Minimum-Variance Coefficients for the Generalized Multivariate Difference Estimator (GMDe)

Technical Report 2020-2

August 4, 2020

DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Raymond L. Czaplewski

Tele. +970-218-5907

Raymond_Czaplewski@Environmetrika.org

A Contribution to the Environmetrika Series on

Simple Robust Multivariate Estimators for Complex Sample Surveys

(Page intentionally left blank)

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

EXTENDED ABSTRACT

The Generalized Multivariate Difference estimator (GMDe) is a broad generalization of the univariate "difference estimator" described by Hansen *et al*. (1953:250-253) and Särndal *et al*. (1992:239-242). Difference estimators use population estimates of auxiliary variables to improve population estimates of correlated study variables. Examples of auxiliary variables include administrative records, remotely sensed measurements, and time-series of predictions from deterministic process models (*e.g.*, econometric models, demographic models, forest-stand projection models).

GMDe is a multivariate alternative to model-assisted GREG regression estimators for finite populations, such as post-stratification, ratio, regression, lasso, ridge, and elastic net estimators. GMDe does not require model-assisted predictions of a proxy variable, nor does GMDe require known population totals for all auxiliary variables.

Much like the composite estimator, GMDe is a simple linear transformation of a vector of population estimates from a probability sample with design-consistent multivariate Horvitz-Thompson "π-estimator". Therefore, GMDe does not directly use the data matrix of study variables and auxiliary variables for each population element included in the probability sample. The following Technical Report derives the $M \times J$ matrix of minimum-variance coefficients for each of M study variables and each of J auxiliary variables for the linear transformation in GMDe.

The degree of variance reduction with GMDe depends, in part, upon the strength of correlations between study variables and auxiliary variables. Substantial improvements of GMDe relative to the prior π-estimate require relatively strong correlations (*e.g.*, ±0.70 and stronger). In a National Forest Inventory (NFI), remotely sensed auxiliary variables are sufficiently correlated with broad groupings of domains. However, predictions from deterministic process models might provide auxiliary variables that are more strongly correlated with detailed study variables that change slowly or more predictably over time; and change-detection with remotely sensed data can post-stratify the population into undisturbed strata for which deterministic process models provide stronger predictors.

This Technical Report includes a simple example of the recursive version of GMDe, which is a relatively simple estimator for complex sample surveys that include longitudinal surveys for time-series of population estimates; interpenetrating panels; multi-phase and multi-stage sampling; and multiple independent surveys. If a design-based π-estimate is feasible for a vector of study variables and correlated auxiliary variables, then GMDe can use those π-estimates for the population to reduce variances of study variables that are correlated with auxiliary variables. The recursive GMDe can also impose equality and inequality constraints on study variables and mitigate influence of outliers. The recursive GMDe replaces inversion of the $J \times J$ partition of the π-covariance matrix for auxiliary residuals with a sequence of J scalar divisions. Therefore, the recursive GMDe can be numerically robust even if there are strong collinearities among numerous auxiliary variables, and dimensionality reduction methods are unnecessary.

GMDe is unbiased for population estimates of the study variables. However, GMDe uses sample survey statistics to compute approximate minimum-variance coefficients. Different hypothetical realizations of the sample would result in different GMDe coefficients. Therefore, the closed-formed GMDe variance estimator with minimum-variance coefficients likely underestimates variance of GMDe population estimates for the study variables. This Technical Report recommends bootstrap resampling a more accurate estimate of the true GMDe covariance matrix.

Keywords: sample survey, difference estimator, multivariate, GMD, GMDe, GREG, model-assisted, minimum-variance, composite estimator, post-sampling estimator, Kalman filter, NFI, National Forest Inventory, FIA, big data, Särndal, Hansen

TABLE OF CONTENTS

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

1 INTRODUCTION

Czaplewski (2020) introduced the Generalized Multivariate Difference (GMDe) estimator for complex sample surveys as a general case of the univariate difference estimator described by Hansen *et al.* (1953) and Särndal *et al.* (1992:239-242). GMDe is a straightforward linear transformation of a vector of design-based population estimates and its covariance matrix. GMDe is an alternative to model-assisted GREG regression estimators for finite populations, such as ratio, regression, lasso, ridge, and elastic net estimators. Unlike model-assisted estimators, GMDe does not directly use data for each sampling unit included in the realized sample; rather, GMDe uses population estimates from a probability sample and its design-based π-estimator. GMDe is more closely related to the multivariate composite estimator and the Kalman filter (Gregoire and Walters, 1988; Lui, 2020).

GMDe starts with a vector estimate of population parameters from a probability sample and a design-consistent estimator, *e.g.*, the multivariate Horvitz-Thompson (HT) estimator, which Särndal *et al.* generalize with notation "π-estimator". That vector estimate has three components:

1. $M\times1$ partition of population estimates for M study variables, which may have sub-partitions for different domains, subpopulations and time periods;
2. $J\times1$ partition of population estimates for J auxiliary variables;
3. A second $J\times1$ partition of population estimates for those same J auxiliary variables.

For example, the Phase-One in a Two-Phase sampling design produces a $J\times1$ partition with population totals for the J auxiliary variables (Component #3). Phase-Two produces a $(M+J)\times1$ partition for the M study variables (Component #1) and those same J auxiliary variables (Component #2). The result is the $(M+2J)\times1$ vector of π-estimates and its $(M+2J)\times(M+2J)$ covariance matrix, a $(M+J)\times J$ partition of which includes covariances caused by joint inclusion probabilities between the Phase-One and Phase-Two. (See Section 3 page 11 for details.)

A simple linear transformation, with coefficients $\{0,1,-1\}$, subtracts one $J\times1$ vector of π-estimates for the auxiliary variables from the other $J\times1$ vector. (See Section 4, page 13, for details.) That "difference" is the $J\times1$ vector of "auxiliary residuals." Thus the term "difference" in "difference estimator." The overall result is a $(M+J)\times1$ vector of transformed π-estimates, in which the leading $M\times1$ partition contains the original population estimates for the M study variables, and the trailing partition contains the J auxiliary residuals. The corresponding $(M+J)\times(M+J)$ covariance matrix accounts for all collinearities among population estimates for the M study variables and the J auxiliary residuals. These are the sufficient "prior" statistics for GMDe, and this formulation simplifies the matrix algebra and implementation of numerical algorithms.

GMDe uses a second linear transformation of that $(M+J)\times1$ vector of π-estimates. The $(M+J)\times(M+J)$ transformation matrix includes an $M\times J$ partition of "arbitrary" coefficients, which are assumed to be known constants (Särndal *et al.*,1992:221-225). Those coefficients represent numerical weights placed on the population estimates for the *J* auxiliary residuals that modify the prior π-estimates for the *M* study variables. A wise choice for those *M·J* coefficients will reduce variance of population estimates for the *M* study variables by "gaining strength" from population estimates for *J* correlated auxiliary residuals. A compelling choice for those *M·J* coefficients is one that approximately minimizes variances of population estimates for all *M* study variables while fully accounting for all collinearities among all (*M*+*J*) π-estimates. This is the topic of Section 6 (page 16).

1.1 Univariate Model-assisted Estimator

For the past 50 years, the univariate model-assisted estimator (*e.g.*, Särndal *et al.*, 1992) has dominated sample survey applications in National Forest Inventories (NFI) and many other programs for official statistics that incorporate auxiliary variables. The univariate model-assisted estimator employs a univariate proxy variable that predicts the value of the study variable for each sampling unit given corresponding measurements of one or more auxiliary variables, which serve as the predictors. Therefore, the structure of the regression model for each auxiliary variable typically differs for categorical v. continuous variables; homoscedastic v. heteroscedastic random errors; and zero v. nonzero intercepts. Regression coefficients are considered random variables that are a nonlinear multivariate function of population parameters.

While the value of the study variable is known only for those population elements selected for the sample, the values of the proxy variable are assumed known for each and every element in the finite population. Therefore, the population total for the proxy variable must be known without sampling error, which somewhat simplifies the stochastic model, just as multiple regression is simpler than errors-in-variables (stochastic) regression. Since the proxy variable is a regression predictand of the auxiliary variables, and the population total for the proxy variable must be a known constant, then the auxiliary variable must be known for each and every population element; and therefore, population totals for all auxiliary variables are known constants. (However, population totals for some auxiliary variables might be sample survey estimates with complex sampling designs, not known constants. This poses challenges to model-assisted estimation with complex sampling designs.)

The model-assisted estimator is fundamentally a univariate approach; it typically estimates the population parameter for a single study variable with a single proxy variable. Population estimates for multiple study variables require separate application of the univariate estimator to each study variable (*i.e.*, a "parallel process"). This provides a vector of population estimates and a vector of variance estimates for all *M* study variables, but the model-assisted estimator is not well suited to estimate the covariance matrix among population estimates for all study variables. The full covariance matrix is useful for synthetic estimators (Särndal *et al.*, 1992:173,205,388,408), such as ratios-of-means, rates of change between time periods, differences between study variables in different domains, table margins, *etc*.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

1.2 Generalized Multivariate Difference (GMDe) Estimator

GMDe does not use proxy variables or an explicit regression model. GMDe does not require known population parameters (*e.g.*, census statistics) for auxiliary variables, although it can use such statistics. GMDe does not require direct observations of variables for each sampling unit in the realized sample. GMDe applies to both finite and infinite population models. Therefore, GMDe can more simply accommodate complex sampling designs compared to model-assisted estimators.

GMDe is a linear transformation of π-estimates from a probability sample and a consistent estimator (*e.g.*, HT estimator). HT and regression estimators use measurements of each population element selected for the sample. On the other hand, GMDe uses estimates of population parameters, although some weighted mean for measurements of population elements in a sample can provide those statistics. Therefore, GMDe is not strongly affected by the functional relationship between measurements of a study variable and an auxiliary variables at the scale of a population element (*e.g.*, zero v. nonzero intercept; continuous v. categorical variables).

GMDe is fundamentally a multivariate approach; it simultaneously estimates the full vector of population totals and the corresponding covariance matrix for random estimation errors (*i.e.*, propagated random sampling errors). A multivariate estimator facilitates closed-form computations and linear approximations for synthetic estimators and their covariance matrices. These support estimates and inference regarding sums of population statistics (*e.g.*, table margins and aggregations of statistics according to a classification hierarchy); differences between population statistics (*e.g.*, changes between different time periods and differences among domains); and ratios (*e.g.*, average biomass per hectare for each type of forest condition, average proportion of tree volume by tree species that is merchantable). Czaplewski (2010) provides examples.

The effectiveness of GMDe variance reduction depends upon three elements: the strength of the correlations between the auxiliary variables and the study variables; the variance of population estimates for the auxiliary variables; and the choice of coefficients in the GMDe linear transformation. It is the latter element that is the topic of this Technical Report.

2 Stochastic Model

Assume each population element κ has a $(M+J)\times 1$ vector measurement

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

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

$$\mathbf{t} = \begin{bmatrix} \mathbf{t}_{\mathbf{y}} \\ \mathbf{t}_{\mathbf{x}} \end{bmatrix} = \sum_{\kappa}^{N} \begin{bmatrix} \mathbf{y}_{\kappa} \\ \mathbf{x}_{\kappa} \end{bmatrix} \tag{1}$$

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, GMDe 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*+2*J*)×1 vector of π population estimates, where the "π" notation follows that of Särndal *et al.* (1992)

$$\hat{\mathbf{t}}_{\pi} = \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hline \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hline \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_{\pi} \tag{2}$$

Partition $\hat{\mathbf{t}}_{\mathbf{y},\pi}$ contains the π-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 π-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 vector estimate from a Two-Phase sample survey, where Phase-One provides the *J*×1 partition $\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$, and Phase-Two provides the (*M*+*J*)×1 partition $\left[\hat{\mathbf{t}}_{\mathbf{y},\pi}^{\mathrm{T}} \mid \hat{\mathbf{t}}_{\mathbf{x}_H,\pi}^{\mathrm{T}}\right]^{\mathrm{T}}$. In a Two-Stage sampling-design, the Primary Sampling Units (PSU) provide the *J*×1 partition $\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}$, and the Secondary Sampling Units (SSU) in the cluster-plot provide the (*M*+*J*)×1 partition $\left[\hat{\mathbf{t}}_{\mathbf{y},\pi}^{\mathrm{T}} \mid \hat{\mathbf{t}}_{\mathbf{x}_H,\pi}^{\mathrm{T}}\right]^{\mathrm{T}}$. Pivot partitions for auxiliary variables such that both share the same expected vector values, *i.e.*, $\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}_H,\pi}\right] = \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}_G,\pi}\right] = \mathbf{t}_{\mathbf{x}}$.

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, GMDe is a simple linear transformation of that stochastic model, where the outcome from a π-estimator guides the choice of coefficients in that transformation.

Represent (*M*+2*J*)×1 vector of π-estimates in Equation (3) as

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hat{\mathbf{t}}_{\mathbf{x}_H} \\ \hat{\mathbf{t}}_{\mathbf{x}_G} \end{bmatrix}_\pi = \begin{bmatrix} \mathbf{t}_{\mathbf{y}} \\ \mathbf{t}_{\mathbf{x}} \\ \mathbf{t}_{\mathbf{x}} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{\mathbf{y}} \\ \mathbf{e}_{\mathbf{x}_H} \\ \mathbf{e}_{\mathbf{x}_G} \end{bmatrix}_\pi \quad (3)$$

where $(M+2J)\times 1$ vector $\left[\mathbf{e}_{\mathbf{y}} \mid \mathbf{e}_{\mathbf{x}_H} \mid \mathbf{e}_{\mathbf{x}_G}\right]_\pi^{\mathrm{T}}$ contains random variables driven by a stochastic process, typically 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 $\mathrm{E}\left[\mathbf{e}_{\mathbf{y}} \mid \mathbf{e}_{\mathbf{x}_H} \mid \mathbf{e}_{\mathbf{x}_G}\right]_\pi^{\mathrm{T}} = \mathbf{0}$; denote its expected $(M+2J)\times(M+2J)$ covariance matrix as

$$\mathbf{V} = \mathrm{E}\left[\begin{bmatrix} \mathbf{e}_{\mathbf{y}} \\ \mathbf{e}_{\mathbf{x}_H} \\ \mathbf{e}_{\mathbf{x}_G} \end{bmatrix}\begin{bmatrix} \mathbf{e}_{\mathbf{y}} \\ \mathbf{e}_{\mathbf{x}_H} \\ \mathbf{e}_{\mathbf{x}_G} \end{bmatrix}^{\mathrm{T}}\right]_\pi = \begin{bmatrix} \mathbf{V}_{\mathbf{y},\mathbf{y}} & \mathbf{V}_{\mathbf{y},\mathbf{x}_H} & \mathbf{V}_{\mathbf{y},\mathbf{x}_G} \\ \mathbf{V}_{\mathbf{x}_H,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_H,\mathbf{x}_G} \\ \mathbf{V}_{\mathbf{x}_G,\mathbf{y}} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H} & \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_G} \end{bmatrix}_\pi \quad (4)$$

GMDe applies to any multivariate π-estimate in the form of Equations (4) and (5) from any probability sampling-design p(·) and its design-consistent π-estimator $\mathbf{\Theta}_\pi$.

Consider Two-Phase sampling as a simple example. Phase-One typically 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 typically 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 $\left[\mathbf{V}_{\mathbf{x}_G,\mathbf{y}} \mid \mathbf{V}_{\mathbf{x}_G,\mathbf{x}_H}\right]$ contains any weak covariances that are caused by joint inclusion probabilities between Phase-One and Phase-Two.

3 DESIGN-BASED SAMPLE SURVEY ESTIMATOR

GMDe is a linear transformation of a "prior" vector of estimated population totals, and its associated covariance matrix, for population *U*. The multivariate HT estimator is an attractive choice as a prior estimate of population parameters as input to GMDe, but any design-consistent π-estimator is suitable, including estimators based on either a finite or infinite population model.

Consider the example of a two phase sample. The Phase-Two sample, denoted with subscript H, estimates a $(M{+}J)\times 1$ vector of population totals $\hat{\mathbf{t}}_{\mathbf{y}H\pi}$ for M study variables and J auxiliary variables $\hat{\mathbf{t}}_{\mathbf{x}H\pi}$. The Phase-One sample, denoted with subscript G, produces a second $J\times 1$ vector estimate $\hat{\mathbf{t}}_{\mathbf{x}G\pi}$ of population totals for J auxiliary variables

$$\hat{\mathbf{t}}_{HG\pi} = \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}H\pi} \\ \hat{\mathbf{t}}_{\mathbf{x}H\pi} \\ \hat{\mathbf{t}}_{\mathbf{x}G\pi} \end{bmatrix} = \sum_U \left(\begin{bmatrix} \left(I_{\kappa H} / \pi_{\kappa H} \right) \begin{bmatrix} \mathbf{y}_\kappa \\ \mathbf{x}_\kappa \end{bmatrix} \\ \left(I_{\kappa G} / \pi_{\kappa G} \right) \ \mathbf{x}_\kappa \end{bmatrix} \right) \tag{5}$$

Where random variables $I_{\kappa H}$ and $I_{\kappa G}$ are indicator variables for population element *κ*. $I_{\kappa H} = 1$ if element *κ* is included in the Phase-Two sample, $I_{\kappa H} = 0$ otherwise; and $I_{\kappa G} = 1$ if element *κ* is included in the Phase-One sample, $I_{\kappa G} = 0$ otherwise. Subscript "*π*" follows notation by Särndal *et al*. (1992) for the HT estimator. Scalar $\pi_{\kappa H}$ is the probability that element κ will be included in Phase-Two sample *H*; and $\pi_{\kappa G}$ is the probability that element κ will be included in Phase-Two sample *H*.

