A Robust Multivariate Estimator with Stepwise Covariate Selection and Inequality Constraints for Complex Sample Surveys: An Initial Concept

Environmetrika Technical Report 2020-1

DOI: 10.13140/RG.2.2.17634.07369

March 29, 2020

Raymond L. Czaplewski

Tele. +970-218-5907

Raymond_Czaplewski@Environmetrika.org

ORCID ID 0000-0001-5259-5593

A Contribution to the Environmetrika Series on

Simple Robust Multivariate Estimators for Complex Sample Surveys

This page left intentionally blank

 DOI: 10.13140/RG.2.2.17634.07369

Extended Abstract: This Technical Report introduces a new multivariate difference-estimator for complex sample-surveys. It is an alternative to conventional model-assisted estimators that use specific inference.

Model-assisted estimators and the new difference-estimator both reduce variance in population estimates for M study-variables by using population statistics for J correlated auxiliary-variables, where M and J can number in the hundreds or thousands. Both are closely related to the difference-estimator offered by Särndal *et al*. (1992), although the new difference-estimator uses a different stochastic model. Both employ linear transformations of design-based estimators (e.g., Horvitz-Thompson). Both choose coefficients for a $M\times(M+J)$ transformation matrix that minimize variance of population estimates for each study-variable, where the degree of variance-reduction depends upon the specific correlation between each study-variable and each auxiliary-variable. Both estimators support expansion factors, which facilitate small-area estimators.

The new difference-estimator introduces a new approach to variance-reduction with auxiliary data. Unlike model-assisted estimators, which require known population parameters for the J auxiliary-variables, the new estimator accommodates sample-survey estimates of those population parameters. Therefore, the new difference-estimator can directly use population estimates for auxiliary-variables from more complex sample-surveys, including components such as multi-phase and multi-stage sampling-designs, cluster plots, interpenetrating panels, and supplemental surveys.

The new difference-estimator introduces numerical advances. The model-assisted estimator with specific inference requires inversion of the $J\times J$ covariance matrix for population estimates of J auxiliary-variables; and that matrix inverse is infeasible or numerically unstable if the covariance matrix is rank-deficient or ill-conditioned. The new difference-estimator incorporates a recursive method; it replaces that $J\times J$ matrix inverse with up to J scalar inverses. The j^{th} step in the recursion $\left(j=1,\ldots,J\right)$ minimizes variances of all M study-variables with the j^{th} scalar auxiliary-residual; and it removes any collinearity between the $1,\ldots,j$ auxiliary-variables and all remaining $\left(j+1\right),\ldots,J$ auxiliary-variables. The recursion ceases at $j<J$ if the j^{th} scalar inverse is numerically unstable (*i.e.*, division by a very small number). This suggests the $J\times J$ covariance matrix has rank $(j-1)$, and all auxiliary information is essentially exhausted after recursions with the first $(j-1)$ auxiliary-variables.

 DOI: 10.13140/RG.2.2.17634.07369

The recursive method used in the new difference-estimator simplifies nonlinear estimation procedures, such as inequality constraints on population estimates for each study-variable and protection from negative variance estimates. The recursive method easily implements procedures that mitigate risks from outliers and overfitting with numerous auxiliary-variables. The recursive method supports stepwise covariate-selection among the auxiliary-variables that reduces variance for the most important study-variables as identified by the analyst.

Internal consistency in statistical tables requires that the sum of population estimates in each row or column equals the population estimate for the corresponding margin in that table. Model-assisted estimators with a generic weight produce internal consistency, but at the cost of statistical efficiency. The new difference-estimator provides an alternative that does not compromise statistical efficiency; it uses recursive raking to sequentially impose equality constraints on each row and column of a statistical table.

 DOI: 10.13140/RG.2.2.17634.07369

Contents

1 Introduction 7
1.1 Model-Assisted Difference-Estimator 9
1.2 New Difference-Estimator 10
1.3 Outline 12
2 New Difference-Estimator 15
2.1 The Stochastic Model 18
2.2 Multivariate π-Estimator 19
2.3 Sufficient Statistics for the Difference-estimator 22
2.4 Difference-Estimator with Arbitrary Coefficients 23
2.5 Minimum-Variance Coefficients for Difference-estimator 25
2.6 Implementation Obstacles 33
3 Recursive Difference-Estimator 35
3.1 Pivot Notation 36
3.2 Recursive Sequence, j=1,…,J 37
3.3 Difference-Estimator for j^{th} Recursion 39
3.4 "Well Chosen" Coefficients in the j^{th} Recursion 43
3.4.1 Minimum-Variance Coefficients 45
3.4.2 Orthogonal Auxiliary-Residuals (Collinearity) 48
4 Concatenated Coefficients from the Recursive Sequence 50
5 Stepwise Covariate-Selection 50
5.1 Modifications to Minimum-Variance Coefficients 52
5.2 Termination of Recursive Sequence 52
6 Constraints for a Robust Difference-Estimator 53
6.1 Outlier Auxiliary-Residuals 55
6.1.1 Constraint for Implausibly Large Auxiliary-Residual 56
6.1.2 Constraint on Interval Coverage of Zero 57
6.1.3 Constraint on the i^{th} Population Estimate 59

 DOI: 10.13140/RG.2.2.17634.07369

6.2 Min-Max Inequality Constraints 60
6.3 Feasible Communalities and Nonnegative Variances 62
6.4 Avoidance of Variance Increase with Difference-Estimator 66
6.5 Minimum Coefficient for Significant Variance-Reduction 67
6.6 Over-Fitting with Numerous Auxiliary-variables 69
7 Improved Covariance Matrix with Bootstrap Estimator 70
8 Computational Efficiency 72
9 Pre-and Post-Processing 73
9.1 Constraints on initial π Estimates 74
9.2 Standardization of Parameter-Space 74
9.3 Reduction in Number of Study-variables 75
10 Merger of π Estimates with Recursive-Estimates 76
11 Statistical Efficiency 77
12 Margin Additivity 82
13 Pseudo-Expansion Factors 83
14 Discussion 86
14.1 Stochastic Model 86
14.2 Constraints 87
14.3 Matrix Structure 88
14.4 Numerics 89
14.5 Forestry Examples 89
14.6 Relationship to the Kalman Filter and Related Estimators 92
14.7 Model-Based Inference 94
15 Summary 97
16 References 99
Notation 104

 DOI: 10.13140/RG.2.2.17634.07369

1 Introduction

Consider a detailed cross-sectional and longitudinal study of an extensive population. For example, the US Forest Service's national Forest Inventory and Analysis (FIA) Program (Bechtold and Patterson, 2005) requires reliable information on the status, changes, and trends in land use, land cover, land ownership, resource management, stand conditions, tree demographics, biomass, and indicators of forest health for large geographic areas (*e.g.*, the State of Utah, 22-million ha.).

The total number (N) of population elements can be large. In the FIA example, treat each population element as a 0.4-ha hexagon, where millions of such elements tessellate this two-dimensional spatial population. The analytical study requires population statistics for multiple "study-variables" (*i.e.*, response variables), which are cross-classified into numerous domains, conditions, sub-populations and time periods. The resulting number of study-variables (M) can in the hundreds (McConville *et al.*, 2020) or even thousands.

Assume initial estimates of those population statistics use a probability sample with an unbiased and design-consistent estimator, which is concisely denoted as the "π-estimator" in the seminal work of Särndal *et al.* (1992). The Horvitz and Thompson (1952) estimator is an example. Furthermore, assume that relatively inexpensive population statistics exist, through census enumeration or sampling or both, for a large number (J) of auxiliary-variables that are measurable, at least in principle, for each population element.

In the current example, auxiliary-variables may be time-series of predictands from statistical models that use multi-temporal data from spaceborne and airborne sensors, collectively known as "remotely-sensed data" (Czaplewski 1999; Næsset et al., 2011; Magnussen and Russo, 2012; Breidt and Opsomer, 2017). Deterministic models that predict the temporal dynamics of the population (*e.g.*, "growth and yield models" and "stand projection models") provide other types of auxiliary-variables (*e.g.*, Weiskittel *et al.*, 2011). Those predictions can be broad groupings of study-variables (*e.g.*, a few categories of land cover) or amalgamations of multiple continuous variables (e.g., total biomass). These aggregate study-variables often serve as margins in statistical tables.

The classification systems for remotely-sensed data (*e.g.*, forest cover) need not exactly match that for the study-variables (*e.g.*, commercial timberland), but the association between the two should be strong.

 DOI: 10.13140/RG.2.2.17634.07369

Certain remotely-sensed predictions, such as the k-nearest neighbor estimator (Crookston and Finley, 2008) correspond to direct measurements (*e.g.*, on-the-ground "field" data) for each study-variable within each population element. Certain deterministic process models can predict the value of each study-variable at time *t* based on direct measurements of those study-variables at previous times $(t-1),(t-2),\ldots,(t-k)$. These can serve as additional auxiliary-variables in an annual forest monitoring system (*e.g.*, Lessard *et al.*, 2001).

At the scale of a population element, assume each auxiliary-variable is well correlated with at least one study-variable, and each study-variable is well correlated with at least one auxiliary-variable. Collinearity among auxiliary-variables can be high, and therefore, much redundant information can exist among auxiliary-variables.

Population estimates for auxiliary-variables may include two general design elements:

- A *population census*, in which some or all auxiliary-variables are measured for every element in the sampled population.

  An example includes administrative census records that fully enumerate the sampled population; another is full-coverage remotely-sensed data measured by spaceborne sensor platforms. Both examples support a single-phase sampling-design, in which population parameters for the auxiliary-variables are known exactly (see Särndal *et al.*, 1992, Chapter 6).

- A *probability sample* of the population that uses a design-consistent π-estimator for population-totals of the auxiliary-variables.

  An example is a sample-survey that uses ultra-high resolution remotely-sensed data from airborne platforms to measure each selected sampling unit. This example includes Two-Phase sampling (see Särndal *et al.*, Chapter 9), where airborne sensors provide auxiliary data for all sampled population elements in the Phase-One and Phase-Two; and field crews measure study-variables for those sampled elements in the Phase-Two sample. Probability sampling and a design-consistent π-estimator provide estimates of the population parameters for the auxiliary-variables.

The following Technical Report uses a simple Two-Phase sampling-design as an example. However, the new estimator accommodates more complex sampling and

 DOI: 10.13140/RG.2.2.17634.07369

estimation situations, including: multi-phase sampling-designs, multi-stage sampling-designs, and blended versions of both; probability sampling with unequal inclusion probabilities; cluster plots; supplemental sampling of rare sub-populations; independent sample-surveys; interpenetrating survey panels with repeated measurements over time; longitudinal surveys with time-series estimation.

Furthermore, the new estimator provides an approximate estimate of the covariance matrix among population estimates for all study-variables. That covariance matrix supports variance estimators for pseudo-estimators (Särndal *et al.*, 1992, Section 5.5); an example is the estimated rate of change over time in a study-variable, which is the ratio of a population estimate at time t and time $t-1$. That covariance matrix also supports approximate variance estimators for other applications (Czaplewski, 2010a), such as: small-area and small-domain estimation; missing data estimators; and estimation with random errors from measurement models and prediction models. The bootstrap method can improve accuracy of variance estimates (Section 7, page 70).

1.1 Model-Assisted Difference-Estimator

Särndal *et al.* illustrate the conventional difference-estimator for a single study-variable. Because of its simplicity, they use the difference-estimator as an introduction to model-assisted regression estimators. Model-assisted estimators reduce variance of the population estimate for a univariate study-variable by "gaining strength" from population estimates for J auxiliary-variables. The difference-estimator for a single study-variable is readily extended into a multivariate estimator for a vector of M study-variables (see Section 2.4 below).

McConville *et al.* (2020) provide a tutorial for model-assisted estimation with a large number (M) of study-variables and a large number (J) of remotely-sensed auxiliary-variables. In order to maximize variance-reduction with numerous auxiliary-variables, regression coefficients in model-assisted estimators must be specific to each of M study-variables. This is termed "specific inference." Furthermore, collinearity can be high among population estimates for auxiliary-variables, and model-assisted estimators require wise and informed selection of a subset of those auxiliary-variables that best reduces variance of the population estimate for each study-variable (*i.e.*, model-selection). Model-assisted estimators assume a census of auxiliary-variables, and application of model-assisted estimators can be challenging if a population-total for an auxiliary-variable is estimated with a sample-survey rather than a census, especially with a complex sample-survey design. Model-assisted estimators can be

 DOI: 10.13140/RG.2.2.17634.07369

complex functions of π-estimators; variance estimates for study-variables may require Taylor-series approximations, which tend to underestimate the true variance, especially with numerous auxiliary-variables. The bootstrap variance estimator can be more accurate.

There are three major challenges in application of the difference-estimator:

- First, computation of minimum-variance coefficients for the difference-estimator requires inversion of a $J \times J$ covariance matrix for sample-survey estimates of population-totals for the J auxiliary-variables. That matrix inverse is numerically unreliable or infeasible if it is ill-conditioned or rank-deficient, necessitating heuristic data reduction methods. The likelihood of numerical obstacles increases as the number of auxiliary-variables increases. The number of potential auxiliary-variables continues to grow in this era of "big data" in official statistics (Scannapieco et al., 2013; Florescu *et al.*, 2014; Struijs *et al.*, 2014; Daas *et al.*, 2015; Kitchin, 2015; Reimsbach-Kounatze, 2015; Tam and Clarke, 2015; Hackl, 2016).
- Second, conventional estimation methods require known population-totals for the auxiliary-variables. Full-coverage remotely-sensed data, censuses, administrative records and official government statistics can supply these population parameters. However, population statistics for some auxiliary-variables require multi-phase and multi-stage sampling-designs, in which case the population-totals are estimates, not known constants.
- Third, the minimum-variance criterion is not always a wise choice for a difference-estimator. Although uncommon, minimum-variance estimates are infeasible yet numerically possible. One example is an estimate of a variance statistic that is negative. Another example is an estimate of the number of individuals in a population that is negative, or a population proportion that exceeds 1.0. Numerical solutions, such as inequality constraints, with a difference-estimator are not apparent.

1.2 New Difference-Estimator

In order to addresses these challenges, the following Technical Report develops a new estimator that is a variation of the conventional difference-estimator (Särndal *et al.*, 1992).

Rather than an inversion of the entire $J \times J$ covariance matrix, the new difference-estimator uses a recursive sequence of J recursions. Each recursion uses

 DOI: 10.13140/RG.2.2.17634.07369

a single auxiliary-variable to improve population estimates for all M study-variables. After the final recursion, all J auxiliary-variables have reduced variance of all M study-variables, at least to the extent possible with a linear estimator and a full-rank and well-conditioned covariance matrix. The following refers to this as the "recursive" difference-estimator.

The recursive difference-estimator replaces inversion of the $J \times J$ covariance matrix for auxiliary-residuals with a sequence of J scalar inversions. If that $J \times J$ covariance matrix is well-conditioned, then the scalar inversion reduces potential for nonrandom numerical errors. If that matrix is rank-deficient, then the recursive difference-estimator proceeds, one auxiliary-variable at a time, until the denominator of the scalar inverse equals, or nearly equals, zero. If the initial estimate of that $J \times J$ covariance matrix is rank-deficient, *i.e.*, rank = $J_r < J$, then the number of recursions will be J_r or fewer. Therefore, the recursive difference-estimator is feasible regardless of matrix rank or condition.

The conventional difference-estimator (Särndal *et al*., pp. 221-225) assumes population parameters for each auxiliary-variable are known prior to sampling through a complete enumeration (census) of the population or equivalent means (see page 10). A linear model predicts the value of a study-variable for each population element, where the predictors are the auxiliary-variables and the predictand is the study-variable. See Breidt and Opsomer (2017) for examples. Assume those predictions are well correlated with the study-variable, and the population-total for the model predictions is a known constant. Assume the population-total for the model predictions is approximately equal to the population-total for the study-variable. The difference-estimator estimates the difference between the unknown population-total for the study-variable and the known total for the auxiliary-variable. Estimate that difference with the probability sample, where the measured value of the study-variable and the predicted value of the study-variable are known for each element in the probability sample. The difference-estimator for the population-total of the study-variable equals the known population-total for the model predictions with the census of auxiliary-variables, plus the sample-survey π estimate of the difference between these two known population parameters.

The conventional difference-estimator is "model-assisted" because it uses a model to predict the study-variable for each population element, where the auxiliary-variables for each corresponding population element are the predictor variables. The stochastic processes include random model prediction errors for each population element and the random sampling errors in the difference-estimate. Särndal *et al*. (pp. 239-242) use population statistics from a

 DOI: 10.13140/RG.2.2.17634.07369

probability sample to derive coefficients for this linear model that minimize the variance of population estimate for the study-variable.

The new difference-estimator considered here uses a different statistical model than that of Särndal *et al*. It begins with the $(M+2J)\times 1$ π vector estimate of population-totals, in which one $M\times 1$ partition is for the study-variables, and the two remaining partitions represent separate population estimates for each of the J auxiliary-variables. It then applies a multivariate linear transformation of that $(M+2J)\times 1$ vector estimate. Assume the coefficients in that transformation are arbitrary but known constants (Särndal *et al*., Section 6.3). Many different sets of coefficients for the auxiliary-variables reduce the variance of the population estimate for the study-variable. One specific set forms the Best Linear Unbiased Estimator, which minimizes the variance of the population estimate for the study-variable. The initial $(M+2J)\times 1$ π vector estimate of population-totals supports computation of those minimum-variance coefficients. Särndal *et al*. (Section 6.8) uses these as an "optimal" choice for the arbitrary coefficients. That specific set is a good provisional choice for coefficients in the linear transformation of the $(M+2J)\times 1$ π vector estimate.

The conventional difference-estimator requires matrix inversion of the $J\times J$ covariance matrix for population estimates of auxiliary-variables; that matrix can be very large, and it is not necessarily full-rank nor well-conditioned. The new estimator uses a recursive sequence of J scalar inversions ($j=1,\ldots,J$). The j^{th} recursion uses the scalar residual difference between two separate, although not necessarily independent, population estimates of the j^{th} auxiliary-variable. Each recursion reduces variances of population estimates for any of the M study-variables that are sufficiently correlated with the j^{th} auxiliary-residual. The sequence of recursions continues until all J auxiliary-variables are processed, or in the case of a rank-deficient $J\times J$ covariance matrix, until all remaining population estimates for unprocessed auxiliary-residuals have zero-variance. In addition, the j^{th} recursion, reduces variances of population estimates for any of the $(J-j-1)$ remaining auxiliary-variables that have a non-zero covariance with the j^{th} auxiliary-residual. This eliminates collinearity (*i.e.*, redundant information) among those remaining auxiliary-variables, analogous to a sequential orthogonalization.

1.3 Outline

Section 2 (page 15) develops the new difference-estimator. Sections 2.1 and 2.2 establish the foundation of that difference estimator, namely the $(M+2J)\times 1$ design-based π vector estimate. Section 2.3 introduces a linear transformation of the design-based π vector estimate. It creates a $J\times 1$ partition that is the difference

 DOI: 10.13140/RG.2.2.17634.07369

between the two π vector estimates of the J auxiliary-variables (*i.e.*, the "difference residual"). The result is the $(M+J)\times 1$ design-based π vector estimate of sufficient statistics for the new difference estimator. Section 2.4 introduces the new difference-estimator for the M study-variables that uses a $(M+J)\times(M+J)$ matrix of arbitrary coefficients in the linear transformation. Section 2.5 derives the minimum-variance coefficients for the linear transformation in the general case of the difference-estimator in Section 2.4, which closely follows the derivation of the difference-estimator in Särndal *et al*. (see their Section 6.3) under a different stochastic model. Section 2.6concludes with a list of numerous obstacles to practical implementation of the new difference-estimator. Solutions to those obstacles appear in Section 3 "Recursive Difference-Estimator" (page 35); Section 5 "Stepwise Covariate-Selection " (page 50); Section 6 "Constraints for a Robust Difference-Estimator" (page 53); and Section 7 "Improved Covariance Matrix with Bootstrap Estimator" (page 70).

Section 3 (page 35) introduces the new multivariate "recursive difference-estimator," which processes one of the J auxiliary residuals at a time in a sequential recursion: $1,\ldots,j,\ldots,J$. Sections 3.1, 3.2, and 3.3 provide preliminaries for the recursive difference-estimator with arbitrary coefficients. Section 3.4 (page 43) derives coefficients for the linear transformation during the j^{th} recursion that achieve two objectives: (1) minimize variance of population estimates for the study-variables; and (2) remove any collinearity between the j^{th} auxiliary residual and all other $(J-1)$ auxiliary residuals. Section 4 (page 50) assembles the recursive sequence of minimum-variance coefficients into a single $(M+J)\times(M+J)$ coefficient matrix, which replaces the arbitrary coefficients for the linear transformation in Section 2.

Section 5 (page 50) offers a stepwise procedure for covariate-selection, which is analogous to model-selection in model-assisted estimators with specific inference. The stepwise procedure selects covariate auxiliary information that reduces variance of population estimates for the most important study-variables as defined by the analyst. At the j^{th} step in the recursive sequence, the stepwise procedure explores variance-reduction for study-variables expected with each and every of the J auxiliary-residuals. It identifies the single most effective scalar auxiliary-residual (*i.e.*, covariate) among the J auxiliary-residuals that is available at the j^{th} step, and selects that one difference-residual that best reduces variance for all M study variables during the j^{th} step.

The new difference-estimator with minimum-variance coefficients is typically the best choice for "arbitrary coefficients" in the linear transformation (see Section 6.3 of Särndal *et al*.). However, minimization of variance is not always best

 DOI: 10.13140/RG.2.2.17634.07369

performance criterion. Section 6 (page 53 below) modifies the minimum-variance coefficients to achieve other objectives, such as inequality constraints on population estimates for study-variables. By definition, those modifications to the minimum-variance coefficients actually increase variance, but the increase in variance can be minor (see *Figure 3*, page 80), and the benefit is improved reliability and credibility of population estimates. Those modifications are relatively simple, and most do not require computationally intensive methods, such as quadratic programming (*e.g.*, Kitanidis, 1985).

It is possible that an auxiliary-residual can be an outlier that strongly affects selection of minimum-variance coefficients. An "outlier" auxiliary-residual has an unexpectedly large value relative to its standard deviation. For example, the population estimate for every auxiliary-residual has an expected value of zero; but if that estimate is more than say 3.0 standard-deviation units from zero, then it may be considered an unexpected outlier. A rare sampling event, numerical errors, or a nonrandom processing error can cause an outlier. Regardless, an outlier can reduce accuracy of the difference-estimator. Section 6.1 (page 55) mitigates this risk by shrinking the minimum-variance coefficients towards zero, which reduces the influence of a suspiciously large auxiliary-residual on population estimates of all correlated study-variables.

An estimate of a study-variable can be analytically infeasible with the new difference-estimator and minimum-variance coefficients. An example is a negative estimate for the number of individuals in a subpopulation, or a proportion that exceeds 100-percent. Section 6.2 (page 60) modifies the minimum-variance coefficients to assure all estimates of study-variables are within minimum and maximum inequality constraints, if any, as specified by the analyst. Section 6.3 modifies the minimum-variance coefficients during each recursion to assure all communalities (*i.e.*, coefficients of variation) are feasible, which precludes negative variance estimates in all remaining recursions $(j+1),\ldots,J$. Sections 6.3, 6.4 and 6.5 further develop methods that mitigate risks from unreliable estimates of covariances between auxiliary-residuals and study-variables, over-fitting with a large number of auxiliary-residuals, and loss of statistical efficiency (*i.e.*, variance increase) with the difference-estimator.

The new difference-estimator with inequality constraints is a nonlinear estimator; and minimum-variance coefficients can differ with different realizations of probability sampling. Therefore, the linear transformation within the new difference-estimator likely underestimates variance of population estimates. An

 DOI: 10.13140/RG.2.2.17634.07369

alternative bootstrap estimator in Section 7 (page 70) is expected to improve accuracy of variance estimates.

Section 8 (page 72) considers the computational burden with the new difference-estimator, and emphasizes the importance of numerically efficient algorithms. Section 9 (page 73) looks at pre- and post-processing steps that assure fidelity to inequality constraints and reduce nonrandom numerical errors. Section 10 (page 76) accommodates study-variables that do not have reliable covariates with the auxiliary-variables. Section 11 (page 77) describes the degree and limitations to which the new difference-estimator can reduce variances of population estimates for the study-variables. Section 12 (page 82) describes a recursive raking procedure with the new difference-estimator that assures, or approximately assures additivity for margins in statistical tables.

Section 13 (page 83) develops "pseudo-expansion factors", which are analogous to adjustments to survey weights that incorporate auxiliary information. Those weights facilitate small-area estimation. Model-assisted estimators with generic inference apply a scalar weight to the values for each selected sampling unit; the new difference-estimator applies a $M\times(M+J)$ matrix of pseudo-expansion factors to each element in the sample, which supports specific inference. Pseudo-expansion factors fully capture nonlinear adjustments to that coefficient matrix, such as inequality constraints (Section 6, page 53).

2 New Difference-Estimator

First, establish the context for the new multivariate difference-estimator. The analyst requires knowledge about population parameters for M study-variables, denoted $y_1,\ldots,y_m,\ldots,y_M$. The following uses the population-total as an example of a population parameter.

Population U consists of N population elements, $\kappa = 1,\ldots,N$. There are insufficient resources to measure all N population elements. Therefore, the study selects survey-sampling to estimate the population parameters. Estimates must be reliable and unbiased, or nearly unbiased relative to their uncertainty. The estimates must have known accuracy, *i.e.*, unbiased, or nearly unbiased, population estimates for the standard deviations of random estimation errors for each study-variable. Probability sampling-design and design-consistent π estimators

 DOI: 10.13140/RG.2.2.17634.07369

(see Särndal *et al.*, 1992, for examples) satisfy these criteria. However, without auxiliary information, assume population estimates are not sufficiently accurate for the analytical objectives of the study.

In addition to study-variables, population parameters include totals for J auxiliary-variables, denoted $x_1,\ldots,x_j,\ldots,x_J$. Assume measurement protocols for auxiliary-variables (*e.g.*, administrative records, censuses, remote sensing, predictions from a deterministic model for population dynamics) are less costly than measurements of the study-variables; there is a strong correlation between each auxiliary-variable and one or more study-variables; estimates of population parameters for auxiliary-variables are typically (although not necessarily) more precise than those for study-variables; but auxiliary-variables, by themselves, do not adequately address objectives of the study.

The goal is to "gain strength" from population estimates for the auxiliary-variables (*i.e.*, covariates) in order to sufficiently improve accuracy of population estimates for study-variables. Särndal *et al.* present model-assisted estimators that can achieve this goal. The finite population model is the foundation for those estimators. To achieve that same goal with the new estimator, employ a different (but closely related) stochastic model than that used by Särndal *et al.* Both estimators start with a stochastic model for a probability sample and its design-consistent π estimator, where the random variables are sampling errors.

- Särndal *et al.* then use that stochastic model with the design-consistent π estimator to estimate the differences between the known population parameters (e.g., censuses) for the auxiliary-variables and the unknown population parameters for the study-variables.
- The new estimator uses a linear transformation of the stochastic model for a probability sample, where coefficients in that linear transformation are known, but arbitrary, constants. Statistical efficiency (*i.e.*, variance reduction) depends on "well chosen" coefficients (see Särndal *et al.*, Remark 6.3.2, page 225). The new difference-estimator uses the minimum-variance criterion as the initial choice for those coefficients, where that criterion is applied using known population estimates from the realized probability sample and its design-consistent π estimator (see Section 2.5, page 25 below). The new estimator modifies those initial values to impose various inequality constraints.

 DOI: 10.13140/RG.2.2.17634.07369

The difference-estimator given by Särndal *et al.* (see their Section 6.3) and the new difference-estimator both share the same foundation. Assume each population element κ has a $(M+J)\times 1$ vector measurement

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}_{\kappa} \tag{1}$$

A $(M+J)\times 1$ vector of parameters for population U is the sum of the values for all N elements

$$\mathbf{t} = \begin{bmatrix} \mathbf{t}_{\mathbf{y}} \\ \mathbf{t}_{\mathbf{x}} \end{bmatrix} = \sum_{\kappa}^{N} \begin{bmatrix} \mathbf{y}_{\kappa} \\ \mathbf{x}_{\kappa} \end{bmatrix} \tag{2}$$

The population-totals in Equation (2) define the population parameter space, which is abbreviated as the "parameter-space" in the following. The values of population-totals for the *study-variables* ($\mathbf{t}_{\mathbf{y}}$) are unknown and inaccessible prior to sampling. The exact population-totals for the *auxiliary-variables* ($\mathbf{t}_{\mathbf{x}}$) can be known (*e.g.*, census enumeration). In addition, the new estimator accommodates estimates of those totals when their exact values are unknown (*e.g.*, from a sample-survey).

Assume some probability sampling-design p(·) of population U and multivariate design-consistent estimator $\boldsymbol{\Theta}_{\pi}$ produce the partitioned $(M+2J)\times 1$ vector of design-based π population estimates

$$\hat{\mathbf{t}}_{\pi} = \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_{\pi} \tag{3}$$

Partition $\hat{\mathbf{t}}_{\mathbf{y},\pi}$ contains population estimates for the M study-variables, where its expected value is vector $\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{y},\pi}\right] = \mathbf{t}_{\mathbf{y}}$. Partition $\hat{\mathbf{t}}_{\mathbf{x}_H,\pi}$ contains population estimates for J auxiliary-variables. Partition $\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$ is a second vector of population estimates for those same J auxiliary-variables. For example, consider a Two-Phase sample-survey vector estimate, where Phase-One provides the $J\times 1$ partition $\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$, and Phase-Two provides the $(M+J)\times 1$ partition $\left[\hat{\mathbf{t}}_{\mathbf{y},\pi}^{\mathrm{T}} \;\vdots\; \hat{\mathbf{t}}_{\mathbf{x}_H,\pi}^{\mathrm{T}}\right]^{\mathrm{T}}$. In a Two-Stage sampling-design, the Primary Sampling Units (PSU) provide the $J\times 1$ partition

 DOI: 10.13140/RG.2.2.17634.07369

$\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$, and the Secondary Sampling Units (SSU) in the cluster-plot provide the $(M+J)\times 1$ partition $\left[\hat{\mathbf{t}}_{\mathbf{y},\pi}^{\mathrm{T}} \;\vdots\; \hat{\mathbf{t}}_{\mathbf{x}_H,\pi}^{\mathrm{T}}\right]^{\mathrm{T}}$. Pivot partitions for auxiliary-variables such that both share the same expected vector values, *i.e.*, $\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}_H,\pi}\right] = \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}\right] = \mathbf{t}_{\mathbf{x}}$.

2.1 The Stochastic Model

Särndal *et al.* (1992) use a prediction model as the foundation for population estimates with a sample-survey, which assumes population parameters for auxiliary-variables are known through a direct census or a prediction model applied to census variables (Särndal, 1980). However, the new difference-estimator is a simple linear transformation of that stochastic model, where the outcome from a design-consistent π estimator guides the choice of coefficients in that transformation.

