). Small area estimation with auxiliary survey data. *The Canadian Journal of Statistics*, *31*(4), 383–396.

Lumley, T. (2005). Survey: analysis of complex survey samples. *R package version 3.0-1, University of Washington: Seattle*.

Lundström, S., & Särndal, C.-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, *15*(2), 305–327.

Nadimpalli, V., Judkins, D., & Chu, A. (2004). Survey calibration to cps household statistics. *Proceedings of the Survey Research Methods Section, American Statistical Association*, (pp. 4090–4094).

National Center for Health Statistics (2006). 2005 National Health Interview Survey (NHIS) Public Use Data Release NHIS Survey Description. Tech. rep., Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, Maryland: U.S.
URL `http://www.cdc.gov/nchs/nhis.htm`

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Development Core Team, Vienna, Austria.
URL `http://www.R-project.org`

Rao, J. (1973). On double sampling for stratification and analytical surveys. *Biometrika*, *60*, 125–133.

Rao, J. (1997). Developments in sample survey theory: An appraisal. *The Canadian Journal of Statistics*, *25*(1), 1–21.

Rao, J. (2003). *Small Area Estimation*. Hoboken, NJ: John Wiley and Sons, Inc.

Rao, J., & Shao, J. (1999). Modified balanced repeated replication for complex survey data. *Biometrika*, *86*(2), 403–415.

Rao, J., & Wu, C. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, *80*(391), 620–630.

Renssen, R. H., & Nieuwenbroek, N. J. (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, *92*(437), 368–374.

Research Triangle Institute (2004). *SUDAAN Language Manual, Release 9.0*. Research Triangle Institute, Research Triangle Park, NC.
URL http://www.rti.org/sudaan/index.cfm

Rust, K. F., & Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, *5*, 283–310.

SAMHSA (2007). Results from the 2006 National Survey on Drug Use and Health: National Findings (NSDUH Series H-32, DHHS Publication No. SMA 07-4293). Tech. rep., Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies, Department of Health and Human Services (DHHS), Rockville, MD.
URL uh/2k6Results.pdf

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, *33*(2), 99–119.

Särndal, C.-E., & Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. England: John Wiley and Sons, Inc.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag, Inc.

Särndal, C.-E., Swensson, B., & Wretman, J. H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, *76*, 527–537.

SAS Institute Inc. (2004). *SAS/STAT 9.1 User's Guide, Volumes 1-7*. SAS Institute Inc, Cary, NC.
URL http://www.sas.com/

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley and Sons, Inc.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, Inc.

Singh, A. C. (1996). Combining information in survey sampling by modified regression. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 120–129).

Singh, A. C., Dever, J. A., & Iannacchione, V. G. (2004). Composite response rates for surveys with nonresponse follow-up. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 4343–4350).

Singh, A. C., Iannacchione, V. G., & Dever, J. A. (2003). Efficient estimation for surveys with nonresponse follow-up using dual-frame calibration. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 3919–3930).

Singh, A. C., & Mohl, C. A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, *22*, 107–115.

Singh, A. C., & Wu, S. (1996). Estimation for multiframe complex surveys by modified regression. *Proceedings of the Statistical Society of Cananda, Survey Methods Section*, (pp. 69–77).

Singh, A. C., & Wu, S. (2003). An extension of generalized regression estimator for dual-frame surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 3911–3918).

Smith, T. M. (1991). Post-stratification. *The Statistician, Special Issue: Survey Design, Methodology and Analysis (2)*, *40*(3), 315–323.

StataCorp (2005). *Stata Statistical Software: Release 9*. StataCorp, College Station, TX.
URL `http://www.stata.com/capabilities/svy.html`

Stukel, D. M., Hidiroglou, M. A., & Särndal, C.-E. (1996). Variance estimation for calibration estimators: A comparison of jackknifing versus taylor linearization. *Survey Methodology*, *22*(2), 117–125.

Sweden, S. (2005). Flow statistics from the swedish labour force survey. Tech. rep., Statistics Sweden, Stockholm: Sweden.

Taylor, M. F., Brice, J., Buck, N., & Prentice-Lane, E. (2007). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. University of Essex., Colchester.

Terhanian, G., Bremer, J., Smith, R., & Thomas, R. (2000). Correcting data from online survey for the effects of nonrandom selection and nonrandom assignment. *Research Paper: Harris Interactive*.

Théberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, *94*(446), 635–644.

Tracy, D., Singh, S., & R., A. (2003). Note on calibration in stratified and double sampling. *Survey Methodology*, *29*(1), 99–104.

UCLA Center for Health Policy Research (2006). California Health Interview Survey (CHIS) - weighting and estimation of variance in the CHIS public use files. Tech. rep., Los Angeles, California: University of California (UCLA).

U.S. Census Bureau (2002). Sources and accuracy of estimates for poverty in the United States: Washington, DC.

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, *88*, 89–96.

Valliant, R., Brick, J. M., & Dever, J. A. (2008). Weight adjustments for the grouped jackknife variance estimator. *Journal of Official Statistics*, (p. in print).

Weinberg, D. (2006). Income data quality issues in the cps. *Monthly Labor Review Online*, *129*(6), 38–45.

West, K., Robinson, G., & Bentley, M. (2005). Did proxy respondents cause age heaping in the census 2000? *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 3658–3665).

Westat (2000). *WesVar 4.0 Users Guide*. Rockville, MD.

Wolter, K. M. (2007). *Introduction to Variance Estimation*. New York: Springer Science+Business Media, LLC.

Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, *66*(334), 411–414.

Wu, C. (1985). Variance estimation for the combined ratio and combined regression estimators. *Journal of the Royal Statistical Society, Series B*, *47*(1), 147–154.

Yung, W., & Rao, J. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, *22*(1), 23–31.

Yung, W., & Rao, J. (2000). Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, *95*(451), 903–915.