The vectors of expected values, denoted E[•], are identical for both partitions of population estimates for the *J* auxiliary variables.

$$\mathrm{E}\begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}H\pi} \\ \hat{\mathbf{t}}_{\mathbf{x}H\pi} \\ \hat{\mathbf{t}}_{\mathbf{x}G\pi} \end{bmatrix} = \sum_U \begin{bmatrix} \mathbf{y}_\kappa \\ \mathbf{x}_\kappa \\ \mathbf{x}_\kappa \end{bmatrix} = \begin{bmatrix} \mathbf{t}_\mathbf{y} \\ \mathbf{t}_\mathbf{x} \\ \mathbf{t}_\mathbf{x} \end{bmatrix} \quad \text{where } \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}H\pi} \right] = \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}G\pi} \right] = \mathbf{t}_\mathbf{x} \tag{6}$$

Denote the corresponding π-estimate of the $(M{+}2J)\times(M{+}2J)$ covariance matrix for $(M{+}2J)\times 1$ vector estimate $\hat{\mathbf{t}}_{HG\pi}$ in Equation (5) as

$$\hat{\mathbf{V}}_{HG\pi} = \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}H\pi} & \hat{\mathbf{V}}_{\mathbf{yx}H\pi} & \hat{\mathbf{V}}_{\mathbf{y}H\mathbf{x}G\pi} \\ \hat{\mathbf{V}}_{\mathbf{yx}H\pi}^{\mathrm{T}} & \hat{\mathbf{V}}_{\mathbf{x}H\pi} & \hat{\mathbf{V}}_{\mathbf{x}H\mathbf{x}G\pi} \\ \hat{\mathbf{V}}_{\mathbf{y}H\mathbf{x}G\pi}^{\mathrm{T}} & \hat{\mathbf{V}}_{\mathbf{x}H\mathbf{x}G\pi}^{\mathrm{T}} & \hat{\mathbf{V}}_{\mathbf{x}G\pi} \end{bmatrix} \tag{7}$$

Other sampling designs can be cast into a similar vector structure. Two stage sampling is the primary example in addition to two phase sampling. Another example is the combination of two separate surveys, in which the one is a general purpose sample with equal inclusion probabilities, and the second is an independent sample with inclusion probabilities that are believed to be

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

higher for rare subpopulations and lower for common subpopulations. The potential designs are numerous.

Expansion of the vector structure in Equation (7) can accommodate more complex probability sampling designs, such as multi-phase sampling, multi-stage sampling and combinations of both. Such complex designs are well suited to monitoring forest ecosystems with multiple sensor technologies, some of which are feasible only for a sample of cluster plots (*e.g.*, Czaplewski 1999).

4 SUFFICIENT GMDE STATISTICS: AUXILIARY RESIDUALS

Define **r** as the $J\times1$ vector of "auxiliary residuals," namely, the residual differences between two vectors of population estimates for the J auxiliary variables in Equation (5), where both vectors share the same expected values, *i.e.*, $\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}H\pi}\right]=\mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{x}G\pi}\right]=\mathbf{t}_{\mathbf{x}}$. A typical source for those two estimates is a Two-Phase sample with design-consistent π-estimator (see Section 3, page 11).

A simple linear transformation of the $(M+2J)\times1$ vector of prior π-estimates in Equation (5) provides the $(M+J)\times1$ vector of sufficient statistics for GMDe

$$\hat{\mathbf{t}}_{\mathbf{yr}}^{-}=\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}}^{-}\\ \mathbf{r}\end{bmatrix}=\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & -\mathbf{I} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}H\pi}\\ \hat{\mathbf{t}}_{\mathbf{x}H\pi}\\ \hat{\mathbf{t}}_{\mathbf{x}G\pi}\end{bmatrix}\quad\text{where } \mathbf{r}=\left(\hat{\mathbf{t}}_{\mathbf{x}H\pi}-\hat{\mathbf{t}}_{\mathbf{x}G\pi}\right) \tag{8}$$

The "–" superscript in Equation (8) and following equations denotes a prior π-estimate before the GMDe linear transformation; this convention follows Maybeck (1979) for the Kalman filter. **I** denotes the conformable identity matrix, and **0** denotes the conformable zero matrix.

Recall from Equation (6) the expected values for the transformation in Equation (8)

$$\begin{aligned}\mathrm{E}\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}}^{-}\\ \mathbf{r}\end{bmatrix}&=\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{-I} & \mathbf{I}\end{bmatrix}\mathrm{E}\begin{bmatrix}\hat{\mathbf{t}}_{\mathbf{y}H\pi}\\ \hat{\mathbf{t}}_{\mathbf{x}H\pi}\\ \hat{\mathbf{t}}_{\mathbf{x}G\pi}\end{bmatrix}\\ &=\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{-I} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{t}_{\mathbf{y}}\\ \mathbf{t}_{\mathbf{x}}\\ \mathbf{t}_{\mathbf{x}}\end{bmatrix}\\ &=\begin{bmatrix}\mathbf{t}_{\mathbf{y}}\\ \mathbf{0}\end{bmatrix}\end{aligned} \tag{9}$$

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Define the $(M+J)\times(M+J)$ covariance matrix for this linear transformation of $(M+2J)\times(M+2J)$ covariance matrix $\hat{\mathbf{V}}_{HG\pi}$ in Equation (7) as

$$\hat{\mathbf{V}}_{\mathbf{yr}} = \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \left[\begin{array}{c|c|c} \hat{\mathbf{V}}_{\mathbf{y}H\pi} & \hat{\mathbf{V}}_{\mathbf{yx}H\pi} & \hat{\mathbf{V}}_{\mathbf{y}H\mathbf{x}G\pi} \\ \hline \hat{\mathbf{V}}^{\mathrm{T}}_{\mathbf{yx}H\pi} & \hat{\mathbf{V}}_{\mathbf{x}H\pi} & \hat{\mathbf{V}}_{\mathbf{x}H\mathbf{x}G\pi} \\ \hline \hat{\mathbf{V}}^{\mathrm{T}}_{\mathbf{y}H\mathbf{x}G\pi} & \hat{\mathbf{V}}^{\mathrm{T}}_{\mathbf{x}H\mathbf{x}G\pi} & \hat{\mathbf{V}}_{\mathbf{x}G\pi} \end{array}\right] \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & -\mathbf{I} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}} & \mathbf{\Gamma} \\ \hline \mathbf{\Gamma}^{\mathrm{T}} & \mathbf{\Lambda} \end{array}\right] \quad \text{where} \begin{cases} \mathbf{\Gamma} = \left(-\hat{\mathbf{V}}_{\mathbf{yx}H\pi} + \hat{\mathbf{V}}_{\mathbf{y}H\mathbf{x}G\pi}\right) \\ \mathbf{\Lambda} = \left(\hat{\mathbf{V}}_{\mathbf{x}H\pi} - \hat{\mathbf{V}}_{\mathbf{x}H\mathbf{x}G\pi} - \hat{\mathbf{V}}^{\mathrm{T}}_{\mathbf{x}H\mathbf{x}G\pi} + \hat{\mathbf{V}}_{\mathbf{x}G\pi}\right) \end{cases} \tag{10}$$

Notation in Equation (10) for $J\times J$ partition $\mathbf{\Lambda}$ of covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}}$ follows that of Särndal *et al.* (1992:240), while $M\times J$ partition $\mathbf{\Gamma}$ is the multivariate extension of their $1\times J$ vector partition.

5 Normalization of Parameter Space

This Technical Report uses normalization to provide intuitive insights to GMDe. The $(M+J)\times(M+J)$ normalized covariance matrix $\hat{\mathbf{V}}^{-}_{\mathbf{yr}*}$ in Equation (11) corresponds to the π-estimator for the $(M+J)\times(M+J)$ covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}}$ in Equation (10).

Normalization refers to a linear transformation such that the vector of estimated population parameters for the study variables and auxiliary residuals in Equation (8) is centered on the zero vector with unit variance. Accomplish the latter with a $(M+J)\times(M+J)$ diagonal matrix $\mathbf{S}^{-1}$, labeled the "normalization matrix." The diagonal elements of $\mathbf{S}^{-1}$ equal the inverse square root of the corresponding diagonal elements of π-estimator for covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}}$ from the π-probability sample in Equation (10), and all off-diagonal elements equal zero.

$$\hat{\mathbf{V}}^{-}_{\mathbf{yr}*} = \mathbf{S}^{-1}\hat{\mathbf{V}}_{\mathbf{yr}}\mathbf{S}^{-1}$$

$$= \left[\begin{array}{c|c} \mathbf{S}^{-1}_{\mathbf{y}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}^{-1}_{\mathbf{r}} \end{array}\right] \left[\begin{array}{c|c} \hat{\mathbf{V}}^{-}_{\mathbf{y}} & \mathbf{\Gamma} \\ \hline \mathbf{\Gamma}^{\mathrm{T}} & \mathbf{\Lambda} \end{array}\right] \left[\begin{array}{c|c} \mathbf{S}^{-1}_{\mathbf{y}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}^{-1}_{\mathbf{r}} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \hat{\mathbf{V}}^{-}_{\mathbf{y}*} & \mathbf{\Gamma}_{*} \\ \hline \mathbf{\Gamma}^{\mathrm{T}}_{*} & \mathbf{\Lambda}_{*} \end{array}\right] \quad \text{where} \begin{cases} \hat{\mathbf{V}}^{-}_{\mathbf{y}*} = \mathbf{S}^{-1}_{\mathbf{y}}\hat{\mathbf{V}}^{-}_{\mathbf{y}}\mathbf{S}^{-1}_{\mathbf{y}} \\ \mathbf{\Gamma}_{*} = \mathbf{S}^{-1}_{\mathbf{y}}\mathbf{\Gamma}\mathbf{S}^{-1}_{\mathbf{r}} \\ \mathbf{\Lambda}_{*} = \mathbf{S}^{-1}_{\mathbf{r}}\mathbf{\Lambda}\,\mathbf{S}^{-1}_{\mathbf{r}} \end{cases} \tag{11}$$

Diagonal elements of the $(M+J)\times(M+J)$ matrix $\mathbf{S}$ equal standard deviation statistics, and all off-diagonal elements of $\mathbf{S}$ equal zero. The normalized covariance matrix $\hat{\mathbf{V}}^{-}_{\mathbf{yr}*}$ is a correlation matrix, in which the diagonal elements all equal 1.0, and all off-diagonal elements are bounded by $\{-1,1\}$. The "*" notation in the subscript denotes the normalized statistics.

The normalized vector $\hat{\mathbf{t}}_{\mathbf{yr}*}$ equals the estimated population vector $\hat{\mathbf{t}}_{\mathbf{yr}\pi}$, minus the best available values for expected values for the population parameters, times the normalization matrix in Equation (11), where $\mathrm{E}\left[\mathbf{r}_{\mathbf{x}\pi}\right]=\mathbf{0}$ in Equation (9).

$$\begin{aligned}
\hat{\mathbf{t}}_{\mathbf{yr}*}^{-} &= \left[\begin{array}{c|c} \mathbf{S}_{\mathbf{y}}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}_{\mathbf{r}}^{-1} \end{array}\right]\left(\left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}} \\ \hline \mathbf{r} \end{array}\right]-\left[\begin{array}{c} \tilde{\mathbf{t}}_{\mathbf{y}} \\ \hline \mathbf{0} \end{array}\right]\right) \quad \text{where } \begin{cases} \mathrm{E}\left[\hat{\mathbf{t}}_{\mathbf{y}}^{-}\right] \approx \hat{\mathbf{t}}_{\mathbf{y}H\pi} \\ \mathrm{E}[\mathbf{r}]=\mathbf{0} \end{cases} \\
&= \left[\begin{array}{c|c} \mathbf{S}_{\mathbf{y}}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}_{\mathbf{r}}^{-1} \end{array}\right]\left(\left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{r} \end{array}\right]\right) \\
&= \left[\begin{array}{c} \hat{\mathbf{t}}_{\mathbf{y}*}^{-} \\ \hline \mathbf{r}_{*} \end{array}\right] \quad \text{where } \begin{cases} \hat{\mathbf{t}}_{\mathbf{y}*}^{-}=\mathbf{0} \\ \mathbf{r}_{*}=\mathbf{S}_{\mathbf{r}}^{-1}\mathbf{r} \end{cases}
\end{aligned} \tag{12}$$

Computations of minimum-variance coefficients in Section 6 (starting on page 16) treat the covariance matrix partitions $\mathbf{\Gamma}_{*}$ and $\mathbf{\Lambda}_{*}$ in Equation (11) and the vector of auxiliary residuals $\mathbf{r}_{*}$ in Equation (12) as known constants, namely, the realized estimates of population parameters.

In addition to providing intuitive insights, normalization provides other benefits:

1. Normalization reduces risk from numerical round-off error if the numerical ranges of the study variables and auxiliary residuals vary greatly.
2. Off-diagonal elements of covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}*}$ are correlation coefficients bounded by $\{-1,1\}$. In rare cases, GMDe numerical estimates of correlations are not so bounded. Czaplewski (2020) introduces inequality constraints in the recursive GMDe that assure estimated correlations are within feasible bounds. Without this constraint, GMDe estimates of variance statistics can be negative.
3. Correlations provide the sufficient statistics for computation of approximate minimum-variance GMDe coefficients in Section 6 (page 16). Sample survey estimates of those correlations may be replaced by other estimates, such as the correlations between direct measurements of study variables and auxiliary predictions of those same of study variables with a deterministic tree- or stand-level projection model as part of an annual NFI (*e.g.*, McRoberts 2000).
4. The correlation statistic is invariant to the slope and intercept of the joint distribution of between the study variables $\mathbf{y}_{\kappa}$ and the auxiliary variables $\mathbf{x}_{\kappa}$. This eliminates the requirement in model-assisted estimators that the study variable and proxy variable share the same metric space.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

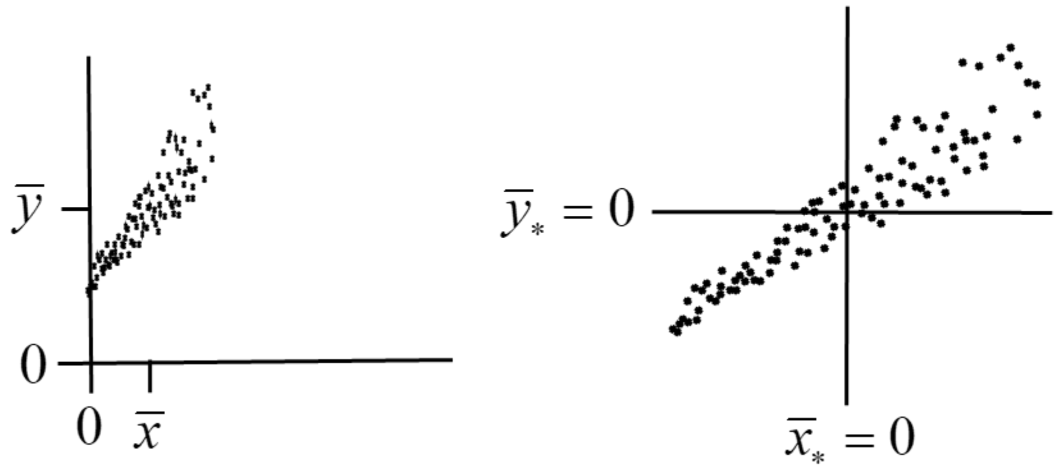


Figure 1 Examples of distributions between a study variable and an auxiliary residual. The left-hand figure is a scatterplot between study variable *y* and auxiliary variable *x*. The right-hand figure portrays the same distribution after normalization.

Final population estimates $\hat{\mathbf{t}}^{+}_{\mathbf{yr}*}$ with GMDe must be retransformed back into the original parameter space of $M\times 1$ vector $\hat{\mathbf{t}}^{-}_{\mathbf{y}}$ in Equation (8)

$$\hat{\mathbf{t}}^{+}_{\mathbf{y}} = \mathbf{S}_{\mathbf{y}}\,\hat{\mathbf{t}}^{+}_{\mathbf{y}*} \tag{13}$$

with the corresponding retransformation of the final $M\times M$ covariance matrix

$$\hat{\mathbf{V}}^{+}_{\mathbf{y}} = \mathbf{S}_{\mathbf{y}}\hat{\mathbf{V}}^{+}_{\mathbf{y}*}\mathbf{S}_{\mathbf{y}} \tag{14}$$

Equations (13) and (14) exclude partitions for the auxiliary residuals, which are assumed no longer relevant to the study after GMDe estimates are complete.

6 Approximate Minimum-variance GMDe Coefficients

GMDe (*i.e.*, vector $\hat{\mathbf{t}}^{+}_{\mathbf{yr}*}$) is a simple linear transformation of normalized π-estimate $\hat{\mathbf{t}}^{-}_{\mathbf{yr}*}$

$$\hat{\mathbf{t}}^{+}_{\mathbf{yr}*} = \mathbf{K}\,\hat{\mathbf{t}}^{-}_{\mathbf{yr}*} \tag{15}$$

where **K** is a $(M+J)\times(M+J)$ coefficient matrix with the following structure

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\mathbf{K} = \mathbf{I} + \left[\mathbf{0} \,\vdots\, \mathbf{A}\right] \tag{16}$$

I is the $(M+J)\times(M+J)$ identify matrix; **0** is the $(M+J)\times M$ zero matrix; and **A** is a $(M+J)\times J$ matrix of known but arbitrary coefficients (*i.e.*, "weights"), which determine the degree to which J correlated auxiliary residuals modify population estimates for the M study variables

$$\begin{aligned}\hat{\mathbf{t}}^{+}_{\mathbf{yr}*} &= \mathbf{K}\,\hat{\mathbf{t}}^{-}_{\mathbf{yr}*} \\ &= \left(\mathbf{I} + \left[\mathbf{0} \,\vdots\, \mathbf{A}\right]\right)\begin{bmatrix}\hat{\mathbf{t}}^{-}_{\mathbf{y}*} \\ \mathbf{r}_* \end{bmatrix} \\ &= \begin{bmatrix}\mathbf{I} & \mathbf{A}_\mathbf{y} \\ \mathbf{0} & \left(\mathbf{I} + \mathbf{A}_\mathbf{r}\right)\end{bmatrix}\begin{bmatrix}\hat{\mathbf{t}}^{-}_{\mathbf{y}*} \\ \mathbf{r}_* \end{bmatrix} \\ &= \begin{bmatrix}\hat{\mathbf{t}}^{-}_{\mathbf{y}*} + \mathbf{A}_\mathbf{y}\mathbf{r}_* \\ \mathbf{r}_* + \mathbf{A}_\mathbf{r}\mathbf{r}_*\end{bmatrix}\end{aligned} \tag{17}$$

The $M\times J$ matrix partition $\mathbf{A}_\mathbf{y}$ contains the weights for the M study variables, and $J\times J$ matrix partition $\mathbf{A}_\mathbf{r}$ contains the weights for the J auxiliary residuals. The objective of the current Section is derivation of the $(M+J)\times J$ matrix of minimum-variance coefficients for matrix **A**, which Särndal *et al.* (1992:240) regard as "optimum."