Assume $(M+2J)\times 1$ vector estimate in Equation (3) of the population parameters in Equation (2) is

$$\begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_{\pi} = \begin{bmatrix} \mathbf{t}_{\mathbf{y}} \\ \mathbf{t}_{\mathbf{x}} \\ \mathbf{t}_{\mathbf{x}} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{\mathbf{y}} \\ \mathbf{e}_{\mathbf{x}_H} \\ \mathbf{e}_{\mathbf{x}_G} \end{bmatrix}_{\pi} \tag{4}$$

where $(M+2J)\times 1$ vector $\left[\mathbf{e}_{\mathbf{y}} \;\vdots\; \mathbf{e}_{\mathbf{x}_H} \;\vdots\; \mathbf{e}_{\mathbf{x}_G}\right]_{\pi}^{\mathrm{T}}$ contains random variables driven by a stochastic process, namely probability sampling. If the population parameters for the auxiliary-variables are known *a priori*, then $\mathbf{e}_{\mathbf{x}_G} = \mathbf{0}$ in Equation (4); this is the model used by Särndal *et al.*; the new estimator accommodates both $\mathbf{e}_{\mathbf{x}_G} = \mathbf{0}$ and $\mathbf{e}_{\mathbf{x}_G} \neq \mathbf{0}$

The $(M+2J)\times 1$ zero vector the is expected value of those random variables $\mathrm{E}\left[\mathbf{e}_{\mathbf{y}} \;\vdots\; \mathbf{e}_{\mathbf{x}_H} \;\vdots\; \mathbf{e}_{\mathbf{x}_G}\right]_{\pi}^{\mathrm{T}} = \mathbf{0}$; denote its expected $(M+2J)\times(M+2J)$ covariance matrix as

 DOI: 10.13140/RG.2.2.17634.07369

$$\mathbf{V} = \mathrm{E}\left[\begin{bmatrix}\mathbf{e}_{\mathbf{y}}\\ \mathbf{e}_{\mathbf{x}_H}\\ \mathbf{e}_{\mathbf{x}_G}\end{bmatrix}\begin{bmatrix}\mathbf{e}_{\mathbf{y}}\\ \mathbf{e}_{\mathbf{x}_H}\\ \mathbf{e}_{\mathbf{x}_G}\end{bmatrix}^{\mathrm{T}}\right]_{\pi} = \left[\begin{array}{c|c|c} \mathbf{V}_{\mathbf{y},\mathbf{y}} & \mathbf{V}_{\mathbf{y},\mathbf{x}_H} & \mathbf{V}_{\mathbf{y},\mathbf{x}_G} \\ \hline \mathbf{V}_{\mathbf{x}_H,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G} \\ \hline \mathbf{V}_{\mathbf{x}_G,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G} \end{array}\right]_{\pi} \tag{5}$$

The new difference-estimator applies to any multivariate design-based population estimate in the form of Equations (4) and (5) from any probability sampling-design p(·) and its design-consistent estimator $\boldsymbol{\Theta}_{\pi}$.

Consider Two-Phase sampling as a simple example. Phase-One uses a large probability sample to precisely estimate the $J\times 1$ vector $\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$ of population-totals for the auxiliary-variables and the associated $J\times J$ covariance matrix $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}\right]_{\pi}$. Phase-Two uses a smaller sample to estimate the $(M+J)\times 1$ vector of population-totals for both the study-variables $\hat{\mathbf{t}}_{\mathbf{y},\pi}$ and auxiliary-variables $\hat{\mathbf{t}}_{\mathbf{x}_H,\pi}$. Phase-Two also estimates the $(M\times J)$ matrix of covariances $\left[\mathbf{V}_{\mathbf{x}_H,\mathbf{y}}\right]_{\pi}$ between the auxiliary-variables and study-variables. In this example, the partition $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{y}} \,\middle|\, \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H}\right]$ contains any weak covariances that are caused by joint inclusion probabilities between Phase-One and Phase-Two.

2.2 Multivariate π-Estimator

Consider two sample-surveys, denoted *H* and *G*. Sample *H* estimates population-totals for *M* study-variables and *J* auxiliary-variables; sample *G* estimates population-totals for *J* auxiliary-variables but none of the study-variables.

 DOI: 10.13140/RG.2.2.17634.07369

Section 2.8 in Särndal *et al.* (1992) expresses the univariate π-estimator for sample *G* as

$$\left(\hat{t}_m\right)_{\pi_H} = \sum_U \frac{1}{\left(\pi_H\right)_i}\left(y_m\right)_i\left(I_H\right)_i \tag{6}$$

where the notation is defined as:

$\left(\hat{t}_m\right)_{\pi_H}$	π-estimator of the population-total for the m^{th} study-variable with survey *H*
$U = \{1, 2, \ldots, i, \ldots, N\}$	Population with total number of elements *N* in population *U*
$\left(\pi_H\right)_i$	First-order inclusion probability for element *i* in sample *H*
$\left(y_m\right)_i$	Value of m^{th} study-variable for element *i*
$\left(I_H\right)_i$	=1 if element *i* is in sample *H* =0 otherwise

A multivariate extension of the univariate estimator in Equation (6) is the π-estimator for the $(M+J)\times 1$ vector of population-totals

$$\begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \end{bmatrix}_{\pi_H} = \sum_U \frac{1}{\left(\pi_H\right)_i}\begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{bmatrix}\left(I_H\right)_i \tag{7}$$

where additional notation is

$\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi_H}$	π-estimator for the $M\times 1$ vector of population-totals for the *M* study-variables with sample *H*
$\left[\hat{\mathbf{t}}_{\mathbf{x}_H}\right]_{\pi_H}$	π-estimator for the $J\times 1$ vector of population-totals for the *J* auxiliary-variables with sample *H*
$\mathbf{y}_i$	$M\times 1$ measurement vector for *M* study-variables in element *i*
$\mathbf{x}_i$	$J\times 1$ measurement vector for *J* auxiliary-variables in element *i*

The corresponding multivariate π-estimator for sample *G* is

$$\left[\hat{\mathbf{t}}_{\mathbf{x}_G}\right]_{\pi_G} = \sum_U \frac{1}{\left(\pi_G\right)_i}\left[\mathbf{x}_i\right]\left(I_G\right)_i \tag{8}$$

 DOI: 10.13140/RG.2.2.17634.07369

where additional notation is

$\left[\hat{\mathbf{t}}_{\mathbf{x}_G}\right]_{\pi_G}$	π-estimator for the M×1 vector of population-totals for the M study-variables with sample G
$\left(I_G\right)_i$	=1 if element i is in sample G =0 otherwise

Combine the π-estimators in Equations (7) and (8)

$$\left[\hat{\mathbf{t}}_{\mathbf{y},\mathbf{x}}\right]_\pi = \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_\pi = \sum_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \tag{9}$$

The difference-estimator requires an estimate of the $(M+J)\times(M+J)$ covariance matrix for the $(M+J)\times 1$ multivariate estimator $\left[\hat{\mathbf{t}}\right]_\pi$ in Equation (9). Sections 2.6 and 5.9 in Särndal *et al.* (1992) express the π-estimator for the scalar covariance between the i^{th} and k^{th} element of that covariance matrix as

$$\begin{aligned} \left[\hat{v}_{ik}\right]_\pi &= \sum_U \sum_U I_{ik} \left(\breve{\Delta}_{ik}\right) z_i z_k \\ &= \sum_U \sum_U I_{ik} \left(1 - \frac{\pi_i \pi_k}{\pi_{ik}}\right) z_i z_k \end{aligned}$$

$$\begin{aligned} \Delta_{ik} &= \pi_{ik} - \pi_i \pi_k \\ \breve{\Delta}_{ik} &= \frac{\Delta_{ik}}{\pi_{ik}} = \frac{\pi_{ik} - \pi_i \pi_k}{\pi_{ik}} = 1 - \frac{\pi_i \pi_k}{\pi_{ik}} \\ &\sum_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \end{aligned} \tag{10}$$

 DOI: 10.13140/RG.2.2.17634.07369

2.3 Sufficient Statistics for the Difference-estimator

Define the "auxiliary-residual" $\mathbf{r}=\left(-\hat{\mathbf{t}}_{\mathbf{x}_H}+\hat{\mathbf{t}}_{\mathbf{x}_G}\right)$ as the $J\times 1$ vector of differences between two separate (although not necessarily independent) population estimates for the J auxiliary-variables. Define $\hat{\mathbf{t}}_\pi$ as the $(M+J)\times 1$ vector of sufficient statistics for the difference-estimator as a linear transformation of the $(M+2J)\times 1$ vector of population estimates in Equation (3).

$$\hat{\mathbf{t}}_\pi=\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}}\\ \hline \mathbf{r}\end{bmatrix}_\pi=\left[\begin{array}{c|c|c}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I}\end{array}\right]\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}}\\ \hline \hat{\mathbf{t}}_{\mathbf{x}_H}\\ \hline \hat{\mathbf{t}}_{\mathbf{x}_G}\end{bmatrix}_\pi \tag{11}$$

where $\mathbf{I}$ represents the conformable identity matrix. Recall from Equation (3) that the two vectors of auxiliary statistics are pivoted such that $\mathrm{E}\left[\hat{\mathbf{t}}_{H,\pi}\right]=\mathrm{E}\left[\hat{\mathbf{t}}_{G,\pi}\right]=\mathbf{t}_{\mathbf{x}}$; and therefore,

$$\begin{aligned}\mathrm{E}[\mathbf{r}]&=\mathrm{E}\left[-\hat{\mathbf{t}}_{\mathbf{x}_H}+\hat{\mathbf{t}}_{\mathbf{x}_G}\right]\\ &=\left(-\mathbf{t}_{\mathbf{x}}+\mathbf{t}_{\mathbf{x}}\right)\\ &=\mathbf{0}\end{aligned} \tag{12}$$

The $(M+J)\times(M+J)$ covariance matrix for the linear transformation in Equation (11) is a simple linear transformation

$$\begin{aligned}\mathbf{V}_\pi&=\left[\begin{array}{c|c|c}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I}\end{array}\right]\left[\begin{array}{c|c|c}\mathbf{V}_{\mathbf{y},\mathbf{y}} & \mathbf{V}_{\mathbf{y},\mathbf{x}_H} & \mathbf{V}_{\mathbf{y},\mathbf{x}_G}\\ \hline \mathbf{V}_{\mathbf{x}_H,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G}\\ \hline \mathbf{V}_{\mathbf{x}_G,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}\end{array}\right]_\pi\left[\begin{array}{c|c}\mathbf{I} & \mathbf{0}\\ \hline \mathbf{0} & -\mathbf{I}\\ \hline \mathbf{0} & \mathbf{I}\end{array}\right]\\ &=\left[\begin{array}{c|c}\mathbf{V}_{\mathbf{y},\mathbf{y}} & \boldsymbol{\Gamma}\\ \hline \boldsymbol{\Gamma}^{\mathrm{T}} & \boldsymbol{\Lambda}\end{array}\right]_\pi \quad \text{where} \begin{cases}\boldsymbol{\Gamma}_\pi=\left[-\mathbf{V}_{\mathbf{y},\mathbf{x}_H}+\mathbf{V}_{\mathbf{y},\mathbf{x}_G}\right]_\pi\\ \qquad=\left[\boldsymbol{\varphi}_1 \quad \cdots \quad \boldsymbol{\varphi}_j \quad \cdots \quad \boldsymbol{\varphi}_J\right]_\pi\\ \boldsymbol{\Lambda}_\pi=\left[\mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H}-\mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G}-\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H}+\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}\right]_\pi\end{cases}\end{aligned} \tag{13}$$

Notation for the $M\times J$ partition $\boldsymbol{\Gamma}$ and $J\times J$ partition $\boldsymbol{\Lambda}$ conform to that in Särndal *et al.* (see their Section 6.8). The recursive version of the new difference-estimator in

 DOI: 10.13140/RG.2.2.17634.07369

Section 3 (page 35 below) uses $M\times 1$ column partitions of $\boldsymbol{\Gamma}=\left[\boldsymbol{\varphi}_1 \quad \cdots \quad \boldsymbol{\varphi}_j \quad \cdots \quad \boldsymbol{\varphi}_J\right]$, one for each of the J auxiliary-residuals.

2.4 Difference-Estimator with Arbitrary Coefficients

Särndal *et al.* 1992 (Section 6.3) present the difference-estimator for a single study-variable (M=1) with J auxiliary-variables. A multivariate extension of that estimator for M>1 study-variables is straightforward. Define $1\times J$ vector $\mathbf{a}_m$ as Särndal's vector of "arbitrary coefficients" for a single study-variable (m). Simply stack those $1,\ldots,m,\ldots,M$ vectors to form the $M\times J$ coefficient matrix $\mathbf{A}_\mathbf{y}$.

$$\mathbf{A}_\mathbf{y}=\begin{bmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_m\\ \vdots\\ \mathbf{a}_M\end{bmatrix}=\begin{bmatrix}a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,J}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ a_{m,1} & \cdots & a_{m,j} & \cdots & a_{m,J}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ a_{M,1} & \cdots & a_{M,j} & \cdots & a_{M,J}\end{bmatrix} \tag{14}$$

where $a_{m,j}$ is the coefficient for the m^{th} study-variable and the j^{th} auxiliary-residual. Although not used by Särndal *et al.*, define a similar $J\times J$ coefficient matrix $\mathbf{A}_\mathbf{y}$ for the J auxiliary-residuals

$$\mathbf{A}_\mathbf{r}=\begin{bmatrix}a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,J}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ a_{j,1} & \cdots & a_{j,j} & \cdots & a_{j,J}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ a_{J,1} & \cdots & a_{J,j} & \cdots & a_{J,J}\end{bmatrix} \tag{15}$$

where $a_{i,j}$ is the coefficient for the i^{th} and j^{th} auxiliary-residual.

The multivariate extension of Särndal's difference-estimator is

$$\left.\begin{aligned}\left[\begin{array}{c}\hat{\mathbf{t}}_\mathbf{y}\\ \hline \mathbf{r}\end{array}\right]_{\text{dif}} &= \left[\begin{array}{c}\hat{\mathbf{t}}_\mathbf{y}\\ \hline \mathbf{r}\end{array}\right]_\pi+\left[\begin{array}{c|c}\mathbf{0} & \mathbf{A}_\mathbf{y}\\ \hline \mathbf{0} & \mathbf{A}_\mathbf{r}\end{array}\right]\left[\begin{array}{c}\hat{\mathbf{t}}_\mathbf{y}\\ \hline \mathbf{r}\end{array}\right]_\pi\\ &= \mathbf{K}\left[\begin{array}{c}\hat{\mathbf{t}}_\mathbf{y}\\ \hline \mathbf{r}\end{array}\right]_\pi\end{aligned}\right\} \quad \text{where } \mathbf{K}=\left[\begin{array}{c|c}\mathbf{I} & \mathbf{A}_\mathbf{y}\\ \hline \mathbf{0} & \left(\mathbf{I}+\mathbf{A}_\mathbf{r}\right)\end{array}\right] \tag{16}$$

 DOI: 10.13140/RG.2.2.17634.07369

The $M\times J$ matrix $\mathbf{A}_{\mathrm{y}}$ and $J\times J$ matrix $\mathbf{A}_{\mathrm{r}}$ contain "arbitrary coefficients" (Särndal *et al.* 1992, Section 6.3) for the study-variables and the vector of auxiliary-residuals respectively; **0** represents the conformable zero matrices; **I** represents the conformable identity matrices; and **K** is a $(M+J)\times(M+J)$ matrix that combines these partitions.

The corresponding covariance matrix for the linear transformation in Equation (16) is

$$
\begin{aligned}
\hat{\mathbf{V}}_{\mathrm{dif}} &= \mathbf{K}\,\hat{\mathbf{V}}_{\pi}\,\mathbf{K}^{\mathrm{T}} \\
&= \left[\begin{array}{c|c} \mathbf{I} & \mathbf{A}_{\mathrm{y}} \\ \hline \mathbf{0} & \left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}\right) \end{array}\right]\left(\left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathrm{y,y}} & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array}\right]_{\pi}\right)\left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{A}_{\mathrm{y}}^{\mathrm{T}} & \left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}\right)^{\mathrm{T}} \end{array}\right] \\
&= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathrm{y,y}}+\hat{\mathbf{\Gamma}}\mathbf{A}_{\mathrm{y}}^{\mathrm{T}}+\mathbf{A}_{\mathrm{y}}\hat{\mathbf{\Gamma}}^{\mathrm{T}}+\mathbf{A}_{\mathrm{y}}\hat{\mathbf{\Lambda}}\mathbf{A}_{\mathrm{y}}^{\mathrm{T}} & \left(\hat{\mathbf{\Gamma}}^{\mathrm{T}}+\mathbf{A}_{\mathrm{y}}\hat{\mathbf{\Lambda}}\right)\left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}^{\mathrm{T}}\right) \\ \hline \left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}\right)\left(\hat{\mathbf{\Gamma}}^{\mathrm{T}}+\hat{\mathbf{\Lambda}}\mathbf{A}_{\mathrm{y}}^{\mathrm{T}}\right) & \left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}\right)\hat{\mathbf{\Lambda}}\left(\mathbf{I}+\mathbf{A}_{\mathrm{r}}\right)^{\mathrm{T}} \end{array}\right]_{\pi} \\
&= \left\{\begin{array}{l} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathrm{y,y}}+ & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array}\right]+\left(\left[\begin{array}{c} \mathbf{A}_{\mathrm{y}} \\ \hline \mathbf{A}_{\mathrm{r}} \end{array}\right]\left[\hat{\mathbf{\Gamma}}^{\mathrm{T}} \;\vdots\; \hat{\mathbf{\Lambda}}\right]\right)+ \\ +\left(\left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array}\right]\left[\mathbf{A}_{\mathrm{y}}^{\mathrm{T}} \;\vdots\; \mathbf{A}_{\mathrm{r}}^{\mathrm{T}}\right]\right)+ \\ +\left(\left[\begin{array}{c} \mathbf{A}_{\mathrm{y}} \\ \hline \mathbf{A}_{\mathrm{r}} \end{array}\right]\hat{\mathbf{\Lambda}}\left[\mathbf{A}_{\mathrm{y}}^{\mathrm{T}} \;\vdots\; \mathbf{A}_{\mathrm{r}}^{\mathrm{T}}\right]\right) \end{array}\right\}_{\pi}
\end{aligned}
\tag{17}
$$

In order for recursive version of the new estimator (Section 3, page 35) to compensate for collinearity among population estimates of the J auxiliary-variables, Equation (16) includes the partition for the $J\times 1$ vector of auxiliary-residuals **r**, even though population estimates for the auxiliary-residuals all have expected values of zero, but it is the $M\times 1$ partition $\left[\hat{\mathbf{t}}_{\mathrm{y}}\right]_{\mathrm{dif}}$ that is the focus of the analyst's attention

$$
\left[\hat{\mathbf{t}}_{\mathrm{y}}\right]_{\mathrm{dif}} = \left[\hat{\mathbf{t}}_{\mathrm{y}}\right]_{\pi} + \left[\mathbf{A}_{\mathrm{y}}\right]\left[\mathbf{r}\right]_{\pi} \tag{18}
$$

Like the difference-estimator given by Särndal *et al*. (see their Section 6.3), the new difference-estimator in Equation (18) equals the vector of prior π estimates for

 DOI: 10.13140/RG.2.2.17634.07369

the study-variables $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi}$ plus and "adjustment" vector $\mathbf{A}_{\mathbf{y}}\mathbf{r}_{\pi}$, which is a linear transformation of the vector of auxiliary-residuals $\mathbf{r}_{\pi}$, that "updates" the π estimate. This estimator is algebraically identical to the static Kalman filter update in digital signal processing (see Section 14.6, page 92).

If the π-estimator $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi}$ in Equation (3) is unbiased, then the difference-estimator with the $M\times J$ matrix $\mathbf{A}_{\mathbf{y}}$ of arbitrary but known coefficients is also unbiased because $\mathrm{E}\left[\mathbf{r}\right]=\mathbf{0}$ from Equation (12), regardless of the value of those coefficients

$$\begin{aligned}\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\mathrm{dif}} &= \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi}+\left[\mathbf{A}_{\mathbf{y}}\right]\mathrm{E}\left[\mathbf{r}\right]_{\pi}\\ &= \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi}\\ &= \mathbf{t}_{\mathbf{y}}\end{aligned} \tag{19}$$

The covariance matrix for this linear transformation from Equation (18) is

$$\left[\hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}}\right]_{\mathrm{dif}}=\left[\hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}}+\left(\hat{\mathbf{\Gamma}}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}}+\mathbf{A}_{\mathbf{y}}\hat{\mathbf{\Gamma}}^{\mathrm{T}}+\mathbf{A}_{\mathbf{y}}\hat{\mathbf{\Lambda}}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}}\right)\right]_{\pi} \tag{20}$$

If the $M\times J$ coefficient matrix $\mathbf{A}_{\mathbf{y}}$ in Equation (20) is "well chosen", then the difference-estimator will reduce variance of population-totals for the study-variables (Särndal *et al.*, Section 6.3). As a minimal criterion, a well-chosen coefficient matrix satisfies the vector inequality

$$\mathrm{diag}\left(\hat{\mathbf{\Gamma}}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}}+\mathbf{A}_{\mathbf{y}}\hat{\mathbf{\Gamma}}^{\mathrm{T}}+\mathbf{A}_{\mathbf{y}}\hat{\mathbf{\Lambda}}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}}\right)_{\pi}<\mathbf{0} \tag{21}$$

where diag() is the matrix operator that extracts the diagonal of a square matrix, which in this context is the variance-reduction for the estimated population-total of each study-variable. Section 2.5 (page 25) derives coefficients that minimize those variances with the linear difference-estimator, and Section 6 (page 53) modifies those coefficients, if necessary, to impose inequality constraints.

2.5 Minimum-Variance Coefficients for Difference-estimator

This Section derives $M\times J$ coefficient matrix $\mathbf{A}_{\mathbf{y}}$ in Equation (21) so that the difference-estimator minimizes variance of the population estimates. The derivation is a multivariate extension that closely follows that for the univariate difference-estimator in Section 6.8 of Särndal *et al.* (1992).

 DOI: 10.13140/RG.2.2.17634.07369

This derivation simply factors out $M\times J$ coefficient matrix $\mathbf{A}_\mathbf{y}$ in Equation (21) to produce the following equality:

$$\hat{\mathbf{V}}_{\text{dif}} = \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} - f_1\left(\hat{\mathbf{\Lambda}}_\pi\right) + f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right)$$
$$\text{where } \begin{cases} f_1\left(\hat{\mathbf{\Lambda}}_\pi\right) = \mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\mathbf{C}^{\mathrm{T}} \\ f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right) = \left(\mathbf{A}_\mathbf{y} + \mathbf{C}\hat{\mathbf{\Lambda}}_\pi^{-1}\right)\hat{\mathbf{\Lambda}}_\pi\left(\mathbf{A}_\mathbf{y} + \mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\right)^{\mathrm{T}} \end{cases} \tag{22}$$

The first matrix function $f_1\left(\hat{\mathbf{\Lambda}}_\pi\right) = \mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\mathbf{C}^{\mathrm{T}}$ in Equation (22) is a linear transformation that *does not* include coefficient matrix $\mathbf{A}_\mathbf{y}$, and $\mathbf{C}$ is a $M\times J$ matrix of known constants. By assumption, $J\times J$ covariance matrix $\hat{\mathbf{\Lambda}}$ with the π estimator is positive-definite in $f_1\left(\hat{\mathbf{\Lambda}}_\pi\right)$; therefore, the matrix inverse $\hat{\mathbf{\Lambda}}^{-1}$ is positive-definite; and the $M\times M$ linear transformation $\mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\mathbf{C}^{\mathrm{T}}$ is also positive-definite. Therefore, all elements of $\text{diag}\left[-\mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\mathbf{C}^{\mathrm{T}}\right]$ are negative; and matrix function $f_1\left(\hat{\mathbf{\Lambda}}_\pi\right)$ decreases variance of the π estimator.

The second $M\times M$ matrix function $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right) = \left(\mathbf{A}_\mathbf{y} + \mathbf{C}\hat{\mathbf{\Lambda}}_\pi^{-1}\right)\hat{\mathbf{\Lambda}}_\pi\left(\mathbf{A}_\mathbf{y} + \mathbf{C}\hat{\mathbf{\Lambda}}^{-1}\right)^{\mathrm{T}}$ in Equation (22) is a linear transformation with known coefficients, but the second function *does* include coefficient matrix $\mathbf{A}_\mathbf{y}$. Since $\hat{\mathbf{\Lambda}}_\pi$ is positive-definite: $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right)$ is also positive-definite. Therefore, all diagonal elements of $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right)$ are positive, and $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right)$ increases variance of the difference estimator.

In order to derive the minimum-variance values for coefficient matrix $\mathbf{A}_\mathbf{y}$, solve for $\mathbf{A}_\mathbf{y}$ such that $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right) = \mathbf{0}$ in Equation (22):

$$\hat{\mathbf{V}}_{\text{dif}} = \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} - f_1\left(\hat{\mathbf{\Lambda}}_\pi\right) \quad \text{where } \begin{cases} \left[\mathbf{A}_\mathbf{y} + \mathbf{C}\hat{\mathbf{\Lambda}}_\pi^{-1}\right] = \mathbf{0} \\ \mathbf{A}_\mathbf{y} = \left[-\mathbf{C}\hat{\mathbf{\Lambda}}_\pi^{-1}\right] \end{cases} \tag{23}$$

Since $f_2\left(\hat{\mathbf{\Lambda}}_\pi, \mathbf{A}_\mathbf{y}\right) = \mathbf{0}$ in Equation (23), the difference-estimator with $M\times J$ coefficient matrix $\left[\mathbf{A}_\mathbf{y}\right]_{\text{opt}} = \left[-\mathbf{C}\hat{\mathbf{\Lambda}}_\pi^{-1}\right]$ is the minimum-variance linear estimator. Recall that Equation (19) establishes that the difference-estimator is unbiased for an arbitrary $M\times J$ coefficient matrix $\mathbf{A}_\mathbf{y}$. Therefore, the difference-estimator with

 DOI: 10.13140/RG.2.2.17634.07369

minimum-variance coefficient matrix $\mathbf{A}_{\mathbf{y}} = \left[\mathbf{A}_{\mathbf{y}}\right]_{\text{opt}}$ in Equation (23) is the Best Linear Unbiased Estimator (BLUE).

Now solve for $\mathbf{A}_{\mathbf{y}} = \left[\mathbf{A}_{\mathbf{y}}\right]_{\text{opt}}$ in Equation (23). Assume the general case of the covariance matrix in Equation (13), in which exact population-totals for the J auxiliary-variables may be unknown, but a π estimate for those parameters is available. Therefore, covariance partition $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}\right]_{\pi} \neq \mathbf{0}$ in Equation (13). Partitions $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{y}}\right]_{\pi} = \mathbf{0}$ and $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H}\right]_{\pi} = \mathbf{0}$ if both $J\times 1$ vectors of population estimates for the auxiliary-variables are mutually independent; otherwise, $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{y}}\right]_{\pi} \neq \mathbf{0}$ and $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H}\right]_{\pi} \neq \mathbf{0}$.

Equation (17) on page 24 presents the difference-estimator with arbitrary coefficients $\left[\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \,\vdots\, \mathbf{A}_{\mathbf{r}}^{\mathrm{T}}\right]$

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Gamma}}^{\mathrm{T}} & \hat{\boldsymbol{\Lambda}} \end{array}\right] + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] \left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right]^{\mathrm{T}} + \left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right] \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right]^{\mathrm{T}} + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] \hat{\boldsymbol{\Lambda}} \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right]^{\mathrm{T}} \right) \right\}_{\pi} \qquad (17)$$

Derive the minimum-variance matrix of coefficients $\left[\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \,\vdots\, \mathbf{A}_{\mathbf{r}}^{\mathrm{T}}\right]_{\text{opt}}$ by factoring out that matrix from Equation (17).

The derivation starts with two identities: $\mathbf{I} = \left(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\right)$ and the $(M+J)\times(M+J)$ zero matrix

$$\mathbf{0} = \left\{ -\left(\left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right] \hat{\boldsymbol{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right]^{\mathrm{T}} \right) + \left(\left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right] \hat{\boldsymbol{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\boldsymbol{\Gamma}} \\ \hline \hat{\boldsymbol{\Lambda}} \end{array}\right]^{\mathrm{T}} \right) \right\}_{\pi} \qquad (24)$$

Add Equation (24) to the difference-estimator for the covariance matrix in Equation (17) with arbitrary coefficients $\left[\mathbf{A}_{\mathbf{y}} \,\vdots\, \mathbf{A}_{\mathbf{r}}\right]$

 DOI: 10.13140/RG.2.2.17634.07369

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \left(\left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \left(- \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \right) + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] \hat{\mathbf{\Lambda}} \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} + \left(+ \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \right) \right) \right\}_{\pi}$$

(25)

Insert identity matrix $\mathbf{I} = \left(\hat{\mathbf{\Lambda}}^{-1} \hat{\mathbf{\Lambda}} \right)$ into Equation (25)

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \left(\left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \left(- \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \right) + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \left(\hat{\mathbf{\Lambda}}^{-1} \hat{\mathbf{\Lambda}} \right) \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] \hat{\mathbf{\Lambda}} \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} + \left(+ \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \right) \right) \right\}_{\pi}$$

(26)

 DOI: 10.13140/RG.2.2.17634.07369

Rearrange Equation (26)

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \left(\begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Gamma}}^{\mathrm{T}} & \hat{\boldsymbol{\Lambda}} \end{bmatrix} + \begin{bmatrix} & \\ & -\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} \end{bmatrix} \right) + \left(\begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} + \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \right) \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix} \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix}^{\mathrm{T}} + \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \right) \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} \right) \right\}_{\pi}$$

$$= \left\{ \left(\begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Gamma}}^{\mathrm{T}} & \hat{\boldsymbol{\Lambda}} \end{bmatrix} + \begin{bmatrix} & \\ & -\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} \end{bmatrix} \right) + \left(\begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix} \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} + \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix}^{\mathrm{T}} \right) + \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \right) \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} + \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix}^{\mathrm{T}} \right) \right) \right\}_{\pi}$$

$$= \left\{ \left(\begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Gamma}}^{\mathrm{T}} & \hat{\boldsymbol{\Lambda}} \end{bmatrix} + \begin{bmatrix} & \\ & -\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} \end{bmatrix} \right) + \left(\begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} \hat{\boldsymbol{\Lambda}}^{-1} \right) \left(\begin{bmatrix} \hat{\boldsymbol{\Gamma}} \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix}^{\mathrm{T}} + \hat{\boldsymbol{\Lambda}} \begin{bmatrix} \mathbf{A}_{\mathbf{y}} \\ \mathbf{A}_{\mathbf{r}} \end{bmatrix}^{\mathrm{T}} \right) \right\}_{\pi} \tag{27}$$

Insert identity matrix $\mathbf{I} = \left(\hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Lambda}}^{-1} \right)$ into Equation (27)

 DOI: 10.13140/RG.2.2.17634.07369

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \left(\begin{array}{c} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \\ - \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \end{array} \right) + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \right) \left(\hat{\mathbf{\Lambda}} \hat{\mathbf{\Lambda}}^{-1} \right) \left(\left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \hat{\mathbf{\Lambda}} \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} \right) \right\}_{\pi}$$

$$= \left\{ \left(\begin{array}{c} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \\ - \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \end{array} \right) + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \right) \hat{\mathbf{\Lambda}} \left(\hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \hat{\mathbf{\Lambda}}^{-1} \hat{\mathbf{\Lambda}} \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} \right) \right\}_{\pi}$$

$$= \left\{ \left(\begin{array}{c} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \\ - \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} \end{array} \right) + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \right) \hat{\mathbf{\Lambda}} \left(\hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right]^{\mathrm{T}} \right) \right\}_{\pi}$$

(28)

And rearrange

$$\hat{\mathbf{V}}_{\text{dif}} = \left\{ \begin{array}{l} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} + & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array} \right] + \\ - \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right]^{\mathrm{T}} + \\ + \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \right) \hat{\mathbf{\Lambda}} \left(\left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array} \right] + \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array} \right] \hat{\mathbf{\Lambda}}^{-1} \right)^{\mathrm{T}} \end{array} \right\}_{\pi}$$

(29)

 DOI: 10.13140/RG.2.2.17634.07369

Equation (29) an equivalent expression for the covariance matrix with the difference-estimator and arbitrary coefficients in Equation (17), and it conforms to the structure in Equation (22).