The corresponding $(M+J)\times(M+J)$ covariance matrix for the linear transformation in Equation (17) is simply

$$\begin{aligned}\hat{\mathbf{V}}^{+}_{\mathbf{yr}*} &= \mathbf{K}\hat{\mathbf{V}}^{-}_{\mathbf{yr}*}\mathbf{K}^\mathrm{T} \\ &= \begin{bmatrix}\mathbf{I} & \mathbf{A}_\mathbf{y} \\ \mathbf{0} & \left(\mathbf{I} + \mathbf{A}_\mathbf{r}\right)\end{bmatrix}\begin{bmatrix}\hat{\mathbf{V}}^{-}_{\mathbf{y}*} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}^\mathrm{T}_* & \mathbf{\Lambda}_*\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{A}^\mathrm{T}_\mathbf{y} & \left(\mathbf{I} + \mathbf{A}^\mathrm{T}_\mathbf{r}\right)\end{bmatrix} \\ &= \begin{bmatrix}\left(\hat{\mathbf{V}}^{-}_{\mathbf{y}*} + \mathbf{\Gamma}_*\mathbf{A}^\mathrm{T}_\mathbf{y}\right) + \mathbf{A}_\mathbf{y}\left(\mathbf{\Gamma}^\mathrm{T}_* + \mathbf{\Lambda}_*\mathbf{A}^\mathrm{T}_\mathbf{y}\right) & \mathbf{\Gamma}_*\left(\mathbf{I} + \mathbf{A}^\mathrm{T}_\mathbf{r}\right) + \mathbf{A}_\mathbf{y}\mathbf{\Lambda}_*\left(\mathbf{I} + \mathbf{A}^\mathrm{T}_\mathbf{r}\right) \\ \left(\mathbf{I} + \mathbf{A}_\mathbf{r}\right)\left(\mathbf{\Gamma}^\mathrm{T}_* + \mathbf{\Lambda}_*\mathbf{A}^\mathrm{T}_\mathbf{y}\right) & \left(\mathbf{I} + \mathbf{A}_\mathbf{r}\right)\mathbf{\Lambda}_*\left(\mathbf{I} + \mathbf{A}^\mathrm{T}_\mathbf{r}\right)\end{bmatrix}\end{aligned} \tag{18}$$

where symmetric partitions $\mathbf{\Lambda} = \mathbf{\Lambda}^\mathrm{T}$ and $\mathbf{I} = \mathbf{I}^\mathrm{T}$.

Computations of minimum-variance coefficients for matrix **K** treat the realized vector of auxiliary residuals $\mathbf{r}_*$ in Equation (17) and covariance matrix partitions $\mathbf{\Gamma}_*$ and $\mathbf{\Lambda}_*$ in Equation (18) as known constants.

Separate the covariance matrix for the new multivariate difference estimator in Equation (18) into its matrix components

$$\begin{aligned}
\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*} \end{array}\right] + \left[\begin{array}{c|c} \boldsymbol{\Gamma}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\Gamma}_{*}^{\mathrm{T}} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \boldsymbol{\Gamma}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*} + \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \\ \hline \boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} + \mathbf{A}_{\mathbf{r}}\boldsymbol{\Gamma}_{*}^{\mathrm{T}} + \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} + \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*} + \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \end{array}\right] \\
&= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*} \end{array}\right] + \left(\left[\begin{array}{c|c} \boldsymbol{\Gamma}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \boldsymbol{\Gamma}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \\ \hline \boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \end{array}\right] + \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{y}}\boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*} \\ \hline \mathbf{A}_{\mathbf{r}}\boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*} \end{array}\right] + \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{A}_{\mathbf{y}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \\ \hline \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{A}_{\mathbf{r}}\boldsymbol{\Lambda}_{*}\mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \end{array}\right] \right) \\
&= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*} \end{array}\right] + \left(\left[\begin{array}{c} \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Lambda}_{*} \end{array}\right] \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \end{array}\right] + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*} \end{array}\right] + \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] \boldsymbol{\Lambda}_{*} \left[\begin{array}{c|c} \mathbf{A}_{\mathbf{y}}^{\mathrm{T}} & \mathbf{A}_{\mathbf{r}}^{\mathrm{T}} \end{array}\right] \right)
\end{aligned} \tag{19}$$

In order to simplify notation for the derivation of the minimum-variance GMDe coefficients, temporarily define $(M+J)\times J$ partition $\mathbf{Q}$ of covariance matrix $\hat{\mathbf{V}}^{-}$ in Equation (11) as

$$\begin{aligned}
\mathbf{Q} &= \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Gamma}_{*}^{\mathrm{T}} & \boldsymbol{\Lambda}_{*} \end{array}\right] \left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{I} \end{array}\right] \\
&= \left[\begin{array}{c} \boldsymbol{\Gamma}_{*} \\ \hline \boldsymbol{\Lambda}_{*} \end{array}\right]
\end{aligned} \tag{20}$$

Matrix $\mathbf{Q}$ is treated as a matrix of known constants, *i.e.*, π-estimates of population parameters from the realized probability sample. Recall the $(M+J)\times J$ matrix of arbitrary coefficients in Equations (17) and (18)

$$\mathbf{A} = \left[\begin{array}{c} \mathbf{A}_{\mathbf{y}} \\ \hline \mathbf{A}_{\mathbf{r}} \end{array}\right] \tag{21}$$

Substitute simplified notation from Equations (20) and (21) into Equation (19)

$$\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} = \hat{\mathbf{V}}_{\mathbf{yr}*}^{-} + \left(\mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{A}\boldsymbol{\Lambda}_{*}\mathbf{A}^{\mathrm{T}} \right) \tag{22}$$

$\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}$ is the prior π-estimate for the $(M+J)\times(M+J)$ covariance matrix, and $\hat{\mathbf{V}}_{\mathbf{yr}*}^{+}$ is the GMDe covariance matrix. Given that matrices $\mathbf{Q}$ and $\boldsymbol{\Lambda}$ are treated as known constants, the covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}*}^{+}$ for GMDe equals the covariance matrix for the prior estimator $\hat{\mathbf{V}}_{\mathbf{yr}*}^{+}$ plus a $(M+J)\times(M+J)$ matrix of constants $\left(\mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{A}\boldsymbol{\Lambda}_{*}\mathbf{A}^{\mathrm{T}} \right)$ where $\mathbf{A}$ is the $(M+J)\times J$ matrix of GMDe coefficients. The minimum-variance GMDe coefficients for matrix $\mathbf{A}$ are chosen so that the diagonal of matrix $\left(\mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{A}\boldsymbol{\Lambda}_{*}\mathbf{A}^{\mathrm{T}} \right)$ is composed of the most negative values possible with a linear transformation.

Derivation of the minimum-variance coefficients for matrix $\mathbf{A}$ in Equation (22) closely follows that for the univariate difference estimator given by Särndal *et al.* (1992:239-242). First, assume

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

the partition of the covariance matrix for the auxiliary variables $\mathbf{\Lambda}_* = \hat{\mathbf{\Lambda}}_{\mathbf{yr}*}$ is full rank and positive definite; therefore, its matrix inverse $\mathbf{\Lambda}_*^{-1}$ exists. Define $(M+J)\times(M+J)$ zero matrix **0** as

$$\mathbf{0} = \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) - \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) \tag{23}$$

Insert the zero matrix from Equation (23) into Equation (22) and rearrange

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \hat{\mathbf{V}}^{-} + \left(\mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right) + \left\{-\left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right)\right\} \\ &= \left(\hat{\mathbf{V}}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\end{aligned} \tag{24}$$

Since $\mathbf{\Lambda}_*$ is a symmetric covariance matrix, its inverse is also symmetric; and therefore, $\left(\mathbf{\Lambda}_*^{-1}\right)^{\mathrm{T}} = \mathbf{\Lambda}_*^{-1}$. Insert identity matrix $\left(\mathbf{I} = \mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_*\right)$ into Equation (24) and rearrange

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \left(\hat{\mathbf{V}}_{\mathbf{yr}*}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \mathbf{A}\mathbf{Q}^{\mathrm{T}} + \left[\mathbf{Q}\left\{\mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_*\right\}\mathbf{A}^{\mathrm{T}}\right] + \mathbf{A}\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}} \\ &= \left(\hat{\mathbf{V}}_{\mathbf{yr}*}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left[\left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\mathbf{Q}^{\mathrm{T}} + \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right] + \left[\mathbf{A}\mathbf{Q}^{\mathrm{T}} + \mathbf{A}\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right] \\ &= \left(\hat{\mathbf{V}}_{\mathbf{yr}*}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\left(\mathbf{Q}^{\mathrm{T}} + \mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right) + \mathbf{A}\left(\mathbf{Q}^{\mathrm{T}} + \mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right) \\ &= \left(\hat{\mathbf{V}}_{\mathbf{yr}*}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left(\mathbf{Q}\mathbf{\Lambda}_*^{-1} + \mathbf{A}\right)\left(\mathbf{Q}^{\mathrm{T}} + \mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}}\right)\end{aligned} \tag{25}$$

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

$$\begin{aligned}\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \left(\hat{\mathbf{V}}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\left\{\mathbf{\Lambda}_*\mathbf{\Lambda}_*^{-1}\right\}\left(\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}} + \mathbf{Q}^{\mathrm{T}}\right) \\ &= \left(\hat{\mathbf{V}}^{-} - \mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) + \left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Lambda}_*\left(\mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_*\mathbf{A}^{\mathrm{T}} + \mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right) \\ &= \left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right] + \left[-\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right] + \left[\left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Lambda}_*\left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)^{\mathrm{T}}\right]\end{aligned} \tag{26}$$

The diagonal elements of the second term in the last expression in Equation (26) are all negative values because $\left[-\mathbf{Q}\mathbf{\Lambda}_*^{-1}\mathbf{Q}^{\mathrm{T}}\right]$ is nonpositive definite; therefore, the second term decreases variances on the diagonal of the prior covariance matrix $\left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right]$. The third term $\left[\left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Lambda}_*\left(\mathbf{A} + \mathbf{Q}\mathbf{\Lambda}_*^{-1}\right)^{\mathrm{T}}\right]$ is positive definite, which increases variances on the diagonal of the prior covariance matrix $\left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right]$. Only this third term includes the matrix of arbitrary coefficients **A**. Therefore, the minimum-variance estimator uses normalized values for the matrix of arbitrary coefficients **A**

$$\mathbf{A}_{\text{opt}} = -\mathbf{Q}\boldsymbol{\Lambda}_*^{-1} \tag{27}$$

To illustrate, substitute $\mathbf{A} = \mathbf{A}_{\text{opt}}$ from Equation (27) into Equation (26)

$$\begin{aligned}
\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right] - \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\mathbf{Q}^{\text{T}} + \left(\mathbf{A}_{\text{opt}} + \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\right)\boldsymbol{\Lambda}_*\left(\mathbf{A}_{\text{opt}} + \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\right)^{\text{T}} \\
&= \left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right] - \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\mathbf{Q}^{\text{T}} + \left(-\mathbf{Q}\boldsymbol{\Lambda}_*^{-1} + \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\right)\boldsymbol{\Lambda}_*\left(-\mathbf{Q}\boldsymbol{\Lambda}_*^{-1} + \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\right)^{\text{T}} \\
&= \left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right] - \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\mathbf{Q}^{\text{T}} + \mathbf{0}\boldsymbol{\Lambda}_*\mathbf{0} \\
&= \left[\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right] - \mathbf{Q}\boldsymbol{\Lambda}_*^{-1}\mathbf{Q}^{\text{T}}
\end{aligned} \tag{28}$$

Express the matrix of minimum-variance coefficients $\mathbf{A}_{*\text{opt}} = -\mathbf{Q}\boldsymbol{\Lambda}_*^{-1}$ in Equation (27) with notation before simplification in Equations (20) and (21)

$$\mathbf{A}_{*\text{opt}} = -\mathbf{Q}\left(\boldsymbol{\Lambda}_*^{-}\right)^{-1} = \begin{bmatrix} \mathbf{A}_{\mathbf{y},\text{opt}} \\ \hdashline \mathbf{A}_{\mathbf{r},\text{opt}} \end{bmatrix} = -\begin{bmatrix} \boldsymbol{\Gamma}_* \\ \hdashline \boldsymbol{\Lambda}_*^{-} \end{bmatrix}\left(\boldsymbol{\Lambda}_*^{-}\right)^{-1} = \begin{bmatrix} -\boldsymbol{\Gamma}_*\left(\boldsymbol{\Lambda}_*^{-}\right)^{-1} \\ \hdashline -\mathbf{I} \end{bmatrix} \tag{29}$$

Therefore, the minimum-variance $(M+J)\times(M+J)$ matrix $\mathbf{K}_{\text{opt}}$ for GMDe is

$$\mathbf{K}_{\text{opt}} = \left[\begin{array}{c:c} \mathbf{I} & \mathbf{A}_{\mathbf{y},\text{opt}} \\ \hdashline \mathbf{0} & \left(\mathbf{I} + \mathbf{A}_{\mathbf{r},\text{opt}}\right) \end{array}\right] = \left[\begin{array}{c:c} \mathbf{I} & -\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*^{-1} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right] \tag{30}$$

The GMDe for the population totals with the minimum-variance coefficients follows from Equations (17) and (30)

$$\begin{aligned}
\hat{\mathbf{t}}_{\mathbf{yr}*,\text{opt}}^{+} &= \mathbf{K}_{\text{opt}}\hat{\mathbf{t}}_{\mathbf{yr}*}^{-} \\
&= \left[\begin{array}{c:c} \mathbf{I} & -\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*^{-1} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right]\begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}*}^{-} \\ \hdashline \hat{\mathbf{t}}_{\mathbf{r}*}^{-} \end{bmatrix} \\
&= \begin{bmatrix} \hat{\mathbf{t}}_{\mathbf{y}*}^{-} - \left(\boldsymbol{\Gamma}_*\boldsymbol{\Lambda}_*^{-1}\right)\mathbf{r}_* \\ \hdashline \mathbf{0} \end{bmatrix}
\end{aligned} \tag{31}$$

At this point, the $J\times1$ partition $\hat{\mathbf{t}}_{\mathbf{r}*,\text{opt}}^{-} = \mathbf{0}$ for the auxiliary residuals might seem perfunctory; however, the purpose of that partition becomes apparent with the recursive GMDe in Section 7 (page 22).

The $M\times1$ partition of $\hat{\mathbf{t}}_{\mathbf{yr}*}^{+}$ for population estimates of the study variables $\hat{\mathbf{t}}_{\mathbf{y}*}^{+}$ is

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\hat{\mathbf{t}}_{\mathbf{y}*}^{+} = \hat{\mathbf{t}}_{\mathbf{y}*}^{-} - \left(\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\right)\mathbf{r}_* \tag{32}$$

Note that elements of the $M \times 1$ vector of adjustments $\left(\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\right)\mathbf{r}_*$ can be either positive or negative depending upon the signs of the corresponding auxiliary residual in vector $\mathbf{r}_*$ and correlations in partition $\mathbf{\Gamma}_*$.

Substitute the $(M+J)\times J$ matrix of minimum-variance coefficients from Equation (29) into the estimator for GMDe covariance matrix in Equation (18)

$$\begin{aligned}
\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} + \left(\begin{bmatrix} \mathbf{\Gamma}_* \\ \mathbf{\Lambda}_* \end{bmatrix} \mathbf{A}_{\mathrm{opt}}^{\mathrm{T}} + \mathbf{A}_{\mathrm{opt}} \begin{bmatrix} \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} + \mathbf{A}_{\mathrm{opt}}\mathbf{\Lambda}_*\mathbf{A}_{\mathrm{opt}}^{\mathrm{T}} \right) \\
&= \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} + \left(\begin{bmatrix} \mathbf{\Gamma}_* \\ \mathbf{\Lambda}_* \end{bmatrix} \begin{bmatrix} -\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{I} \end{bmatrix} + \begin{bmatrix} -\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1} \\ -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} + \begin{bmatrix} -\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1} \\ -\mathbf{I} \end{bmatrix} \mathbf{\Lambda}_* \begin{bmatrix} -\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{I} \end{bmatrix} \right) \\
&= \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} + \left(\begin{bmatrix} -\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{\Gamma}_* \\ -\left(\mathbf{\Lambda}_*\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{\Lambda}_* \end{bmatrix} + \begin{bmatrix} -\mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{\Gamma}_*\left(\mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_*\right) \\ -\mathbf{\Gamma}_*^{\mathrm{T}} & -\mathbf{\Lambda}_* \end{bmatrix} + \begin{bmatrix} \mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\left(\mathbf{\Lambda}_*\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Gamma}_*\left(\mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_*\right) \\ \left(\mathbf{\Lambda}_*\mathbf{\Lambda}_*^{-1}\right)\mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} \right) \\
&= \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} - \begin{bmatrix} \mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Gamma}_* \\ \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{bmatrix} \\
&= \begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} - \mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
\end{aligned} \tag{33}$$

Notice from Equations (31) and (33) that the linear transformation of the vector of auxiliary residuals exactly equals zero with the minimum-variance coefficients. In a heuristic sense, GMDe with minimum-variance coefficients extracts all relevant information from the auxiliary variables and transfers that information to population estimates for the study variables. The $M\times M$ partition in Equation (33) for population estimates of the study variables in Equation (32) is

$$\hat{\mathbf{V}}_{\mathbf{y}*}^{+} = \hat{\mathbf{V}}_{\mathbf{y}*}^{-} - \mathbf{\Gamma}_*\mathbf{\Lambda}_*^{-1}\mathbf{\Gamma}_*^{\mathrm{T}} \tag{34}$$

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Equation (32) and (34) agree with a special case considered by Särndal *et al.* (1992:240) and Montanari (1987), in which M=1 and the population totals for all auxiliary (*i.e.*, proxy) variables are known constants, *i.e.*, the covariance matrix in Equation (10) equals

$$\hat{\mathbf{V}}_{\mathbf{yr}\pi} = \left[\begin{array}{c:c:c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & -\mathbf{I} & \mathbf{I} \end{array}\right] \left[\begin{array}{c:c:c} \hat{\mathbf{V}}_{\mathbf{y}H\pi} & \hat{\mathbf{V}}_{\mathbf{yx}H\pi} & \mathbf{0} \\ \hdashline \hat{\mathbf{V}}_{\mathbf{yx}H\pi}^{\mathrm{T}} & \hat{\mathbf{V}}_{\mathbf{x}H\pi} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c:c} \mathbf{I} & \mathbf{0} \\ \hdashline \mathbf{0} & -\mathbf{I} \\ \hdashline \mathbf{0} & \mathbf{I} \end{array}\right]$$

$$= \left[\begin{array}{c:c} \hat{\mathbf{V}}_{\mathbf{y}\pi} & \mathbf{\Gamma}_{\mathbf{yr}\pi} \\ \hdashline \mathbf{\Gamma}_{\mathbf{yr}\pi}^{\mathrm{T}} & \mathbf{\Lambda}_{\mathbf{yr}\pi} \end{array}\right] \quad \text{where} \quad \begin{cases} \mathbf{\Gamma} = \mathbf{\Gamma}_{\mathbf{yr}\pi} = -\hat{\mathbf{V}}_{\mathbf{yx}H\pi} \\ \mathbf{\Lambda} = \mathbf{\Lambda}_{\mathbf{yr}\pi} = \hat{\mathbf{V}}_{\mathbf{x}H\pi} \end{cases} \tag{35}$$

To be clear, Equation (35) applies only to the aforementioned special case, which is the foundational assumption for conventional model-assisted regression estimators.

7 BIVARIATE EXAMPLE

The following considers the special case of the bivariate difference estimator, in which two auxiliary residuals (J=2) reduce variance of population estimates for two study variables (M=2). The purpose is two-fold: gain intuitive insights into GMDe; and illustrate the recursive version of GMDe introduced by Czaplewski (2020).

The following bivariate example of the difference estimator uses notation for the normalized vector of population estimates from Equation (12)

$$\hat{\mathbf{t}}_{\mathbf{yr}*} = \left[\begin{array}{c} \hat{t}_{i*} \\ \hat{t}_{k*} \\ \hdashline r_{1*} \\ r_{2*} \end{array}\right] = \mathbf{S}^{-1}\left(\hat{\mathbf{t}}_{\mathbf{yr}} - \hat{\mathrm{E}}\left[\hat{\mathbf{t}}_{\mathbf{yr}}\right]\right) = \left[\begin{array}{c} \hat{t}_i^- \big/ \sqrt{\hat{v}_i} \\ \hat{t}_k^- \big/ \sqrt{\hat{v}_k} \\ \hdashline r_1^- \big/ \sqrt{\Lambda_1} \\ r_2^- \big/ \sqrt{\Lambda_2} \end{array}\right]$$

$$= \left[\begin{array}{cc:cc} 1\big/\sqrt{\hat{v}_i} & 0 & 0 & 0 \\ 0 & 1\big/\sqrt{\hat{v}_k} & 0 & 0 \\ \hdashline 0 & 0 & 1\big/\sqrt{\Lambda_1} & 0 \\ 0 & 0 & 0 & 1\big/\sqrt{\Lambda_2} \end{array}\right] \left(\left[\begin{array}{c} \hat{t}_i^- \\ \hat{t}_k^- \\ \hdashline r_1^- \\ r_2^- \end{array}\right] - \left[\begin{array}{c} \hat{t}_i^- \\ \hat{t}_k^- \\ \hdashline 0 \\ 0 \end{array}\right] \right) \tag{36}$$

The diagonal nominalization matrix $\mathbf{S}^{-1}$ uses the prior π-variances for the two study variables ($\hat{v}_i^-$ and $\hat{v}_k^-$), which are on the leading diagonal of covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}}$ in Equation (11); and the trailing diagonal elements $(\Lambda_1 \text{ and } \Lambda_2)$ of $\mathbf{S}^{-1}$, which are the transformed π-variances for the two auxiliary residuals on the trailing diagonal of covariance matrix $\hat{\mathbf{V}}_{\mathbf{yr}}$ from Equation (10).

The normalized covariance matrix for the prior population estimates in Equation (11) is the correlation matrix

$$\hat{\mathbf{V}}_{\mathbf{yr}*} = \mathbf{S}^{-1}\,\hat{\mathbf{V}}_{\mathbf{yr}}^{-}\,\mathbf{S}^{-1}$$
$$= \left[\begin{array}{cc|cc} \hat{v}_i^{-}\big/\sqrt{\hat{v}_i^{-}\hat{v}_i^{-}} & \hat{v}_{ik}^{-}\big/\sqrt{\hat{v}_i^{-}\hat{v}_k^{-}} & \hat{\Gamma}_{ih}^{-}\big/\sqrt{\hat{v}_i^{-}\Lambda_1^{-}} & \hat{\Gamma}_{ig}^{-}\big/\sqrt{\hat{v}_i^{-}\Lambda_2^{-}} \\ \hat{v}_{ki}^{-}\big/\sqrt{\hat{v}_k^{-}\hat{v}_i^{-}} & \hat{v}_k^{-}\big/\sqrt{\hat{v}_k^{-}\hat{v}_k^{-}} & \hat{\Gamma}_{kh}^{-}\big/\sqrt{\hat{v}_k^{-}\Lambda_1^{-}} & \hat{\Gamma}_{kg}^{-}\big/\sqrt{\hat{v}_k^{-}\Lambda_2^{-}} \\ \hline \hat{\Gamma}_{1i}^{-}\big/\sqrt{\Lambda_1^{-}\hat{v}_i^{-}} & \hat{\Gamma}_{1k}^{-}\big/\sqrt{\Lambda_1^{-}\hat{v}_k^{-}} & \Lambda_1^{-}\big/\sqrt{\Lambda_1^{-}\Lambda_1^{-}} & \hat{\Gamma}_{12}^{-}\big/\sqrt{\Lambda_1^{-}\Lambda_2^{-}} \\ \hat{\Gamma}_{2i}^{-}\big/\sqrt{\Lambda_2^{-}\hat{v}_i^{-}} & \hat{\Gamma}_{2k}^{-}\big/\sqrt{\Lambda_2^{-}\hat{v}_k^{-}} & \hat{\Gamma}_{21}^{-}\big/\sqrt{\Lambda_2^{-}\Lambda_1^{-}} & \Lambda_2^{-}\big/\sqrt{\Lambda_2^{-}\Lambda_2^{-}} \end{array}\right]$$
$$= \left[\begin{array}{cc|cc} 1 & \hat{\rho}_{ik}^{-} & \hat{\rho}_{i1}^{-} & \hat{\rho}_{i2}^{-} \\ \hat{\rho}_{ik}^{-} & 1 & \hat{\rho}_{k1}^{-} & \hat{\rho}_{k2}^{-} \\ \hline \hat{\rho}_{i1}^{-} & \hat{\rho}_{k1}^{-} & 1 & \hat{\rho}_{12}^{-} \\ \hat{\rho}_{i2}^{-} & \hat{\rho}_{k2}^{-} & \hat{\rho}_{12}^{-} & 1 \end{array}\right] \tag{37}$$

where element $-1 \le \hat{\rho}_{12}^{-} \le 1$ is the estimated correlation between auxiliary residuals r_1 and r_2 in the prior normalized vector of population estimates $\hat{\mathbf{t}}_{\mathbf{yr}*}$ in Equation (36), and all normalized variances on the diagonal equal 1.

7.1 "BATCH" GMDE ESTIMATOR

Express the normalized prior π-correlation matrix $\left(\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}\right)_1$ in Equation (37) with the partition notation in Equation (11)

$$\hat{\mathbf{V}}_{\mathbf{yr}*}^{-} = \left[\begin{array}{cc|cc} 1 & \hat{\rho}_{ik}^{-} & \hat{\rho}_{i1}^{-} & \hat{\rho}_{i2}^{-} \\ \hat{\rho}_{ik}^{-} & 1 & \hat{\rho}_{k1}^{-} & \hat{\rho}_{k2}^{-} \\ \hline \hat{\rho}_{i1}^{-} & \hat{\rho}_{k1}^{-} & 1 & \hat{\rho}_{12}^{-} \\ \hat{\rho}_{i2}^{-} & \hat{\rho}_{k2}^{-} & \hat{\rho}_{12}^{-} & 1 \end{array}\right] = \left[\begin{array}{c|c} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} & \mathbf{\Gamma}_* \\ \hline \mathbf{\Gamma}_*^{\mathrm{T}} & \mathbf{\Lambda}_* \end{array}\right] \quad \text{where} \left\{\begin{array}{l} \hat{\mathbf{V}}_{\mathbf{y}*}^{-} = \left[\begin{array}{c|c} 1 & \hat{\rho}_{ik}^{-} \\ \hline \hat{\rho}_{ik}^{-} & 1 \end{array}\right] \\ \mathbf{\Lambda}_* = \left[\begin{array}{c|c} 1 & \hat{\rho}_{12}^{-} \\ \hline \hat{\rho}_{12}^{-} & 1 \end{array}\right] \\ \mathbf{\Gamma}_* = \left[\begin{array}{c|c} \hat{\rho}_{i1}^{-} & \hat{\rho}_{i2}^{-} \\ \hline \hat{\rho}_{k1}^{-} & \hat{\rho}_{k2}^{-} \end{array}\right] \end{array}\right. \tag{38}$$

Recall that $\hat{\mathbf{V}}_{\mathbf{y}*}^{-}$ is the correlation matrix between normalized π-estimates for the two study variables; $\mathbf{\Lambda}_*$ is the normalized π-correlation matrix between π-estimates for the two auxiliary residuals, where $\mathbf{\Lambda}_*$ is assumed to be full-rank and well-conditioned, and therefore $\mathbf{\Lambda}_*^{-1}$ exists. $\mathbf{\Lambda}_*$; and $\mathbf{\Gamma}_*$ contains the correlations between normalized π-estimates for the study variables and the auxiliary residuals.

The matrix inverse $\mathbf{\Lambda}_*^{-1}$ of correlation partition $\mathbf{\Lambda}_*$ for the two auxiliary residuals equals

$$\mathbf{\Lambda}_*^{-1} = \frac{1}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\left[\begin{array}{c:c} 1 & -\hat{\rho}_{12}^{-} \\ \hdashline -\hat{\rho}_{12}^{-} & 1 \end{array}\right] \tag{39}$$

where $\mathbf{\Lambda}_*^{-1}\mathbf{\Lambda}_* = \mathbf{I}$. The sufficient partition of the minimum-variance coefficient matrix $\mathbf{A}_{*\text{opt}}$ in Equation (29) equals

$$\begin{aligned}
\mathbf{A}_{*\text{opt}} &= -\left[\begin{array}{c} \mathbf{\Gamma}_* \\ \hdashline \mathbf{\Lambda}_*^{-} \end{array}\right]\left(\mathbf{\Lambda}_*^{-}\right)^{-1} \\
&= \left[\begin{array}{c:c} -\hat{\rho}_{i1}^{-} & -\hat{\rho}_{i2}^{-} \\ \hdashline -\hat{\rho}_{k1}^{-} & -\hat{\rho}_{k2}^{-} \\ \hdashline -1 & -\hat{\rho}_{12}^{-} \\ \hdashline -\hat{\rho}_{12}^{-} & -1 \end{array}\right]\left(\frac{1}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\left[\begin{array}{c:c} 1 & -\hat{\rho}_{12}^{-} \\ \hdashline -\hat{\rho}_{12}^{-} & 1 \end{array}\right]\right) \\
&= \frac{1}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\left[\begin{array}{c:c} -\hat{\rho}_{i1}^{-} + \hat{\rho}_{i2}^{-}\hat{\rho}_{12}^{-} & \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-} - \hat{\rho}_{i2}^{-} \\ \hdashline -\hat{\rho}_{k1}^{-} + \hat{\rho}_{k2}^{-}\hat{\rho}_{12}^{-} & \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-} - \hat{\rho}_{k2}^{-} \\ \hdashline -1+\left(\hat{\rho}_{12}^{-}\right)^2 & \hat{\rho}_{12}^{-} - \hat{\rho}_{12}^{-} \\ \hdashline \hat{\rho}_{12}^{-} - \hat{\rho}_{12}^{-} & -1+\left(\hat{\rho}_{12}^{-}\right)^2 \end{array}\right] \\
&= \left[\begin{array}{c:c} \left(\dfrac{-\hat{\rho}_{i1}^{-} + \hat{\rho}_{i2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ \hdashline \left(\dfrac{-\hat{\rho}_{k1}^{-} + \hat{\rho}_{k2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ \hdashline -1 & 0 \\ \hdashline 0 & -1 \end{array}\right]
\end{aligned} \tag{40}$$

The minimum-variance coefficient matrix $\mathbf{K}_{\text{opt}}$ from Equations (30) and (40) in this example is

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\mathbf{K}_{\text{opt}} = \mathbf{I} + \left[\mathbf{0} \;\vdots\; \mathbf{A}_{*\text{opt}}\right]$$
$$= \left[\begin{array}{cc|cc} 1 & 0 & \left(\dfrac{-\hat{\rho}_{i1}^{-} + \hat{\rho}_{i2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ 0 & 1 & \left(\dfrac{-\hat{\rho}_{k1}^{-} + \hat{\rho}_{k2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{41}$$

GMDe for the vector of population estimates $\hat{\mathbf{t}}_{\mathbf{yr}*}^{+}$ in Equation (17) with minimum-variance coefficient matrix $\mathbf{K}_{\text{opt}}$ in Equation (41) and the normalized vector of prior π-estimates $\hat{\mathbf{t}}_{\mathbf{yr}*}^{-}$ in Equation (36) equals

$$\hat{\mathbf{t}}_{\mathbf{yr}*}^{+} = \mathbf{K}_{\text{opt}}\,\hat{\mathbf{t}}_{\mathbf{yr}*}^{-} = \left[\begin{array}{cc|cc} 1 & 0 & \left(\dfrac{-\hat{\rho}_{i1}^{-} + \hat{\rho}_{i2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ 0 & 1 & \left(\dfrac{-\hat{\rho}_{k1}^{-} + \hat{\rho}_{k2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]^{-} \left[\begin{array}{c} \hat{t}_{i*} \\ \hat{t}_{k*} \\ \hline r_{1*} \\ r_{2*} \end{array}\right]^{-} = \left[\begin{array}{c} \hat{t}_{i*} + \dfrac{r_{1*} - \hat{\rho}_{12} r_{2*}}{1-\hat{\rho}_{12}^2} \\ \hat{t}_{k*} + \dfrac{r_{2*} - \hat{\rho}_{12} r_{1*}}{1-\hat{\rho}_{12}^2} \\ \hline 0 \\ 0 \end{array}\right]^{-} \tag{42}$$

Recall the "–" superscript denotes the prior π-estimate, and superscript "+" denotes the GMDe population estimate.