Equation (29) exhibits the following characteristics

1. The first term in Equation (29) equals the $(M+J)\times(M+J)$ covariance matrix $\hat{\mathbf{\Lambda}}$ with the initial design-based π-estimator, which is positive-definite by assumption. The diagonal of that first term is the $(M+J)\times 1$ vector of variances with the π-estimator. The purpose of the difference-estimator is to reduce those variances from the π-estimator. This first term is unaffected by coefficient matrix $\left[\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \,\vdots\, \mathbf{A}_{\mathbf{r}}^{\mathrm{T}}\right]$ in the difference-estimator in Equation (29).
2. All diagonal elements of the second term in Equation (29) are negative because $\hat{\mathbf{\Lambda}}^{-1}$ is positive-definite. Therefore, the matrix diagonal of the second term decreases variances on the matrix diagonal of the first term (i.e., the π estimate), which is the purpose of the difference-estimator. This second term is also unaffected by coefficient matrix $\left[\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \,\vdots\, \mathbf{A}_{\mathbf{r}}^{\mathrm{T}}\right]$ in the difference-estimator in Equation (29).
3. All diagonal elements of the third term in Equation (29) are positive because the covariance matrix for the auxiliary-residuals $\hat{\mathbf{\Lambda}}$ and its inverse $\hat{\mathbf{\Lambda}}^{-1}$ are both positive-definite. The diagonal of the matrix in the third term actually increases variances with the difference-estimator, which is contrary to the purpose of difference. Therefore, choose coefficient matrix $\left[\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \,\vdots\, \mathbf{A}_{\mathbf{r}}^{\mathrm{T}}\right]$ that minimizes the third term in Equation (29).

$$\hat{\mathbf{V}}_{\mathrm{dif}} = \left\{ \begin{array}{l} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} & \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Gamma}}^{\mathrm{T}} & \hat{\mathbf{\Lambda}} \end{array}\right] + \\ -\left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array}\right] \hat{\mathbf{\Lambda}}^{-1} \left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array}\right]^{\mathrm{T}} + \\ +\left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{0} \end{array}\right] \hat{\mathbf{\Lambda}} \left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{0} \end{array}\right]^{\mathrm{T}} \end{array} \right\}_{\pi} \quad \text{for} \quad \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] = \left(-\left[\begin{array}{c} \hat{\mathbf{\Gamma}} \\ \hline \hat{\mathbf{\Lambda}} \end{array}\right] \hat{\mathbf{\Lambda}}^{-1} \right) \tag{30}$$

Therefore, the $(M+J)\times J$ matrix of minimum-variance coefficients in Equation (30) equals

 DOI: 10.13140/RG.2.2.17634.07369

$$\mathbf{A}_{\text{opt}} = \left[\begin{array}{c} \mathbf{A}_{\mathbf{y},\text{opt}} \\ \hline \mathbf{A}_{\mathbf{r},\text{opt}} \end{array}\right] = \left[\begin{array}{c} \left(-\hat{\boldsymbol{\Gamma}}\,\hat{\boldsymbol{\Lambda}}^{-1}\right)_{\pi} \\ \hline -\mathbf{I} \end{array}\right] \tag{31}$$

The minimum-variance estimator for the $(M+J)\times 1$ vector of population-totals uses the coefficient matrix in Equation (31) with the difference-estimator in Equation (16)

$$\left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hline \mathbf{0} \end{array}\right]_{\text{opt}} = \mathbf{K}_{\text{opt}} \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hline \mathbf{r} \end{array}\right]_{\pi}$$

$$\text{where} \begin{cases} \mathbf{K}_{\text{opt}} = \left[\begin{array}{c|c} \mathbf{I} & \left(\mathbf{A}_{\mathbf{y},\text{opt}}\right) \\ \hline \mathbf{0} & \left(\mathbf{I} + \mathbf{A}_{\mathbf{r},\text{opt}}\right) \end{array}\right] = \left[\begin{array}{c|c} \mathbf{I} & -\left(\hat{\boldsymbol{\Gamma}}\,\hat{\boldsymbol{\Lambda}}^{-1}\right)_{\pi} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \\ \mathbf{r}_{\text{opt}} = \mathbf{0} \end{cases} \tag{32}$$

The corresponding covariance matrix estimator for linear transformation in Equation (32) uses coefficient matrix $\mathbf{K}_{\text{opt}}$ from Equation (32) with the general difference-estimator from Equation (17)

$$\hat{\mathbf{V}}_{\text{opt}} = \mathbf{K}_{\text{opt}} \hat{\mathbf{V}}_{\pi} \mathbf{K}_{\text{opt}}^{\mathrm{T}} = \left[\begin{array}{c|c} \left(\hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} - \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{\Gamma}}^{\mathrm{T}}\right)_{\pi} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \tag{33}$$

Equations (32) and (33) retain partitions for the auxiliary-residuals, even though they equal the conformable zero or identity matrices with the minimum-variance coefficients. The purpose becomes apparent with the recursive version of the new difference-estimator in Section 3 (page 35).

The difference-estimator given by Särndal *et al.* is algebraically identical to a special case of the difference-estimator in Equations (31) and (32), even though they do not share the same stochastic model. Furthermore, the difference-estimator in Equation (33) for the vector of auxiliary-residuals **r** with the minimum-variance coefficients $\mathbf{A}_{\mathbf{r},\text{opt}}$ equals exactly zero, and the covariances between the minimum-variance estimates of the auxiliary-residuals and the study-variables exactly equal zero (*i.e.*, random estimation errors for the vectors of population estimates $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\text{opt}}$ and $\left[\hat{\mathbf{t}}_{\mathbf{r}}\right]_{\text{opt}}$ are orthogonal). The importance of this property becomes apparent with recursive methods presented in Section 3 below. In a heuristic sense, the difference-estimator with minimum-variance coefficients

 DOI: 10.13140/RG.2.2.17634.07369

"filters out" all relevant information from the population estimates of the auxiliary-variables, and uses that information to improve population estimates for correlated study-variables.

2.6 Implementation Obstacles

The multivariate difference-estimator with minimum-variance coefficients (Section 2.5, page 25) is relatively simple, and its potential for variance-reduction can be substantial. However, there are major obstacles to successful implementation:

1. Computation of the minimum-variance coefficients $\mathbf{A}_{\text{opt}}$ in Equation (31) requires the inverse of $J \times J$ matrix $\hat{\mathbf{\Lambda}}_{\pi}^{-1}$, and a reliable numerical solution requires covariance matrix partition $\hat{\mathbf{\Lambda}}_{\pi}$ be positive-definite and well-conditioned. This becomes unlikely as the number of auxiliary-variables J becomes large, especially if correlations among random estimation errors for some auxiliary-variables are relatively strong.
2. Some elements of the estimated population vector $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\text{opt}}$ in Equation (32) can reside outside of bounds that are known *a prior*, such as negative estimates of nonnegative population parameters.
3. Some variance estimates on the diagonal of covariance matrix $\text{diag}\left(\hat{\mathbf{V}}_{\text{opt}}\right)$ in Equation (33) can be negative, which is numerically possible but infeasible by definition.
4. Some elements of covariance matrix $\hat{\mathbf{V}}_{\text{opt}}$ can exhibit correlations ρ not bounded by ± 1.0, *i.e.*, communalities, or coefficients of determination $\rho^2 > 1.0$, which is numerically possible but infeasible by definition.
5. If the variance for an auxiliary-residual on the diagonal of covariance matrix $\left[\hat{\mathbf{\Lambda}}\right]_{\pi}$ is very small, its inverse on the diagonal of $\left[\hat{\mathbf{\Lambda}}^{-1}\right]_{\pi}$ can suffer from substantial numerical error, and computation of the associated minimum-variance coefficients might not be reliable. In extreme cases, deterministic numerical error can exceed random estimation error.
6. If the numeric values vary greatly among population parameters, numerical errors can become substantial in matrix operations.
7. The sample size might be inadequate to support reliable population estimates for some covariances in the $(M+J) \times J$ matrix $\left[\hat{\mathbf{\Gamma}}\right]_{\pi}$ in Equation (13).

   Unreliable covariance estimates can decrease accuracy of population estimates with the difference-estimator.

 DOI: 10.13140/RG.2.2.17634.07369

8. Some covariances in the $M \times J$ partition $\boldsymbol{\Gamma}$ in Equation (13) can be truly zero, but random sampling error can cause their population estimates in partition $\hat{\boldsymbol{\Gamma}}$ to be nonzero. Those spurious covariances will be small because their expected value is zero. The effected minimum-variance coefficients in Equation (31) will also be nonzero, which will slightly reduce the corresponding population variances with the difference-estimator in Equation (33). If the number of spurious nonzero coefficients in coefficient matrix $\mathbf{K}_{\text{opt}}$, is large, then the cumulative effect on the population estimators in Equations (32) and (33) can be substantial. In other words, large numbers of spurious covariances can cause "overfitting" with the difference-estimator.
9. Some elements of auxiliary-residual vector $\left[\mathbf{r}\right]_{\pi}$ from the design-based π-estimator in Equation (11) can be unexpectedly large compared to their estimated standard deviations $\sqrt{\text{diag}\left[\hat{\boldsymbol{\Lambda}}\right]_{\pi}}$ from Equation (13). These are likely "outlier" residuals, which can have undue influence on the vector of difference-estimates $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\text{opt}}$ in Equation (32).
10. The difference between the initial vector of design-based π population estimates $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\pi}$ in Equation (4) and that with the minimum-variance difference-estimator $\left[\hat{\mathbf{t}}_{\mathbf{y}}\right]_{\text{opt}}$ in Equation (32) can be implausibly large relative to the initial standard deviations with the π-estimator in Equation (5); these outliers can result from unexpectedly large random errors or nonrandom errors, the latter of which can be caused by data-base anomalies, unquantified errors from auxiliary models, and numerical errors. Outliers can cause unreliable population estimates with the difference-estimator.
11. The difference-estimator in Equations (32) and (33) treat the minimum-variance coefficients $\mathbf{A}_{\text{opt}}$ in Equation (31) as known constants; however, they are not constants because they are computed from sample-survey estimates. Hypothetically, replicate surveys would produce slightly different matrices of minimum-variance coefficients $\mathbf{A}_{\text{opt}}$. Equations (32) and (33) ignore this source of endogenous variability; therefore, the estimated covariance matrix in Equation (33) might systematically underestimate the variances of difference-estimates, the magnitude of which varies by application.

Sections 3 to 9 (below) develop methods that solve, reduce, or mitigate these obstacles with the minimum-variance difference-estimator.

 DOI: 10.13140/RG.2.2.17634.07369

3 Recursive Difference-Estimator

The difference-estimator in Section 2.5 (page 25) multiplies the $(M+J)\times(M+J)$ coefficient matrix $\mathbf{K}$ times the $(M+J)\times 1$ vector of population estimates. Computation of the minimum-variance coefficients in $\mathbf{K}_{opt}$ requires inversion of covariance matrix $\hat{\mathbf{\Lambda}}_{\pi}$. The recursive difference-estimator replaces inversion of that $J\times J$ matrix with a recursive sequence of J scalar inverses. This directly addresses Obstacle 1 in Section 2.6 (page 33).

During each step in that recursion, the difference-estimator minimizes variance of population estimates for M study-variables that are sufficiently correlated with the j^{th} auxiliary-residual; and the difference-estimator removes any collinearity between the j^{th} auxiliary-residual and all remaining $(j+1),\ldots,J$ auxiliary-residuals (see Section 3.4.2 below). Furthermore, the recursive difference-estimator allows relatively simple modifications to the minimum-variance coefficients within each of the J steps (Sections 6.1 to 6.5 below), and those modifications address obstacles 2 through 10 in Section 2.6 (page 33).

The initial step in the recursion (j = 0) starts with the $(M+2J)\times 1$ vector of π population estimates from the design-based sample-survey in Section 2.1. That vector of estimated population parameters is transformed in Section 2.3 (page 22) into the $(M+J)\times 1$ vector of sufficient population statistics, which is composed of M study-variables and J *auxiliary-residuals* (not to be confused with the $2\cdot J$ *auxiliary-variables*). During each step $j=\{1,\ldots,j,\ldots,J\}$ in the recursion, the difference-estimator uses a matrix of coefficients and the j^{th} scalar auxiliary-residual to reduce variances and improve accuracy of population estimates for the M study-variables.

The degree of variance-reduction depends on three components: the magnituce of the correlation between each study-variable and the j^{th} residual; the variance of j^{th} residual; and the choice of coefficients in the difference-estimator. In the following, the initial choice for those coefficients uses the minimum-variance criterion. Modifications to the minimum-variance coefficients within each of the J

 DOI: 10.13140/RG.2.2.17634.07369

reclusions address the implementation obstacles in Section 2.6 (page 33). This sacrifices variance-reduction for the sake of robust and numerically reliable population estimates with the difference-estimator. The sacrifice is generally expected to be minor.

3.1 Pivot Notation

In an attempt to simplify notation, pivot parameter-space at the beginning of each j[th] step in the recursion so that the last column/row represents the j^{th} auxiliary-residual

$$\begin{aligned}\hat{\mathbf{t}}_{\mathrm{j}} &= \left[\begin{array}{c} \hat{\mathbf{z}}_j \\ \hline r_j \end{array}\right]_{\mathrm{j}} \\ &= \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hdashline \mathbf{r}_{\notin j} \\ \hline r_j^- \end{array}\right]_{\mathrm{j}}\end{aligned} \tag{34}$$

The corresponding $(M+J)\times(M+J)$ covariance matrix in this pivoted space becomes

$$\begin{aligned}\hat{\mathbf{V}}_{\mathrm{j}} &= \left[\begin{array}{c:c} \hat{\mathbf{V}}_{\mathbf{z}} & \hat{\boldsymbol{\varphi}} \\ \hline \hat{\boldsymbol{\varphi}}^{\mathrm{T}} & \hat{\Lambda} \end{array}\right]_{\mathrm{j}} \\ &= \left[\begin{array}{c:c|c} \hat{\mathbf{V}}_{\mathbf{y},\mathbf{y}} & \hat{\mathbf{V}}_{\mathbf{y},\notin j} & \hat{\boldsymbol{\varphi}}_{\mathbf{y}} \\ \hline \hat{\mathbf{V}}_{\notin j,\mathbf{y}} & \hat{\mathbf{V}}_{\notin j} & \hat{\boldsymbol{\varphi}}_{\notin j} \\ \hline \hat{\boldsymbol{\varphi}}_{\mathbf{y}}^{\mathrm{T}} & \hat{\boldsymbol{\varphi}}_{\notin j}^{\mathrm{T}} & \hat{\Lambda}_j \end{array}\right]_{\mathrm{j}}\end{aligned} \tag{35}$$

The $(M+J)\times(M+J)$ matrix of arbitrary coefficients $\mathbf{K}_j$ for the difference-estimator in this pivoted space becomes

 DOI: 10.13140/RG.2.2.17634.07369

$$\mathbf{K}_{\mathrm{j}} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{a} \\ \hline \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} = \left[\begin{array}{cc|c} \mathbf{I} & \mathbf{0} & \mathbf{a}_{\mathrm{y}} \\ \mathbf{0} & \mathbf{I} & \mathbf{a}_{\notin j} \\ \hline \mathbf{0} & \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \tag{36}$$

Note that coefficient matrix **K** is sparse and highly structured, with only the last column populated with more than a single nonzero element. This offers opportunities to improve numerical accuracy and reduce processing demands with algorithms that implement the new difference-estimator.

3.2 Recursive Sequence, *j*=1,…,*J*

The recursive difference-estimator begins with population estimates from the design-based π sample-survey and the first auxiliary-residual (*j*=1). The difference-estimator uses that first auxiliary-residual to minimize variance for population estimates of all correlated study-variables (subject to constraints and criteria in Section 3.4 below). At the same time, the difference-estimator reduces variance for population estimates of the remaining auxiliary-residuals (although not subject to the inequality constraints in Section 3.4), which in turn, reduces the covariances between those remaining auxiliary-residuals and the population estimates for the study-variables. The difference-estimator removes any collinearities between the jth auxiliary-residual and all other auxiliary-residuals. The next step in the recursion (j+1=2) uses the second auxiliary-residual (*j*=2) to further minimize variance for population estimates of all correlated study-variables. The recursive sequence proceeds until *j*=*J*, or sooner if with an ill-conditioned or rank-deficient covariance matrix (see Section 5.2, page 52).

 DOI: 10.13140/RG.2.2.17634.07369

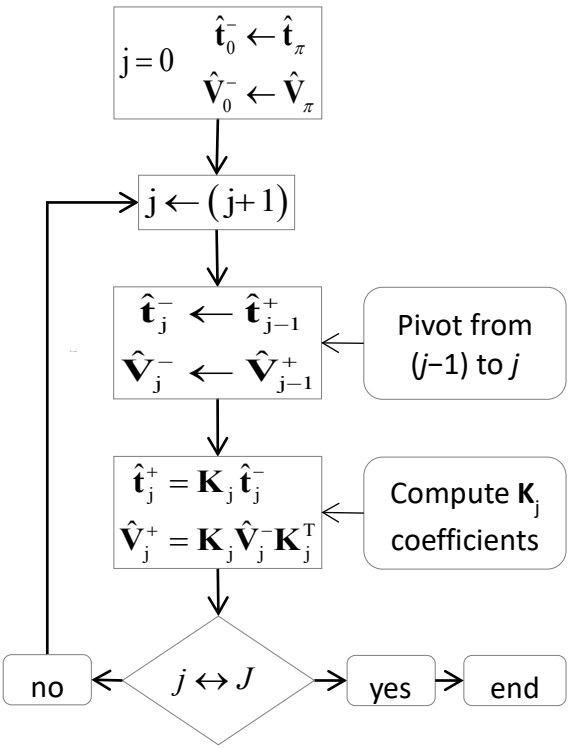

Figure 1. Flow chart for the recursive sequence

 DOI: 10.13140/RG.2.2.17634.07369

3.3 Difference-Estimator for j[th] Recursion

Consider the j[th] recursion within the flow chart in Figure 1 (page 38). Superscript "$^-$" denotes the estimate at the beginning of the j[th] step in the recursion, and superscript "$^+$"denotes the estimate at the end of the j[th] step. Combine these two components into a more complete matrix equation

$$\hat{\mathbf{t}}_{\mathrm{j}}^{+} = \mathbf{K}_{\mathrm{j}}\hat{\mathbf{t}}_{\mathrm{j}}^{-}$$

$$\left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}}^{+} \\ \hline \mathbf{r}_{\notin j}^{+} \\ \hline r_j^{+} \end{array}\right]_{\mathrm{j}} = \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{a}_{\mathbf{y}} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{a}_{\notin j} \\ \hline \mathbf{0} & \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}}^{-} \\ \hline \mathbf{r}_{\notin j}^{-} \\ \hline r_j^{-} \end{array}\right]_{\mathrm{j}} \tag{37}$$

The corresponding covariance matrix for population estimates in Equation (39) is

$$\hat{\mathbf{V}}_{\mathrm{j}}^{+} = \mathbf{K}_{\mathrm{j}}\hat{\mathbf{V}}_{\mathrm{j}}^{-}\mathbf{K}_{\mathrm{j}}^{\mathrm{T}}$$

$$= \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{a}_{\mathbf{y}} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{a}_{\notin j} \\ \hline \mathbf{0} & \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c|c|c} \hat{\mathbf{V}}_{\mathbf{y}} & \hat{\mathbf{V}}_{\mathbf{y},\notin j} & \hat{\boldsymbol{\varphi}}_{\mathbf{y}}^{-} \\ \hline \hat{\mathbf{V}}_{\notin j,\mathbf{y}} & \hat{\mathbf{V}}_{\notin j} & \hat{\boldsymbol{\varphi}}_{\notin j}^{-} \\ \hline \left[\hat{\boldsymbol{\varphi}}_{\mathbf{y}}^{-}\right]^{\mathrm{T}} & \left[\hat{\boldsymbol{\varphi}}_{\notin j}^{-}\right]^{\mathrm{T}} & \hat{\Lambda}_j^{-} \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \hline \mathbf{a}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{a}_{\notin j}^{\mathrm{T}} & \left(1+a_j\right) \end{array}\right] \tag{38}$$

A more concise version of Equation (37) becomes useful later

$$\left.\begin{aligned} \hat{\mathbf{t}}_{\mathrm{j}}^{+} &= \left[\begin{array}{c|c} \mathbf{I} & \mathbf{a} \\ \hline \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c} \hat{\mathbf{z}}_j^{-} \\ \hline r_j^{-} \end{array}\right]_{\mathrm{j}} \\ &= \mathbf{K}_{\mathrm{j}} \left[\begin{array}{c} \hat{\mathbf{z}}_j^{-} \\ \hline r_j^{-} \end{array}\right]_{\mathrm{j}} \end{aligned}\right\} \quad \text{where} \left\{\begin{aligned} \mathbf{a} &= \left[\begin{array}{c} \mathbf{a}_{\mathbf{y}} \\ \hline \mathbf{a}_{\notin j} \end{array}\right]_{\mathrm{j}} \\ \mathbf{K}_{\mathrm{j}} &= \left[\begin{array}{c|c} \mathbf{I} & \mathbf{a} \\ \hline \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \end{aligned}\right. \tag{39}$$

In order to set the stage for the inequality constraints on coefficients in Sections 6.1 and 6.2 (below), define individual elements of the vector of population

 DOI: 10.13140/RG.2.2.17634.07369

estimates in Equation (39). The i^{th} element $\left(\hat{\mathbf{t}}_{\text{j}}^{+}\right)_i$ of $(M+J-1)\times 1$ vector estimate $\hat{\mathbf{t}}_{\text{j}}^{+}$ is

$$\left(\hat{t}_i^{+}\right)_{\text{j}} = \left\{\left(\hat{t}_i^{-}\right)_{\text{j}} + a_i\, r_j^{-}\right\}_{\text{j}} \tag{40}$$

The covariance matrix corresponding to the estimator in Equation (39) is

$$\begin{aligned}
\hat{\mathbf{V}}_{\text{j}}^{+} &= \mathbf{K}_{\text{j}}\left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{z}}^{-} & \hat{\boldsymbol{\varphi}}^{-} \\ \hline \left[\hat{\boldsymbol{\varphi}}^{-}\right]^{\text{T}} & \hat{\Lambda}^{-} \end{array}\right]_{\text{j}} \mathbf{K}_{\text{j}}^{\text{T}} \\
&= \left[\begin{array}{c|c} \mathbf{I} & \mathbf{a} \\ \hline \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\text{j}} \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{z}}^{-} & \hat{\boldsymbol{\varphi}}^{-} \\ \hline \left[\hat{\boldsymbol{\varphi}}^{-}\right]^{\text{T}} & \hat{\Lambda}^{-} \end{array}\right]_{\text{j}} \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{a}^{\text{T}} & \left(1+a_j\right) \end{array}\right]_{\text{j}} \\
&= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{z}} & \hat{\boldsymbol{\varphi}} \\ \hline \hat{\boldsymbol{\varphi}}^{\text{T}} & \hat{\Lambda} \end{array}\right]_{\text{j}}^{-} + \left[\begin{array}{c|c} \hat{\boldsymbol{\varphi}}\mathbf{a}_j^{\text{T}} + \mathbf{a}_j\hat{\boldsymbol{\varphi}}^{\text{T}} + \mathbf{a}_j\hat{\Lambda}\mathbf{a}_j^{\text{T}} & \hat{\boldsymbol{\varphi}}a_j + \mathbf{a}_j\hat{\Lambda}\left(1+a_j\right) \\ \hline a_j\hat{\boldsymbol{\varphi}}^{\text{T}} + \left(1+a_j\right)\hat{\Lambda}\mathbf{a}_j^{\text{T}} & 2\hat{\Lambda}a_j + \hat{\Lambda}a_j^2 \end{array}\right]_{\text{j}}^{-}
\end{aligned} \tag{41}$$

The $(i,k)^{\text{th}}$ element $\left\{\hat{v}_{i,k}^{+}\right\}_{\text{j}}$ is the scalar covariance with the difference-estimator between population estimates for the i^{th} and k^{th} variables in the j^{th} step equals

$$\left\{\hat{v}_{i,k}^{+}\right\}_{\text{j}} = \left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_k^{-}a_i + a_k\hat{\varphi}_i^{-} + a_i\hat{\Lambda}^{-}a_k\right\}_{\text{j}} \tag{42}$$

$\hat{\varphi}_k^{-}$ is the i^{th} element of covariance vector $\hat{\boldsymbol{\varphi}}^{-}$, which is the covariance between the population estimate for the i^{th} variable and the j^{th} auxiliary-residual at the beginning of the j^{th} step; and a_i is the i^{th} element of coefficient vector $\mathbf{a}_j$.

The scalar variance for the population estimate of the i^{th} variable is a special case of Equation (42)

$$\left\{\hat{v}_{i,i}^{+}\right\}_j = \left\{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_i^{-}a_i + \hat{\Lambda}^{-}a_i^2\right\}_j^{-} \quad \text{where} \begin{cases} \hat{\varphi}_i^{-} = \left(-\hat{v}_{y_i,x_H}^{-} + \hat{v}_{y_i,x_G}^{-}\right)_i \\ \hat{\Lambda}_{\text{j}} = \left(\hat{v}_{\mathbf{x}_H,\mathbf{x}_H}^{-} - 2\hat{v}_{\mathbf{x}_H,\mathbf{x}_G}^{-} + \hat{v}_{\mathbf{x}_H,\mathbf{x}_G}^{-}\right) \end{cases} \tag{43}$$

 DOI: 10.13140/RG.2.2.17634.07369

The corresponding difference-estimator for the scalar variance $\hat{\Lambda}$ of the j^{th} auxiliary-residual in Equations (38) and (41) is

$$\begin{aligned}\left\{\hat{\Lambda}^{+}\right\}_{j} &= \left\{\hat{\Lambda}^{-} + 2\hat{\Lambda}^{-}a_{j} + \hat{\Lambda}^{-}a_{j}^{2}\right\}_{j} \\ &= \hat{\Lambda}_{j}^{-}\left\{1 + 2a_{j} + a_{j}^{2}\right\}_{j}\end{aligned} \tag{44}$$

Equation (43) reveals several properties of the new difference-estimator. To illustrate, rescale parameter-space so that the variances in Equation (43) at the beginning of the j^{th} step are $\left\{\hat{v}_{i,i}^{-}\right\}_{\text{j}} = \left\{\hat{\Lambda}^{-}\right\}_{\text{j}} = 1$. Figure 2 plots the variance $\left\{\hat{v}_{i,i}\right\}_{\text{j}}^{+}$ of the difference-estimator at the end of the j^{th} step relative to the initial variance $\left\{\hat{v}_{i,i}\right\}_{\text{j}}^{-} = 1$ in this particular parameter-space. The variance with the new difference-estimator $\left\{\hat{v}_{i,i}\right\}_{\text{j}}^{+}$ is a function of three components: the correlation between population estimates for the i^{th} variable and the j^{th} auxiliary-residual, which equals $\rho_{i} = \hat{\varphi}_{i}$ in this parameter-space; the variance of the j^{th} auxiliary-residual, where $\left\{\hat{v}_{i,i}\right\}_{\text{j}}^{-} = 1$ in this parameter-space; and the value of arbitrary coefficient a_i, which is illustrated in Figure 2.

 DOI: 10.13140/RG.2.2.17634.07369

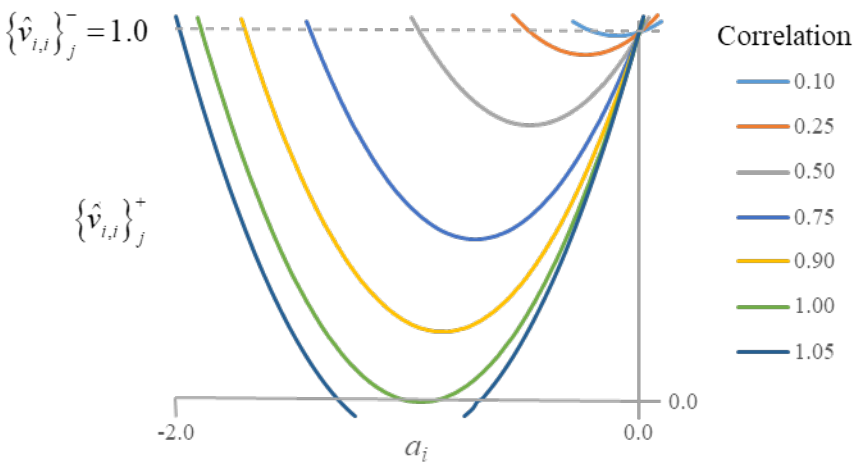

Figure 2. An example of the variance of the population estimate for the i^{th} variable $\{\hat{v}_{i,i}^+\}_j$ with the new difference estimator as a function of arbitrary coefficient a_i. There is unique value of arbitrary coefficient a_i that minimizes the variance. Also the correlation ρ_{ij} between the i^{th} variable and the j^{th} auxiliary-residual influences the magnitude of variance reduction $\{\hat{v}_{i,i}^+\}_j < \{\hat{v}_{i,i}^-\}_j$. However, if the correlation is infeasible (*e.g.*, $\rho_{ij} > 1.0$), then the variance estimate is negative $\{\hat{v}_{i,i}^+\}_j < 0$, *i.e.*, infeasible. Feasibility constraints on correlation coefficients in prior recursions $0,1,\ldots,(j-1)$ preclude infeasible variance estimates. Furthermore, the difference estimator can actually increase variances, *e.g.*, $\{\hat{v}_{i,i}^+\}_j > \{\hat{v}_{i,i}^-\}_j$, if the value of arbitrary coefficient a_i is too large or too small; inequality constraints on arbitrary coefficient a_i avoid this mistake.

 DOI: 10.13140/RG.2.2.17634.07369

Observations based on Figure 2:

A. The arbitrary coefficient a_i is "wisely chosen" (*i.e.*, $\left\{\hat{v}_{i,i}\right\}_{\mathrm{j}}^{+} < \left\{\hat{v}_{i,i}\right\}_{\mathrm{j}}^{-}$) if its value is within a certain numerical interval, although the reduction is not necessarily "optimal" and not necessarily substantial.

B. The difference-estimator will increase variance, *i.e.*, $\left\{\hat{v}_{i,i}\right\}_{\mathrm{j}}^{+} > \left\{\hat{v}_{i,i}\right\}_{\mathrm{j}}^{-}$, if the choice for the arbitrary coefficient a_i is "unwise", *i.e.*, not within the interval described in Observation A.

C. As the correlation ρ_i between population estimates for the i^{th} variable and the j^{th} auxiliary-residual increases, the variance-reduction improves. The improvement is minor, even negligible, if the correlation is weak.

D. The difference-estimator reduces variance even if correlation ρ_i is nearly zero. A spurious condition can occur through random sampling error, even if the true correlation is exactly zero. If spurious correlations are numerous, then the new difference-estimator can substantially "overfit" coefficients in the difference-estimator. This illustrates Obstacle 8 in Section 2.6 (above). This condition has also been observed as $j \to J$ in the recursion, in which case the variance of auxiliary-residuals $(j+1),\ldots,J$ approach zero.

E. The estimated variance with the new difference-estimator is negative only if the correlation is not bounded by $-1.0 \le \rho_i \le 1.0$ (*i.e.*, the coefficient of determination $\rho_i^2 < 1.0$). This aberration is not feasible by definition, but it is numerically possible, especially as the variance of the j^{th} auxiliary-residual approaches zero as $j \to J$ in the recursion.