The covariance matrix for GMDe in Equation (18) with minimum-variance coefficient matrix $\mathbf{K}_{\text{opt}}$ in Equation (41) and correlation matrix $\hat{\mathbf{V}}_{\mathbf{yr}*}^{-}$ in Equation (37) equals

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$
\begin{aligned}
\hat{\mathbf{V}}_{\mathbf{yr}*}^{+} &= \mathbf{K}_{\mathrm{opt}} \hat{\mathbf{V}}_{\mathbf{yr}*}^{-} \mathbf{K}_{\mathrm{opt}}^{\mathrm{T}} \\
&= \hat{\mathbf{V}}_{\mathbf{y}*}^{-} + \mathbf{A}_{\mathbf{y}} \boldsymbol{\Gamma}_{*}^{\mathrm{T}} + \boldsymbol{\Gamma}_{*} \mathbf{A}_{\mathbf{y}}^{\mathrm{T}} + \mathbf{A}_{\mathbf{y}} \boldsymbol{\Lambda}_{*} \mathbf{A}_{\mathbf{y}}^{\mathrm{T}} \\
&= \left\{
\begin{array}{l}
\begin{bmatrix} 1 & \hat{\rho}_{ik}^{-} \\ \hat{\rho}_{ik}^{-} & 1 \end{bmatrix} + \\[2ex]
+\left(\dfrac{2}{1-\hat{\rho}_{12}^{2}}\right)
\begin{bmatrix}
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right)\hat{\rho}_{i1} + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right)\hat{\rho}_{i2} \end{array}\right) &
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{k1}\hat{\rho}_{12}-\hat{\rho}_{i1}\hat{\rho}_{k1}\right) + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{k2}\hat{\rho}_{12}-\hat{\rho}_{i2}\hat{\rho}_{k2}\right) \end{array}\right) \\
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{k1}\hat{\rho}_{12}-\hat{\rho}_{i1}\hat{\rho}_{k1}\right) + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{k2}\hat{\rho}_{12}-\hat{\rho}_{i2}\hat{\rho}_{k2}\right) \end{array}\right) &
\left(\begin{array}{l} \left(\hat{\rho}_{k2}\hat{\rho}_{12}-\hat{\rho}_{k1}\right)\hat{\rho}_{k1} + \\ \left(\hat{\rho}_{k1}\hat{\rho}_{12}-\hat{\rho}_{k2}\right)\hat{\rho}_{k2} \end{array}\right)
\end{bmatrix} + \\[2ex]
+\left(\dfrac{1}{1-\hat{\rho}_{12}^{2}}\right)^{2}
\begin{bmatrix}
\begin{array}{l}
\left(\begin{array}{l} \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right) + \\ \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{i1}\hat{\rho}_{12} + \\ -\hat{\rho}_{i2} \end{array}\right) + \\
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right) + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{i2}\hat{\rho}_{12} + \\ -\hat{\rho}_{i1} \end{array}\right)
\end{array} &
\begin{array}{l}
\left(\begin{array}{l} \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right) + \\ \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k1}\hat{\rho}_{12} + \\ -\hat{\rho}_{k2} \end{array}\right) + \\
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right) + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k2}\hat{\rho}_{12} + \\ -\hat{\rho}_{k1} \end{array}\right)
\end{array} \\
\begin{array}{l}
\left(\begin{array}{l} \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right) + \\ \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k1}\hat{\rho}_{12} + \\ -\hat{\rho}_{k2} \end{array}\right) + \\
\left(\begin{array}{l} \left(\hat{\rho}_{i2}\hat{\rho}_{12}-\hat{\rho}_{i1}\right) + \\ \left(\hat{\rho}_{i1}\hat{\rho}_{12}-\hat{\rho}_{i2}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k2}\hat{\rho}_{12} + \\ -\hat{\rho}_{k1} \end{array}\right)
\end{array} &
\begin{array}{l}
\left(\begin{array}{l} \left(\hat{\rho}_{k1}\hat{\rho}_{12}-\hat{\rho}_{k2}\right) + \\ \left(\hat{\rho}_{k2}\hat{\rho}_{12}-\hat{\rho}_{k1}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k1}\hat{\rho}_{12} + \\ -\hat{\rho}_{k2} \end{array}\right) + \\
\left(\begin{array}{l} \left(\hat{\rho}_{k2}\hat{\rho}_{12}-\hat{\rho}_{k1}\right) + \\ \left(\hat{\rho}_{k1}\hat{\rho}_{12}-\hat{\rho}_{k2}\right)\hat{\rho}_{12}^{-} \end{array}\right)
\left(\begin{array}{l} \hat{\rho}_{k2}\hat{\rho}_{12} + \\ -\hat{\rho}_{k1} \end{array}\right)
\end{array}
\end{bmatrix}
\end{array}
\right\}
\end{aligned}
\tag{43}
$$

7.2 RECURSIVE GMD ESTIMATOR

Czaplewski (2020) introduced the recursive GMDe to solve numerical challenges with the "batch" version GMDe in Sections 6 (page 16) and 7.1 (page 23). The primary challenge is the matrix inverse of the correlation matrix $\boldsymbol{\Lambda}_{*}^{-1}$ in Equations (23) and (27) for the auxiliary residuals. A numerically reliable matrix inverse exists only if partition $\boldsymbol{\Lambda}_{*}$ is full-rank and well-conditioned. Rank-deficient matrices become more common as the number of auxiliary variables J increases. Model-assisted estimators face a similar challenge because they require a matrix inverse in the regression model to predict the proxy variable. In the current example, J=2, and matrix $\boldsymbol{\Lambda}_{*}$ is always full-rank. Regardless, the following section provides an introduction to the recursive GMDe, and it demonstrates that the recursive GMDe ultimately produces the same estimates as does the regular GMDe, at least for the bivariate case.

The first GMDe recursion uses a single auxiliary residual to reduce π-variance of all M study variable and remove any collinearity between the remaining (J−1) auxiliary residuals and the M study variables. The second recursion uses the second auxiliary residual to further reduce variance of the first GMDe recursion for those same *M* study variables; and so forth, until the final J^{th} recursion with the J^{th} auxiliary residual. Each recursion uses a scalar inverse rather than a matrix inverse. If the $J\times J$ correlation matrix among the *J* auxiliary residuals is rank deficient, then the recursion halts if the scalar inverse is infeasible, *i.e.*, division by zero in Equations (48), (49), (59), and (60) below. Details follow.

7.2.1 GMDe Recursion #1

The bivariate example starts with the transformation matrix of arbitrary coefficients $(\mathbf{K})_1$, where subscript "1" designates the first GMDe recursion with auxiliary residual $\left(r_{1*}^{-}\right)_1$ in this example.

$$\left(\hat{\mathbf{t}}_*^{+}\right)_1 = (\mathbf{K})_1 \left(\hat{\mathbf{t}}_*^{-}\right)_1 \quad \text{where } (\mathbf{K})_1 = \mathbf{I} + \left[\mathbf{0} \quad \mathbf{0} \;\vdots\; \mathbf{a}_1 \;\vdots\; \mathbf{0}\right] \qquad \begin{bmatrix} \hat{t}_{i*}^{+} \\ \hat{t}_{k*}^{+} \\ \hline r_{1*}^{+} \\ r_{2*}^{+} \end{bmatrix}_1 = \begin{bmatrix} 1 & 0 & a_{i1} & 0 \\ 0 & 1 & a_{k1} & 0 \\ \hline 0 & 0 & 1+a_1 & 0 \\ 0 & 0 & a_{21} & 1 \end{bmatrix}_1 \begin{bmatrix} \hat{t}_{i*}^{-} \\ \hat{t}_{k*}^{-} \\ \hline r_{1*}^{-} \\ r_{2*}^{-} \end{bmatrix}_1 = \begin{bmatrix} \hat{t}_{i*}^{-} + a_{i1} \\ \hat{t}_{k*}^{-} + a_{k1} \\ \hline r_{1*}^{-} + a_1 \\ r_{2*}^{-} + a_{21} \end{bmatrix}_1 \tag{44}$$

Equation (44) illustrates several properties of the recursive GMDe:

- The second auxiliary residual $\left(r_2^{-}\right)_1$ in this first recursion does not directly reduce variances of the estimates for the two study variables $\left(\hat{t}_i^{-}\right)_1$ and $\left(\hat{t}_k^{-}\right)_1$. In that sense, the second auxiliary residual is treated in the same way as the two study variables during this first recursion.
- The recursive GMDe can easily impose inequality constraints on population estimates for study variables through inequality constraints on GMDe coefficients. For example, assume through *a priori* knowledge that normalized population parameter t_{1*} is bounded by $t_{1*,\min} \le t_{1*} \le t_{1*,\max}$; and further assume those conditions are true for the prior π-estimates. The range of consistent GMDe coefficients is $\left(a_{i1,\min} \le a_{i1} \le a_{i1,\max}\right)_1$ where $\left(a_{i1,\min}\right)_1 = t_{1*,\min} - \left(\hat{t}_{i*}^{-}\right)_1$ and $\left(a_{i1,\max}\right)_1 = t_{1*,\max} - \left(\hat{t}_{i*}^{-}\right)_1$. See Czaplewski (2020:58-60) for details.

The corresponding linear transformation of covariance matrix $\left(\hat{\mathbf{V}}_*^-\right)_1$ in Equation (37) is

$$\left(\hat{\mathbf{V}}_*^+\right)_1 = \left[\begin{array}{c:c:c:c} a_{i1}^2\hat{\rho}_1^- + 2a_{i1}\hat{\rho}_{i1}^- + \hat{\rho}_i^- & a_{i1}a_{k1}\hat{\rho}_1^- + a_{k1}\hat{\rho}_{i1}^- + a_{i1}\hat{\rho}_{k1}^- + \hat{\rho}_{ik}^- & (1+a_1)\left(a_{i1}\hat{\rho}_1^- + \hat{\rho}_{i1}^-\right) & a_{i1}a_{21}\hat{\rho}_1^- + a_{21}\hat{\rho}_{i1}^- + a_{i1}\hat{\rho}_{12}^- + \hat{\rho}_{i2}^- \\ \hdashline a_{k1}a_{i1}\hat{\rho}_1^- + a_{i1}\hat{\rho}_{k1}^- + a_{k1}\hat{\rho}_{i1}^- + \hat{\rho}_{ik}^- & a_{k1}^2\hat{\rho}_1^- + 2a_{k1}\hat{\rho}_{k1}^- + \hat{\rho}_k^- & (1+a_1)\left(a_{k1}\hat{\rho}_1^- + \hat{\rho}_{k1}^-\right) & a_{k1}a_{21}\hat{\rho}_1^- + a_{21}\hat{\rho}_{k1}^- + a_{k1}\hat{\rho}_{12}^- + \hat{\rho}_{k2}^- \\ \hdashline (1+a_1)\left(a_{i1}\hat{\rho}_1^- + \hat{\rho}_{i1}^-\right) & (1+a_1)\left(a_{k1}\hat{\rho}_1^- + \hat{\rho}_{k1}^-\right) & a_1^2\hat{\rho}_1^- + 2a_1\hat{\rho}_1^- + \hat{\rho}_1^- & (1+a_1)\left(a_{21}\hat{\rho}_1^- + \hat{\rho}_{12}^-\right) \\ \hdashline a_{21}a_{i1}\hat{\rho}_1^- + a_{i1}\hat{\rho}_{12}^- + a_{21}\hat{\rho}_{i1}^- + \hat{\rho}_{i2}^- & a_{21}a_{k1}\hat{\rho}_1^- + a_{k1}\hat{\rho}_{12}^- + a_{21}\hat{\rho}_{k1}^- + \hat{\rho}_{k2}^- & (1+a_1)\left(a_{21}\hat{\rho}_1^- + \hat{\rho}_{12}^-\right) & a_{21}^2\hat{\rho}_1^- + 2a_{21}\hat{\rho}_{12}^- + \hat{\rho}_2^- \end{array}\right]_1 \tag{45}$$

Recall the normalization in Equation (37) and substitute $\left(\hat{\rho}_i^-\right)_1 = \left(\hat{\rho}_k^-\right)_1 = \left(\hat{\rho}_1^-\right)_1 = \left(\hat{\rho}_2^-\right)_1 = 1$ into Equation (45)

$$\left(\hat{\mathbf{V}}_*^+\right)_1 = \left[\begin{array}{c:c:c:c} a_{i1}^2 + 2a_{i1}\hat{\rho}_{i1}^- + 1 & a_{i1}a_{k1} + a_{k1}\hat{\rho}_{i1}^- + a_{i1}\hat{\rho}_{k1}^- + \hat{\rho}_{ik}^- & (1+a_1)\left(a_{i1} + \hat{\rho}_{i1}^-\right) & a_{i1}a_{21} + a_{21}\hat{\rho}_{i1}^- + a_{i1}\hat{\rho}_{12}^- + \hat{\rho}_{i2}^- \\ \hdashline a_{k1}a_{i1} + a_{i1}\hat{\rho}_{k1}^- + a_{k1}\hat{\rho}_{i1}^- + \hat{\rho}_{ik}^- & a_{k1}^2 + 2a_{k1}\hat{\rho}_{k1}^- + 1 & (1+a_1)\left(a_{k1} + \hat{\rho}_{k1}^-\right) & a_{k1}a_{21} + a_{21}\hat{\rho}_{k1}^- + a_{k1}\hat{\rho}_{12}^- + \hat{\rho}_{k2}^- \\ \hdashline (1+a_1)\left(a_{i1} + \hat{\rho}_{i1}^-\right) & (1+a_1)\left(a_{k1} + \hat{\rho}_{k1}^-\right) & a_1^2 + 2a_1 + 1 & (1+a_1)\left(a_{21} + \hat{\rho}_{12}^-\right) \\ \hdashline a_{21}a_{i1} + a_{i1}\hat{\rho}_{12}^- + a_{21}\hat{\rho}_{i1}^- + \hat{\rho}_{i2}^- & a_{21}a_{k1} + a_{k1}\hat{\rho}_{12}^- + a_{21}\hat{\rho}_{k1}^- + \hat{\rho}_{k2}^- & (1+a_1)\left(a_{21} + \hat{\rho}_{12}^-\right) & a_{21}^2 + 2a_{21}\hat{\rho}_{12}^- + 1 \end{array}\right]_1 \tag{46}$$

7.2.1.1 Variances with Arbitrary Coefficients

Särndal *et al.* (1992:240) use the minimum-variance criterion as the "optimal choice" for the arbitrary coefficients. In this example, this criterion means selection of coefficients a_{i1} and a_{k1} that minimize the variances of the study variables in the first recursion; those variances are the leading diagonal elements $\left(\hat{v}_{i*}^+\right)_1 = \left(a_{i1}^2 + 2a_{i1}\hat{\rho}_{1i}^- + 1\right)_1$ and $\left(\hat{v}_{k*}^+\right)_1 = \left(a_{k1}^2 + 2a_{k1}\hat{\rho}_{1k}^- + 1\right)_1$ in Equation (46). The difference estimators for these two population statistics are quadratic functions of the transformation coefficients $\left(a_{i1}\right)_1$ and $\left(a_{k1}\right)_1$ in the first recursion. Figure 2 illustrates variance reduction $\left(\hat{v}_{k*}^+\right)_1 / \left(\hat{v}_{k*}^-\right)_1$ for study variable *i* during the first recursion as a function of arbitrary GMDe coefficient $\left(a_{i1}\right)_1$ at seven different levels of correlation $\left(\hat{\rho}_{1i}^-\right)_1$ between the first auxiliary residual and the study variable

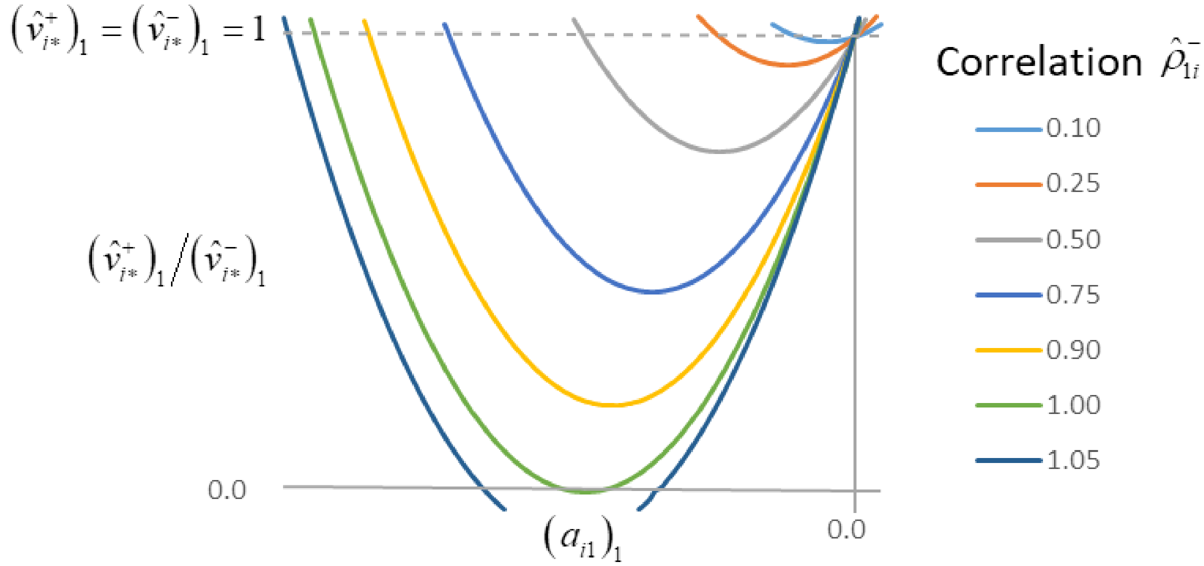


Figure 2. Variance reduction $(\hat{v}_{k*}^{+})_1 / (\hat{v}_{k*}^{-})_1$ as a function of GMDe coefficient $(a_{i1})_1$ and degree of correlation between a study variable and the auxiliary residual $(\hat{\rho}_{1i}^{-})_1$.

Figure 2 illustrates several properties of the recursive GMDe

- GMDe variance reduction is minor if the correlation between a study variable and the auxiliary residual is small. Because of random errors, prior estimates of the correlation can be nonzero, even if that correlation is truly zero. If the number of auxiliary variables is large, then the cumulative effect of insignificantly small correlations can be large. Särndal *et al.* (1992:250) and Czaplewski (2020:68) recommend that correlation $(\hat{\rho}_{1i}^{-})_1 > 0.5$ to avoid "overfitting" with chance correlations. An alternative criterion may be selected given Section 7.2.1.5 (page 33)
- GMDe can reduce variance for a population estimate of a study variable even if the minimum-variance value of GMDe coefficient $(a_{i1})_1$ does not equal its minimum-variance value. Czaplewski (2020:53-64) uses a suboptimal value of GMDe coefficient $(a_{i1})_1$ to impose inequality constraints on population estimates, mitigate likely outliers, and preclude negative estimates of variances. The variance of the constrained GMDe estimate can be nearly as efficient as the minimum-variance estimate given the quadratic relationships in Figure 2.
- GMDe can actually increase variance of the estimate of a study variable if GMDe coefficient $(a_{i1})_1$ is outside a certain range of values.
- GMDe variance is positive only if the correlation $(\hat{\rho}_{1i}^{-})_1$ between the auxiliary residual and the study variable is bounded by ±1.0. Therefore, Czaplewski (2020) precludes

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

negative variance estimates through inequality constraints on GMDe estimates of correlation coefficient $\left(\hat{\rho}_{1i}^{+}\right)_1$.