F. There is a unique value of the arbitrary coefficient a_i that minimizes the variance with the difference-estimator; Section 3.4.1 (page 45) develops this further. (However, the minimum-variance coefficient is not always a wise choice; see "Constraints for a Robust Difference-Estimator" in Section 6, page 53).

3.4 "Well Chosen" Coefficients in the j[th] Recursion

The multivariate difference-estimator in Equations (16) and (17) is simply a linear transformation of the design-based π-estimator with a $(M+J)\times(M+J)$ matrix **K** of arbitrary constants. If those arbitrary coefficients are "well chosen", then the difference-estimator will reduce variance of population-totals for the

 DOI: 10.13140/RG.2.2.17634.07369

study-variables (Särndal *et al.*, Section 6.3). The section that follows below identifies criteria for well-chosen coefficients that are robust, efficient and feasible, and implements those criteria.

- Section 3.4.1 solves for coefficients that minimize the variance of population estimates for each the M study-variables with the j^{th} auxiliary-residual. This relates to Observation A (page 43), and it assures efficient estimates.
- Section 3.4.2 solves for coefficients such that covariance between the population estimate of the j^{th} auxiliary-residual and those for all other $(J-1)$ auxiliary-residuals equals zero. This criterion eliminates the problem of collinearity among auxiliary-residuals. In a heuristic sense, this implements a sequential orthogonalization. This assures estimates are not "over-fit" to the initial π estimates, and it addresses Observation E (page 43).

Within each j^{th} recursion, the new difference-estimator begins with minimum-variance coefficients from Sections 3.4.1 and 3.4.2. The remaining Sections impose inequality constraints on those coefficients that assure feasible and credible population estimates, although not necessarily minimum-variance estimates.

- Section 6.1 imposes inequality constraints on the coefficients that mitigate auxiliary-residual "outliers," which is an auxiliary-residual whose population is suspiciously or implausibly larger than expected given its estimated standard deviation. For example, define a plausible auxiliary-residual as one that is within ±2 standard deviation units from zero.
- Section 6.2 can impose minimum-maximum inequality constraints on population estimates for each of the M study-variables, such as an estimated population-total must be nonnegative. It does so by solving for the minimum and maximum interval for the m^{th} coefficient in the difference-estimator. If the minimum-variance coefficient in Section 3.4.1 does not reside within that interval, then the coefficient is changed to the appropriate limit within the min-max interval. In addition, Section 6.3 can impose inequality constraints on population estimates for all study-variables such that the estimates are bounded by prior confidence intervals.
- Section 6.3 imposes inequality constraints that assure feasible population estimates for the coefficients of determination (also known as "communalities"), denoted ρ^2. In the context here, the coefficient of determination ρ^2 equals the squared value of Pearson's correlation coefficient between two population estimates. Therefore, the maximum feasible value of the coefficient of determination is 1.0. A population estimate of a coefficient of

 DOI: 10.13140/RG.2.2.17634.07369

determination can exceed 1.0 if coefficients in the difference-estimator exceed a certain value, the ultimate consequence of which is the possibility of an estimated variance that is negative. This Section solves for constraints on coefficients in the difference-estimator that assure all coefficients of determination are ≤1.0.

- Section 6.4 imposes inequality constraints on coefficients that assure every variance estimate with the difference-estimator does not exceed the variance estimate without the difference-estimator. This addresses Observation B (above), and it precludes inefficient population estimates with the difference-estimator.
- Section 6.5 uses an inequality constraint on the minimum value of coefficient in the difference-estimator. A value that is relatively near zero will reduce variance of population estimates for study-variables, but the reduction will be small, and the computational burden might not be worth a miniscule gain in statistical efficiency. Also small coefficients can be symptomatic of "over fitting" with too many auxiliary-residuals.

Sequential implementation of these constraints on coefficients in the new difference-estimator is relatively easy within each of the J recursions. If all population estimates after the $(j\text{-}1)^{th}$ recursion are feasible and efficient, and all estimates with the j^{th} auxiliary-residual are feasible and efficient, then all population estimates after the final J^{th} recursion will be feasible and efficient.

3.4.1 Minimum-Variance Coefficients

Figure 2 (page 42) demonstrates existence of a unique choice for coefficient a_m that minimizes variance of the population estimate for the m^{th} study-variable with the j^{th} auxiliary-residual. Särndal *et al.* refer to this choice as the "optimal" coefficient. Section 3.4.1 derives that minimum-variance coefficient. These results directly extend to the $M\times 1$ vector of coefficients $\mathbf{a}_y$ that minimizes variance for all M study-variables $m = \{1,\ldots,M\}$ within the j^{th} recursion. The derivation is surprisingly simple.

Equation (43) expresses the variance of the difference-estimator $\left\{\hat{v}_{m,m}\right\}_j^+$ as a function of coefficient a_m. Derive the minimum-variance coefficient with the first derivative of that function

 DOI: 10.13140/RG.2.2.17634.07369

$$\frac{d\left\{\hat{v}_{m.m}\right\}_{\mathrm{j}}^{+}}{d\left(a_m\right)}=\frac{d}{d\left(a_m\right)}\left\{\hat{v}_{m,m}^{-}+2\hat{\varphi}_m^{-}a_m+\hat{\Lambda}_j^{-}a_m^2\right\}_{\mathrm{j}}$$
$$=\left\{2\hat{\varphi}_m^{-}+2\hat{\Lambda}_j^{-}a_m\right\}_{\mathrm{j}} \tag{45}$$

Solve Equation (45) for coefficient a_m such that derivative $\left\{2\hat{\varphi}_m^{-}+2\hat{\Lambda}_j^{-}a_m\right\}_{\mathrm{j}}=0$

$$\left\{a_m\right\}_{\mathrm{j,opt}}=\left\{\frac{-\hat{\varphi}_m^{-}}{\hat{\Lambda}_j^{-}}\right\}_{\mathrm{j}} \quad \text{where} \begin{cases}\hat{\varphi}_m^{-}=\left(-\hat{v}_{y_m,x_H}+\hat{v}_{y_m,x_G}\right)\\ \hat{\Lambda}_j^{-}=\left(\hat{v}_{\mathbf{x}_H,\mathbf{x}_H}-2\hat{v}_{\mathbf{x}_H,\mathbf{x}_G}+\hat{v}_{\mathbf{x}_G,\mathbf{x}_G}\right)\end{cases} \tag{46}$$

Equations (40) and (46) provide the minimum-variance difference-estimator of the population estimate for the m^{th} variable with the j^{th} auxiliary-residual

$$\left(\hat{t}_m^{+}\right)_{\mathrm{j}}=\left(\hat{t}_m^{-}\right)_{\mathrm{j}}+\left\{a_m\right\}_{\mathrm{j,opt}}r_{\mathrm{j}}^{-}$$
$$=\left(\hat{t}_m^{-}\right)_{\mathrm{j}}+\left(\frac{-\hat{\varphi}_m^{-}}{\hat{\Lambda}_j^{-}}\right)r_{\mathrm{j}}^{-} \tag{47}$$

The corresponding variance estimator $\left\{\hat{v}_{m,m}^{+}\right\}_j=\left\{\hat{v}_{m,m}^{-}+2\hat{\varphi}_m^{-}a_m+\hat{\Lambda}^{-}a_m^2\right\}_j^{-}$ in Equation (43) with the minimum-variance coefficient $\left\{a_m\right\}_{\mathrm{opt,j}}$ in Equation (46) equals the initial variance estimate $\left\{\hat{v}_{m,m}^{-}\right\}_{\mathrm{j}}$ at the beginning of the j^{th} step minus a nonnegative term

 DOI: 10.13140/RG.2.2.17634.07369

$$\left\{\hat{v}^{+}_{m,m}\right\}_{\mathrm{j,opt}} = \left\{\hat{v}^{-}_{m,m} + a_{m,\mathrm{opt}}\, 2\hat{\varphi}^{-}_{m} + a^{2}_{m,\mathrm{opt}}\hat{\Lambda}^{-}\right\}_{\mathrm{j}} \quad \text{where } \left\{a_{m}\right\}_{\mathrm{j,opt}} = \left\{\frac{-\hat{\varphi}^{-}_{m}}{\hat{\Lambda}^{-}_{j}}\right\}_{\mathrm{j}}$$

$$= \left\{\hat{v}^{-}_{m,m} - \left(\frac{\left(\hat{\varphi}^{-}_{m}\right)^{2}}{\hat{\Lambda}^{-}_{j}}\right)2 + \left(\frac{\left(\hat{\varphi}^{-}_{m}\right)^{2}}{\hat{\Lambda}^{-}}\right)\right\}_{\mathrm{j}}$$

$$= \left\{\hat{v}^{-}_{m,m} - \left(\frac{\left(\hat{\varphi}^{-}_{m}\right)^{2}}{\hat{\Lambda}^{-}_{j}}\right)\right\}_{\mathrm{j}} \tag{48}$$

Express Equation (48) in terms of the coefficient of determination $\left(\hat{\rho}^{-}_{m}\right)^{2}_{j}$ between the m^{th} study-variable and the j^{th} auxiliary-residual at the beginning of the j$^{\text{th}}$ recursion

$$\left\{\hat{v}^{+}_{m,m}\right\}_{\mathrm{j,opt}} = \left(\hat{v}^{-}_{m,m}\right)_{\mathrm{j}}\left\{1 - \left(\hat{\rho}^{-}_{m}\right)^{2}_{j}\right\} \quad \text{where } \left(\hat{\rho}^{-}_{m}\right)^{2}_{j} = \left(\frac{\left(\hat{\varphi}^{-}_{m}\right)^{2}}{\hat{v}^{-}_{m,m}\hat{\Lambda}^{-}_{j}}\right)_{\mathrm{j}} \tag{49}$$

If the estimate for the coefficient of determination $\left(\hat{\rho}^{-}_{m}\right)^{2}_{j} > 1.0$ in Equation (49), then the variance of the population estimate for the m^{th} variable will be negative, *i.e.*, infeasible. The inequality constraints in Section 6.3 (page 62) assure $\left(\hat{\rho}^{-}_{m}\right)^{2}_{\mathrm{q}} \le 1.0$ for $\left(1 \le \mathrm{q} < \mathrm{j}\right)$.

The difference-estimator in Equations (42) with the minimum-variance coefficients in Equation and (46) estimates the covariance between population estimates for the i^{th} and k^{th} with the j^{th} auxiliary-residual

 DOI: 10.13140/RG.2.2.17634.07369

$$\left\{\hat{v}_{i,k}^{+}\right\}_{\mathrm{j,opt}} = \left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_{k}^{-}a_{i,\mathrm{opt}} + a_{k,\mathrm{opt}}\hat{\varphi}_{i}^{-} + a_{i,\mathrm{opt}}\hat{\Lambda}^{-}a_{k,\mathrm{opt}}\right\}_{\mathrm{j}} \quad \text{where } a_{m,\mathrm{opt}} = \left(\frac{-\hat{\varphi}_{m}^{-}}{\hat{\Lambda}_{j}^{-}}\right)_{\mathrm{j}}$$

$$= \left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_{k}^{-}\left(\frac{-\hat{\varphi}_{i}^{-}}{\hat{\Lambda}_{j}^{-}}\right) + \left(\frac{-\hat{\varphi}_{k}^{-}}{\hat{\Lambda}_{j}^{-}}\right)\hat{\varphi}_{i}^{-} + \left(\frac{-\hat{\varphi}_{i}^{-}}{\hat{\Lambda}_{j}^{-}}\right)\hat{\Lambda}^{-}\left(\frac{-\hat{\varphi}_{k}^{-}}{\hat{\Lambda}_{j}^{-}}\right)\right\}_{\mathrm{j}}$$

$$= \left\{\hat{v}_{i,k}^{-} + \left(\frac{-\hat{\varphi}_{i}^{-}\hat{\varphi}_{k}^{-}}{\hat{\Lambda}_{j}^{-}}\right)\right\}_{\mathrm{j}} \tag{50}$$

3.4.2 Orthogonal Auxiliary-Residuals (Collinearity)

The stochastic model in Section 2.1, and the linear transformation for the auxiliary-residuals in Section 2.4, include a collinearity assumption, *i.e.*, nonzero covariances among auxiliary-residuals are possible. This section derives coefficients for the new difference-estimator such that the covariances between the difference-estimator for the j^{th} auxiliary-residual and any other (q^{th}) auxiliary-residual $\left(r_{q\neq j}^{-}\right)_{\mathrm{j}}$ are zero, *i.e.*, $\left\{\hat{v}_{j,q}^{+}\right\}_{\mathrm{j}} = 0$.

Substitute notation in Equation (42) to express the covariance of population estimates between the q^{th} and j^{th} auxiliary-residuals with the difference-estimator

$$\left\{\hat{\varphi}_{q}^{+}\right\}_{j} = \left\{\hat{\varphi}_{q}^{-} + \hat{\Lambda}^{-}a_{q} + a_{j}\hat{\varphi}_{q}^{-} + a_{q}\hat{\Lambda}^{-}a_{j}\right\}_{j} \tag{51}$$

where $\hat{\varphi}_{j}^{-} = \hat{\Lambda}^{-}$ and $\hat{v}_{q,j}^{-} = \hat{\varphi}_{q}^{-}$. By definition, $\left\{\hat{\varphi}_{q}^{+}\right\}_{\mathrm{j}}$ is the covariance between the q^{th} and j^{th} auxiliary-residuals. After the j$^{\text{th}}$ recursion, the q^{th} auxiliary-residual is independent of (*i.e.*, orthogonal to) the j^{th} auxiliary-residual if the covariance between these two auxiliary-residuals equals zero. Solve for coefficient a_q such that covariance $\left\{\hat{\varphi}_{q}^{+}\right\}_{\mathrm{j}} = 0$ in Equation (51)

$$0 = \left\{\hat{\varphi}_{q}^{-} + \hat{\Lambda}^{-}a_{q} + a_{j}\hat{\varphi}_{q}^{-} + a_{q}\hat{\Lambda}^{-}a_{j}\right\}_{j}$$

$$= \left\{a_{q}\hat{\Lambda}^{-}\left(1+a_{j}\right) + \hat{\varphi}_{q}^{-}\left(1+a_{j}\right)\right\}_{j} \tag{52}$$

 DOI: 10.13140/RG.2.2.17634.07369

Solve Equation (52) for coefficient $\{a_q\}_j$ such that covariance $\{\hat{\varphi}_q^+\}_j = 0$

$$\begin{aligned}\{a_q\}_j &= \left\{\frac{-\hat{\varphi}_q^-\left(1+a_j\right)}{\hat{\Lambda}^-\left(1+a_j\right)}\right\}_j \\ &= \left\{\frac{-\hat{\varphi}_q^-}{\hat{\Lambda}^-}\right\}_j\end{aligned} \tag{53}$$

This result equals the minimum-variance coefficient in Equation (46).

To verify that covariance $\{\hat{\varphi}_q^+\}_j = 0$, substitute coefficient $\{a_q\}_j$ from Equation (53) into the covariance estimator in Equation (51)

$$\begin{aligned}\{\hat{\varphi}_q^+\}_j &= \left\{\hat{\varphi}_q^- + \hat{\Lambda}^- a_q + a_j\hat{\varphi}_q^- + a_q\hat{\Lambda}^- a_j\right\}_j \\ &= \left\{\hat{\varphi}_q^- + \hat{\Lambda}^-\left(\frac{-\hat{\varphi}_i^-}{\hat{\Lambda}^-}\right) + a_j\hat{\varphi}_q^- + \left(\frac{-\hat{\varphi}_q^-}{\hat{\Lambda}^-}\right)\hat{\Lambda}^- a_j\right\}_j \\ &= \left\{\hat{\varphi}_q^- + \left(-\hat{\varphi}_q^-\right) + a_j\hat{\varphi}_q^- + \left(-\hat{\varphi}_q^-\right)a_j\right\}_j \\ &= 0\end{aligned} \tag{54}$$

Finally, determine coefficient a_j such that the population variance of the j^{th} residual equals exactly zero with Equation (44).

$$\begin{aligned}0 &= \hat{\Lambda}_j^-\left\{1+2a_j+a_j^2\right\}_j \\ a_j &= \left(-1\right)\end{aligned} \tag{55}$$

Results in this Section agree with Equations (31) and (33).

 DOI: 10.13140/RG.2.2.17634.07369

4 Concatenated Coefficients from the Recursive Sequence

Equation (37) from page 39 is the estimator for the $(M+J)\times 1$ vector of population statistics with the during the j^{th} recursion.

$$\hat{\mathbf{t}}_j^+ = \mathbf{K}_j \hat{\mathbf{t}}_j^- \quad \text{where } \hat{\mathbf{t}}_j^- = \hat{\mathbf{t}}_{j-1}^+ \tag{37}$$

After the J recursions, the new difference-estimator becomes the sequential product of each recursion

$$\begin{aligned} \hat{\mathbf{t}}_J^+ &= \left\{ \mathbf{K}_J, \ldots, \left[\mathbf{K}_2 \left(\mathbf{K}_1 \hat{\mathbf{t}}_1^- \right) \right] \right\} \quad \text{where } \hat{\mathbf{t}}_1^- = \hat{\mathbf{t}}_\pi \\ &= \mathbf{K}_J^+ \hat{\mathbf{t}}_\pi \quad \text{where } \mathbf{K}_J^+ = \prod_{j=1}^{J} \mathbf{K}_j \end{aligned} \tag{56}$$

The corresponding estimator for the $(M+J)\times(M+J)$ covariance matrix is simply

$$\hat{\mathbf{V}}_J^+ = \mathbf{K}_J^+ \hat{\mathbf{V}}_\pi \left[\mathbf{K}_J^+ \right]^T \tag{57}$$

Coefficient matrix **K** can serve as a $M\times(M+J)$ matrix of expansion factors (Czaplewski, 2010a, pp. 57-82) if the inclusion probabilities are known for each sampling element in the design-based π estimate from Section 2.3 (above). Section 12 (page 82) provides details. Furthermore, matrix **K** supports certain types of small-area and small-domain estimators in Czaplewski (2010a, pages 68-76, and 87-94).

5 Stepwise Covariate-Selection

Recall that the recursive difference-estimator sequentially processes one scalar auxiliary-residual at a time. Stepwise covariate-selection refers to the order in which each of those J scalar auxiliary-residuals is selected during the recursive

 DOI: 10.13140/RG.2.2.17634.07369

sequence. Since the new difference-estimator is model-free, covariate-selection is analogous to model-selection in model-assisted estimators (Särndal *et al.,* 1992).

If the $J \times J$ covariance matrix $\hat{\mathbf{\Lambda}}_{\pi}$ is full-rank and well-conditioned, then all J scalar auxiliary-residuals will be processed. However, such a well-behaved covariance matrix is unlikely if the number of auxiliary-variables is large, *i.e.*, $J > 20$. If covariance matrix $\hat{\mathbf{\Lambda}}_{\pi}$ is rank-deficient or ill-conditioned, then the recursive difference-estimator will terminate the recursion with fewer than J scalar auxiliary-residuals (see Section 5.2 below).

The final coefficient matrix with the recursive difference-estimator $\mathbf{K}_{\mathrm{J}}^{+} = \prod_{j=1}^{J} \mathbf{K}_{\mathrm{j}}$ in Equation (56) can differ with different sequences in which the scalar auxiliary-residuals are processed. The procedures here in Section 5 select one, and only one, of those possible sequences.

At the beginning of the j[th] recursion, the stepwise difference-estimator selects the one auxiliary-residual that "best" reduces variances of those study-variables. This is a multivariate analogy to stepwise-regression. In a heuristic sense, the stepwise difference-estimator allocates a limited amount of auxiliary information to improve population estimates for study-variables that are most important to the analyst. Details follow here.

Consider the beginning of the j[th] recursion. At this point in the recursive sequence, $(j-1)$ residuals have been processed, and $(J-j)$ residuals remain available to the recursive difference-estimator for further variance-reduction. The degree of variance-reduction for the M study-variables is one wise criterion for choosing among the remaining $(J-j)$ residuals.

Equation (48) on page 47, which is reproduced here, is the difference-estimator for the variance of the population estimate for the m^{th} study-variable with the q^{th} auxiliary-residual

$$\left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{q,opt}} = \left\{\hat{v}_{m,m}^{-} + a_{m,\mathrm{opt}}\, 2\hat{\varphi}_{m}^{-} + a_{m,\mathrm{opt}}^{2}\hat{\Lambda}_{q}^{-}\right\}_{\mathrm{q}} \quad \text{where } \left\{a_{m}\right\}_{\mathrm{q,opt}} = \left\{\frac{-\hat{\varphi}_{m}^{-}}{\hat{\Lambda}_{q}^{-}}\right\}_{\mathrm{q}} \qquad (48)$$

Equations (48) and (58) provide a numerically efficient preview for the degree of variance-reduction with each of the remaining residuals $j \le q \le J$. At the

 DOI: 10.13140/RG.2.2.17634.07369

beginning of the j[th] recursion, compute the trace of covariance matrix $\mathrm{tr}\left(\hat{\mathbf{\Lambda}}_q^-\right)_{\mathrm{j}}$ for each of the remaining auxiliary-residuals $j \le q \le J$

$$\begin{aligned}\tau_q &= \mathrm{tr}\left(\hat{\mathbf{\Lambda}}_q^-\right)_{\mathrm{j}} \\ &= \sum_{m=1}^{M}\left\{\hat{v}_{m,m}^+\right\}_{\mathrm{q,opt}} \quad \text{where } j \le q \le J\end{aligned} \tag{58}$$

If all study-variables are equally important to the analyst, then the auxiliary-residual (*i.e.*, the q^{th} auxiliary residual) with the smallest trace τ_q among these $(J - j + 1)$ traces in Equation (58) is the most efficient auxiliary-residual for variance-reduction during the j[th] recursion. If some study-variables are more important to the analyst than other study-variables, then the variances for each study variable can be weighted by an *ad hoc* "importance" coefficient c_m

$$\tau_q = \sum_{m=1}^{M} c_m \left\{\hat{v}_{m,m}^+\right\}_{\mathrm{q,opt}} \quad \text{where } j \le q \le J \tag{59}$$

5.1 Modifications to Minimum-Variance Coefficients

Section 6 (below) includes constraints that modify the minimum-variance coefficients $\{a_m\}_{\mathrm{q,opt}}$ in Equation (48). These constraints mitigate risk from outliers, numerical errors, and overfitting. With one exception, those modifications require little extra complexity or numerical overhead if applied to Equations (48) and (58). The exception is from Section 6.3 (page 62), which assures feasible communalities and nonnegative variances. To better anticipate the most important auxiliary-residual for variance-reduction, modify Equation (48) so that the minimum-variance coefficients a_m are constrained with criteria in Sections 6.1, 6.2 and 6.5 (below), and substitute into the covariate-selection criterion in Equation (59).

5.2 Termination of Recursive Sequence

If $J \times J$ covariance matrix $\hat{\mathbf{\Lambda}}_\pi$ is rank-deficient or ill-conditioned, then the maximum number of numerically feasible recursions will be less than J at some point during the recursive sequence. Therefore, define a criterion that terminates the recursive sequence.

 DOI: 10.13140/RG.2.2.17634.07369

Variance $\left(\hat{\Lambda}_q^-\right)_{\mathrm{j}}$ of the q^{th} auxiliary-residual r_q is a special case of Equation (49), in which $m = q$.

$$\left\{\hat{\Lambda}_q^+\right\}_{\mathrm{j,opt}} = \left(\hat{\Lambda}_q^-\right)_{\mathrm{j}} \left\{1 - \left(\hat{\rho}_q^-\right)_j^2\right\} \quad \text{where} \begin{cases} \hat{\Lambda}_q^- = \hat{v}_{q,q}^- \\ \left(\hat{\rho}_q^-\right)_j^2 = \left(\dfrac{\left(\hat{\varphi}_q^-\right)^2}{\hat{\Lambda}_q^- \hat{\Lambda}_j^-}\right)_{\mathrm{j}} \end{cases} \tag{60}$$

If the q^{th} auxiliary-residual is correlated with the j^{th} auxiliary-residual, *i.e.*, $0 < \left(\hat{\rho}_q^-\right)_j^2$, then the j^{th} auxiliary-residual with the minimum-variance coefficient for the q^{th} auxiliary-residual in Equation (49) will reduce the variance of the q^{th} auxiliary-residual, *i.e.*, $\left\{\hat{\Lambda}_q^+\right\}_{\mathrm{j,opt}} < \left\{\hat{\Lambda}_q^-\right\}_{\mathrm{j}}$ for $q > j$. As the recursion progresses $j \to J$, the variances of the auxiliary-residuals will decrease in the presence of collinearity, *i.e.*, $\mathrm{tr}\left(\hat{\mathbf{\Lambda}}_{\mathrm{j}}^-\right) \to 0$, and the inverse variance for the minimum-variance coefficient in Equation (46) becomes numerically unstable as $1 \big/ \hat{\Lambda}_q^- \to \infty$ for $j < q < J$. This is consistent with results in Section 2.5 (page 25).

Section 6 (below) includes inequality constraints and other conventions that mitigate numerical errors, especially those related to the variance-reduction of auxiliary-residuals during the recursive sequence as $\mathrm{tr}\left(\hat{\mathbf{\Lambda}}_{\mathrm{j}}^-\right) \to 0$. At some point in that sequence, none of the remaining auxiliary-residuals support numerically reliable results. That is the point to terminate the stepwise recursion.

6 Constraints for a Robust Difference-Estimator

Recall that the new difference-estimator in Section 2.5 (page 25) is the best linear estimator that minimizes variances of population estimates for the M study-variables. However, computation of minimum-variance coefficients in Equation (31) requires an inverse of $J \times J$ covariance matrix $\mathbf{\Lambda}_\pi$ in Equation (13). That matrix inverse is not feasible if $\mathbf{\Lambda}_\pi$ is rank-deficient.

 DOI: 10.13140/RG.2.2.17634.07369

The recursive version of the new difference-estimator in Section 3.4 (above) also reduces variances of population estimates for study-variables that are correlated with one or more auxiliary-residuals. However, an auxiliary-residual can be a linear combination of other auxiliary-variables, which reduces the rank of covariance matrix $\hat{\mathbf{\Lambda}}_j^-$. If $\mathbf{\Lambda}_\pi$ is rank-deficient, then the scalar inverse $1/\hat{\Lambda}_j^-$ within some, but not all, recursions can be feasible. Other numerical issues can arise within each recursion as $j \to J$; $\mathrm{tr}\left(\hat{\mathbf{\Lambda}}_j^-\right) \to 0$; the J eigenvalues of $\hat{\mathbf{\Lambda}}_j^-$ shrink towards zero; and off-diagonal covariances in $\hat{\mathbf{\Lambda}}_j^-$ approach zero. Section 6 that follows addresses these implementation problems.

Section 6.1 (page 55) mitigates the leverage imposed by an outlier residual, which is an auxiliary-residual that is unexpectedly large compared to the best estimate of its standard deviation. Three difference constraints reduce the influence of an outlier residual on population estimates for all other variables.

Section 6.2 (page 60) imposes minimum and maximum inequality constraints on population estimates for study-variables. For example, an estimate of the number of population elements in a subpopulation cannot be negative, and an estimate of a proportion of a population cannot exceed 1.0.

Section 6.3 (page 62) assures that variance estimates are nonnegative. However, a negative variance is actually a symptom of an infeasible estimate of the coefficient of determination between an auxiliary-residual and a study-variable. The coefficient of determination is the squared correlation between these two variables, and it is concisely referred to as a "communality." The estimate is infeasible if the communality exceeds 1.0, and this aberration is equivalent to a correlation coefficient that is not bounded by ±1.0.

Section 6.4 (page 66) assures that the difference-estimator does not increase variance of population estimates, which is numerically possible, but obviously an unwise choice.

Section 6.5 (page 67) reduces risk from a certain type of numerical instability. The minimum-variance coefficient for the m^{th} study-variable is the ratio between two population statistics: the covariance between the m^{th} study-variable and the j^{th} auxiliary-residual. As the recursion proceeds, those covariances approach zero while the inverse variance of the auxiliary-residual approaches infinity. Their ratio can become numerically unstable.

 DOI: 10.13140/RG.2.2.17634.07369

Section 6.6 (page 69) reduces risk of 'overfitting" population estimates for a large number of study-variables with a large number of auxiliary-residuals, many of which have small but nonzero correlations with the study-variables.

6.1 Outlier Auxiliary-Residuals

Recall from Equation (47), which is reproduced here, that the noew difference-estimator for $\left(\hat{t}_m^+\right)_{\mathrm{j}}$ equals the prior population estimate $\left(\hat{t}_m^-\right)_{\mathrm{j}}$ plus an adjustment term $\left\{a_m\right\}_{\mathrm{j,opt}} r_{\mathrm{j}}^-$, where $\left\{a_m\right\}_{\mathrm{j,opt}}$ is the minimum-variance coefficient and r_{j}^- is the j^{th} auxiliary-residual

$$\left(\hat{t}_m^+\right)_{\mathrm{j}} = \left(\hat{t}_m^-\right)_{\mathrm{j}} + \left\{a_m\right\}_{\mathrm{j,opt}} r_{\mathrm{j}}^- \tag{47}$$

If the value of residual r_{j}^- is relatively large, then the difference between the prior estimate $\left(\hat{t}_m^-\right)_{\mathrm{j}}$ and the difference-estimate $\left(\hat{t}_m^+\right)_{\mathrm{j}}$ will also be relatively large. If the population estimate for auxiliary-residual r_{j}^- is suspiciously inaccurate, *i.e.*, an "outlier", then the difference-estimate is strongly affected by that outlier. The source of an outlier residual can be an extreme, but rare, random sampling error; or it can be an unquantified systematic error, such as deterministic numerical errors or database errors. Regardless, neither circumstance is desirable, and inequality constraints on coefficient $\left\{a_m\right\}_{\mathrm{j}}$ can mitigate risk from an outlier residual.

Mitigation requires objective *a prior* criteria that classify an auxiliary-residual as an outlier within each of the J recursions. Three criteria are available here, and all three support inequality constraints that reduce the leverage of an outlier residual. All three adapt the Bayesian concept of a "credibility interval" (Edwards *et al.*, 1963). The first two criteria use a credibility interval for the initial auxiliary-residual r_j^-, while the third criterion uses a credibility interval for the difference between the population estimates at the beginning and end of the jth recursion $\left(r_m\right)_{\mathrm{j}} = \left(\hat{t}_m^- + \hat{t}_m^+\right)_{\mathrm{j}}$.

All three mitigation methods identify an outlier based on the maximum number of standard deviation units $\pm\sigma_{\max}$ between the expected and realized population estimates. All three methods use Equation (12), which assumes the residuals have

 DOI: 10.13140/RG.2.2.17634.07369

an expected value of zero, *i.e.*, $\mathrm{E}\left[r_{\mathrm{j}}\right]=0$ and $\mathrm{E}\left[\left(r_m\right)_{\mathrm{j}}\right]=0$; and all three assume a symmetric credibility interval. All three impose inequality constraints on the arbitrary coefficient $\left\{a_i\right\}_{\mathrm{j}}$ in the difference-estimator; and none affect the initial estimate of the variance for the j^{th} auxiliary-residual $\hat{\Lambda}_{\mathrm{j}}^{-}$.

The analyst's choice for parameter $\pm\sigma_{\max}$ reflects the analyst's tolerance for risk from outlier residuals. For example, a standard deviation unit of $\left|\sigma_{\max}\right|=1.0$ is a risk-adverse choice; if the auxiliary-residual is not an outlier, and random error for the estimated residual has a normal distribution, then approximately 32-percent of residuals are treated as outliers. If the auxiliary-residual is not truly an outlier, then that auxiliary-residual will not maximize variance reduction. A risk-adverse choice is appropriate if the auxiliary-variables are predictions from a deterministic model for population dynamics, and the prediction error is not well quantified. If the analyst's choice is less conservative, *e.g.*, $\left|\sigma_{\max}\right|=2.0$, then approximately 5-percent of residuals will be treated as outliers. Outlier mitigation is a tradeoff between statistical efficiency and risk reduction.

The following three methods reduce the magnitude of coefficient $\left|a_m\right|_{\mathrm{j}}$ such that $\left|a_m\right|_{\mathrm{j}}<\left|a_m\right|_{\mathrm{j,opt}}$ in Equation (47). This reduces the influence of a suspected outlier residual r_{j}^{-} in Equation (47) on population estimates for all study-variables with the j$^{\text{th}}$ residual. Variance is still reduced with the constrained coefficient, but not to the degree to which it would be reduced with the minimum-variance coefficient.

6.1.1 Constraint for Implausibly Large Auxiliary-Residual

The first inequality constraint uses a credibility interval around the initial auxiliary-residual r_j^{-},

$$\left(-\sigma_{\max}\sqrt{\hat{\Lambda}_{\mathrm{j}}^{-}}\right)\le r_j^{-}\le\left(+\sigma_{\max}\sqrt{\hat{\Lambda}_{\mathrm{j}}^{-}}\right) \tag{61}$$

where $\sqrt{\hat{\Lambda}_{\mathrm{j}}^{-}}$ is an estimate of standard deviation of the j^{th} auxiliary-residual at the beginning of the j$^{\text{th}}$ recursion, which is defined in Equation (35). Therefore, standard deviation units define the credibility interval. Declare auxiliary-residual

 DOI: 10.13140/RG.2.2.17634.07369

r_j^- as a suspected outlier if it is not bounded by the credibility interval in Equation (61).

If the residual is not credible because $r_j^- > \left(+\sigma_{\max}\sqrt{\hat{\Lambda}_{\mathrm{j}}^-}\right)$, then assume the estimate for the standard deviation $\sqrt{\hat{\Lambda}_{\mathrm{j}}^-}$ is too small. Define the credible standard deviation $\sqrt{\hat{\Lambda}_{\mathrm{j}\sigma}^-}$ as

$$\sqrt{\hat{\Lambda}_{\mathrm{j},\sigma}^-} = \frac{r_j^-}{+\sigma_{\max}} \tag{62}$$

If the residual is not credible because $r_j^- < \left(-\sigma_{\max}\sqrt{\hat{\Lambda}_{\mathrm{j}}^-}\right)$, then define the credible standard deviation $\sqrt{\hat{\Lambda}_{\mathrm{j}\sigma}^-}$ as

$$\sqrt{\hat{\Lambda}_{\mathrm{j}\sigma}^-} = \frac{r_j^-}{-\sigma_{\max}} \tag{63}$$

Use the "credible standard deviation" in Equation (62) or (63) to compute the constrained coefficients for population estimate of all (*M*+*J*) variables

$$\left\{a_i\right\}_{\mathrm{j}\sigma} = \left\{\frac{-\hat{\varphi}_i^-}{\hat{\Lambda}_{\mathrm{j}\sigma}^-}\right\}_{\mathrm{j}} \quad \text{for } i = 1,\ldots,(M+J) \tag{64}$$

6.1.2 Constraint on Interval Coverage of Zero

From Equation (12), the auxiliary-residual has an expected value of zero, *i.e.*, $\mathrm{E}\left[r_{\mathrm{j}}\right] = 0$. Therefore, the credibility interval around the residual in Equation (61) should cover zero, *i.e.*,

$$\left(+\sigma_{\max}\sqrt{\hat{\Lambda}_{\mathrm{j,opt}}^-}\right) < 0.0 \tag{65}$$

 DOI: 10.13140/RG.2.2.17634.07369

$$\left(-\sigma_{\max}\sqrt{\hat{\Lambda}^{-}_{\mathrm{j,opt}}}\right) > 0.0 \tag{66}$$

If the condition in Equation (65) exists, then then assume the estimate for the standard deviation $\sqrt{\hat{\Lambda}^{-}_{\mathrm{j}}}$ is too small. Given Equation (61), define the credible standard deviation $\sqrt{\hat{\Lambda}^{-}_{\mathrm{j}\sigma}}$ as one in which

$$r_j^{-} + \sigma_{\max}\sqrt{\hat{\Lambda}^{-}_{\mathrm{j}\sigma}} = 0 \tag{67}$$

and solve for $\hat{\Lambda}^{-}_{\mathrm{j}\sigma}$

$$\hat{\Lambda}^{-}_{\mathrm{j}\sigma} = \left(\frac{-r_j^{-}}{\sigma_{\max}}\right)^2 \quad \text{for } \left(+\sigma_{\max}\sqrt{\hat{\Lambda}^{-}_{\mathrm{j,opt}}}\right) < 0.0 \tag{68}$$

Likewise, solve for $\hat{\Lambda}^{-}_{\mathrm{j}\sigma}$ if the condition in Equation (66) exists

$$\hat{\Lambda}^{-}_{\mathrm{j}\sigma} = \left(\frac{+r_j^{-}}{\sigma_{\max}}\right)^2 \quad \text{for } \left(-\sigma_{\max}\sqrt{\hat{\Lambda}^{-}_{\mathrm{j,opt}}}\right) > 0.0 \tag{69}$$

Use the "credible standard deviation" $\hat{\Lambda}^{-}_{\mathrm{j}\sigma}$ in Equation (68) or (69) to compute the constrained coefficient for population estimate of all (M+J) variables

$$\{a_i\}_{\mathrm{j}\sigma} = \left\{\frac{-\hat{\varphi}_i^{-}}{\hat{\Lambda}^{-}_{\mathrm{j}\sigma}}\right\}_{\mathrm{j}} \quad \text{for } i = 1,\ldots,(M+J) \tag{70}$$

 DOI: 10.13140/RG.2.2.17634.07369

6.1.3 Constraint on the i^{th} Population Estimate

The second inequality constraint uses the credibility interval for the difference $\left(r_i\right)_{\mathrm{j}}$ between the initial population estimate $\left(\hat{t}_i^{-}\right)_{\mathrm{j}}$ and its difference-estimate $\left(\hat{t}_i^{+}\right)_{\mathrm{j}}$ in the j[th] recursion

$$\begin{aligned}\left(r_i\right)_{\mathrm{j}} &= \left(-\left(\hat{t}_i^{-}\right)_{\mathrm{j}} + \left(\hat{t}_i^{+}\right)_{\mathrm{j}}\right) \quad \text{for } i = 1,\ldots,(M+J) \\ &= \left[-1 \;\vdots\; 1\right]\left(\left[\begin{array}{c:c} 1 & 0 \\ \hdashline 1 & a_i \end{array}\right]\left[\begin{array}{c} \hat{t}_i^{-} \\ \hdashline r \end{array}\right]\right)_{\mathrm{j}} \\ &= \left(\left[0 \;\vdots\; a_i\right]\left[\begin{array}{c} \hat{t}_i^{-} \\ \hdashline r \end{array}\right]\right)_{\mathrm{j}} \\ &= a_i r_{\mathrm{j}}\end{aligned} \tag{71}$$

The corresponding variance of the linear transformation in Equation (71) is

$$\begin{aligned}\left\{\hat{v}_{r_i}\right\}_{\mathrm{j}} &= \left\{\left[0 \;\vdots\; a_i\right]\left[\begin{array}{c:c} \hat{v}_{i,i} & \hat{\varphi}_i \\ \hdashline \hat{\varphi}_i & \hat{\Lambda}^{-} \end{array}\right]\left[\begin{array}{c} 0 \\ \hdashline a_i \end{array}\right]\right\}_{\mathrm{j}} \quad \text{for } i = 1,\ldots,(M+J) \\ &= \left\{a_i^2 \hat{\Lambda}^{-}\right\}_{\mathrm{j}}\end{aligned} \tag{72}$$

The credibility interval that defines a plausible variance $\left\{\hat{v}_{r_i}\right\}_{\mathrm{j}}$ of the j^{th} auxiliary-residual $\left(r_i\right)_{\mathrm{j}}$ is

$$-\sigma_{\max}^2 \le \left\{\hat{v}_{r_i}\right\}_{\mathrm{j}} \le +\sigma_{\max}^2 \tag{73}$$

If the difference between estimates in Equation (71) not credible because $\left\{\hat{v}_{r_i}\right\}_{\mathrm{j}} > +\sigma_{\max}^2$ in Equation (73), then constrain the i^{th} coefficient such that

$$\left\{a_{i,\sigma}^2 \hat{\Lambda}^{-}\right\}_{\mathrm{j}} = \left(\pm\sigma_{\max}\right)^2 \tag{74}$$

 DOI: 10.13140/RG.2.2.17634.07369

Solve Equation (74) for the constrained coefficient $\left\{a_{i,\sigma}\right\}_{\mathrm{j}}$

$$\left\{a_{i,\sigma}\right\}_{\mathrm{j}} = \left(-\sqrt{\left\{ \frac{\sigma^2_{\max}}{\hat{\Lambda}^{-}} \right\}_{\mathrm{j}}} \right) \quad \text{for } \left\{a_i\right\} < 0$$

$$\left\{a_{i,\sigma}\right\}_{\mathrm{j}} = \left(+\sqrt{\left\{ \frac{\sigma^2_{\max}}{\hat{\Lambda}^{-}} \right\}_{\mathrm{j}}} \right) \quad \text{for } \left\{a_i\right\} > 0 \tag{75}$$

This constraint will not be effective if the variance of the auxiliary-residual is very small, i.e., $\hat{\Lambda}^{-}_{\mathrm{j}} \approx 0$, which is expected as the recursive sequence proceeds.

6.2 Min-Max Inequality Constraints

Prior knowledge about the feasible values of population parameters might be available. For example, the number of timberland hectares in a spatial population cannot be negative, nor can that number exceed the total number of hectares in the sampled population. Therefore, impose an inequality constraint on each of the arbitrary coefficients in the difference-estimator such that its corresponding population estimate is bounded by the *a priori* minimum and maximum values whenever available. Constraints in this Section separately apply to each population estimate of a study-variable, regardless of the credibility of the j^{th} auxiliary-residual; while constraints in Section 6.1 (above) apply to each and every study-variable if the estimated value of the j^{th} auxiliary-residual is not credible.

Equation (40) on page 40, which is reproduced here, gives the difference-estimator with arbitrary coefficient a_m for the population estimate of the m^{th} study-variable (m=i)

$$\left(\hat{t}^{+}_{m}\right)_{\mathrm{j}} = \left\{ \left(\hat{t}^{-}_{m}\right)_{\mathrm{j}} + a_m \, r^{-}_{j} \right\}_{\mathrm{j}} \tag{40}$$

Constrain the difference-estimator $\left(\hat{t}^{+}_{m}\right)_{\mathrm{j}}$ such that its estimated value is within its *a priori* minimum and maximum values

$$\left(t_m\right)_{\min} \;\le\; \left(\hat{t}^{-}_{m}\right)_{\mathrm{j}} + a_m \, r^{-}_{j} \;\le\; \left(t_m\right)_{\max} \tag{76}$$

 DOI: 10.13140/RG.2.2.17634.07369

If the constraint in Equation (76) is not satisfied because $\left(t_m\right)_{\min} > \left(\hat{t}_m^+\right)_{\mathrm{j}}$, then solve for the arbitrary coefficient a_m such that the difference-estimator equals the minimum inequality constraint for the m^{th} study-variable

$$\left\{a_m\right\}_{\mathrm{j,min}} = \left\{\frac{\left(t_m\right)_{\min} - \left(\hat{t}_m^-\right)_{\mathrm{j}}}{r_j^-}\right\}_{\mathrm{j}} \quad \text{for } \left(\hat{t}_m^+\right)_{\mathrm{j}} < \left(t_m\right)_{\min} \tag{77}$$

Likewise, if $\left(\hat{t}_m^+\right)_{\mathrm{j}} > \left(t_m\right)_{\max}$, then solve Equation (76) for the arbitrary coefficient a_m such that the difference-estimator equals the maximum inequality constraint for the m^{th} study-variable

$$\left\{a_m\right\}_{\mathrm{j,max}} = \left\{\frac{\left(t_m\right)_{\max} - \left(\hat{t}_{\mathrm{j}}^-\right)_m}{r_j^-}\right\}_{\mathrm{j}} \quad \text{for } \left(\hat{t}_m^+\right)_{\mathrm{j}} > \left(t_m\right)_{\max} \tag{78}$$

Therefore, impose inequality constraints on the population estimate for the m^{th} study-variable by constraining coefficient a_m within the interval

$$\left\{a_m\right\}_{\mathrm{j,min}} \le \left\{a_m\right\}_{\mathrm{j}} \le \left\{a_m\right\}_{\mathrm{j,max}} \tag{79}$$

A credibility interval might further define the minimum and maximum constraints on the population estimate of a study-variable. For example, assume a normal distribution of random estimation errors for the m^{th} study-variable. Define the credibility interval relative to the estimated standard deviation of the population estimate.

$$\left\{\left(\hat{t}_m^-\right)_{\mathrm{j}} - \sigma_{\max}\sqrt{\left(\hat{v}_{m,m}^-\right)_{\mathrm{j}}}\right\} \le \left\{\left(\hat{t}_m^-\right)_{\mathrm{j}} + a_m\, r_j^-\right\} \le \left\{\left(\hat{t}_m^-\right)_{\mathrm{j}} + \sigma_{\max}\sqrt{\left(\hat{v}_{m,m}^-\right)_{\mathrm{j}}}\right\} \tag{80}$$

and solve for the constraints on coefficient $\left\{a_m\right\}_{\mathrm{j}}$

$$\left\{\frac{-\sigma_{\max}\sqrt{\hat{v}_{m,m}^-}}{r_j^-}\right\}_{\mathrm{j}} \le \left\{a_m\right\}_{\mathrm{j}} \le \left\{\frac{\sigma_{\max}\sqrt{\hat{v}_{m,m}^-}}{r_j^-}\right\}_{\mathrm{j}} \tag{81}$$

 DOI: 10.13140/RG.2.2.17634.07369

Other assumed distributions might be more appropriate than the normal distribution. For example, if a population parameter must have nonnegative minimum value (e.g., zero hectares of timberland), then the inverted beta distribution might be a good choice. If the population parameter must be nonnegative and less than some maximum value (the estimate of a population proportion must be between 0.0 and 1.0), then the beta distribution might be a good choice. This type of inequality constraint complements those in Section 6.1 (above) to further reduce risk from outliers.

6.3 Feasible Communalities and Nonnegative Variances

A communality (*i.e.*, coefficient of determination) between population estimates for the i^{th} and k^{th} variables, denoted $\hat{\rho}_{i,k}^2$, is the squared value of the corresponding Pearson's correlation coefficient, denoted $\pm\hat{\rho}_{i,k}$. Therefore, a feasible estimate with the difference-estimate must be bounded by $0 \le \hat{\rho}_{i,k}^2 \le 1$. However, it is numerically possible for an estimated communality to exceed 1.0. Figure 2 offers an example of a negative variance estimate if the correlation is 1.05. Of course, this correlation is infeasible because, by definition, a valid correlation coefficient cannot exceed 1.00.

A difference-estimate for a correlation that is infeasible can cause negative variance estimates in a subsequent $\mathrm{j} = (j+1),\ldots,J$ recursion, which is also infeasible. This pathological condition in factor analysis is termed an "ultra-Heywood case" (SaS, 1999, Chapter 26, Section 21). A negative estimate of a variance is a symptom of an infeasible estimate of a communality. Constraints that preclude an infeasible communality will also preclude a negative variance estimate with the difference-estimator.

To be more precise, consider the minimum-variance difference-estimator for the m^{th} population variance in Equation (49) from page 47 in subsequent recursion *n*.

$$\left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{n,opt}} = \left(\hat{v}_{m,m}^{-}\right)_{\mathrm{n}} \left\{1 - \left(\hat{\rho}_{m}^{-}\right)_{\mathrm{n}}^{2}\right\} \quad \text{for } \mathrm{n} < \mathrm{j} \tag{49}$$

Clearly, $\left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{n,opt}} < 0$ if $\left(\hat{\rho}_{m}^{-}\right)_{\mathrm{n}}^{2} > 1.0$. The following solves for constraints on coefficients in the difference-estimator during the j^{th} recursion that assure each and

 DOI: 10.13140/RG.2.2.17634.07369

every communality is constrained such that $\left(\hat{\rho}_{i,n}^{+}\right)_{\mathrm{j}}^{2} \le 1.0$, and therefore, all variance estimates in subsequent recursions q > j are nonnegative.

The communality between population estimates for the i^{th} and k^{th} variables with the j^{th} auxiliary-residual is, by definition, equal to

$$\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2} = \left(\frac{\left\{\hat{v}_{i,k}^{+}\right\}^{2}}{\left\{\hat{v}_{i,i}^{+}\right\}\left\{\hat{v}_{k,k}^{+}\right\}} \right)_{\mathrm{j}} \tag{82}$$

Expand Equation (82) with the difference-estimators in Equations (42) and (43)

$$\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2} = \left(\frac{\left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_{k}^{-}a_{i} + a_{k}\hat{\varphi}_{i}^{-} + a_{i}\hat{\Lambda}^{-}a_{k}\right\}^{2}}{\left\{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_{i}^{-}a_{i} + \hat{\Lambda}^{-}a_{i}^{2}\right\}\left\{\hat{v}_{k,k}^{-} + 2\hat{\varphi}_{k}^{-}a_{k} + \hat{\Lambda}^{-}a_{k}^{2}\right\}} \right)_{\mathrm{j}} \tag{83}$$

Constrain the communality $\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2}$ in Equation (86)

$$0.0 \le \left(\frac{\left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_{k}^{-}a_{i} + a_{k}\hat{\varphi}_{i}^{-} + a_{i}\hat{\Lambda}^{-}a_{k}\right\}^{2}}{\left\{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_{i}^{-}a_{i} + \hat{\Lambda}^{-}a_{i}^{2}\right\}\left\{\hat{v}_{k,k}^{-} + 2\hat{\varphi}_{k}^{-}a_{k} + \hat{\Lambda}^{-}a_{k}^{2}\right\}} \right)_{\mathrm{j}} \le 1.0 \tag{84}$$

Since the numerator and the denominator in Equation (84) are both nonnegative, the lower inequality constraint is always satisfied. However, if $\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2} > 1.0$, then constrain coefficients a_i and/or a_k such that $\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2} = 1.0$.

$$1.0 = \left(\frac{\left\{\hat{v}_{i,k}^{-} + \hat{\varphi}_{k}^{-}a_{i} + a_{k}\hat{\varphi}_{i}^{-} + a_{i}\hat{\Lambda}^{-}a_{k}\right\}^{2}}{\left\{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_{i}^{-}a_{i} + \hat{\Lambda}^{-}a_{i}^{2}\right\}\left\{\hat{v}_{k,k}^{-} + 2\hat{\varphi}_{k}^{-}a_{k} + \hat{\Lambda}^{-}a_{k}^{2}\right\}} \right)_{\mathrm{j}} \tag{85}$$

One solution is to treat coefficient a_k in Equation (85) as a constant, and solve the quadratic in Equation (86) for coefficient a_i

 DOI: 10.13140/RG.2.2.17634.07369

$$0=\left(\begin{array}{c} +\left[\left(\hat{\varphi}_k^-\right)^2-\hat{v}_{k,k}^-\hat{\Lambda}^-\right]a_i^2+ \\ +\left[\begin{array}{c} 2\hat{v}_{i,k}^-\hat{\varphi}_k^-+2\hat{v}_{i,k}^-\hat{\Lambda}^-a_k+ \\ -2\hat{v}_{k,k}^-\hat{\varphi}_i^--2\hat{\varphi}_i^-\hat{\varphi}_k^-a_k \end{array}\right]a_i+ \\ +\left[\begin{array}{c} \left(\hat{v}_{i,k}^-\right)^2+2\hat{v}_{i,k}^-\hat{\varphi}_i^-a_k+\left(\hat{\varphi}_i^-\right)^2a_k^2+ \\ -\hat{v}_{i,i}^-\hat{v}_{k,k}^--2\hat{v}_{i,i}^-\hat{\varphi}_k^-a_k-\hat{v}_{i,i}^-\hat{\Lambda}^-a_k^2 \end{array}\right] \end{array}\right) \tag{86}$$

If the i^{th} or k^{th} variable is an auxiliary-residual, then there are other considerations. Recall from Section 3.4.2 that the j^{th} auxiliary-residual is orthogonal to all other auxiliary-residuals with the minimum-variance coefficients. If coefficient a_i corresponds to an auxiliary-residual ($i\neq j$) and a_k corresponds to a study-variable, *i.e.*, $M\times J$ partition $\mathbf{\Gamma}$ in Equation (13), then retain the minimum-variance coefficient $\left\{a_i\right\}_{\text{j,opt}}=\left\{-\hat{\varphi}_i^-\middle/\hat{\Lambda}_j^-\right\}_{\text{j}}$ and constrain coefficient a_k in order to assure a feasible communality $\left(\hat{\rho}_{i,k}^+\right)_j^2\leq 1.0$ while maintaining orthogonality. Likewise if coefficient a_i corresponds to a study-variable and coefficient a_k corresponds to an auxiliary-residual ($i\neq j$). If the i^{th} and k^{th} variables are both auxiliary-variables ($i\neq j$ and $k\neq j$), *i.e.*, $J\times J$ partition $\mathbf{\Lambda}$ in Equation (13), then maintain orthogonality with coefficient values a_i=0.0 and a_k=0.0.

Express the right-hand side of Equation (83) in terms of a standardized parameter-space

 DOI: 10.13140/RG.2.2.17634.07369

$$\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j}}^{2}=\left(\frac{\left\{\hat{v}_{i,k}^{-}+\hat{\varphi}_{k}^{-}a_{i}+a_{k}\hat{\varphi}_{i}^{-}+a_{i}a_{k}\right\}^{2}}{\left\{1+2\hat{\varphi}_{i}^{-}a_{i}+a_{i}^{2}\right\}\left\{1+2\hat{\varphi}_{k}^{-}a_{k}+a_{k}^{2}\right\}}\right)_{\mathrm{j}} \quad \text{for} \begin{cases}\hat{v}_{i,i}^{-}=1\\ \hat{v}_{k,k}^{-}=1\\ \hat{\Lambda}^{-}=1\end{cases}$$

$$=\left(\frac{\left\{\hat{\rho}_{i,k}^{-}+\hat{\rho}_{k,j}^{-}a_{i}+a_{k}\hat{\rho}_{i,j}^{-}+a_{i}a_{k}\right\}^{2}}{\left\{1+2\hat{\rho}_{i,j}^{-}a_{i}+a_{i}^{2}\right\}\left\{1+2\hat{\rho}_{k,j}^{-}a_{k}+a_{k}^{2}\right\}}\right)_{\mathrm{j}} \quad \text{for} \begin{cases}\hat{\rho}_{i,j}^{-}=\dfrac{\hat{\varphi}_{i}^{-}}{\sqrt{\hat{v}_{i,i}^{-}\hat{\Lambda}^{-}}}=\hat{\varphi}_{i}^{-}\\ \hat{\rho}_{k,j}^{-}=\dfrac{\hat{\varphi}_{k}^{-}}{\sqrt{\hat{v}_{k,k}^{-}\hat{\Lambda}^{-}}}=\hat{\varphi}_{k}^{-}\\ \hat{\rho}_{i,k}^{-}=\dfrac{\hat{v}_{i,k}^{-}}{\sqrt{\hat{v}_{i,i}^{-}\hat{v}_{k,k}^{-}}}=\hat{v}_{i,k}^{-}\end{cases} \tag{87}$$

Substitute the minimum-variance coefficients from Equation (46)

$$\left(\hat{\rho}_{i,k}^{+}\right)_{\mathrm{j,opt}}^{2}=\left(\frac{\left\{\hat{\rho}_{i,k}^{-}-\hat{\rho}_{i,j}^{-}\,\hat{\rho}_{k,j}^{-}\right\}^{2}}{\left\{1-\left(\hat{\rho}_{i,j}^{-}\right)^{2}\right\}\left\{1-\left(\hat{\rho}_{k,j}^{-}\right)^{2}\right\}}\right)_{\mathrm{j}}$$

$$\text{for} \begin{cases}a_{i}=\left\{a_{i}\right\}_{\mathrm{j,opt}}=\dfrac{-\hat{\varphi}_{i}^{-}}{\hat{\Lambda}_{j}^{-}}=\left(-\hat{\rho}_{i,j}^{-}\right)\\ a_{k}=\left\{a_{k}\right\}_{\mathrm{j,opt}}=\dfrac{-\hat{\varphi}_{k}^{-}}{\hat{\Lambda}_{j}^{-}}=\left(-\hat{\rho}_{k,j}^{-}\right)\end{cases} \tag{88}$$

An infeasible communality between population estimates for the i^{th} ($i\neq j$) and k^{th} ($k\neq j$) study-variables, *i.e.*, elements of the $M\times M$ partition $\mathbf{V}_{\mathbf{yy}}$ in Equation (13), with the j^{th} auxiliary-residual often occurs if the following conditions are all true:

- Strong communality between the population estimate for the i^{th} study-variable and the j^{th} auxiliary-residual
- Strong communality between the population estimate for the k^{th} study-variable and the j^{th} auxiliary-residual
- Weak communality between population estimates for the i^{th} and k^{th} variables.

 DOI: 10.13140/RG.2.2.17634.07369

These conditions, although uncommon, offer substantial gains in statistical efficiency for population estimates of study-variables. Therefore, motivation is compelling for a numerical solution that imposes all constraints in Sections 6.1, 6.2 and 6.5 plus the communality constraint considered here, while minimizing variance of population estimates for the study-variables subject to all of these constraint. A closed form solution is not apparent. In order to simultaneously constrain both coefficients a_i and a_k, see Rios and Sahinidis (2013) review of relevant optimization algorithms.

An even simpler alternative is the choice of coefficients $a_i = a_k = 0$ with the j^{th} auxiliary-residual, the result of which will always be $\left(\hat{\rho}^{-}_{i,k}\right)^2_{\text{j}} \le 1.0$ assuming the communality constraint is successfully imposed in all previous recursions $\left\{\left(\hat{\rho}^{+}_{i,k}\right)^2_{\text{j}-1} \le 1\right\},\ldots,\left\{\left(\hat{\rho}^{+}_{i,k}\right)^2_{0} \le 1\right\}$. However, this convenient solution can forgo improvements in statistical efficiency for population estimates of the i^{th} and k^{th} variables, the degree to which depends on the specific application.

6.4 Avoidance of Variance Increase with Difference-Estimator

Figure 2 (page 42) demonstrates that the new difference-estimator reduces variance of population estimates if the coefficient values are bounded within a certain interval. Determine that interval with Equation (43) on page 40, which is the variance estimator for the i^{th} population estimate with the difference-estimator and j^{th} auxiliary-residual and arbitrary coefficient a_i.

$$\left\{\hat{v}^{+}_{i,i}\right\}_j = \left\{\hat{v}^{-}_{i,i} + 2\hat{\varphi}^{-}_i a_i + \hat{\Lambda}^{-} a_i^2\right\}^{-}_j \tag{43}$$

To determine the first boundary for the interval in which an arbitrary coefficient reduces variance with the difference-estimator, solve Equation (43) for $\left\{\hat{v}^{+}_{i,i}\right\}_{\text{j}} = \left\{\hat{v}^{-}_{i,i}\right\}_{\text{j}}$. The first boundary value is simply a_i=0. The second boundary value equals

 DOI: 10.13140/RG.2.2.17634.07369

$$
\begin{aligned}
0 &= \left\{ 2\hat{\varphi}_i^- a_i + \hat{\Lambda}^- a_i^2 \right\}_{\mathrm{j}} \quad \text{for } \left\{ \hat{v}_{i,i}^+ \right\}_{\mathrm{j}} = \left\{ \hat{v}_{i,i}^- \right\}_{\mathrm{j}} \\
&= \left\{ 2\hat{\varphi}_i^- + \hat{\Lambda}^- a_i \right\}_{\mathrm{j}} \\
a_i &= \left\{ \frac{-2\hat{\varphi}_i^-}{\hat{\Lambda}^-} \right\}_{\mathrm{j}} \\
&= 2\left\{ a_i \right\}_{\mathrm{j,opt}}
\end{aligned}
\tag{89}
$$

Note: The second boundary value equals 2-times the value of the minimum-variance coefficient $\{a_m\}_{\mathrm{j,opt}}$ in Equation (46).

The boundary also depends on the sign of covariance $\left\{ \hat{\varphi}_i^- \right\}_{\mathrm{j}}$

$$
\begin{aligned}
\frac{-2\hat{\varphi}_i}{\hat{\Lambda}} < a_i < 0 \quad &\text{for } \hat{\varphi}_i > 0 \\
0 < a_i < \frac{-2\hat{\varphi}_i}{\hat{\Lambda}} \quad &\text{for } \hat{\varphi}_i < 0
\end{aligned}
\tag{90}
$$

This inequality constraint is satisfied if $g_{\min} \geq 0$ in Section 6.5 (below)

$$
\begin{aligned}
\left\{ \left(\frac{-2\hat{\varphi}_i^-}{\hat{\Lambda}^-} \right) \right\}_{\mathrm{j}} \leq \left(a_i \right)_{\mathrm{j}} \leq 0 \quad &\text{for } \begin{cases} g_{\min} = 0 \\ \hat{\varphi}_i^- > 0 \end{cases} \\
0 \leq \left(a_i \right)_{\mathrm{j}} \leq \left\{ \left(\frac{-2\hat{\varphi}_i^-}{\hat{\Lambda}^-} \right) \right\}_{\mathrm{j}} \quad &\text{for } \begin{cases} g_{\min} = 0 \\ \hat{\varphi}_i^- < 0 \end{cases}
\end{aligned}
\tag{91}
$$

6.5 Minimum Coefficient for Significant Variance-Reduction

A small value of arbitrary coefficient a_i reduces variance of the i^{th} population estimate, but the variance-reduction can be miniscule. Also, the minimum-variance coefficient $\{a_m\}_{\mathrm{j,opt}}$ can be small because the covariance $\left\{ \hat{\varphi}_i^- \right\}_{\mathrm{j}}$ in Equation (43) between the i^{th} population estimate and the j^{th} auxiliary-residual is weak, especially as recursion $j \to J$. Computation and application of a small value for coefficient

 DOI: 10.13140/RG.2.2.17634.07369

a_i can require substantial computational resources but without substantial variance-reduction. The cumulative effect of small by spurious correlations can induce "over-fitting". Criteria in this Section address the minimum value of $\left|a_m\right|_{\mathrm{j}}$.

Define statistic g as the ratio of the variance with the difference-estimate $\left(\hat{v}_{i,i}^{+}\right)_{\mathrm{j}}$ and without $\left(\hat{v}_{i,i}^{-}\right)_{\mathrm{j}}$ the difference-estimator

$$\begin{aligned} g &= \left(\frac{\hat{v}_{i,i}^{+}}{\hat{v}_{i,i}^{-}}\right)_{\mathrm{j}} \\ &= \left(\frac{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_i^{-}a_i + \hat{\Lambda}^{-}a_i^2}{\hat{v}_{i,i}^{-}}\right)_{\mathrm{j}} \end{aligned} \tag{92}$$

Define analyst's parameter $g_{\min}$ as the minimum criterion for reduction in statistical efficiency with arbitrary coefficient $\left(a_i\right)_{\mathrm{j,min}}$

$$\left(1 - g_{\min}\right) \le \left(\frac{\hat{v}_{i,i}^{-} + 2\hat{\varphi}_i^{-}a_i + \hat{\Lambda}^{-}a_i^2}{\hat{v}_{i,i}^{-}}\right)_{\mathrm{j}} \quad \text{where } 0 \le g_{\min} < 1 \tag{93}$$

For example, if the analyst choses a minimum criterion of at least a 20-percent gain in statistical efficiency, which is approximately a 10-percent reduction in the standard deviation, then $g_{\min}$=0.20.