7.2.1.2 Minimum-variance Coefficients for Study Variables

The minimum-variance values within the first recursion of GMDe are determined by solving for $\left(a_{i1,\text{opt}}\right)_1$ and $\left(a_{k1,\text{opt}}\right)_1$ such that their first derivatives equal zero in Figure 2.

$$\begin{aligned} 0 &= \frac{d}{d\left(a_{i1}\right)}\left(a_{i1}^2\hat{\rho}_1^- + 2a_{i1}\hat{\rho}_{i1}^- + \hat{\rho}_i^-\right)_1 \\ &= \left(2a_{i1}\hat{\rho}_1^- + \hat{\rho}_{i1}^-\right)_1 \\ \left(a_{i1,\text{opt}}\right)_1 &= \left(\frac{-\hat{\rho}_{1i}^-}{\hat{\rho}_1^-}\right)_1 \end{aligned} \qquad \begin{aligned} 0 &= \frac{d}{d\left(a_{k1}\right)}\left(a_{k1}^2\hat{\rho}_1^- + 2a_{k1}\hat{\rho}_{k1}^- + \hat{\rho}_k^-\right)_1 \\ &= \left(2a_{k1}\hat{\rho}_1^- + 2\hat{\rho}_{k1}^-\right)_1 \\ \left(a_{k1,\text{opt}}\right)_1 &= \left(\frac{-\hat{\rho}_{1k}^-}{\hat{\rho}_1^-}\right)_1 \end{aligned} \qquad (47)$$

Recall the normalization in Equation (37), and substitute $\hat{\rho}_i^- = 1$ into Equation (47)

$$\begin{aligned} 0 &= \frac{d}{d\left(a_{i1}\right)}\left(a_{i1}^2 + 2a_{i1}\hat{\rho}_{1i}^- + 1\right)_1 \\ &= \left(2a_{i1} + 2\hat{\rho}_{1i}^-\right)_1 \\ \left(a_{i1,\text{opt}}\right)_1 &= \left(-\hat{\rho}_{1i}^-\right)_1 \end{aligned} \qquad \begin{aligned} 0 &= \frac{d}{d\left(a_{k1}\right)}\left(a_{k1}^2 + 2a_{k1}\hat{\rho}_{1k}^- + 1\right)_1 \\ &= \left(2a_{k1} + 2\hat{\rho}_{1k}^-\right)_1 \\ \left(a_{k1,\text{opt}}\right)_1 &= \left(-\hat{\rho}_{1k}^-\right)_1 \end{aligned} \qquad (48)$$

The minimum-variance coefficients in Equation (48) for the first recursion use only the first auxiliary residual $\left(r_1^-\right)_1$ to minimize variances of both study variables plus and remove any collinearity between the first and second auxiliary residuals (see Section 7.2.1.3 below). The second recursion (page 35) uses second auxiliary residual $\left(r_2^-\right)_2 = \left(r_2^+\right)_1$ for the same purpose.

7.2.1.3 Orthogonal Coefficients for Auxiliary Residuals

However, the two auxiliary residuals are not independent if their covariance is nonzero, *i.e.*, $\left|\hat{\rho}_{21}^-\right|_1 > 0$. To avoid collinearity in subsequent recursions, the second auxiliary residual must be orthogonal to the GMDe of the first auxiliary residual after the first recursion and before the second recursion. Therefore, select coefficients $\left(a_{11,\text{opt}}\right)_1$ and $\left(a_{21,\text{opt}}\right)_1$ so that the auxiliary residuals are mutually orthogonal, *i.e.*, $\left(\hat{\rho}_{21}^+\right)_1 = \left(1 + a_{11,\text{opt}}\right)_1\left(a_{21,\text{opt}} + \hat{\rho}_{12}^-\right)_1 = 0$ in Equation (46)

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\left(1+a_{11}\right)_1 = 0 \qquad \left(a_{21}\hat{\rho}_1^- + \hat{\rho}_{12}^-\right)_1 = 0$$

$$\left(a_{11,\text{opt}}\right)_1 = -1 \qquad \left(a_{21,\text{opt}}\right)_1 = \frac{-\hat{\rho}_{12}^-}{\hat{\rho}_1^-}$$

$$\text{for } \left(\hat{\rho}_{21}^+\right)_1 = \left(1+a_{11,\text{opt}}\right)_1 \left(a_{21}\hat{\rho}_1^- + \hat{\rho}_{12}^-\right)_1 = 0 \tag{49}$$

Recall the normalization in Equation (37), and substitute $\hat{\rho}_i^- = 1$ into Equation (49)

$$\left(1+a_{11}\right)_1 = 0 \qquad \left(a_{21} + \hat{\rho}_{12}^-\right)_1 = 0$$

$$\left(a_{11,\text{opt}}\right)_1 = -1 \qquad \left(a_{21,\text{opt}}\right)_1 = -\hat{\rho}_{12}^-$$

$$\begin{aligned} \text{for } \left(\hat{\rho}_{21}^+\right)_1 &= \left(1+a_{11,\text{opt}}\right)_1 \left(a_{21}\hat{\rho}_1^- + \hat{\rho}_{12}^-\right)_1 = 0 \\ \left(\hat{\rho}_{21}^+\right)_1 &= \left(1+a_{11,\text{opt}}\right)_1 \left(a_{21} + \hat{\rho}_{12}^-\right)_1 = 0 \quad \text{where normalized } \left(\hat{\rho}_1^-\right)_1 = 1 \end{aligned} \tag{50}$$

Given the minimum-variance coefficients in Equation (48) and orthogonal coefficients in Equation (50), the covariances between the first auxiliary residual and the remaining elements all equal zero after the first recursion

$$\begin{aligned} \left(\hat{\rho}_{i1}^+\right)_1 &= \begin{pmatrix} \hat{\rho}_{i1}^- + a_{i1,\text{opt}} a_{11,\text{opt}} + \\ +\hat{\rho}_{i1}^- a_{11,\text{opt}} + a_{i1,\text{opt}} \end{pmatrix}_1 \\ &= \begin{pmatrix} \hat{\rho}_{i1}^- + \left(-\hat{\rho}_{1i}^-\right)(-1) + \\ +\hat{\rho}_{i1}^-(-1) + \left(-\hat{\rho}_{1i}^-\right) \end{pmatrix}_1 \\ &= \left(2\hat{\rho}_{i1}^- - 2\hat{\rho}_{i1}^-\right)_1 \\ &= 0 \end{aligned} \qquad \begin{aligned} \hat{\rho}_{k2}^+ &= \begin{pmatrix} \hat{\rho}_{k1}^- + \left(a_{k1,\text{opt}}\right)\left(a_{11,\text{opt}}\right) + \\ +\hat{\rho}_{k1}^-\left(a_{11,\text{opt}}\right) + \left(a_{k1,\text{opt}}\right) \end{pmatrix} \\ &= \begin{pmatrix} \hat{\rho}_{k1}^- + \left(-\hat{\rho}_{k1}^-\right)(-1) + \\ +\hat{\rho}_{k1}^-(-1) + \left(-\hat{\rho}_{k1}^-\right) \end{pmatrix} \\ &= 2\hat{\rho}_{k1}^- - 2\hat{\rho}_{k1}^- \\ &= 0 \end{aligned} \tag{51}$$

7.2.1.4 GMDe with Minimum-variance Coefficients

Insert the four minimum-variance coefficients in this example into the minimum-variance GMDe transformation matrix $\left(\mathbf{K}_{\text{opt}}\right)_1$

$$\left(\mathbf{K}_{*\text{opt}}\right)_1 = \left[\begin{array}{cc:cc} 1 & 0 & a_{i1,\text{opt}} & 0 \\ 0 & 1 & a_{k1,\text{opt}} & 0 \\ \hdashline 0 & 0 & \left(1+a_{1,\text{opt}}\right) & 0 \\ 0 & 0 & a_{21,\text{opt}} & 1 \end{array}\right]_1 = \left[\begin{array}{cc:cc} 1 & 0 & -\hat{\rho}_{i1}^- & 0 \\ 0 & 1 & -\hat{\rho}_{k1}^- & 0 \\ \hdashline 0 & 0 & 0 & 0 \\ 0 & 0 & -\hat{\rho}_{12}^- & 1 \end{array}\right]_1 \tag{52}$$

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

GMDe with minimum-variance coefficients after the first recursion with auxiliary residual $\left(r_1^-\right)_1$ equals Equation (44) with the linear transformation matrix $\left(\mathbf{K}_{\text{opt}}\right)_1$ in Equation (52)

$$\begin{aligned}\left(\hat{\mathbf{t}}_{*\text{opt}}^{+}\right)_1 &= \left(\mathbf{K}_{*\text{opt}}\right)_1 \left(\hat{\mathbf{t}}^{-}\right)_1 \\ &= \left[\begin{array}{cc|cc} 1 & 0 & -\hat{\rho}_{i1}^{-} & 0 \\ 0 & 1 & -\hat{\rho}_{k1}^{-} & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & -\hat{\rho}_{12}^{-} & 1 \end{array}\right]_1 \left[\begin{array}{c} \hat{t}_{i*}^{-} \\ \hat{t}_{k*}^{-} \\ \hline r_{1*}^{-} \\ r_{2*}^{-} \end{array}\right]_1 \\ &= \left[\begin{array}{c} \hat{t}_{i*}^{-} - \hat{\rho}_{1i}^{-}\, r_{1*}^{-} \\ \hat{t}_{k*}^{-} - \hat{\rho}_{k1}^{-}\, r_{1*}^{-} \\ \hline 0 \\ r_{2*}^{-} - \hat{\rho}_{12}^{-}\, r_{1*}^{-} \end{array}\right]_1 \end{aligned} \tag{53}$$

The first recursion of GMDe in Equation (53) with the first auxiliary residual and the minimum-variance coefficients in Equation (48) reveals several outcomes:

- The recursive GMDe estimates for the normalized study variables with the first auxiliary residual and after the first recursion are affected by:
  - The strength of the correlations $\left(\hat{\rho}_{i1}^-\right)_1$ and $\left(\hat{\rho}_{k1}^-\right)_1$ between the study variables and the first auxiliary residual $\left(r_{1*}^-\right)_1$;
  - The value of the first auxiliary residual $\left(r_{1*}^-\right)_1$, which can be either positive or negative.
- The recursive GMDe estimate of the first auxiliary residual $\left(r_1^+\right)_1 = 0$ after first recursion, which means it can no longer affect subsequent recursions. In other words, the minimum-variance GMDe coefficients in Equation (48), by design, remove collinearity between the first auxiliary residual and the remaining auxiliary residual.
- The recursive GMDe estimate for the second auxiliary residual $\left(r_2^+\right)_1$ contracts towards zero, the magnitude of which depends upon the strength of the correlation between the two auxiliary residuals. This is consistent with the previous outcome: the minimum-variance GMDe coefficients in Equation (48) remove collinearity between the first auxiliary residual and the remaining auxiliary residuals during subsequent recursions. In a heuristic sense, the recursive GMDe performs a sequential orthogonalization.
- The minimum-variance estimates can use either negative or positive correlations.
- GMDe coefficient matrix $(\mathbf{K})_1$ is sparse and highly structured. Therefore, optimization of matrix routines can substantially improve GMDe numerics and computational speed.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

The covariance matrix for the minimum-variance GMDe after the first recursion with auxiliary residual $\left(r_1^-\right)_1$ equals Equation (46) with the linear transformation matrix $\left(\mathbf{K}_{\text{opt}}\right)_1$ in Equation (52)

$$
\begin{aligned}
\left(\hat{\mathbf{V}}_{*\text{opt}}^+\right)_1 &= \left(\mathbf{K}_{*\text{opt}}\right)_1 \left(\hat{\mathbf{V}}_*^-\right)_1 \left(\mathbf{K}_{*\text{opt}}\right)_1^{\text{T}} \\
&= \begin{bmatrix}
1-\left(\hat{\rho}_{i1}^-\right)^2 & \hat{\rho}_{ik}^- - \hat{\rho}_{i1}^-\hat{\rho}_{k1}^- & 0 & \hat{\rho}_{i2}^- - \hat{\rho}_{i1}^-\hat{\rho}_{12}^- \\
\hat{\rho}_{ik}^- - \hat{\rho}_{i1}^-\hat{\rho}_{k1}^- & 1-\left(\hat{\rho}_{k1}^-\right)^2 & 0 & \hat{\rho}_{k2}^- - \hat{\rho}_{k1}^-\hat{\rho}_{12}^- \\
0 & 0 & 0 & 0 \\
\hat{\rho}_{i2}^- - \hat{\rho}_{i1}^-\hat{\rho}_{12}^- & \hat{\rho}_{k2}^- - \hat{\rho}_{k1}^-\hat{\rho}_{12}^- & 0 & 1-\left(\hat{\rho}_{12}^-\right)^2
\end{bmatrix}_1 \\
\left(\hat{v}_{*\text{opt}}^+\right)_{i1} &= 1-\left(\hat{\rho}_{i1}^-\right)^2 \\
\left(\hat{v}_{*\text{opt}}^+\right)_{k1} &= 1-\left(\hat{\rho}_{k1}^-\right)^2
\end{aligned}
\tag{54}
$$

GMDe variance estimator in Equation (54) demonstrates several other outcomes with the minimum-variance coefficients in Equation (48).

- Variance reduction after the first recursion of GMDe is related to the strength of the correlation $\left(\hat{\rho}_{1i}^-\right)_1$ and $\left(\hat{\rho}_{1k}^-\right)_1$ between the study variables and the first auxiliary residual
- GMDe of the first auxiliary residual is a constant (zero variance) with GMDe; therefore, the other auxiliary residuals are independent of the first auxiliary residual during subsequent recursions.
- After the first GMDe recursion, correlation between a study variable and the second auxiliary residual is reduced (assuming $\hat{\rho}_{i1}^-\hat{\rho}_{12}^- > 0$), which supports previous observations that the recursive GMDe fully compensates for collinearities among auxiliary residuals.

Covariance matrix $\left(\hat{\mathbf{V}}_{*\text{opt}}^+\right)_1$ is no longer a correlation matrix because variances on its diagonal do not equal 1.0. For purpose of numerical accuracy, renormalize vector $\left(\mathbf{S}_{*\text{opt}}^{-1}\hat{\mathbf{t}}_{*\text{opt}}^+\right)_1 \rightarrow \left(\hat{\mathbf{t}}_*^-\right)_2$ and covariance matrix $\left(\mathbf{S}_{*\text{opt}}^{-1}\hat{\mathbf{V}}_{*\text{opt}}^+\mathbf{S}_{*\text{opt}}^{-1}\right)_1 \rightarrow \left(\hat{\mathbf{V}}_*^-\right)_2$ as in Equations (11) and (12). Recursive normalization maintains dynamic range of GMDe variances and standard deviations into the interval $\{-1,1\}$, which can improved numerics. The following section on Recursion #2 omits that renormalization to simplify notation.

7.2.1.5 Variance Reduction

The essential purpose of GMDe is reduction of variances in population estimates for the study variables by “gaining strength” from population estimates for auxiliary variables. Equation (54) illustrates the strength of variance reduction with the leading two diagonal elements from the GMDe covariance matrix

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\frac{\hat{v}^{+}_{i*\text{opt}}}{\hat{v}^{-}_{i*}} = 1-\left(\hat{\rho}^{-}_{i1}\right)^{2} \qquad \frac{\hat{v}^{+}_{k*\text{opt}}}{\hat{v}^{-}_{k*}} = 1-\left(\hat{\rho}^{-}_{k1}\right)^{2} \tag{55}$$

where $\hat{v}^{-}_{i*} = \hat{v}^{-}_{k*} = 1$ after initial normalization (Section 5, page 14) and $\hat{\rho}^{-}_{i1}$ and $\hat{\rho}^{-}_{k1}$ are the correlations between the first auxiliary residual and the i^{th} and k^{th} study variables respectively. Table 1 quantifies normalized variance reduction with GMDe as a function of the correlations between the study variables and the auxiliary residual, *i.e.*, $\hat{\rho}^{-}_{i1}$ and $\hat{\rho}^{-}_{k1}$.

Table 1 Normalized GMDe variance as a function of correlation between population estimates for the study variables and the auxiliary residual. For example, if the correlation is 0.50, then GMDe variance is 75% and standard deviation 87% that of that with the prior π-estimate (*e.g.*, HT).