Solve Equation (93) for the minimum value of the i^{th} arbitrary coefficient $\left(a_i\right)_{\mathrm{j,min}}$

$$\begin{aligned} 0 &= \left\{\hat{\Lambda}^{-}\left(a_i\right)_{\min}^{2} + 2\hat{\varphi}_i^{-}\left(a_i\right)_{\min} + g_{\min}\hat{v}_{i,i}^{-}\right\}_{\mathrm{j}} \\ \left(a_i\right)_{\mathrm{j,min}} &= \left\{\frac{-\hat{\varphi}_i^{-} \pm \sqrt{\left(\hat{\varphi}_i^{-}\right)^2 - \hat{\Lambda}^{-}\left(g_{\min}\hat{v}_{i,i}^{-}\right)}}{\hat{\Lambda}^{-}}\right\}_{\mathrm{j}} \\ &= \left\{\left(\frac{-\hat{\varphi}_i^{-}}{\hat{\Lambda}^{-}}\right) \pm \sqrt{\left(\frac{\hat{\varphi}_i^{-}}{\hat{\Lambda}^{-}}\right)^2 - \frac{g_{\min}\hat{v}_{i,i}^{-}}{\hat{\Lambda}^{-}}}\right\}_{\mathrm{j}} \end{aligned} \tag{94}$$

 DOI: 10.13140/RG.2.2.17634.07369

Therefore, the interval within which the i^{th} arbitrary coefficient a_i contributes a minimally sufficient reduction in variance is

$$\left\{\left(\frac{-\hat{\varphi}_i^-}{\hat{\Lambda}^-}\right)-\sqrt{\left(\frac{\hat{\varphi}_i^-}{\hat{\Lambda}^-}\right)^2-\frac{g_{\min}\hat{v}_{i,i}^-}{\hat{\Lambda}^-}}\right\}_{\text{j}} \le \left(a_i\right)_{\text{j}} \le \left\{\left(\frac{-\hat{\varphi}_i^-}{\hat{\Lambda}^-}\right)+\sqrt{\left(\frac{\hat{\varphi}_i^-}{\hat{\Lambda}^-}\right)^2-\frac{g_{\min}\hat{v}_{i,i}^-}{\hat{\Lambda}^-}}\right\}_{\text{j}}$$

$$-\sqrt{1-\frac{g_{\min}}{\left(\hat{\rho}_{i,j}^-\right)_{\text{j}}^2}} \le \left(a_i\right)_{\text{j}} \le \sqrt{1-\frac{g_{\min}}{\left(\hat{\rho}_{i,j}^-\right)_{\text{j}}^2}} \qquad (95)$$

If arbitrary coefficient a_i, is not included in this interval, then constrain that coefficient to $a_i = 0.0$.

6.6 Over-Fitting with Numerous Auxiliary-variables

The reliability of population statistics with the new difference-estimator depends strongly on the reliability of estimated covariances between the study-variables and auxiliary-residuals, *i.e.*, the $M \times J$ partition $\mathbf{\Gamma}$ of the covariance matrix in Equation (13). A design-based π-estimator provides the initial values (j=0) of those covariances, and the reliability of those estimates depends on sufficient sample size. In order to reduce risk from unreliable estimates for spurious covariances between study-variables and auxiliary-residuals, use the following heuristic to censor difference-estimates that are not supported by reliable empirical estimates of those covariances.

Consider a subpopulation for which both the m^{th} study-variable and the j^{th} auxiliary-residual have nonzero values. This subpopulation supports the population estimate for the covariance $\hat{v}_{m,j}$ between the m^{th} study-variable and the j^{th} auxiliary-residual. Chose a minimum sample size that supports a reliable estimate of subpopulation parameters, such as a minimum sample size of 20 (Särndal *et al.*, 1992 p.251). If $\hat{v}_{m,j}$ is not supported by a minimum number of nonzero values for both the m^{th} study-variable and the j^{th} auxiliary-residual, then impose the equality constraint $\left\{a_m\right\}_{\text{j}} = 0$.

This simple convention has two consequences.

- Fortunately, this convention reduces the risk of "overfitting", in which a large number of weak and spurious but nonzero covariances between study-variables

 DOI: 10.13140/RG.2.2.17634.07369

and auxiliary-residuals produce estimates with unrealistically small variances. This risk might be further reduced by "outlier mitigation" in Sections 6.1 and 6.2 (page 55).

- Unfortunately, this same convention precludes variance-reduction with the difference-estimator for exceptionally rare domains and subpopulations. In this case, estimates for those rare study-variables might solely rely on the initial design-based π-estimator (see Section 10, page 76). However, the new difference-estimator can be used to assure additivity in statistical tables, in which the sum of row or column elements equals separate estimates of the margin totals for that row or column (see Section 12, page 82). This type of equality constraint can reduce variances of all row or column elements, even those for extremely rare conditions, although the reduction might be inconsequential.

A large number of estimated covariances can be reliable, given the convention above, but many can be relatively weak. Again, the cumulative effect of variance-reduction with large numbers of weak correlations can become consequential. To further mitigate this risk, constrain coefficient $\left(a_m\right)_{\mathrm{j}} = 0$ if the correlation $\left|\hat{\rho}_{m,j}\right|_{\mathrm{j}}$ between the m^{th} study-variable and the j^{th} auxiliary-residual during the j^{th} recursion is smaller than some criterion. As illustrated in Figure 2, substantial variance-reduction requires correlations of approximately $\left|\hat{\rho}_{m,j}\right|_{\mathrm{j}} > 0.50$. Särndal *et al.* (1992, page 250) offer similar suggestions for the model-assisted regression estimator. If the communality is less than some criterion, *e.g.*, $\left(\hat{\rho}^{-}_{m,j}\right)^2_{\mathrm{j}} < 0.25$, then constrain coefficient $\left(a_m\right)_{\mathrm{j}} = 0$. If this simple convention does not sensibly fit a specific application, then formulate an alternative convention that achieves the same objectives.

7 Improved Covariance Matrix with Bootstrap Estimator

The new difference-estimator, which is summarized in Equations (56) and (57), assumes coefficients in Equation (34) are known constants, and in this sense, this new difference-estimator is "model-free". Those procedures permit approximate estimates of the covariance for the linear transformations with the new difference-estimator. However, those coefficients are computed, in part, with

 DOI: 10.13140/RG.2.2.17634.07369

nonlinear procedures (*e.g.*, inequality constraints, stepwise covariate-selection) from the sample-survey estimates. Furthermore, replicate realizations of the probability sample can produce slightly different matrices of minimum-variance coefficients, much like a stepwise-regression. Therefore, the new difference-estimator tends to underestimate the true variance vector, although the magnitude might be minor (see Breidt and Opsomer, 2008, for conclusions regarding model-assisted estimator). However, Equation (19) demonstrates that the new difference-estimator remains unbiased for population estimates of the study-variables.

If more accurate variance estimates are important to the study, then a nonparametric resampling method, such as the bootstrap and jackknife (Wu and Deng 1983; Politis and Romano, 1994; Särndal *et al.*, 2003), can improve the accuracy of the estimated covariance matrix from the difference-estimator. This addresses Obstacle 11 in Section 2.6 (page 34). McConville *et al.* (2020) describe the bootstrap method for variance estimates with model-assisted estimators. The same general method applies to the new difference-estimator.

The new difference-estimator assumes the $(M+J)\times(M+J)$ coefficient matrix $\mathbf{K}$ in Equation (16) is a known constant rather than a population estimate. Therefore, the bootstrap method retains this assumption. The bootstrap method simulates different realizations of the probability sample, but it does not simulate different realizations of coefficient matrix $\mathbf{K}$.

McConville *et al.*, implement the bootstrap method as follows:

1. Draw a sample with replacement from the single realized sample in Section 2.2 (page 19), with inclusion probabilities $\left(\pi_H\right)_i$ in Equation (7) and $\left(\pi_G\right)_i$ in Equation (8).
2. Compute the $(M+J)\times 1$ design-based vector of difference-estimates $\left[\hat{\mathbf{t}}\right]_b$ from Section 2.2.
3. Compute the $(M+J)\times 1$ design-based vector $\left[\hat{\mathbf{t}}\right]_b$ with the b^{th} bootstrap sample from Step 2 and the final $(M+J)\times(M+J)$ concatenated $\mathbf{K}_{\mathrm{J}}^{+}$ matrix of constants from Equation (56).
4. Repeat Steps 1 to 3 $(B-1)$ times until some convergence criterion is met.

For sample size n much smaller than population size N, the approximate bootstrap estimator for the $(M+J)\times(M+J)$ covariance matrix $\hat{\mathbf{V}}_B$ is

 DOI: 10.13140/RG.2.2.17634.07369

$$\hat{\mathbf{V}}_B \simeq \frac{1}{B-1}\sum_{b=1}^{B}\left([\hat{\mathbf{t}}]_b - \hat{\mathbf{t}}_{\mathrm{J}}^{+}\right)\left([\hat{\mathbf{t}}]_b - \hat{\mathbf{t}}_{\mathrm{J}}^{+}\right)^{\mathrm{T}} \tag{96}$$

The rate of convergence, and differences between the concatenated covariance matrix $\hat{\mathbf{V}}_{\mathrm{J}}^{+}$ in Equation (57) and the bootstrapped covariance matrix $\hat{\mathbf{V}}_B$ in Equation (96) might provide evidence for over- or under-fitting with numerous auxiliary-variables.

Recall that Equations (38) and (57) provide the estimator, which is a simple linear transformation, for the covariance matrix with the new difference-estimator, but that approximation likely underestimates the variances. The bootstrap estimator is computationally more demanding, but more accurate. Rapid estimation in a database report-writer could use the approximate variance estimator, while the bootstrap estimator could support more rigorous analyses.

8 Computational Efficiency

Covariance matrix $\hat{\mathbf{V}}_{\mathrm{j}}^{-}$ and coefficient matrix $\mathbf{K}_{\mathrm{j}}$ in Equation (38) have dimensions $(M+J)\times(M+J)$, and Equation (38) from page 39 requires $2\cdot(M+J)^3$ multiplications

$$\begin{aligned}\hat{\mathbf{V}}_{\mathrm{j}}^{+} &= \mathbf{K}_{\mathrm{j}}\hat{\mathbf{V}}_{\mathrm{j}}^{-}\mathbf{K}_{\mathrm{j}}^{\mathrm{T}} \\ &= \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{a}_{\mathbf{y}} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{a}_{\notin j} \\ \hline \mathbf{0} & \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c|c|c} \hat{\mathbf{V}}_{\mathbf{y}} & \hat{\mathbf{V}}_{\mathbf{y},\notin j} & \hat{\boldsymbol{\varphi}}_{\mathbf{y}}^{-} \\ \hline \hat{\mathbf{V}}_{\notin j,\mathbf{y}} & \hat{\mathbf{V}}_{\notin j} & \hat{\boldsymbol{\varphi}}_{\notin j}^{-} \\ \hline \left[\hat{\boldsymbol{\varphi}}_{\mathbf{y}}^{-}\right]^{\mathrm{T}} & \left[\hat{\boldsymbol{\varphi}}_{\notin j}^{-}\right]^{\mathrm{T}} & \hat{\Lambda}_j^{-} \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \hline \mathbf{a}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{a}_{\notin j}^{\mathrm{T}} & \left(1+a_j\right) \end{array}\right] \end{aligned} \tag{38}$$

Furthermore, Equation (38) repeats within each and every of the J recursions, with up to $2\cdot(M+J)^3\cdot J$ multiplications.

If the number of study-variables M is large, and the number of auxiliary-variables J is large, then the estimator in Equation (38) can require a very large number of

 DOI: 10.13140/RG.2.2.17634.07369

calculations. For $M=10^3$ and $J=10^3$, the number of multiplications is over 10^{13}. The bootstrap estimator in Section 7 would require an additional 10^2 to 10^4 iterations, and each iteration requires over 10^{13} multiplications. Therefore, computational efficiency is essential for studies with numerous variables.

Covariance matrix $\hat{\mathbf{V}}_j^-$ is dense, but it is symmetric, and there is an opportunity to reduce the number of computations by almost 50-percent.

More important, coefficient matrix $\mathbf{K}_j$ within each of the j recursions is sparse and highly structured; all elements equal zero except for the diagonal and the J^{th} column vector (after the pivot in Section 3.1). There are a maximum of only $(M+J)$ elements out of the $(M+J)^2$ total elements coefficient matrix $\mathbf{K}_j$ that do not equal zero or one. Replacement of the matrix operators in Equation (38) with the scalar operators in Equation (42) reduces the number of multiplications to approximately $2\cdot(M+J)^2\cdot J$, which in this example is approximately $1.6\cdot10^{10}$ multiplications. The stepwise methods in Section 5 (above) might further reduce the number of multiplications by half. The conventions in Section 6.5 (above), which reduce risk from overfitting with numerous auxiliary-variables, might further reduce multiplications by 25- to 75-percent.

The constraints in Sections 6.1, 6.5 and 6.3 (above) require computation of the diagonal of covariance matrix $\hat{\mathbf{V}}_j^+$, but Equation (43) efficiently computes that $(M+J)\times1$ vector without computation of the entire $(M+J)\times(M+J)$ covariance matrix.

There are opportunities to decrease the number of computations by 75-percent or more compared to conventional matrix operators, but that requires algorithms that are specially designed for the structure of the new difference-estimator. Memory requirements can also be minimized by careful algorithm design. Regardless, the maximum number of variables is limited by computational capability.

9 Pre-and Post-Processing

Numerical concerns are endemic to matrix algebra, especially as matrix dimensions are large. For example, computations should use double precision if practical. Solutions to other concerns follow here.

 DOI: 10.13140/RG.2.2.17634.07369

9.1 Constraints on initial π Estimates

The recursive version of the new difference-estimator assumes that the constraints in Section 6 (page 53) are fully satisfied at the beginning of each recursion, and modifications to the minimum-variance coefficients assure those same constraints are satisfied at the end of each recursion. However, this further assumes that the initial π estimates satisfy all of those same constraints.

Therefore, the analyst must assure the constraints in Section 6.1 are met, namely, all initial π estimates in the vector of auxiliary-residuals $[\mathbf{r}]_{\pi}$ in Equation (11) are not outliers given the initial π estimates for variances of the auxiliary-residuals, which are on the diagonal of covariance matrix $\hat{\mathbf{\Lambda}}_{\pi}$ in Equation (13). Also, the analyst must assure that the initial π estimates in vector $\left[\hat{\mathbf{t}}_{\mathrm{y}}\right]_{\pi}$ from Equation (11) satisfy all minimum and maximum inequality constraints in Section 6.2. Furthermore, the analyst must assure that all communalities in the initial π estimate for covariance matrix $\hat{\mathbf{V}}_{\pi}$ in Equation (13) are feasible, i.e., $0.0 \le \left(\hat{\rho}_{i,k}^{+}\right)_{\pi}^{2} \le 1.0$ from Section 6.3.

9.2 Standardization of Parameter-Space

The new difference-estimator makes extensive use of the addition operator, and as such, it is more vulnerable to numerical round-off error if the numeric values of the population estimates have very difference numerical scales. Standardize population parameter-space (page 17) so that all population estimates share a similar scale.

1. Subtract the mean vector $\hat{\mathbf{t}}_{\pi}$ in Equation (11);
2. Divide the mean vector in Step 1 by its estimated standard deviation, i.e., the square root of the diagonal of $(M+J)\times(M+J)$ covariance matrix $\hat{\mathbf{V}}_{\pi}$ in Equation (13)
3. Apply the transformation in Step 1 to covariance matrix $\hat{\mathbf{V}}_{\pi}$ in Equation (13), i.e., $\mathbf{S}\hat{\mathbf{V}}_{\pi}\mathbf{S}^{\mathrm{T}}$, where diagonal $(M+J)\times(M+J)$ matrix $\mathbf{S}$ has the inverse square-roots of the diagonal of $\hat{\mathbf{V}}_{\pi}$.

The resulting linear transformation forces the initial population estimates for all of the study-variables and all of the auxiliary-residuals to have zero mean and unit variance. Apply the same transformation to any minimum and maximum inequality constraints in Section 6.2.

 DOI: 10.13140/RG.2.2.17634.07369

After the final recursion, perform the retransformation back to the original parameter-space:

1. Multiply the final vector $\hat{\mathbf{t}}_{\mathrm{J}}^{+}$ of difference-estimates in Equation (56) by $\left(\mathbf{S}^{-1}\,\hat{\mathbf{t}}_{\mathrm{J}}^{+}\right)$;
2. Apply retransformation $\mathbf{S}^{-1}\,\hat{\mathbf{V}}_{\mathrm{J}}^{+}\left(\mathbf{S}^{-1}\right)^{\mathrm{T}}$ to the final covariance matrix with the difference-estimator in Equation (57); and
3. Add $\hat{\mathbf{t}}_{\pi}$ to the final vector $\hat{\mathbf{t}}_{\mathrm{J}}^{+}$

Do not standardize parameter-space if the recursive difference-estimator is used to assure additivity in statistical tables (see Section 12, page 82), for which the row and column margins agree exactly with the sum of table element in those same rows and margins.

9.3 Reduction in Number of Study-variables

Section 6.5 includes criteria that reduce risk from overfitting. Those same criteria can suggest preprocessing procedures that reduce the number of variables, the importance of which is covered in Section 8 (page 72). Section 6.5 assumes that a reliable estimate is available for each covariance between a study-variable and an auxiliary-residual, where the number of nonzero observations is a recommended criterion for reliability. If the estimate for the covariance between the m^{th} study-variable and the j^{th} auxiliary-residual is not reliable, then the coefficient $\left\{a_m\right\}_{\mathrm{j}} = 0$. Section 6.5 also assumes that communality between the m^{th} study-variable and the j^{th} auxiliary-residual is sufficiently strong, *e.g.*, $\left(\hat{\rho}_{m,j}^{-}\right)_{\mathrm{j}}^{2} < 0.25$; otherwise, $\left(a_m\right)_{\mathrm{j}} = 0$.

If there are no reliable estimates of the covariances between the j^{th} auxiliary-residual and any of the M study-variables, and none of the reliable estimates are sufficiently strong, then omit the j^{th} auxiliary-residual. Likewise, if there are no reliable estimates of the covariances between the m^{th} study-variable and any of the J auxiliary-residuals, and if none of the reliable estimates are sufficiently strong, then omit the m^{th} study-variable. In the latter case, enlarge the $M\times(M+J)$ matrix of final coefficients $\mathbf{K}$ to include all study-variables that were omitted, where the final estimate for each of the omitted variables is its original π estimate (see Section 10, page 76).

 DOI: 10.13140/RG.2.2.17634.07369

10 Merger of π Estimates with Recursive-Estimates

Section 9.3 (page 75) recommends a pre-processing step that omits certain study-variables because they are not sufficiently and reliably correlated with any auxiliary-variable. However, variance estimators for linear transformations of population statistics (*i.e.*, pseudo-estimators, such as sums and ratio approximations) require a covariance matrix that includes all variables in those transformations (Czaplewski 2010a). Therefore, collate the π estimates for the omitted study-variables with the final recursive population estimates.

Recall the multivariate π-estimator for the $(M+J)\times 1$vector of sufficient population statistics $\hat{\mathbf{t}}_\pi$ in Equation (11), which is the starting set of population estimates in the new difference-estimator. Assume there are M_1 study-variables that are sufficiently and reliably correlated with at least one auxiliary-residual, with $M_2=(M-M_1)$ remaining study-variables; and assume there are J_1 auxiliary-residuals that are sufficiently and reliably correlated with at least one study-variable, with $J_2=(J-J_1)$ remaining auxiliary-residuals. Define the rectangular $(M_1+J_1)\times(M+J)$ indicator matrix $\mathbf{U}_1$ as the operator that selects those sufficiently and reliably correlated study-variables and auxiliary-residuals, and rectangular $(M_2+J_2)\times(M+J)$ indicator matrix $\mathbf{U}_2$ as the operator that selects all remaining variables. The $(M+J)\times(M+J)$ matrix $\begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}$ is a pivoted version of the conformable identity matrix.

Define the initial $(M_1+J_1)\times 1$ vector of π population estimates as $\hat{\mathbf{t}}_{\pi,U}=\mathbf{U}_1\hat{\mathbf{t}}_\pi$. The final recursive difference-estimate is the $(M_1+J_1)\times 1$ vector $\hat{\mathbf{t}}_{\mathrm{J}}^{+}$ from Equation (56). The $(M+J)\times 1$vector of merged population estimates equals

$$\hat{\mathbf{t}}_{\mathrm{dif}}=\begin{bmatrix}\mathbf{K}_{\mathrm{J}}^{+}\\ \mathbf{U}_2\end{bmatrix}\hat{\mathbf{t}}_\pi \tag{97}$$

The corresponding $(M+J)\times(M+J)$ covariance matrix equals

$$\hat{\mathbf{V}}_{\mathrm{dif}}=\begin{bmatrix}\mathbf{K}_{\mathrm{J}}^{+}\\ \mathbf{U}_2\end{bmatrix}\hat{\mathbf{V}}_\pi\begin{bmatrix}\mathbf{K}_{\mathrm{J}}^{+}\\ \mathbf{U}_2\end{bmatrix}^{\mathrm{T}} \tag{98}$$

 DOI: 10.13140/RG.2.2.17634.07369

11 Statistical Efficiency

This section uses a simplified special case to investigate statistical efficiency (*i.e.*, variance-reduction) of the new difference-estimator. In the following special case, assume a probability sample produces a design unbiased population estimate for a single auxiliary-variable (j=J=1), denoted $\left[\hat{t}_{x_G}\right]_\pi$. A second independent probability sample produces design unbiased population $M\times 1$ vector estimate of M study-variables and a one auxiliary-variable.

An example of this special case consists of two independent sample-surveys that produce two sets of population estimates: (1) estimate $\hat{t}_1 = \left[\hat{t}_{x_G}\right]$ for the j^{th} auxiliary-variable; and (2) a $(M+1)\times 1$ vector of population estimates $\hat{\mathbf{t}}_2 = \left[\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{T}} \;\vdots\; \hat{t}_{x_H}\right]_\pi^{\mathrm{T}}$ for the M study-variables and the auxiliary-variable. By definition, the scalar auxiliary-residual is $r_{\mathrm{j}} = \left(-\hat{t}_{x_H} + \hat{t}_{x_G}\right)_{\mathrm{j}}$. The consequence of the independence assumption in this special case is $\hat{v}^{-}_{\mathbf{x}_H,\mathbf{x}_G} = \hat{v}^{-}_{y_m,x_G} = 0$.

The degree of variance-reduction with the new difference-estimator depends on three Components:

1. Choice of arbitrary coefficients
2. Variance of the population estimate for the j^{th} auxiliary-residual
3. Correlation between population estimates for the m^{th} study-variable and the j^{th} auxiliary-residual

Section 3.4 (page 43) uses the minimum-variance criterion to address Component 1. The current section examines the effects of the other two Components.

First, set the stage for investigation of Components 2 and 3. Equation (48) provides the estimator for the variance of population estimate for the mth study-variable, where Equation (35) defines the scalar variances Λ^{-}_{j} and $\left(\hat{\varphi}^{-}_{m}\right)_{\mathrm{j}}$

 DOI: 10.13140/RG.2.2.17634.07369

$$\left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{j,opt}} = \left\{\hat{v}_{m,m}^{-} - \frac{\left(\hat{\varphi}_{m}^{-}\right)^{2}}{\Lambda_{\mathrm{j}}^{-}}\right\}_{\mathrm{j}}$$
$$= \left\{\hat{v}_{m,m}^{-} - \left(\frac{\left(\hat{v}_{y_m,x_H}^{-}\right)^{2} - 2\hat{v}_{y_m,x_H}^{-}\hat{v}_{y_m,x_G}^{-} + \left(\hat{v}_{y_m,x_G}^{-}\right)^{2}}{v_{\mathbf{x}_H,\mathbf{x}_H}^{-} - 2\hat{v}_{\mathbf{x}_H,\mathbf{x}_G}^{-} + v_{\mathbf{x}_G,\mathbf{x}_G}^{-}}\right)\right\}_{\mathrm{j}} \quad (99)$$

Given the assumptions at the beginning of this Section, Equation (99) simplifies to

$$\left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{j,opt}} = \left\{\hat{v}_{m,m}^{-} - \left(\frac{\left(\hat{v}_{y_m,x_H}^{-}\right)^{2}}{v_{\mathbf{x}_H,\mathbf{x}_H}^{-} + v_{\mathbf{x}_G,\mathbf{x}_G}^{-}}\right)\right\}_{\mathrm{j}} \quad \text{for} \begin{cases} \left\{\hat{v}_{m,m}^{+}\right\}_{\mathrm{j,opt}} = \left\{\hat{v}_{m,m}^{-} - \dfrac{\left(\hat{\varphi}_{m}^{-}\right)^{2}}{\Lambda_{\mathrm{j}}^{-}}\right\}_{\mathrm{j}} \\ \hat{v}_{\mathbf{x}_H,\mathbf{x}_G}^{-} = 0 \\ \hat{v}_{y_m,x_G}^{-} = 0 \end{cases} \quad (100)$$

where $v_{\mathbf{x}_G,\mathbf{x}_G}^{-}$ is the variance of the first population estimate for the j^{th} auxiliary-variable $\hat{t}_1 = \left[\hat{t}_{x_G}\right]$, and $v_{\mathbf{x}_H,\mathbf{x}_H}^{-}$ is the variance of the second population estimate for the same auxiliary-variable $\hat{t}_{x_H}$.

Furthermore, express variance $v_{\mathbf{x}_G,\mathbf{x}_G}^{-}$ as a fraction of the variance $v_{\mathbf{x}_G,\mathbf{x}_G}^{-} = v_{\mathbf{x}_H,\mathbf{x}_H}^{-} \big/ c$ and substitute into Equation (100)

 DOI: 10.13140/RG.2.2.17634.07369

$$\left\{\hat{v}^{+}_{m,m}\right\}_{\text{j,opt}} = \left\{\begin{array}{l} \left\{\hat{v}^{-}_{m,m} - \left(\dfrac{\left(\hat{v}^{-}_{y_m,x_H}\right)^2}{v^{-}_{\mathbf{x}_H,\mathbf{x}_H} + v^{-}_{\mathbf{x}_H,\mathbf{x}_H}/c}\right)\right\}_{\text{j}} \\ = \left\{\hat{v}^{-}_{m,m}\left(1 - \dfrac{1}{(1+1/c)}\left(\dfrac{\left(\hat{v}^{-}_{y_m,x_H}\right)^2}{\hat{v}^{-}_{m,m}v^{-}_{\mathbf{x}_H,\mathbf{x}_H}}\right)\right)\right\}_{\text{j}} \\ = \left\{\hat{v}^{-}_{m,m}\left(1 - \dfrac{c}{(c+1)}\left(\hat{\rho}^{-}_{y_m,x_H}\right)^2\right)\right\}_{\text{j}} \end{array}\right\}$$

$$\text{where}\left\{\begin{array}{l} v^{-}_{\mathbf{x}_G,\mathbf{x}_G} = v^{-}_{\mathbf{x}_H,\mathbf{x}_H}/c \\ \left(\hat{\rho}^{-}_{y_m,x_H}\right)^2 = \dfrac{\left(\hat{v}^{-}_{y_m,x_H}\right)^2}{\hat{v}^{-}_{m,m}v^{-}_{\mathbf{x}_H,\mathbf{x}_H}} \\ \pm\hat{\rho}^{-}_{y_m,x_H} = \dfrac{\hat{v}^{-}_{y_m,x_H}}{\sqrt{\hat{v}^{-}_{m,m}v^{-}_{\mathbf{x}_H,\mathbf{x}_H}}} \end{array}\right. \tag{101}$$

$\left(\hat{\rho}^{-}_{y_m,x_H}\right)^2$ is the coefficient of determination between population estimates for the m^{th} study-variable $\hat{t}_{\mathbf{y}_m}$ and the j^{th} auxiliary-variable $\hat{t}_{x_H}$ in the second sample-survey $\hat{\mathbf{t}}_2 = \left[\hat{\mathbf{t}}^{\mathrm{T}}_{\mathbf{y}} \;\vdots\; \hat{t}_{x_H}\right]^{\mathrm{T}}$, and $\hat{\rho}^{-}_{y_m,x_H}$ is the corresponding correlation coefficient;.
Equation (101) allows several generalizations, at least given the simplifications in this special case.

Variance of the population estimate for the j^{th} auxiliary-residual (Component 2, page 77) affects statistical efficiency with the new difference-estimator.

 DOI: 10.13140/RG.2.2.17634.07369

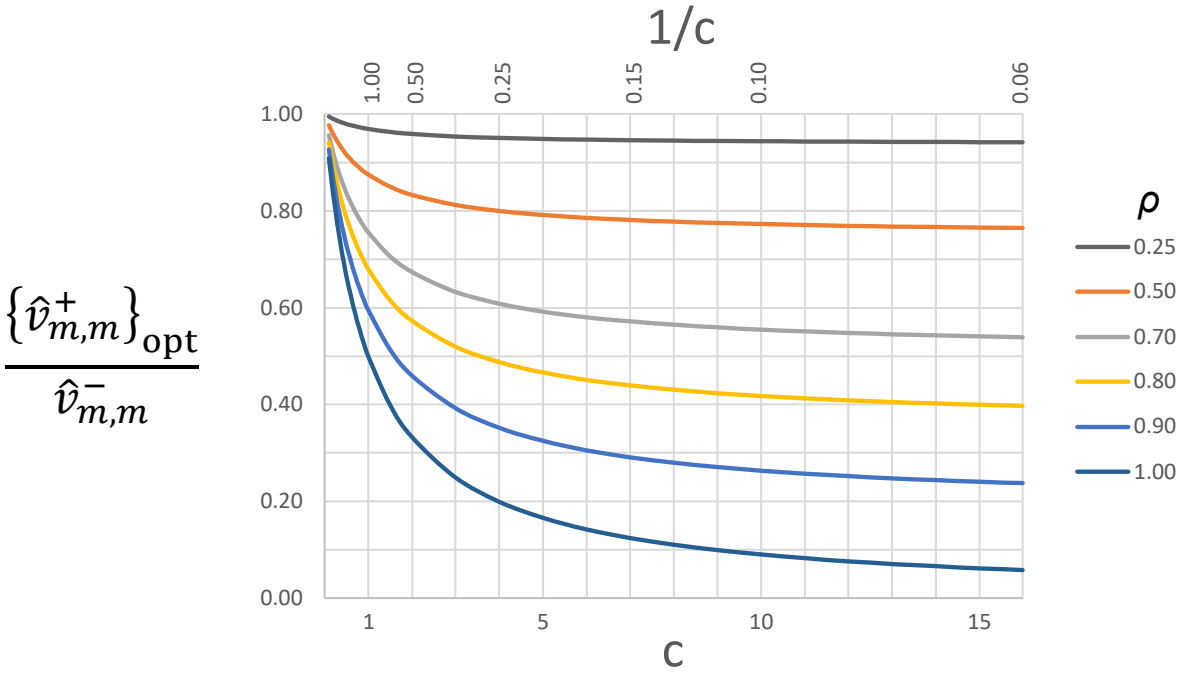

Figure 3 Variance-reduction in the j[th] recursion with the new difference-estimator compared to initial π variance as a function of the variance of the j^{th} auxiliary-residual and the correlation between the mth study-variable.