Correlation	Normalized GMDe Estimates	
$\hat{\rho}^{-}_{i1}$ and $\hat{\rho}^{-}_{ik}$	Variance	Standard Deviation
0.00	1.00	1.00
0.10	0.99	0.99
0.20	0.96	0.98
0.30	0.91	0.95
0.40	0.84	0.92
0.50	0.75	0.87
0.60	0.64	0.80
0.70	0.51	0.71
0.80	0.36	0.60
0.90	0.19	0.44
1.00	0.00	0.00

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Figure 3 illustrates the relationships in Table 1

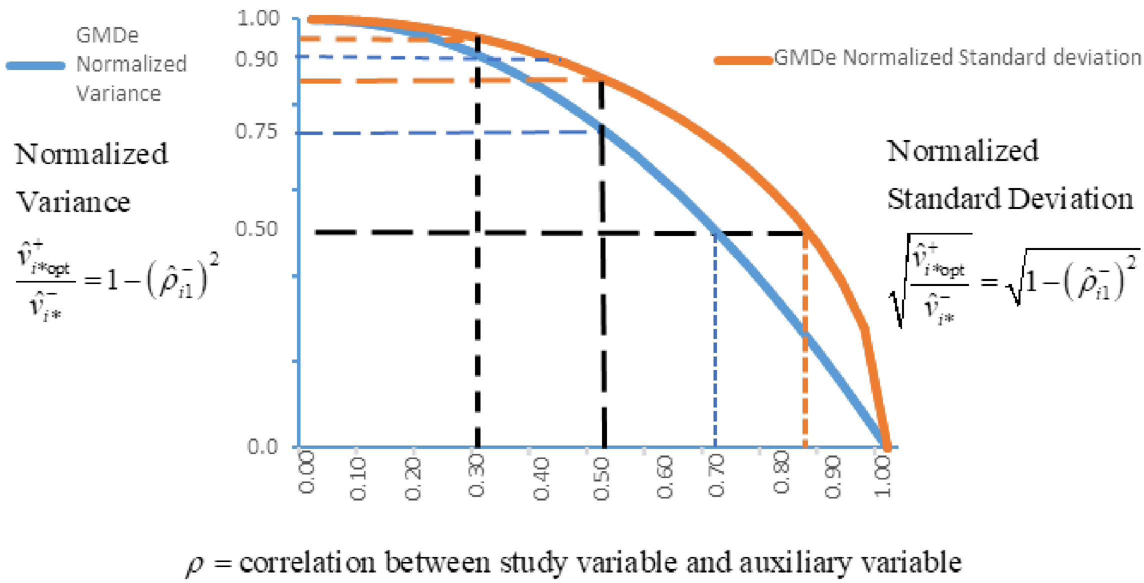


Figure 3 Normalized GMDe variance (*i.e.*, statistical efficiency) and standard deviation (*i.e.*, confidence interval) as a function of correlation between population estimates for the study variables and the auxiliary residual.

Table 1 and Figure 3 offer several observations:

- Variance reduction with GMDe is less than 10% if the correlation between population estimate for the study variable and the auxiliary residual is $|\rho|<0.30$, and the reduction in standard deviation is even less.
- The recommendation of $|\rho|>0.50$ for minimum correlation by Särndal *et al.* (1992) and Czaplewski (2020) avoids undue influence from spurious correlations, but that recommendation forgoes GMDe variance reduction of 25% and reduction in standard deviation by 13%. This is a cautious recommendation, and a less risk adverse recommendation might be $|\rho|>0.30$.
- A 50% reduction or more with GMDe in variance for a population estimate of a study variable requires a relatively strong correlation of $|\rho|>0.70$, and a 50% reduction in standard deviation or more requires an even stronger correlation of $|\rho|=0.85$

7.2.2 GMDe Recursion #2

The second recursion begins with GMDe population estimates from the first recursion in Equations (53) and (54). In other words, the prior estimate for the second recursion is the final GMDe from the first recursion.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

The GMDe for the 4×1 vector of population statistics with the second auxiliary residual and arbitrary coefficients $\begin{bmatrix} a_{i2} & a_{k2} & \vdots & a_{12} & a_{22} \end{bmatrix}_2^{\mathrm{T}}$ corresponds to that with the first auxiliary residual in Equation (44)

$$\begin{aligned}\left(\hat{\mathbf{t}}_*^+\right)_2 &= \mathbf{K}_2\left(\hat{\mathbf{t}}_*^-\right)_2 \\ &= \mathbf{K}_2\left(\hat{\mathbf{t}}_*^+\right)_1 \\ &= \mathbf{K}_2\left[\mathbf{K}_1\left(\hat{\mathbf{t}}_*^-\right)_1\right]\end{aligned} \qquad \left(\hat{\mathbf{t}}_*^+\right)_2 = \mathbf{K}_2\left(\hat{\mathbf{t}}_*^-\right)_2 = \begin{bmatrix} \hat{t}_{i*}^+ \\ \hat{t}_{k*}^+ \\ r_{1*}^+ \\ r_{2*}^+ \end{bmatrix}_2 = \begin{bmatrix} 1 & 0 & 0 & a_{i2} \\ 0 & 1 & 0 & a_{k2} \\ 0 & 0 & 1 & a_{12} \\ 0 & 0 & 0 & 1+a_{22} \end{bmatrix}_2 \begin{bmatrix} \hat{t}_{i*}^- \\ \hat{t}_{k*}^- \\ r_{1*}^- \\ r_{2*}^- \end{bmatrix}_2 \tag{56}$$

with the corresponding covariance matrix as with Equation (46)

$$\begin{aligned}\left(\hat{\mathbf{V}}_*^-\right)_2 &= \mathbf{K}_2\left(\hat{\mathbf{V}}_*^-\right)_2\mathbf{K}_2^{\mathrm{T}} = \mathbf{K}_2\left(\hat{\mathbf{V}}_*^+\right)_1\mathbf{K}_2^{\mathrm{T}} = \mathbf{K}_2\left[\mathbf{K}_1\left(\hat{\mathbf{V}}_*^-\right)_1\mathbf{K}_1\right]\mathbf{K}_2 \\ &= \begin{bmatrix} 1 & 0 & 0 & a_{i2} \\ 0 & 1 & 0 & a_{k2} \\ 0 & 0 & 1 & a_{12} \\ 0 & 0 & 0 & 1+a_{22} \end{bmatrix}_2 \begin{bmatrix} \hat{\rho}_i^- & \hat{\rho}_{ik}^- & \hat{\rho}_{i1}^- & \hat{\rho}_{i2}^- \\ \hat{\rho}_{ik}^- & \hat{\rho}_k^- & \hat{\rho}_{k1}^- & \hat{\rho}_{k2}^- \\ \hat{\rho}_{i1}^- & \hat{\rho}_{k1}^- & \hat{\rho}_1^- & \hat{\rho}_{12}^- \\ \hat{\rho}_{i2}^- & \hat{\rho}_{k2}^- & \hat{\rho}_{12}^- & \hat{\rho}_2^- \end{bmatrix}_2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a_{i2} & a_{k2} & a_{12} & 1+a_{22} \end{bmatrix}_2\end{aligned} \tag{57}$$

Expand the 4×4 GMDe covariance matrix in Equation (57)

$$\begin{aligned}&\left(\hat{\mathbf{V}}_*^+\right)_2 = \mathbf{K}_2\left(\hat{\mathbf{V}}_*^-\right)_2\mathbf{K}_2^{\mathrm{T}} \\ &= \mathbf{A}_2\left(\hat{\mathbf{V}}_*^-\right)_2\mathbf{A}_2^{\mathrm{T}} + \left[\left(\hat{\mathbf{V}}_*^-\right)_2\mathbf{A}_2^{\mathrm{T}} + \mathbf{A}_2\left(\hat{\mathbf{V}}_*^-\right)_2\right] + \left(\hat{\mathbf{V}}_*^-\right)_2 \\ &\begin{bmatrix} a_{i2}^2\hat{\rho}_2^- + 2a_{i2}\hat{\rho}_{i2}^- + \hat{\rho}_i^- & \begin{gathered} a_{i2}a_{k2}\hat{\rho}_2^- + a_{i2}\hat{\rho}_{k2}^- \\ +a_{k2}\hat{\rho}_{i2}^- + \hat{\rho}_{ik}^- \end{gathered} & \begin{gathered} a_{i2}a_{12}\hat{\rho}_2^- + a_{i2}\hat{\rho}_{12}^- \\ +a_{12}\hat{\rho}_{i2}^- + \hat{\rho}_{i1}^- \end{gathered} & \left(1+a_2\right)\left(a_{i2}\hat{\rho}_2^- + \hat{\rho}_{i2}^-\right) \\ \begin{gathered} a_{k2}a_{i2}\hat{\rho}_2^- + a_{i2}\hat{\rho}_{k2}^- \\ +a_{k2}\hat{\rho}_{i2}^- + \hat{\rho}_{ik}^- \end{gathered} & a_{k2}^2\hat{\rho}_2^- + 2a_{k2}\hat{\rho}_{k2}^- + \hat{\rho}_k^- & \begin{gathered} a_{k2}a_{12}\hat{\rho}_2^- + a_{k2}\hat{\rho}_{12}^- \\ +a_{12}\hat{\rho}_{k2}^- + \hat{\rho}_{k1}^- \end{gathered} & \left(1+a_2\right)\left(a_{k2}\hat{\rho}_2^- + \hat{\rho}_{k2}^-\right) \\ \begin{gathered} a_{12} + a_{i2}\hat{\rho}_{12}^- \\ +a_{12}\hat{\rho}_{i2}^- + \hat{\rho}_{i1}^- \end{gathered} & \begin{gathered} a_{12}a_{k2}\hat{\rho}_2^- + a_{k2}\hat{\rho}_{12}^- \\ +a_{12}\hat{\rho}_{k2}^- + \hat{\rho}_{k1}^- \end{gathered} & a_{12}^2\hat{\rho}_2^- + 2a_{12}\hat{\rho}_{12}^- + \hat{\rho}_1^- & \left(1+a_2\right)\left(a_{12}\hat{\rho}_2^- + \hat{\rho}_{12}^-\right) \\ \left(1+a_2\right)\left(a_{i2}\hat{\rho}_2^- + \hat{\rho}_{i2}^-\right) & \left(1+a_2\right)\left(a_{k2}\hat{\rho}_2^- + \hat{\rho}_{k2}^-\right) & \left(1+a_2\right)\left(a_{12}\hat{\rho}_2^- + \hat{\rho}_{12}^-\right) & a_2^2\hat{\rho}_2^- + 2a_2\hat{\rho}_2^- + \hat{\rho}_2^- \end{bmatrix}_2\end{aligned} \tag{58}$$

7.2.2.1 Minimum-variance Coefficients for Study Variables

The minimum-variance values of the arbitrary coefficients in Equation (58) follow those in Equation (48) for the first auxiliary residual. Solve for the minimum-variance coefficients for the study variables by setting the first derivative of the variance estimator, *i.e.*, the leading two elements on the diagonal of GMDe covariance in Equation (58), to equal zero:

$$0 = \frac{d}{d\left(a_{i2}\right)_2}\left(a_{i2}^2\hat{\rho}_2^- + 2a_{i2}\hat{\rho}_{i2}^- + \hat{\rho}_i^-\right)_2 \qquad 0 = \frac{d}{d\left(a_{k2}\right)_2}\left(a_{k2}^2\hat{\rho}_2^- + 2a_{k2}\hat{\rho}_{k2}^- + \hat{\rho}_k^-\right)_2$$

$$= \left(2a_{i2}\hat{\rho}_2^- + 2\hat{\rho}_{i2}^-\right)_2 \qquad = \left(2a_{k2}\hat{\rho}_2^- + 2\hat{\rho}_{k2}^-\right)_2$$

$$\left(a_{i2,\text{opt}}\right)_2 = \left(\frac{-\hat{\rho}_{i2}^-}{\hat{\rho}_2^-}\right)_2 \qquad \left(a_{k2,\text{opt}}\right)_2 = \left(\frac{-\hat{\rho}_{k2}^-}{\hat{\rho}_2^-}\right)_2 \tag{59}$$

The minimum-variance coefficients in Equation (59) for Recursion #2 directly correspond to those in Equation (47) for Recursion #1, where $\left(\hat{\rho}_i^-\right)_1 = 1$, $\left(\hat{\rho}_k^-\right)_1 = 1$, and $\left(\hat{\rho}_2^-\right)_1 = 1$.

7.2.2.2 Orthogonal Coefficients for Auxiliary Residuals

To avoid collinearity, the second auxiliary residual must be orthogonal to the first auxiliary residual. Therefore, as in Equation (49), select coefficients $\left(a_{2,\text{opt}}\right)_2$ and $\left(a_{12,\text{opt}}\right)_2$ so that the auxiliary residuals are mutually orthogonal, *i.e.*, $\left(\hat{\rho}_{21}^+\right)_2 = \left(1 + a_{22}\right)_2\left(a_{12}\hat{\rho}_2^- + \hat{\rho}_{12}^-\right)_2 = 0$ in Equation (58):

$$\left(a_{12}\hat{\rho}_2^- + \hat{\rho}_{12}^-\right)_2 = 0 \qquad \left(1 + a_2\right)_2 = 0$$

$$\left(a_{12,\text{opt}}\right)_2 = \left(\frac{-\hat{\rho}_{12}^-}{\hat{\rho}_2^-}\right)_2 \qquad \left(a_{2,\text{opt}}\right)_2 = -1$$

$$\text{for } \left(\hat{\rho}_{21}^+\right)_2 = \left(1 + a_2\right)_2\left(a_{12}\hat{\rho}_2^- + \hat{\rho}_{12}^-\right)_2 = 0 \tag{60}$$

The orthogonal coefficients in Equation (60) for GMDe Recursion #2 directly correspond to those in Equation (48) for GMDe Recursion #1.

7.2.2.3 GMDe with Minimum-variance Coefficients

Insert the optimum coefficients from Equations (59) and (60) into the minimum-variance GMDe transformation matrix $\mathbf{K}_{2,\text{opt}}$

$$\mathbf{K}_{2,\text{opt}} = \left[\begin{array}{cc:cc} 1 & 0 & 0 & a_{i2,\text{opt}} \\ 0 & 1 & 0 & a_{k2,\text{opt}} \\ \hdashline 0 & 0 & 1 & a_{12,\text{opt}} \\ 0 & 0 & 0 & 1 + a_{2,\text{opt}} \end{array}\right]_2 = \left[\begin{array}{cc:cc} 1 & 0 & 0 & -\hat{\rho}_{i2}^-/\hat{\rho}_2^- \\ 0 & 1 & 0 & -\hat{\rho}_{k2}^-/\hat{\rho}_2^- \\ \hdashline 0 & 0 & 1 & .-\hat{\rho}_{12}^-/\hat{\rho}_2^- \\ 0 & 0 & 0 & 0 \end{array}\right]_2 \tag{61}$$

GMDe with the minimum-variance coefficients in Equation (61) and the vector estimate after Recursion #1 in Equation (56) equals

$$\left(\hat{\mathbf{t}}^{+}_{*\text{opt}}\right)_2 = \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{t}}^{-}\right)_2 = \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{t}}^{+}\right)_{1,\text{opt}} = \begin{bmatrix} \hat{t}^{-}_{i*} + \left(-\hat{\rho}^{-}_{i2}/\hat{\rho}^{-}_{2}\right) r^{-}_{2*} \\ \hat{t}^{-}_{k*} + \left(-\hat{\rho}^{-}_{k2}/\hat{\rho}^{-}_{2}\right) r^{-}_{2*} \\ \hdashline \hat{t}^{-}_{1*} + \left(-\hat{\rho}^{-}_{12}/\hat{\rho}^{-}_{2}\right) r^{-}_{2*} \\ 0 \end{bmatrix}_2 \tag{62}$$

The covariance matrix for GMDe in Equation (62) with the minimum-variance coefficients in Equation (61) equals

$$\left(\hat{\mathbf{V}}^{+}_{*\text{opt}}\right)_2 = \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{V}}^{-}_{*}\right)_2 \mathbf{K}^{\mathrm{T}}_{2,\text{opt}} = \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{V}}^{+}_{*}\right)_{1,\text{opt}} \mathbf{K}^{\mathrm{T}}_{2,\text{opt}} = \begin{bmatrix} \hat{\rho}^{-}_{i} - \frac{\left(\hat{\rho}^{-}_{i2}\right)^2}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{ik} - \frac{\hat{\rho}^{-}_{i2}\hat{\rho}^{-}_{k2}}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{i1} - \frac{\hat{\rho}^{-}_{i2}\hat{\rho}^{-}_{12}}{\hat{\rho}^{-}_{2}} & 0 \\ \hat{\rho}^{-}_{ik} - \frac{\hat{\rho}^{-}_{i2}\hat{\rho}^{-}_{k2}}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{k} - \frac{\left(\hat{\rho}^{-}_{k2}\right)^2}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{k1} - \frac{\hat{\rho}^{-}_{k2}\hat{\rho}^{-}_{12}}{\hat{\rho}^{-}_{2}} & 0 \\ \hdashline \hat{\rho}^{-}_{i1} - \frac{\hat{\rho}^{-}_{i2}\hat{\rho}^{-}_{12}}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{k1} - \frac{\hat{\rho}^{-}_{k2}\hat{\rho}^{-}_{12}}{\hat{\rho}^{-}_{2}} & \hat{\rho}^{-}_{1} - \frac{\left(\hat{\rho}^{-}_{12}\right)^2}{\hat{\rho}^{-}_{2}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_2 \tag{63}$$

This corresponds to the covariance matrix after GMDe Recursion #1 in Equation (54).