 DOI: 10.13140/RG.2.2.17634.07369

ρ =0.80		
c	Variance Reduction	Standard Deviation Reduction
1.00	32%	18%
2.00	43%	24%
5.00	53%	32%
10.00	58%	35%

Recall the example of a Two-Phase surveys, where each Phase uses simple random sampling. The First-Phase produces population estimates for the j^{th} auxiliary-variable. The Second-Phase produces population estimates for both the m^{th} study-variable and the j^{th} auxiliary-variable, and this Second-Phase estimates the covariance between population estimates for the m^{th} study-variable and the j^{th} auxiliary-variable. The lower horizontal axis in *Figure 3* represents the sample size with the First-Phase divided by the sample size of the Second-Phase. If both Phases have the same sample size, then c=1; if the First-Phase has 10-times the sample size as the Second-Phase, then c=10. The table provides an example of reduction in variance and standard deviation for the population of the m^{th} study-variable with the difference-estimator assuming a correlation of ρ=0.80 between population estimates for the m^{th} study-variable and the j^{th} auxiliary-variable.

For example, consider a simple case of Two-Phase sampling. The Phase-One uses a large sample size with equal inclusion probabilities to accurately estimate population parameters for the J auxiliary-variables. Phase-Two uses a smaller sample size with equal inclusion probabilities to estimate population parameters for the M study-variables, the J auxiliary-variables, and the $M \times J$ matrix of correlations between population estimates for the study-variables and the auxiliary-variables. Joint inclusion probabilities between the Phase-One and Phase-Two cause nonzero covariances between population estimates from the two phases, but those covariances are weak.

In the special case considered in this example, the horizontal axis in *Figure 3* represents the number of Phase-One sample units relative to the number of Phase-Two sample units; if c=10, then there are ten times more sampling units in the Phase-One sample compared to the Phase-Two sample. The vertical axis represents the variance of the difference-estimator relative to the variance of the π estimator; small values represent strong gains in variance-reduction with the difference-estimator. Each line represents a different degree of correlation between the m^{th} study-variable and the j^{th} auxiliary-variable with the Phase-Two sample. If that correlation is weak, then variance-reduction is minimal, regardless of the Phase-One sample size. If that correlation is strong, then variance-reduction is substantial, although the gain diminishes as the Phase-One sample size exceeds 10-times the Phase-Two sample size.

 DOI: 10.13140/RG.2.2.17634.07369

12 Margin Additivity

A large number of study-variables appear as tables in official government statistics. If a table has R rows and C columns, with margin totals for each row and each column, and a grand total for all table elements, then the statistical table requires population estimates for $(RC+R+C+1)$ elements in that table.

Credibility of statistical tables improves with internal consistency; for example, the sum of all table elements in the r^{th} row should exactly equal the corresponding r^{th} margin of that table. The scalar weights with generic inference and model-assisted methods provide that consistency, but at the cost of statistical efficiency. The new difference-estimator, with its specific inference, is more efficient for variance-reduction. However, the new difference-estimator does not necessarily produce statistical tables with internal consistency. A post-estimation procedure with the multivariate recursive version of the new difference-estimator can improve, if not achieve, internal consistency for statistical tables.

Without an attempt at even more notation, consider two population estimates: the margin total for the r^{th} row in the table, and the sum of estimates for every table element in the r^{th} row (excluding the margin). Those two population estimates share the same expected value, and their estimates should be equal, but their realized estimates can differ. The difference between these two estimates is structurally identical to the scalar residual r_j^- in Equation (37) on page 39.

$$\left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}}^{+} \\ \hdashline \mathbf{r}_{\notin j}^{+} \\ \hdashline r_j^{+} \end{array}\right]_{\mathrm{j}} = \left[\begin{array}{c:c:c} \mathbf{I} & \mathbf{0} & \mathbf{a}_{\mathbf{y}} \\ \hdashline \mathbf{0} & \mathbf{I} & \mathbf{a}_{\notin j} \\ \hdashline \mathbf{0} & \mathbf{0} & \left(1+a_j\right) \end{array}\right]_{\mathrm{j}} \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}}^{-} \\ \hdashline \mathbf{r}_{\notin j}^{-} \\ \hdashline r_j^{-} \end{array}\right]_{\mathrm{j}} \tag{37}$$

The new difference-estimator, with the minimum-variance coefficient a_j in Section 3 (page 35), produces the updated residual $r_j^+ = 0$, meaning there is no difference between estimates for the r^{th} row margin and the sum of all table elements in the r^{th} row. In other words, the new difference-estimator can impose equality constraints, and recursive difference-estimates for the r^{th} row are internally consistent. Similar to post-processing of survey weights (*e.g.*, Deville, 1993), a recursive raking

 DOI: 10.13140/RG.2.2.17634.07369

estimator, sequentially applied to each row and each column, can converge towards internal consistency within a statistical table (convergence is not guaranteed). This type of application of the new difference-estimator requires suspension of criteria in Section 6.5 (page 67) and Section 6.5 (page 67).

The raking version of the new difference-estimator is not compatible with standardization of parameter-space in Section 9.2 (page 74). Standardization of parameter-space is not a compelling concern because the sum of table elements in each row and each column share approximately the same numerical scales as their corresponding table margins.

Internal consistency has the potential to reduce variance for rare domains that do not satisfy the sample size criterion in Section 6.5 (page 67). In these cases, population estimates for those rare study-variables equal their π estimates (Section 11, page 77). Consider a rare study-variable in the r^{th} row of a statistical table. The new difference-estimator can constrain the sum of population estimates in the r^{th} row to exactly equal the corresponding r^{th} table margin. This reduces the variance for each and every study-variable in the r^{th} row, including any π estimates for rare study-variables. However, the reduction might be small.

13 Pseudo-Expansion Factors

Model-assisted estimation uses survey weights to reduce variance of population estimates for study-variables with auxiliary information (see McConville *et al.*, 2020 for a tutorial on model-assisted estimation). A straightforward matrix operation provides a heuristic alternative with the new difference-estimator.

Recall from Equation (9) on page 21 the initial π-estimator provides the $(M+2J)\times 1$ vector of population-totals

$$\left[\hat{\mathbf{t}}_{\mathbf{y},\mathbf{x}}\right]_{\pi} = \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_{\pi} = \sum_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \tag{9}$$

 DOI: 10.13140/RG.2.2.17634.07369

Equation (11) is a linear transformation of Equation (9) that provides the initial $(M+J)\times 1$ vector of sufficient statistics for the new difference-estimator.

$$\begin{aligned}\hat{\mathbf{t}}_{\pi} &= \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hline \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hline \hat{\mathbf{t}}_{\mathbf{x}_G} \end{array}\right]_{\pi} \\ &= \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \sum\nolimits_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \end{aligned} \tag{11}$$

Equation (56) applies the final $(M+J)\times (M+J)$ matrix of coefficients from the new difference-estimator

$$\begin{aligned}\hat{\mathbf{t}}_{\mathrm{J}}^{+} &= \mathbf{K}_{\mathrm{J}}^{+} \hat{\mathbf{t}}_{\pi} \\ &= \mathbf{K}_{\mathrm{J}}^{+} \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \sum\nolimits_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \end{aligned} \tag{102}$$

Use an $M\times(M+J)$ indicator matrix to extract the population estimates for the study-variables, thus omitting the auxiliary-residuals. The result is the $M\times(M+J)$ weight matrix $\mathbf{K}_{\mathrm{J,y}}^{+}$

$$\begin{aligned}\hat{\mathbf{t}}_{\mathrm{J,y}}^{+} &= \mathbf{K}_{\mathrm{J,y}}^{+} \sum\nolimits_U \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \\ \text{where } \mathbf{K}_{\mathrm{J,y}}^{+} &= \left(\left[\mathbf{I} \;\vdots\; \mathbf{0}\right] \mathbf{K}_{\mathrm{J}}^{+} \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \right) \end{aligned} \tag{103}$$

Therefore, a straightforward transformation of the new difference-estimator provides the $M\times(M+J)$ matrix $\mathbf{K}_{\mathrm{J,y}}^{+}$ of coefficients that can be applied to each and every population element in the sample. That coefficient matrix is invariant for each and every population element, which makes the untested assumption that the

 DOI: 10.13140/RG.2.2.17634.07369

relationships between study-variables and auxiliary-residuals are the same for all sub-populations.

For example, consider an analysis that requires estimates for subpopulation U_s. Define the 0-1 indicator value as $\left(I_s\right)_i = 1$ if the i^{th} population element is in subpopulation s, and $\left(I_s\right)_i = 0$ otherwise. The difference-estimator for subpopulation U_s is

$$\hat{\mathbf{t}}^{+}_{\mathrm{J},\mathbf{y},s} = \mathbf{K}^{+}_{\mathrm{J},\mathbf{y}} \sum\nolimits_U \left(\left(I_s\right)_i \begin{bmatrix} \mathbf{y}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_H\right)_i / \left(\pi_H\right)_i \\ \mathbf{x}_i \left(I_G\right)_i / \left(\pi_G\right)_i \end{bmatrix} \right) \tag{104}$$

If Equation (104) is applied to each and every subpopulation, $s = 1, \ldots, S$, then the vector sum of all subpopulation estimates will exactly agree with the population estimates in Equation (56).

The estimator for subpopulation s in Equation (104) uses a single $M\times(M+J)$ matrix of known coefficients (weights) that are invariant across all S subpopulations. Equation (104) supports specific inference, which does not sacrifice statistical efficiency in order to gain simplicity. However, Equation (104) assumes the correlations between study-variables and auxiliary-residuals are invariant across all S subpopulations, and this assumption cannot be validated without sufficient sample size in each subpopulation.

Equation (104) estimates the $M\times1$ vector of population estimates for the M study-variables for subpopulation s as the $M\times1$ vector of π-estimator for subpopulation s (*i.e.*, the design-based estimator that uses only those sampled population elements within subpopulation s) plus an $M\times1$ vector adjustment vector that uses the $J\times1$ vector of subpopulation statistics for the J auxiliary-variables, where those auxiliary statistics are computed only with those sampled or enumerated population elements in subpopulation s. This assumes that subpopulation s has a sufficiently large number of sampled population elements to reliably support the π-estimator for that subpopulation. Czaplewski (2010a) explores a model-based estimator for smaller subpopulations.

 DOI: 10.13140/RG.2.2.17634.07369

14 Discussion

14.1 Stochastic Model

The difference-estimator given by Särndal *et al.* (Sections 6.3 and 6.8) assumes the value for each of *J* auxiliary-variables is known before sampling for each and every population element. Therefore, the population parameters for all *J* auxiliary-variables are known constants, as with a census enumeration. Furthermore, that difference-estimator assumes a prediction for each of the *M* study-variables is known for each and every population element. Therefore, the population parameters for predictions with each of the *M* models are known constants through census enumeration. However, the true values for each of the *M* study-variables are known only for those population elements in the probability sample. Therefore, the true population parameters for the study-variables are unknown, but the sample supports design-consistent estimates of those population parameters for the *M* study-variables. Since the value for each model prediction for every study-variable is known for all population elements, those values are known for each population element in the sample. Therefore, the sample-survey provides a population estimate for the difference between the true value of each study-variable and its predicted value.

Population parameters for the *J* auxiliary-variables are analogous to independent variables in a regression model. They can improve the estimated difference between the *M* known population parameters for the model predictions and the estimated population parameters of the true values for the *M* study-variables. The difference-estimator presented by Särndal *et al.* adjusts known population parameters for predictions each of the *M* study-variables. The adjustments are population estimates for the differences between population estimates for the true and predicted values of the *M* study-variables.

The new difference-estimator depends upon a different formulation. Like the Särndal *et al.* difference-estimator, the new difference-estimator uses the same population statistics for the *M* study-variables and the *J* auxiliary-variables. The latter may be census constants, as in Chapter 6 of Särndal *et al*. However, unlike the difference-estimator given by Särndal *et al.*, the new difference-estimator accommodates sample-survey estimates for some or all of those *J* auxiliary-variables.

The new difference-estimator is simply a linear transformation of the $(M+J)\times 1$ vector of sufficient statistics. The $(M+J)\times(M+J)$ transformation matrix contains

 DOI: 10.13140/RG.2.2.17634.07369

known but arbitrary constants, just as in Section 6.3 of Särndal *et al.* Coefficients in that transformation matrix are selected to reduce variances of population estimates for the *M* study-variables given the single realized probability sample. The Best Linear Unbiased Estimator (BLUE) criterion provides initial values for those arbitrary coefficients, just as in Section 6.8 of Särndal *et al.* In addition, the new difference-estimator imposes inequality constraints that shrink those BLUE coefficients towards zero. Those constrained coefficients differ from the minimum-variance coefficients; therefore, constraints reduce statistical efficiency, although the corresponding increase in variance is often minor. However, inequality constraints do assure plausible and robust estimates, as discussed in Section 14.2.

14.2 Constraints

Unlike the difference-estimator given by Särndal *et al.*, the new difference-estimator modifies its minimum-variance coefficients to assure robust, feasible, plausible, efficient and numerically reliable population estimates within each j^{th} recursion.

Inequality constraints on those constraints perform the following functions in the recursive implementation of the new difference-estimator:

- Mitigate undue leverage caused by implausible differences between two separate population estimates of an auxiliary-variable, i.e., an "outlier" residual (Section 6.1, page 55);
- Bound population estimates by minimum and maximum inequality constraints (Section 6.2, 60);
- Assure variance estimates are non-negative and correlation coefficients are bounded by ±1 (Section 6.3, page 62);
- Assure the new difference estimator does not increase variance of population estimates (Section 6.4, page 66);
- Improve numerical precision and reduce computation burden by omission of auxiliary-residuals that do not substantially reduce variances of population estimates (Section 6.5, page 67);
- Avoid over-fitting with numerous population estimates for auxiliary-variables that are weakly correlated with population estimates for study-variables (Section 6.6, page 69).

In addition to *inequality* constraints, the new difference-estimator can impose *equality* constraints. Additive margins (Section 12, page 82) is an example. There are other opportunities. Consider a general-purpose sample survey with equal

 DOI: 10.13140/RG.2.2.17634.07369

inclusion probabilities, with which the sample size for rare subpopulations is relatively small; the difference-estimator might not improve population estimates for small-areas and small-domains relative to the π estimators (Section 6.6, page 69; Section 10, page 76). One or more independent surveys with inclusion probabilities proportional to remotely-sensed predictions for those small-areas and small-domains can improve population estimates. The new difference-estimator can combine population estimates from two surveys, much like a multivariate composite estimator, by constraining the two population estimates for each study-variable to be equal.

14.3 Matrix Structure

In order to simplify the mathematics, Sections 2 and 3 (above) use matrix algebra to develop the new difference-estimator. They begin with concatenation of the $M\times1$vector of population estimates for the M study-variables with two $J\times1$ vectors of population estimates of the population parameters for the J auxiliary-variables, resulting in a $(M+2J)\times1$ vector of π-estimates and its corresponding $(M+2J)\times(M+2J)$ covariance matrix. A $J\times J$ partition of that covariance matrix captures any collinearity between the two population estimates of the auxiliary-variables; such as those caused by nonzero joint inclusion probabilities. The assumption is that this matrix structure accommodates population estimates from any multivariate design-based sample-survey, either from a finite population perspective with unequal inclusion probabilities, or an infinite population with known expansion factors; or even some prior use of the new difference-estimator.

The new difference-estimator performs a relatively simple linear transformation of that prior population estimate into a $(M+J)\times1$ vector (and its $(M+J)\times(M+J)$ covariance matrix), where the $J\times1$ partition is the vector difference between the two separate population estimates for the J auxiliary-variables. That vector of auxiliary-residuals contains all sufficient statistics required by the new difference-estimator. The $M\times M$ diagonal partition for the M study-variables is unchanged by this initial linear transformation; but the $J\times J$ diagonal partition contains the covariances among the differences between the two separate vectors of population statistics for the auxiliary-variables, *i.e.*, the auxiliary-residuals; and the $M\times J$ off-diagonal partition (and its corresponding $J\times M$ off-diagonal partition) contains the covariances between the M study-variables and the J auxiliary-residuals. If covariances between study-variables and auxiliary-residuals are relatively strong, then the new difference-estimator can substantially reduce the variance of population estimates for the M study-variables.

 DOI: 10.13140/RG.2.2.17634.07369

14.4 Numerics

Särndal *et al.* assume the $J \times J$ diagonal partition of the covariance matrix for the π-estimates of the auxiliary-residuals is full-rank and well-conditioned. Section 2 (page 15) and Section 3 (page 35) make the same assumption for the new difference-estimator. However, this happy circumstance is unlikely if the number of auxiliary-variables is large (e.g., $J > 20$).

Fortunately, the recursive version of the new difference-estimator can be numerically well-behaved given a rank-deficient or ill-conditioned $J \times J$ covariance matrix for the J auxiliary-residuals. Section 5 (above) offers a stepwise approach to the recursive difference-estimator. At the beginning of each recursion, the recursive difference-estimator uses minimum-variance coefficients for each of the J auxiliary-residuals to compute the estimated variance for each of the M study-variables. It then selects one of those J auxiliary-residuals for that recursion. This is similar to multivariate stepwise-regression, in which an automatic procedure sequentially choses a single predictor variable that best improves the fit of the regression model. Since the j^{th} recursion only uses one of the J auxiliary-residuals, computation of the minimum-variance coefficients with the new difference-estimator requires only a scalar inverse, which is less prone to numerical aberrations than a matrix inverse.

As the stepwise recursion proceeds $\{1,\ldots,j\}$, variances of the remaining auxiliary-residuals $\{(j+1),\ldots,J\}$ approach zero, and the scalar inverse for each of the latter residuals approaches infinity, especially with a rank-deficient or ill-conditioned covariance matrix for the auxiliary-residuals. Section 5 (page 50) offers an algorithmic solution.

Random sampling error can cause small estimates for covariances between the study-variables and the auxiliary-residuals, even if the true population parameter is exactly 0.0. The cumulative effect of numerous weak (presumably spurious) correlations can result in "over-fitting." Section 6.5 (page 67) offers a criterion that omits certain auxiliary-residuals that are very small; in a heuristic sense, a relative small auxiliary-residual contains relatively little auxiliary information.

14.5 Forestry Examples

There is a long history of proposals and pilot studies that attempted to improve monitoring of the condition, status, changes and trends in extensive forest conditions with models and multiple types of remotely sensed data, some of which are feasible for full coverage of the population (i.e., census); some of which are

 DOI: 10.13140/RG.2.2.17634.07369

feasible only for a sample of that population; and some of which can be repeated over time to monitor changes. A few examples follow; there are many more examples elsewhere.

The purpose of this section is to portray characteristics of potential sources of auxiliary data, and the challenges in integration of population estimates for auxiliary-variables with population estimates for study-variables. The constant goal to "do more with less" regularly motivates attempts to more fully use remotely-sensed measurements of auxiliary-variables. Sensor technologies shape choices for the geometry of sampling units and their re-measurement over time; and those technologies change over time. Likewise, there is strong motivation to use model-based estimators that improve population estimates for study-variables at one point it time based on direct measurements of those study-variables at prior times, which assume current conditions in a population are closely correlated to prior conditions in that population.

Schreuder *et al.* (1995) explored a Four-Phase sampling-design with multiple sources and resolutions of remotely-sensed data to improve monitoring forest conditions in Alaska, most of which are remote and expensive to sample. Phase-One used full-coverage (*i.e.*, census), low resolution, remotely-sensed data (AVHRR); Phase-Two used full-coverage (*i.e.*, census), moderate-resolution, remotely-sensed data (Landsat); Phase-Three used a probability sample of high-resolution remotely-sensed data (aerial photography); and the Phase-Four used on-the-ground measurements by field crews. They developed univariate estimators for this multiphase design. Results were discouraging because the correlation was unexpectedly weak among data from different remotely-sensed sensors and field data. The suspected reason is poor registration of sampling units between field plots and different sources of remotely sensed data. "Registration" means identification of the remotely sensed pixels that measure the exact same site on the ground, which can be a difficult and inexact task.

Lessard et al. (2001) summarize a multiyear pilot test in the State of Minnesota for an Annual Forest Inventory System (AFIS). It was a partnership between the Minnesota Department of Natural Resources and the Forest Inventory and Analysis (FIA) program (Bechtold and Patterson, 2005) in the US Forest Service. All FIA field plots in Minnesota had previously been measured by field crews only once every 10-years, but the value of those data to analysts becomes suspect 5-years after on-the-ground measurement. AFIS envisioned an annual inventory system, in which 10-percent of all FIA plots are re-measured each year (rather than all plots re-measured once every 10-years). AFIS used full-coverage census data from a spaceborne Landsat sensors as auxiliary-variables, with on-the-ground FIA

 DOI: 10.13140/RG.2.2.17634.07369

field data as the study-variables. A deterministic model predicted annual changes in detailed study-variables for each plot, where the most recent field measurements served as initial conditions; those served in a model-based estimator, and a subsample of new field data adjusted for model prediction errors. Each year, plots where prioritized for re-measurement and selected with probability sampling. If a plot was "undisturbed", then AFIS assigned a low probability of re-measurement during that year. If remotely-sensed data revealed evidence of episodic changes, such as wildfire damage, windthrow, timber harvest, and land use changes, then AFIS assigned a high probability of re-measurement during that year. Orbital sensor data are imperfectly correlated with episodic changes, and the geo-registration of the exact location of each field plot with orbital sensor data is not perfect. As a solution to mis-registration error, each FIA field plot was assigned an index that approximately represented the likelihood that the plot experienced episodic change; that index allowed for minor geo-registration errors. Each year, a selection probability was assigned to each and every FIA plot, regardless of probability of episodic change, with higher probability to those plots with higher index for episodic change. Each year, a 10-percent subsample of all FIA plots were selected for re-measurement, with probability proportional to the index of episodic change. Towards the end of the AFIS pilot study, an airborne remote sensing subsystem was added to better validate the type and degree (if any) of change for each FIA plot that had a high spaceborne index for episodic change, and the subsampling probability was reduced for any plot that was nonstocked with trees (*e.g.*, a recent clearcut tree-harvest). AFIS was not implemented at the national level for at least two reasons: it was very complicated, especially with the time-series of subsampling probabilities; and reliable sample-survey estimators where not available.

Gillis *et al.* (2005) Russo (2015) describe the three phase sampling-design for the Canadian National Forest Inventory (NFI). Phase-One produces auxiliary data from full-coverage spaceborne sensors (*i.e.*, Landsat census); Phase-Two produces auxiliary data with a sample of higher resolution airborne sensors (*i.e.*, aerial photography); and Phase-Three uses on-the-ground field measurements of the study-variables. The effectiveness of the auxiliary-variables is less than expected because of registration errors and variability among photo-interpreter with Phase-Two measurements (Magnussen and Russo, 2012.).

Tomppo *et al.* (2002) proposed a sample-survey of forest status, changes and trends for the Food and Agricultural Organization of the United Nations. The objective of the project is a practical and scalable global monitoring system of forest cover that could be consistently applied by all nations, with subpopulations

 DOI: 10.13140/RG.2.2.17634.07369

defined as large multinational geographic areas. It uses a two types of spaceborne orbital sensors for cluster-plots. Phase-One uses full-coverage medium-resolution remotely-sensed data (*i.e.*, Landsat census); and Phase-Two uses probability sampling with high-resolution orbital sensors. The design accommodates Phase-Three, in which each member nation collects the field data.

Nelson *et al.* (2009) and Holm *et al.*, (2017) deployed Three-Phase sampling-designs for population estimates of above ground forest biomass. The spaceborne lidar system (ICESat/GLAS) performed measurements at the Phase-One, an airborne laser system performed measurements at the Phase-Two, and on-the-ground field measurements (FIA) supplied data at Phase-Three. Saarela et al. (2016) used model-based estimators for growing stock volume and its uncertainty estimation, combining a sparse sample of field plots, a sample of laser (LiDAR) data, and full-coverage Landsat data. Næsset *et al.* (2013) and Gregoire *et al.* (2017) explored the statistical properties of such surveys. The LiDAR measurements used linear cluster plots that works well with orbital paths and flight-lines with airborne sensors. The cost of extensive surveys is high, emerging sensor technologies provide correlated auxiliary data at less cost, but the combination of different sources of auxiliary data creates difficulties for credible estimation.

14.6 Relationship to the Kalman Filter and Related Estimators

The multivariate difference-estimator in Section 2.4 (page 23) is algebraically identical to the static Kalman filter (e.g., Maybeck 1979, page 114), which is widely used by electrical engineers for digital signal processing, with applications such as global positioning systems (GPS), navigation, avionics, target-tracking, robotics, and others.

The Kalman filter is a statistical method to estimate the multivariate state of a system (*i.e.*, the "state vector"), which is analogous to the vector of population parameters in Equation (2) on page 17. The Kalman filter combines a vector of predicted measurements (observations) of that system with separate multivariate sensor measurements of that same system to reduce variance of the estimated state vector. Maybeck (1979) refers to the residual difference between the predicted measurements and the sensor measurements as the "innovation residual," while the new difference-estimator uses the term "auxiliary-residual" for the same statistic (Section 2.3 , page 22). The Kalman "gain matrix" is equivalent to the matrix of minimum-variance weights with the difference-estimator in Section 2.5 (page 25). Maybeck (1979) illustrates use of the matrix inversion lemma to derive the minimum-variance gain matrix for the Kalman filter; as an alternative, Section 3.4

 DOI: 10.13140/RG.2.2.17634.07369

(page 43) uses a derivation for the new difference-estimator that is similar to that given by Särndal *et al.* (see their Section 6.8).

The Kalman filter is a recursive formulation of the Bayes estimator (e.g., Jazwinski 1970, p.145; Maybeck 1979, p.205). It is also the minimum variance BLUP predictor (*e.g.*, Jazwinski 1970, p.149; Maybeck, 1977, p.232). Furthermore, it is the maximum likelihood estimator when joint densities are Gaussian (e.g., Jazwinski 1970, p.207; Maybeck 1977, p.234). Duncan and Horn (1972), Diderrich (1985), Gregoire and Walters (1988), and Piepho and Ogutu (2007) demonstrate the close relationship between the Kalman filter and Theil and Goldberger's (1961) mixed estimator.

Knottnerus (2003:315, 326) uses the perspective of Pythagorean regression to make the strong connection between the Kalman filter and estimators for complex sample-surveys. This includes constraints on the estimands. De Gruijter *et al.* (2006) give a contemporaneous exposition on the Kalman filter in the context of sampling in time and space for monitoring of natural resources. Czaplewski (2010a, 2010b, 2015, 2017a, 2017b) developed the Kalman filter as an estimator for complex sample-surveys in National Forest Inventories (NFI).

In the example of Two-Phase sampling considered above, the $(M+J)\times 1$ vector of population estimates from the Phase-Two sample is analogous to the model predictions for the state vector in the Kalman filter, and the $J\times 1$ vector of population estimates for the auxiliary-variables with the Phase-One sample is analogous to the senor measurement of that system (Czaplewski, 2010a).

Practical applications of the Kalman filter in digital signal processing provide lessons for application of the new difference-estimator to sample-surveys. "Divergence" occurs when the Kalman filter places too much weight on a model prediction and too little weight on the sensor measurements (Maybeck 1979, p.338). Divergence is typically caused by underestimates for the variance of model prediction errors, and perhaps overestimates of covariances between model predictions and sensor measurements. Divergence is detected when the residual difference between the predicted measurements and the sensor measurements is suspiciously large relative to the expected standard deviation of that residual. In the case of the new difference-estimator, Section 6.1 (page 55) applies this lesson to mitigate outliers, which can result from rare cases of large random sampling error, and database anomalies or nonrandom numerical error. If model predictions serve as auxiliary-variables with the new difference-estimator (Section 14.7, page 94), then outlier mitigation reduces risk from biased model predictions.

 DOI: 10.13140/RG.2.2.17634.07369

The conventional multivariate Kalman filter requires the same $J \times J$ matrix inversion as does the new difference-estimator in Section 2.5 (page 25). The square-root filter is a numerically more robust variation of the Kalman filter. For example, Bar-Shalom *et al.*, (2001) applies a linear orthogonalization transformation to the $J \times J$ covariance matrix (assuming it is full rank and well-conditioned); since all off-diagonal elements of the orthogonalized covariance matrix are zeros, each orthogonalized innovation residual may be treated independently from all other orthogonalized innovation residuals. The square-root filter of Bar-Shalom *et al.* sequentially processes each of the J scalar orthogonal innovation residuals, where J scalar inverses replace the $J \times J$ matrix inverse. The recursive version of the new difference-estimator in Section 3 (page 35) uses the same concept, except it develops a sequential orthogonalization that is feasible with a rank-deficient covariance matrix.

14.7 Model-Based Inference

Model-based estimators strongly depend on models for inference and statistical efficiency. Model-assisted estimators use a different stochastic model, but use probability sampling and a π estimator to adjust for any difference between model predictions and direct measurements of sampling units. In this sense, the new difference-estimator uses the same stochastic model as does model-assisted estimators, but the new difference-estimator is model-free because it is simply a linear transformation, with arbitrary and known coefficients, of a multivariate design-based π-estimator.

For example, the Interior West Unit of the Forest Inventory and Analysis (FIA) Program (Bechtold and Patterson 2005) conducts on-the-ground field measurements on an equal-probability, 10-percent, interpenetrating subsample (*i.e.*, 10 independent "panels") of FIA permanent field plots each year. Any individual FIA plot is measured on-the-ground only once every 10-years. After 10-years, all FIA plots have been measured by a field crew. The re-measurement cycle repeats at the 11^{th}, 21^{st} … years.

FIA study-variables include different metrics for wood biomass and forest health that are categorized by tree species, tree size, and merchantability criteria. Other study-variables characterize the condition of the vegetative communities (i.e., forest stands) in which the field plot resides. Stand conditions are cross-classified by the predominant mixture of tree species and tree sizes, forest management objectives, land ownership, land use, accessibility, *etc*. These study-variables require field measurements. In the absence of an episodic disturbance, such as wildfire damage, windthrow, timber harvest, and land use changes, these tree and

 DOI: 10.13140/RG.2.2.17634.07369

stand conditions change slowly over time, and the 10-year re-measurement cycle is well suited for such gradual changes. However, episodic disturbances can change more rapidly, and measuring those changes with remotely-sensed data is less expensive than field measures. Episodic change is often an important topic for analysis.

Because the number of FIA plots in each annual panel is relatively small, sampling error is high for π population estimates for any single panel. FIA reduces variance with a five-year moving average that combines all FIA plots measured during the five most recently re-measured panels into a single population estimate. Each of the five panels has equal weight. In other words, the FIA model assumes all FIA panels are measured at the same point in time. This is a simple solution, and simplicity in a large government program for official statistics has value. The new difference-estimator in the current Technical Report offers an alternative to the five-year moving average.

One type of model-based application with the new difference-estimator assumes a static model in which the population is invariant over time, *i.e.*, a dynamic system at a steady state equilibrium. This is the assumption behind the five-year moving average used by FIA. Under such a model, population estimates from different panels are perfectly correlated. In the context of the difference-estimator, there are four sets of population estimates for auxiliary-variables (one for each of the four prior years) to reduce variances of population estimates for study-variables during the current year. As with the previous example, if the population is not truly invariant over time, then these estimates will be biased, the degree to which varies with the rate of change in the population. The procedures in Sections 6.1 (page 55) and Section 6.2 (page 60) can reduce risk of substantial bias. If procedures in Sections 6.1 (page 55) for outliers, and procedures in Section 6.2 (page 60) for inequality constraints on population estimates are not necessary or omitted, then the difference-estimator described here will produce the same estimates as the five year moving average.

In addition, the difference-estimator can support model-based inference with a different stochastic model. If time-series predictions from deterministic models (*e.g.*, tree-level growth and yield models in forestry, see page 7) are available for each sampled population element, and if the vector of study-variables includes partitions for different time periods, then those predictions can serve auxiliary-variables, and the procedures in Section 3 (page 35), Section 4 (page 50), Section 5 (page 50), Section 6 (page 53) and Section 7 (page 70) are relevant.

 DOI: 10.13140/RG.2.2.17634.07369

Start with the design-based π estimates for the *M* study-variables measured with the Year-1 panel. Use a deterministic growth and yield model to estimate the condition of those *M* study-variables for each FIA plot during Year-2, where conditions measured during Year-1 serve as the initial conditions in the growth and yield model. Use those predictions for design-based π population estimates for the *M* study-variables during Year-2. Now there are two sources of population estimates for Year-2: the FIA plots in the panel measured on-the-ground during Year-2; and the growth and yield model predictions for Year-2 that use on-the-ground measurements during Year-1. The former are the study-variables for Year-2; the latter serve as auxiliary-variables for Year-2.

The difference-estimator combines these two $M\times1$ vectors of population estimates with a linear transformation, where computation of the constrained minimum-variance coefficients use the correlations between the two population estimates. Regrettably, there are no sample-survey estimates for those correlations because no FIA plots are re-measured during both Year-1 and Year-2. However, a model-based estimate of those correlations is available from only those plots measured during Year-1. Using just those plots measured during Year-1, estimate the $M\times M$ covariance matrix between plot measurements during Year-1 and predictions with the growth and yield model for those same *M* study-variables during Year-2. Strong correlations are expected between the two population estimates. The resulting covariance matrix provides sufficient statistics for coefficients in the transition matrix.

With any dynamic system and an imperfect deterministic model, the expected values of the measurements during Year-2 will not likely equal the expected values for the corresponding model predictions, and the true correlations between population estimates for Year-1 and Year-2 are not as strong as correlations between measurements during Year-1 and model predictions for Year-2. The difference-estimator can combins the two population estimates for Year-2, which reduces variance but introduces model prediction bias. This is a common situation in avionics applications with the Kalman filter, leading to divergence (Section 14.6, page 10). Estimation bias can be mitigated with a conservative criterion for the outliers in Section 6.1.1 (page 17). For example, a reasonable criterion for an unbiased population estimate for an auxiliary-residual might be ±2.0 standard deviation units; a more conservative criterion for a biased estimate might be ±1.0 standard deviation units. Section 6.1 (page 55) has other methods for outliers that can further mitigate bias in model predictions. The minimum and maximum inequality constraints in Section 6.2 (page 60) offer insurance against extreme bias in the time series of difference-estimates.

 DOI: 10.13140/RG.2.2.17634.07369

Engineering applications of the Kalman filter benefit from analyses of the "innovation sequence," which is the time series of innovation residuals. If the Kalman filter fully utilize all auxiliary information, then the innovation sequence should be characterized as "white noise," meaning there are no nonrandom patterns. The scalar auxiliary-residual with the recursive difference-estimator is equivalent to the innovation residual with the Kalman filter. Equation (37) defines the j^{th} auxiliary-residual as the scalar r_j^- in Equation (37). Its estimated variance is the scalar $\hat{\Lambda}_j^-$ in Equation (38). Standardize each of the j auxiliary-residuals as $s_j^- = r_j^- \Big/ \sqrt{\hat{\Lambda}_j^-}$. In the current context, the auxiliary-residual is the difference between a measurement of the population and its value as predicted by a deterministic model-based on previous measurements. If all parameters in the deterministic model are approximately true, then the time series of auxiliary-residuals should have mean that is approximately equal to 0.0, the standard deviation of that time services should approximately equal 1.0, and there should be no obvious temporal patterns in the time series. Exploratory analyses of the residual sequence can evaluate these expectations. For example, a mean residual that is significantly greater than 0.0 suggests that the deterministic model tends to overestimate a population parameter, and therefore, the population is not behaving as expected with the deterministic model. This information might reveal changes in population dynamics that are of interest to the analyst.

15 Summary

This Technical Report develops a new difference-estimator that has the potential to improve accuracy (*i.e.*, reduce variance) of detailed official statistics. Analysts in government agencies use these statistics to support strategic policy decisions and promote sustainable economic development. Analysts in the private sector use these statistics to guide capital investments and other strategic business decisions. Analysts in nongovernmental organizations use these statistics to access consequences of strategic decisions that impact their constituencies.

The new difference-estimator uses population estimates of auxiliary-variables to improve population estimates for the study-variables. Examples of auxiliary data include remotely-sensed data, administrative records, predictions from deterministic models of population changes over time, and emerging sources of

 DOI: 10.13140/RG.2.2.17634.07369

"Big Data." Auxiliary-variables are correlated with study-variables, but population statistics for auxiliary-variables are less expensive and more precise than population statistics for the study-variables. In a heuristic sense, the difference-estimator extracts information about the study-variables from the auxiliary-variables, and transfers that information into more accurate population estimates for the study-variables.

The recursive procedures in the new difference-estimator simplify computations required to transfer that information. The recursive procedures are robust because they protect against outliers, numerical errors and infeasible estimates (e.g., negative variances); and they constrain population estimates within known bounds (*e.g.*, the number of individuals in a subpopulation cannot be negative). The recursive procedures simplify matrix computations so that they are less vulnerable to numerical errors, and so that they can be implemented within institutional database software. The recursive procedures are designed to remain numerically stable with numerous study-variables (*e.g.*, $J>1000$) and numerous auxiliary-variables (*e.g.*, $M>1000$). The new difference-estimator can improve population estimates from conventional finite population estimates, regardless of the complexity of the sampling design. The recursive procedures can use raking methods to assure margins of statistical tables are additive. The stepwise difference-estimator uses information from auxiliary-variables to improve specific population estimates that are most important to the objectives of a study, thus automating covariate-selection. The multivariate difference-estimator includes the $M \times M$ covariance matrix for the M study-variables, which supports pseudo-estimators and synthetic-estimators, such as sums, differences, and ratios of population statistics (Särndal *et al.*, 1992, pp.173, 205, 388, 408). If the vector of population estimates for the study-variables includes partitions for different time periods and different subpopulations, then pseudo-estimators can produce population statistics for changes over time and differences among subpopulations. Bootstrap methods improve variance estimates for complex or nonlinear operators.

The model-assisted methods summarized by Särndal *et al.* are extensively developed, familiar, and widely used with confidence. The new difference-estimator introduced in the current Technical Report offers potential improvements relative to conventional model-assisted estimators. Future research will test these hypotheses.

 DOI: 10.13140/RG.2.2.17634.07369

16 References

Bar-Shalom, Y., Li, X.R., and Kirubarajan, T., 2001. *Estimation with applications to tracking and navigation.* New York, NY: John Wiley & Sons. 584 p.

Bechtold, W.A. and Patterson, P.L., 2005. *The enhanced forest inventory and analysis program-national sampling-design and estimation procedures*. Gen. Tech. Rep. SRS-80. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station. 85 p., 80.

Breidt, F.J. and Opsomer, J.D., 2008. Endogenous post-stratification in surveys: classifying with a sample-fitted model. *The Annals of Statistics*, 36(1), pp.403-427.

Crookston, N.L. and Finley, A.O., 2008. yaImpute: an R package for kNN imputation. *Journal of Statistical Software*. 23 (10). 16 p.

Czaplewski, R.L., 2010a. *Complex sample-survey estimation in static state-space*. Gen. Tech. Rep. RMRS-GTR-239. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 124 p., 239.

Czaplewski, R.L., 2010b. *Recursive restriction estimation: an alternative to post-stratification in surveys of land and forest cover*. Res. Pap. RMRS-RP-81. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 32 p., 81.

Czaplewski, R., 2015. Novel Kalman filter algorithm for statistical monitoring of extensive landscapes with synoptic sensor data. *Sensors*, 15(9), pp.23589-23617.

Czaplewski, R.L., 2017a. *KFz Algorithm: A Sequential Kalman Filter Estimator for Large State and Measurement Vectors and Rank-deficient Innovation Covariance Matrix*. Tech. Rep. 2017-1, Environmetrika, DOI: 10.13140/RG. 2.2. 32002.58565.

Czaplewski, R.L., 2017b. *Taming the Kalman Filter: A New Algorithm for Large Vector Estimates in Sample-surveys*. Tech. Rep. 2017-2, Environmetrika, DOI: 10.13140/RG.2.2.31163.72481.

Daas, P.J., Puts, M.J., Buelens, B. and van den Hurk, P.A., 2015. Big data as a source for official statistics. *Journal of Official Statistics*, *31*(2), pp.249-262.

de Gruijter, J., Brus, D., Bierkens, M., and Knotters, M, 2006. *Sampling for natural source monitoring*. Berlin: Springer-Verlag. 332 p.

Deville, Jean-Claude, Carl-Erik Särndal, and Olivier Sautory, 1993. Generalized raking procedures in survey-sampling. Journal of the American Statistical Association 88: 1013-1020.

Diderrich, G.T., 1985. The Kalman filter from the perspective of Goldberger-Theil estimators. *The American Statistician* 39:193-198.

Duncan, D.B., and Horn, S.D., 1972. Linear dynamic recursive-estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association* 67(340):815-821

Edwards, W., Lindman, H. and Savage, L.J., 1963. Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), p.193.

Florescu, D., Karlberg, M., Reis, F., Del Castillo, P.R., Skaliotis, M. and Wirthmann, A., 2014, June. Will 'big data'transform official statistics. In *European Conference on the Quality of Official Statistics. Vienna, Austria* (pp. 2-5).

Gillis, M.D., Omule, A.Y. and Brierley, T., 2005. Monitoring Canada's forests: the national forest inventory. *The Forestry Chronicle*, 81(2), pp.214-221.

Gregoire, T. and D. Walters. 1988. Composite vector estimators derived by weighting inversely proportional to variance. *Canadian Journal of Forest Research* 18(2):282-284.

Gregoire, T.G., E., Næsset, R.E., McRoberts, G., Ståhl, H.-E., Andersen, T., Gobakken, L.,Ene, R., Nelson, 2016. Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment* 173:98–108.

Hackl, P., 2016. Big Data: What can official statistics expect? *Statistical Journal of the IAOS*, *32*(1), pp.43-52.

Holm, S., Nelson, R. and Ståhl, G., 2017. Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne lidar, and space lidar. *Remote Sensing of Environment*, 197, pp.85-97.

Jazwinski, A. 1970. *Stochastic processes and filtering*. San Diego, CA: Academic Press. 376pp.

 DOI: 10.13140/RG.2.2.17634.07369

Kitanidis, P.K., 1985. Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association for Mathematical Geology*, 17(2), pp.195-208.

Kitchin, R., 2015. The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*(3), pp.471-481.

Knottnerus, P., 2003. *Sample-survey theory: some Pythagorean perspectives*. Springer Science & Business Media.

Lessard, V.C., McRoberts, R.E. and Holdaway, M.R., 2001. Diameter growth models using Minnesota Forest Inventory and Analysis data. *Forest Science*, 47(3), pp.301-310.

Magnussen, S. and Russo, G., 2012. Uncertainty in photo-interpreted forest inventory variables and effects on estimates of error in Canada's National Forest Inventory. *The Forestry Chronicle*, 88(4), pp.439-447.

Maybeck, P.S., 1979. *Stochastic models, estimation, and control*, ser. Mathematics in Science and Engineering, 141.

McConville, K.S. Moisen, G.G., and Frescino, T.S., 2020. A Tutorial on Model-Assisted Estimation with Application to Forest Inventory. *Forests*, 11(2), 24.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G. and Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. *Remote Sensing of Environment*, *115*(12), pp.3599-3614.

Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., Ståhl, G., 2013. Comparison of precision of biomass estimates in regional field sample-surveys and air-borne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sensing of Environment* 130 (2013):108–120.

Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H., Næsset, E., Gobakken, T. and Ståhl, G., 2009. Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39(4), pp.862-881.

Piepho, H.P., and J. Ogutu, J., 2007. Simple state-space models in a mixed model framework. *The American Statistician* 61(3):224-232.

Politis, D.N. and Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical association, 89(428), pp.1303-1313.

 DOI: 10.13140/RG.2.2.17634.07369

Reimsbach-Kounatze, C., 2015. *The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies*. OECD Digital Economy Papers. ISSN: 20716826

Rios, L.M. and Sahinidis, N.V., 2013. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), pp.1247-1293.

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T.G., Nelson, R.F. and Ståhl, G., 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. *Annals of Forest Science*, 73(4), pp.895-910.

Särndal, C.E., 1980. Two Model-Based Inference Arguments in Survey-sampling. *Australian Journal of Statistics*, 22(3), pp.341-348.

Särndal, C.E., Swensson, B. and Wretman, J., 1992. *Model-assisted survey-sampling Springer*. New York.

Särndal, C.E., Swensson, B. and Wretman, J., 2003. *Model-assisted survey-sampling*. Springer Science & Business Media.

Scannapieco, M., Virgillito, A. and Zardetto, D., 2013. Placing big data in official statistics: a big challenge. *New Techniques and Technologies in Statistics*.

Schreuder, H.T., LaBau, V.J., and Hazard, J.W., 1995. The Alaska Four-Phase Forest Inventory Sampling-design Using Remote Sensing and Ground Sampling. *Photogrammetric Engineering & Remote Sensing*, 61(3), pp. 291.-297.

Struijs, P., Braaksma, B. and Daas, P.J., 2014. Official statistics and big data. *Big Data & Society*, *1*(1), p.2053951714538417.

Tam, S.M. and Clarke, F., 2015. Big data, official statistics and some initiatives by the Australian Bureau of Statistics. *International Statistical Review*, *83*(3), pp.436-448.

Theil, H., and Goldberger, A.S., 1961. On pure and mixed statistical estimation in economics. *International Economic Review* 2(1):65-78.

Tomppo, E., Czaplewski, R.L., and Makisara, K. 2002. *The role of remote sensing in global forest assessment: A remote sensing background paper for Kotka IV expert consultation*; 01.07-05.07.2002, Kotka, Finland. Forest Resources Assessment (FRA) Working Paper 61. Rome, Italy: Food and Agriculture

Organization of the United Nations, Forest Resources Division, Forest Resources Assessment Programme. 72 p.

Weiskittel, A.R., Hann, D.W., Kershaw Jr, J.A. and Vanclay, J.K., 2011. *Forest growth and yield modeling*. John Wiley & Sons.

Wu, C.J. and Deng, L.Y., 1983. Estimation of variance of the ratio estimator: an empirical study. In *Scientific Inference, Data Analysis, and Robustness* (pp. 245-277). Academic Press.

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\frac{d\{\hat{v}_{m.m}\}_{\mathrm{j}}^{+}}{d(a_m)}$	Change in variance $\{\hat{v}_{m.m}\}_{\mathrm{j}}^{+}$ of the population estimate for the m^{th} study-variable with the difference-estimator as a function of the m^{th} arbitrary coefficient a_m, used to determine the value of a_m that minimizes variance $\{\hat{v}_{m.m}\}_{\mathrm{j}}^{+}$	46	scalar
$\mathbf{a}=\begin{bmatrix}\mathbf{a}_{\mathrm{y}}\\ \mathbf{a}_{\notin j}\end{bmatrix}$	Coefficient vector for population estimates for all M study-variables and all other J auxiliary-residuals except the j^{th} residual r_j	37	$(M+J-1)\times 1$
$\mathbf{a}_{\notin j}$	Coefficient vector for population estimates for all other J auxiliary-residuals other than the j^{th} residual r_j	37	$(J-1)\times 1$
a_j	Coefficient for the j^{th} auxiliary-residual	37	scalar
$\{a_m\}_{\mathrm{j,opt}}$	Value of the difference-estimator coefficient for the population estimate for the mth study-variable with the jth auxiliary-variable	46	scalar
$\mathbf{A_r}$	Matrix of arbitrary coefficients in the difference-estimator for the J auxiliary-residuals	23	$J\times J$
a_r	The coefficient in the difference-estimator for the auxiliary-residual when $J=1$		scalar

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\mathbf{A}_{\mathbf{y}}$	Matrix of arbitrary coefficients in the difference-estimator for the M study-variables	23	$M \times J$
$\mathbf{a}_{\mathbf{y}}$	Coefficient vector for population estimates for all M study-variables	37	$M \times 1$
diag()	Vector operator that extracts the diagonal of a square matrix	25	
$\left[\mathbf{e}_{\mathbf{x}_G}\right]_{\pi}$	Random error in population estimate for the J auxiliary-variables with sample G	18	$J \times 1$
$\left[\mathbf{e}_{\mathbf{x}_H}\right]_{\pi}$	Random error in population estimate for the J auxiliary-variables with sample H	18	$J \times 1$
$\left[\mathbf{e}_{\mathbf{y}}\right]_{\pi}$	Random error in population estimate for the M study-variables	18	$M \times 1$
$\mathbf{I}$	Conformable identity matrix		See context
j	Index for auxiliary-variable, $\mathrm{j} = \{1, \cdots, j, \ldots J\}$		scalar

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
J	Total number of auxiliary-variables		scalar
j	Index for the jth recursion, $\mathrm{j}=\{1,\cdots,j,\ldots J\}$	37	scalar
k	Index for k^{th} population element $k \in \{1,\ldots,N\}$	15	scalar
$\mathbf{K}$	Matrix of arbitrary coefficients in the difference-estimator for $J>1$	23	$(M+J)\times(M+J)$
$\mathbf{K}_{\mathrm{j}}$	Matrix of arbitrary coefficients in the difference-estimator for $J=1$	36	$(M+1)\times(M+1)$
m	Index for the study-variable $m \in \{1,\ldots,M\}$		scalar
M	Total number of study-variables		scalar
N	Number of elements in the sampled population	15	scalar

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\mathbf{r}$	The "auxiliary-residual" $\mathbf{r}=\left(-\hat{\mathbf{t}}_{\mathbf{x}_H}+\hat{\mathbf{t}}_{\mathbf{x}_G}\right)$; vector of differences between two population estimates for the J auxiliary-variables. $\mathrm{E}[\mathbf{r}]=\mathrm{E}\left[-\hat{\mathbf{t}}_{\mathbf{x}_H}+\hat{\mathbf{t}}_{\mathbf{x}_G}\right]=\mathbf{0}$	22	$J\times 1$
r	Sole auxiliary-residual when $J=1$, $r=\left(-\hat{t}_{x_H}+\hat{t}_{x_G}\right)$		scalar
$\mathbf{r}_{\notin j}$	Population estimates for all auxiliary-residuals except the j^{th} residual r_j	36	$(J-1)\times 1$
r_j	Population estimate for the j^{th} residual		scalar
$\hat{\mathbf{t}}$	Population estimates for all M study-variables and all J auxiliary-residuals	17	$(M+J)\times 1$
$\mathbf{t}_{\mathbf{x}}$	Vector of true population-totals (population parameters) for the J auxiliary-variables; $\mathrm{E}\left[\hat{\mathbf{t}}_{H,\pi}\right]=\mathrm{E}\left[\hat{\mathbf{t}}_{G,\pi}\right]=\mathbf{t}_{\mathbf{x}}$	17	$J\times 1$
t_x	Population parameter value for the auxiliary-variable when $J=1$		scalar
$\left[\hat{\mathbf{t}}_{\mathbf{x}_G}\right]_{\pi}$	Vector of π estimates of population-totals (population parameters) for J auxiliary-variables from sample G, where $\mathrm{E}\left[\hat{\mathbf{t}}_{H,\pi}\right]=\mathrm{E}\left[\hat{\mathbf{t}}_{G,\pi}\right]=\mathbf{t}_{\mathbf{x}}$	17	$J\times 1$

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\left[\hat{t}_{x_G}\right]_\pi$	Second population estimate for the auxiliary-variable when J=1, $\mathrm{E}\left[\hat{t}_{x_H}\right] = \mathrm{E}\left[\hat{t}_{x_G}\right] = t_x$		scalar
$\left[\hat{\mathbf{t}}_{\mathbf{x}_H}\right]_\pi$	Vector of π estimates of population-totals (population parameters) for J auxiliary-variables from sample H, where $\mathrm{E}\left[\hat{\mathbf{t}}_{H,\pi}\right] = \mathrm{E}\left[\hat{\mathbf{t}}_{G,\pi}\right] = \mathbf{t}_\mathbf{x}$	17	J×1
$\left[\hat{t}_{x_H}\right]_\pi$	First population estimate for the auxiliary-variable when J=1, $\mathrm{E}\left[\hat{t}_{x_H}\right] = \mathrm{E}\left[\hat{t}_{x_G}\right] = t_x$		scalar
$\mathbf{t}_\mathbf{y}$	Vector of true population-totals (population parameters) for the M study-variables	17	M×1
$\hat{\mathbf{t}}_\mathbf{y}$	Population estimates for all M study-variables	22	M×1
$\hat{\mathbf{t}}_\pi = \left[\begin{array}{c} \hat{\mathbf{t}}_\mathbf{y} \\ \hline \mathbf{r} \end{array}\right]_\pi$	Vector of sufficient statistics for the difference-estimator; a linear transformation of the (M+2J)×1 vector of population estimates in Equation (3).	22	(M+J)×1
U	Sampled population $U = (1, 2, \ldots, k, \ldots, N)$	15	
U_s	Subpopulation s, $U = \sum_{s=1}^{S} U_s$	85	

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\hat{\mathbf{V}}$	Covariance matrix for estimates of M study-variables and J auxiliary-residuals	19	$(M+J)\times(M+J)$
$\hat{v}_{i,k}$	$(i,k)^{\text{th}}$ element of covariance matrix $\hat{\mathbf{V}}$		scalar
$\hat{\mathbf{V}}_{\notin j}$	Partition of covariance matrix among population estimates for all J auxiliary-residuals except the j^{th} residual	36	$(J-1)\times(J-1)$
$\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}$	Partition of covariance matrix among the M study-variables estimated with sample s_G. $\mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G}=\mathbf{0}$ if all population parameters are known (e.g., census)	19	$M\times M$
$v_{\mathbf{x}_G,\mathbf{x}_G}$	Variance of the second population estimate of auxiliary-variable when J=1, $\Lambda=\left(v_{\mathbf{x}_H,\mathbf{x}_H}-2v_{\mathbf{x}_H,\mathbf{x}_G}+v_{\mathbf{x}_G,\mathbf{x}_G}\right)$		scalar
$\mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G}$	Partition of covariance matrix among the M study-variables estimated with samples s_H and s_G	19	$M\times M$
$v_{\mathbf{x}_H,\mathbf{x}_G}$	Covariance between two population estimates of auxiliary-variable when J=1, $\Lambda=\left(v_{\mathbf{x}_H,\mathbf{x}_H}-2v_{\mathbf{x}_H,\mathbf{x}_G}+v_{\mathbf{x}_G,\mathbf{x}_G}\right)$		scalar
$\mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H}$	Partition of covariance matrix among the M study-variables estimated with sample s_H	19	$M\times M$

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$v_{\mathbf{x}_H,\mathbf{x}_H}$	Variance of the first population estimate of auxiliary-variable when J=1, $\Lambda = \left(v_{\mathbf{x}_H,\mathbf{x}_H} - 2v_{\mathbf{x}_H,\mathbf{x}_G} + v_{\mathbf{x}_G,\mathbf{x}_G}\right)$		scalar
$\hat{\mathbf{V}}_{\mathbf{y},\notin j}$	Partition of covariance matrix between population estimates for the M study-variables and all J auxiliary-residuals other than the j^{th} residual	36	$M\times(J-1)$
$\mathbf{v}_{\mathbf{y},\mathbf{x}_G}$	Covariances between population estimates for the M study-variables and the second estimate (G) of the auxiliary-variable when J=1 , $\boldsymbol{\varphi} = \left(-\mathbf{v}_{\mathbf{y},\mathbf{x}_H} + \mathbf{v}_{\mathbf{y},\mathbf{x}_G}\right)$		$M\times1$
$\mathbf{V}_{\mathbf{y},\mathbf{x}_H}$	Partition of covariance matrix between the M study-variables and the J auxiliary-variables, the latter of which are estimated with sample s_H	19	$M\times J$
$\mathbf{V}_{\mathbf{y},\mathbf{x}_G}$	Partition of covariance matrix between the M study-variables and the J auxiliary-variables, the latter of which are estimated with sample s_G	19	$M\times J$
$\mathbf{v}_{\mathbf{y},\mathbf{x}_H}$	Covariances between population estimates for the M study-variables and the first estimate (H) of the auxiliary-variable when J=1 , $\boldsymbol{\varphi} = \left(-\mathbf{v}_{\mathbf{y},\mathbf{x}_H} + \mathbf{v}_{\mathbf{y},\mathbf{x}_G}\right)$		$M\times1$
$\mathbf{V}_{\mathbf{y},\mathbf{y}}$	Partition of covariance matrix among the M study-variables	19	$M\times M$

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\hat{\mathbf{V}}_{\mathbf{z}}$	Partition of covariance matrix among population estimates for all M study-variables and all other J auxiliary-residuals except the j^{th} residual r_j		$(M+J-1)\times(M+J-1)$
$\mathbf{V}_{\pi}$	Covariance matrix for the sufficient population statistics in Equation (11), where $\mathbf{V}_{\pi}=\left[\begin{array}{c:c}\mathbf{V}_{\mathbf{y},\mathbf{y}} & \boldsymbol{\Gamma} \\ \hdashline \boldsymbol{\Gamma}^{\mathrm{T}} & \boldsymbol{\Lambda}\end{array}\right]_{\pi}$	22	$(M+J)\times(M+J)$
$\mathbf{V}_{\pi}$	Covariance matrix for the sufficient population statistics for $J=1$, where $\mathbf{V}_{\pi}=\left[\begin{array}{c:c}\mathbf{V}_{\mathbf{y},\mathbf{y}} & \boldsymbol{\varphi} \\ \hdashline \boldsymbol{\varphi}^{\mathrm{T}} & \Lambda\end{array}\right]_{\pi}$		$(M+1)\times(M+1)$
$\left\{\hat{v}_{i,k}^{-}\right\}_{\mathrm{j}}$	Covariance between the i^{th} and k^{th} elements of covariance matrix $\hat{\mathbf{V}}_{\mathrm{j}}^{+}$ at the beginning of the j^{th} step in the recursion $\left\{\hat{v}_{i,k}^{-}\right\}_{\mathrm{j}}=\left\{\hat{v}_{i,k}^{+}\right\}_{\mathrm{j}-1}$	40	scalar
$\left\{\hat{v}_{i,k}^{+}\right\}_{\mathrm{j}}$	Covariance between the i^{th} and k^{th} elements of covariance matrix $\hat{\mathbf{V}}_{\mathrm{j}}^{+}$ at the end of the j^{th} step in the recursion $\left\{\hat{v}_{i,k}^{-}\right\}_{\mathrm{j}+1}=\left\{\hat{v}_{i,k}^{+}\right\}_{\mathrm{j}}$	40	scalar
x_j	j^{th} auxiliary-variable, $x_1,\ldots,x_j,\ldots,x_J$	15	scalar
$\mathbf{x}_k$	vector of measurements of study-variables for the k^{th} population element	17	$M\times1$

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
y_m	Value of study-variable y for the m^{th} population element	15	scalar
$\mathbf{y}_k$	vector of values for auxiliary-variables for the kth population element	17	$J\times 1$
y_m	m^{th} study-variable $y_1,\ldots,y_m,\ldots,y_M$	15	
$\hat{\mathbf{z}}=\begin{bmatrix}\hat{\mathbf{t}}_{\mathrm{y}}\\ \hline \mathbf{r}_{\notin j}\end{bmatrix}$	Population estimates for all M study-variables and all other $(J-1)$ auxiliary-residuals except the jth residual r_j	36	$(M+J-1)\times 1$

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\boldsymbol{\Gamma}_{\pi}$	Partition of covariance matrix $\mathbf{V}_{\pi}$ that contains the covariances between the vector of M study-variables and the vector of J auxiliary-residuals: $\boldsymbol{\Gamma}_{\pi} = \left[-\mathbf{V}_{\mathbf{y},\mathbf{x}_H} + \mathbf{V}_{\mathbf{y},\mathbf{x}_G} \right]_{\pi}$; see Särndal *et al.* 1992 (Section 6.8)	22	$M \times J$
π	Population estimator based on probability sampling		
$[\boldsymbol{\Theta}]_{\text{dif}}$	Difference-estimator with arbitrary coefficients	23	
$[\boldsymbol{\Theta}]_{\text{opt}}$	Difference-estimator with minimum-variance ("optimal") coefficients		
$\boldsymbol{\Theta}_{\pi}$	design-consistent π-estimator	19	

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
Λ	Variance of auxiliary-residual when J=1, $\Lambda = \left(v_{\mathbf{x}_H,\mathbf{x}_H} - 2v_{\mathbf{x}_H,\mathbf{x}_G} + v_{\mathbf{x}_G,\mathbf{x}_G} \right)$		scalar
$\hat{\Lambda}_j$	Variance for the population estimate for the j^{th} residual r_j, where $\Lambda = \left(v_{\mathbf{x}_H,\mathbf{x}_H} - 2v_{\mathbf{x}_H,\mathbf{x}_G} + v_{\mathbf{x}_G,\mathbf{x}_G} \right)$; see Särndal *et al.* 1992 (Section 6.8)	36	scalar
$\mathbf{\Lambda}_\pi$	Partition of covariance matrix $\mathbf{v}_\pi$ that contains the covariances among the J auxiliary-residuals; see Särndal *et al.* 1992 (Section 6.8)	22	$J \times J$
$\mathbf{\Lambda}_\pi$	Covariance matrix among J auxiliary-residuals, estimated with initial π sample $\mathbf{\Lambda}_\pi = \left[\mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H} - \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G} - \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H} + \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G} \right]_\pi$; see Särndal *et al.* 1992 (Section 6.8)	22	$J \times J$
$\boldsymbol{\pi}$	design-consistent Horvitz-Thompson estimator	7	

 DOI: 10.13140/RG.2.2.17634.07369

	Notation		
Variable	Definition	page	Dimensions
$\hat{\boldsymbol{\varphi}} = \begin{bmatrix} \hat{\boldsymbol{\varphi}}_{\mathbf{y}} \\ \hat{\boldsymbol{\varphi}}_{\notin j} \end{bmatrix}$	Vector partition of covariance matrix between population estimates for the j^{th} auxiliary-residual and all remaining variables: M study-variables and all other $(J-1)$ auxiliary-residuals, $\boldsymbol{\varphi} = \left(-\mathbf{v}_{\mathbf{y},\mathbf{x}_H} + \mathbf{v}_{\mathbf{y},\mathbf{x}_G}\right)$		$(M+J-1)\times 1$
$\hat{\varphi}_i$	The i^{th} element of vector $\hat{\boldsymbol{\varphi}}$, which contains the covariances between the population estimates for the j^{th} auxiliary-residual and the i^{th} variable, $\left(\hat{\varphi}_m\right)_{\text{j}} = \left(-\hat{v}_{y_m,\mathbf{x}_H} + \hat{v}_{y_m,\mathbf{x}_G}\right)_{\text{j}}$	40	scalar
$\hat{\boldsymbol{\varphi}}_{\notin j}$	Vector partition of covariance matrix between population estimates for the j^{th} auxiliary-residual and all other $(J-1)$ auxiliary-residuals	36	$(J-1)\times 1$
$\hat{\boldsymbol{\varphi}}_{\mathbf{y}}$	Vector partition of covariance matrix between population estimates for the M study-variables and the j^{th} auxiliary-residual	36	$M\times 1$
$\mathbf{0}$	Conformable zero matrix		See context

 DOI: 10.13140/RG.2.2.17634.07369