7.2.3 Combine Recursions #1 and #2

The two recursions in this example may be combined into a simple linear model:

$$\begin{aligned} \left(\hat{\mathbf{t}}^{+}_{*}\right)_{2,\text{opt}} &= \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{t}}^{-}_{*}\right)_2 & \left(\hat{\mathbf{V}}^{-}_{*}\right)_{2,\text{opt}} &= \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{V}}^{-}_{*}\right)_{2,\text{opt}} \mathbf{K}^{\mathrm{T}}_{2,\text{opt}} \\ &= \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{t}}^{+}_{*}\right)_{1,\text{opt}} & &= \mathbf{K}_{2,\text{opt}}\left(\hat{\mathbf{V}}^{+}_{*}\right)_{1,\text{opt}} \mathbf{K}^{\mathrm{T}}_{2,\text{opt}} \\ &= \mathbf{K}_{2,\text{opt}}\left[\mathbf{K}_{1,\text{opt}}\left(\hat{\mathbf{t}}^{-}_{*}\right)_1\right] & &= \mathbf{K}_{2,\text{opt}}\left[\mathbf{K}_{1,\text{opt}}\left(\hat{\mathbf{V}}^{-}_{*}\right)_1 \mathbf{K}^{\mathrm{T}}_{1,\text{opt}}\right]\mathbf{K}^{\mathrm{T}}_{2,\text{opt}} \\ &= \left[\mathbf{K}_{2,\text{opt}}\mathbf{K}_{1,\text{opt}}\right]\left(\hat{\mathbf{t}}^{-}_{*}\right)_1 & &= \left[\mathbf{K}_{2,\text{opt}}\mathbf{K}_{1,\text{opt}}\right]\left(\hat{\mathbf{V}}^{-}_{*}\right)_1\left[\mathbf{K}_{2,\text{opt}}\mathbf{K}_{1,\text{opt}}\right]^{\mathrm{T}} \end{aligned} \tag{64}$$

The combined minimum-variance GMDe coefficient matrices from the Recursion #1 in Equation (52) and Recursion #2 in Equation (61) is

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$
\begin{aligned}
\mathbf{K}_{\text{opt}} &= \left[\mathbf{K}_{2,\text{opt}}\mathbf{K}_{1,\text{opt}}\right] \\
&= \begin{bmatrix}
1 & 0 & 0 & \left(-\hat{\rho}_{i2}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 1 & 0 & \left(-\hat{\rho}_{k2}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 0 & 1 & \left(-\hat{\rho}_{12}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 & \left(-\hat{\rho}_{i1}^{-}\right)_1 & 0 \\
0 & 1 & \left(-\hat{\rho}_{k1}^{-}\right)_1 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & \left(-\hat{\rho}_{12}^{-}\right)_1 & 1
\end{bmatrix} \\
&= \begin{bmatrix}
1 & 0 & \left(-\hat{\rho}_{i1}^{-}\right)_1 + \left(-\hat{\rho}_{i2}^{-}/\hat{\rho}_{2}^{-}\right)_2\left(-\hat{\rho}_{12}^{-}\right)_1 & \left(-\hat{\rho}_{i2}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 1 & \left(-\hat{\rho}_{k1}^{-}\right)_1 + \left(-\hat{\rho}_{k2}^{-}/\hat{\rho}_{2}^{-}\right)_2\left(-\hat{\rho}_{12}^{-}\right)_1 & \left(-\hat{\rho}_{k2}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 0 & \left(-\hat{\rho}_{12}^{-}/\hat{\rho}_{2}^{-}\right)_2\left(-\hat{\rho}_{12}^{-}\right)_1 & \left(-\hat{\rho}_{12}^{-}/\hat{\rho}_{2}^{-}\right)_2 \\
0 & 0 & 0 & 0
\end{bmatrix}
\end{aligned}
\tag{65}
$$

Equation (65) includes a mixture of minimum-variance coefficients from GMDe Recursion #1 (subscript "1") and Recursion #2 (subscript "2"). In order to express the coefficients in Recursion #2 in terms of the prior coefficients for Recursion #1, substitute the minimum-variance GMDe estimates from Recursion #1 in Equation (54) into the minimum-variance coefficients from Recursion #2

$$
\begin{aligned}
&\left(\hat{\rho}_{i2}^{-}\right)_2 \leftarrow \left(\hat{\rho}_{i2}^{+}\right)_1 = \left(\hat{\rho}_{i2}^{-} - \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}\right)_1
&& \left(-\hat{\rho}_{i2}^{-}/\hat{\rho}_{2}^{-}\right)_2 = \frac{\left(-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2} \\
&\left(\hat{\rho}_{k2}^{-}\right)_2 \leftarrow \left(\hat{\rho}_{k2}^{+}\right)_1 = \left(\hat{\rho}_{k2}^{-} - \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}\right)_1
&& \left(-\hat{\rho}_{k2}^{-}/\hat{\rho}_{2}^{-}\right)_2 = \frac{\left(-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2} \\
&\left(\hat{\rho}_{12}^{-}\right)_2 \leftarrow \left(\hat{\rho}_{12}^{+}\right)_1 = 0
&& \left(-\hat{\rho}_{12}^{-}/\hat{\rho}_{2}^{-}\right)_2 = 0 \\
&\left(\hat{\rho}_{2}^{-}\right)_2 \leftarrow \left(\hat{\rho}_{2}^{+}\right)_1 = 1-\left(\hat{\rho}_{12}^{-}\right)_1^2 &&
\end{aligned}
\tag{66}
$$

Replace terms in Equation (65) with the identities in Equation (66)

$$
\begin{aligned}
\mathbf{K}_{\text{opt}} &= \left[\mathbf{K}_{2,\text{opt}} \mathbf{K}_{1,\text{opt}}\right] \\
&= \left[\begin{array}{cc|cc}
1 & 0 & \left(-\hat{\rho}_{i1}^{-}\right)_1 + \dfrac{\left(-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2}\left(-\hat{\rho}_{12}^{-}\right)_1 & \dfrac{\left(-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2} \\
0 & 1 & \left(-\hat{\rho}_{k1}^{-}\right)_1 + \dfrac{\left(-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2}\left(-\hat{\rho}_{12}^{-}\right)_1 & \dfrac{\left(-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}\right)_1}{1-\left(\hat{\rho}_{12}^{-}\right)_1^2} \\
\hline
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right] \\
&= \left[\begin{array}{cc|cc}
1 & 0 & \left(\dfrac{-\hat{\rho}_{i1}^{-} + \hat{\rho}_{i2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) & \left(\dfrac{\left(-\hat{\rho}_{i2}^{-} + \hat{\rho}_{i1}^{-}\hat{\rho}_{12}^{-}\right)}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\
0 & 1 & \dfrac{-\hat{\rho}_{k1}^{-} + \hat{\rho}_{k2}^{-}\hat{\rho}_{12}^{-}}{1-\left(\hat{\rho}_{12}^{-}\right)^2} & \left(\dfrac{\left(-\hat{\rho}_{k2}^{-} + \hat{\rho}_{k1}^{-}\hat{\rho}_{12}^{-}\right)}{1-\left(\hat{\rho}_{12}^{-}\right)^2}\right) \\
\hline
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right]_1
\end{aligned}
\tag{67}
$$

The minimum-variance coefficient matrix with the cumulative recursive GMDe in Equation (67) agrees with the minimum-variance coefficient matrix with the regular GMDe in Equation (41). Therefore, the recursive GMDe in Sections 7.2.1 and 7.2.2 agree with the batch GMDe in Section 7.1.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

8 MITIGATION OF BIAS IN GMDE COVARIANCE MATRIX

GMDe is a simple linear transformation of the $(M+J)\times 1$ vector of π-estimates for the M study variables and the J auxiliary residuals with a $(M+J)\times(M+J)$ matrix $\mathbf{K}$ of known constants, where $\mathbf{A}$ is a $J\times(M+J)$ partition of arbitrary coefficients.

$$\hat{\mathbf{t}}^{+} = \mathbf{K}\,\hat{\mathbf{t}}^{-} \tag{15}$$

$$\mathbf{K} = \mathbf{I} + \left[\mathbf{0} \mid \mathbf{A}\right] \tag{16}$$

$$\hat{\mathbf{V}}^{+} = \mathbf{K}\hat{\mathbf{V}}^{-}\mathbf{K}^{\mathrm{T}} \tag{18}$$

Regardless of the analyst's choice for the arbitrary coefficients in Equations (15) to (18), GMDe is unbiased (Hansen *et al.*, 1953, Särndal *et al.*, 1992). However, the choice of arbitrary coefficients strongly affects statistical efficiency (*i.e.*, variance reduction) of GMDe. With a poor choice, GMDe can increase variance relative to the prior π-estimate.

Särndal *et al.* (1992) chose the minimum-variance criterion to select "optimal" coefficients for GMDe, which requires the prior $(M+J)\times(M+J)$ covariance matrix $\mathbf{V}^{-}$. They assume $\mathbf{V}^{-}$ is known exactly, whereas GMDe computes approximately minimum-variance coefficients with the π-estimate $\hat{\mathbf{V}}^{-}$ for covariance matrix $\mathbf{V}^{-}$

$$\mathbf{A}_{\mathrm{opt}} = \left[\begin{array}{c} -\hat{\boldsymbol{\Gamma}}^{-}\left(\hat{\boldsymbol{\Lambda}}^{-}\right)^{-1} \\ \hline -\mathbf{I} \end{array}\right] \tag{29}$$

$$\mathbf{K}_{\mathrm{opt}} = \left[\begin{array}{c|c} \mathbf{I} & -\hat{\boldsymbol{\Gamma}}^{-}\left(\hat{\boldsymbol{\Lambda}}^{-}\right)^{-1} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \tag{30}$$

where $M\times J$ matrix $\hat{\boldsymbol{\Gamma}}^{-}$ and $J\times J$ matrix $\hat{\boldsymbol{\Lambda}}^{-}$ are partitions of the prior π-estimate for $(M+J)\times(M+J)$ covariance matrix $\hat{\mathbf{V}}^{-}$.

Herein lies a concern with GMDe coefficients that use π-estimates of the covariance matrix $\hat{\mathbf{V}}^{-}$.

If GMDe coefficients do not exactly equal their true (but unknown) minimum-variance values, then the variances for population estimates of the study variables will be greater than the minimum possible variances (see Figure 2, page 29), and this can cause GMDe to overestimate variances of population estimates for the study variables.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

On the other hand, different realizations of the probability sample would produce different π-estimates of covariance matrix $\hat{\mathbf{V}}^{-}$, which would produce different coefficient matrices $\mathbf{K}_{\text{opt}}$. GMDe does not account for the variability of coefficient matrix $\mathbf{K}_{\text{opt}}$ with different hypothetical realizations of the design-based sample survey, and this can cause GMDe to underestimate variances of population estimates for the study variables. The net effect of random sampling error in covariance matrix $\hat{\mathbf{V}}^{-}$ on approximate minimum-variance GMDe population estimates is confounded.

There are at least three different mitigation alternatives to resolve this concern.

1. Accept biased GMDe variances, and assume net bias is too small to be of practical concern. Typical applications of the composite estimator and the Kalman filter rely on this same assumption. This approach is more palatable if the random error in π-estimates are small, *i.e.*, partitions $\hat{\boldsymbol{\Gamma}}^{-}$ and $\hat{\boldsymbol{\Lambda}}^{-}$ of covariance matrix $\hat{\mathbf{V}}^{-}$ in Equations (29) and (30). Such a well-behaved covariance matrix $\hat{\mathbf{V}}^{-}$ is less likely if the number of study variables (M) is large.
2. Chose arbitrary coefficients that are weaker (closer to zero) than the approximate minimum-variance coefficients such that they would not be affected by minimum-variance values with different hypothetical realizations of the π-estimates. The Appendix starting on page 47 explores this alternative with a special case, in which both the study variable and auxiliary variable are binary categorical variables. This alternative reduces influence of random sampling error on potential bias in GMDe variance estimates, but at the cost of reduction in statistical efficiency (*i.e.*, larger variances in population estimates for study variables). This cost is relatively high for rare categories in this example (see Table 3 and Figure 5, page 54). In the special case of official statistics, with numerous and detailed statistical tables, most, if not all, study variables are relatively rare. Therefore, this alternative is not recommended.
3. Apply bootstrap resampling methods to the realized sample data to estimate GMDe covariance matrix. McConville *et al*. (2020) used the bootstrap variance estimator in a simulation study for GREG regression estimators, for which post-stratification, ratio, regression, lasso, ridge, and elastic net estimators are special cases. They found that closed form variance computations for these model-assisted estimators are prone to underestimation of the true variance of population estimates. They recommend the bootstrap variance estimator as a more accurate method for GREG variance estimators. Those same methods apply to GMDe.

Therefore, the preliminary recommendation is to use the bootstrap variance estimator with GMDe for rigorous analyses (Alternative #3), although the assumption of inconsequential biased (Alternative #1) can be useful for preliminary analyses because it is computationally less burdensome than (Alternative #3).

9 DISCUSSION

GMDe is especially relevant in government programs for official statistics. One example is an annual National Forest Inventory (NFI), *e.g.*, McRoberts, 2000; Bechtold and Patterson, 2005; Gillis *et al*., 2005; and Gschwantner *et al*., 2009. Study variables require expensive measurements of sampled forest stands and trees by field crews for thousands of sampling units. Auxiliary variables include relatively inexpensive multi-sensor measurements with various spaceborne and airborne platforms, each of which can require different types of sampling units (*e.g.*, Rodriguez-Veiga *et al*., 2014; Lu *et al*., 2016). Predictions from deterministic process models provide another source of auxiliary variables; growth and yield models and stand projection models are examples in forest science. Assume each auxiliary variable is correlated with at least one study variable. There may be much collinearity among auxiliary variables.

The degree of GMDe variance reduction for each study variable strongly depends on the strength of the correlations between the study variable and auxiliary variables. Table 1 (page 34) and Figure 3 (page 35) illustrate those relationships. Variance reduction is inconsequential of the correlation is less than ±0.30. If the correlation is 0.50, variance reduction is 25-percent, and reduction in standard deviation is 13-percent. In order to reduce variance by 50-percent, the correlation between a study variable and an auxiliary variable must be approximately ±0.70. In order to reduce standard deviation (*e.g.*, confidence interval) by 50-percent, the correlation must be approximately ±0.85. Therefore, the correlations between study variables and auxiliary variable should be relatively strong for substantial improvements with GMDe.

GMDe can combine time-series of remotely sensed measurements with time-series of predictions from deterministic models to improve time-series of detailed population estimates (*e.g.*, McRoberts, 2000). Broadly defined study variables, such as total aboveground biomass and stand-altering disturbances, are more strongly correlated with remotely sensed measurements than very specific study variables, such as merchantable wood volume by tree-species. Other auxiliary variables may include time-series of predictions from a deterministic tree- or stand-level projection model (*e.g.*, Sullivan and Clutter, 1972; Stage, 1973; Belcher *et al*., 1982; Lessard *et al*., 2001). Model predictions of current values for detailed study variables based on past measurements can be strongly correlated with current measurements by field crews, especially in strata that are not strongly affected by land management activities and catastrophic disturbances. If the objective is more accurate population estimates for detailed study variables in a NFI, then change detection with remotely sensed data might be most valuable for post-stratification into “disturbed” and “undisturbed” strata, and model predictions might be most valuable for variance reduction in the undisturbed stratum.

GMDe is a broad generalization of the univariate difference estimator described by Hansen *et al*. (1953:250-253) and Särndal *et al*. (1992:221-225). Unlike their univariate estimators, GMDe does not require a proxy variable that is correlated with the study variable(s) and predicted with the auxiliary variables; nor does GMDe require known population totals (*e.g.*, full coverage censuses) for all auxiliary variables and the proxy variable. Rather, GMDe directly incorporates

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

sampling error for population estimates of auxiliary variables into GMDe estimation error for correlated study variables. Unlike those univariate estimators, GMDe estimates the $M \times M$ covariance matrix for the $M \times 1$ vector of population estimates for the M study variables, which further supports population estimates for synthetic variables (*e.g.*, margin totals, ratios, rates of change over time). Furthermore normalization of the design-based population estimates permit computation of minimum-variance GMDe coefficients from correlation statistics between the study variables and auxiliary variables, which, unlike a proxy variable, obviates the need for auxiliary variables to share the same metric space as the study variables.

GMDe is closely related to the multivariate composite estimator and the Kalman filter: both are linear transformations of a vector of population estimates from a probability sample with a design-consistent estimator.

This Technical Report includes a simple example that provides intuitive insights. The example also serves as an introduction to a recursive version of GMDe, which can impose equality and inequality constraints, mitigate affects from outliers, and manage numerical errors that are endemic to computations with large matrices. The recursive GMDe replaces inversion of the $J \times J$ partition of the covariance matrix for auxiliary residuals with a sequence of J scalar divisions; the recursive GMDe can be numerically robust even if there are strong collinearities among numerous auxiliary variables, and data reduction methods are unnecessary. Therefore, the recursive GMDe does not require preprocessing with dimensionality reduction techniques, which has become more relevant as sources of “big data” as auxiliary variables become more common (*e.g.*, Reddy *et al.*, 2020).

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

LITERATURE CITED

Bechtold, W.A. and Patterson, P.L., 2005. *The enhanced forest inventory and analysis program--national sampling design and estimation procedures* (Vol. 80). USDA Forest Service, Southern Research Station.

Belcher, D.M., Holdaway, M.R. and Brand, G.J., 1982. A description of STEMS--the stand and tree evaluation and modeling system.

Czaplewski, R.L., 1999. Multistage remote sensing: toward an annual national inventory. *Journal of Forestry*, *97*(12), pp.44-48.

Czaplewski, R.L., 2010. 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.*

Czaplewski, R.L., 2020. A Robust Multivariate Estimator with Stepwise Covariate Selection and Inequality Constraints for Complex Sample Surveys: An Initial Concept. *Tech. Rep. 2020-1. Fort Collins, CO: Environmetrika. 115 p*. DOI: 10.13140/RG.2.2.17634.07369

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

Gschwantner, T., Lawrence, M., McRoberts, R.E. and Tomppo, E. eds., 2009. *National Forest Inventories: Pathways for Common Reporting*. Springer.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G., 1953. *Sample survey methods and theory. V. 1. Methods and applications*. Wiley.

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G. and Moran, E., 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth*, *9*(1), pp.63-105.

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.

Lui, K.J., 2020. Notes on Use of the Composite Estimator: an Improvement of the Ratio Estimator. *Journal of Official Statistics, 36(1)*, pp.137-149.

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), p.244.

McRoberts, Ronald E. "Background for AFIS, the Annual Forrest Inventory System." In *In: McRoberts, Ronald E.; Reams, Gregory A.; Van Deusen, Paul C., eds. Proceedings of the First Annual Forest Inventory and Analysis Symposium; Gen. Tech. Rep. NC-213. St. Paul, MN: US Department of Agriculture, Forest Service, North Central Research Station: 1-4*, vol. 213. 2000.

Montanari, G.E., 1987. Post-sampling efficient QR-prediction in large-sample surveys. *International Statistical Review/Revue Internationale de Statistique*, pp.191-202.

Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. and Baker, T., 2020. Analysis of dimensionality reduction techniques on big data. *IEEE Access*, *8*, pp.54776-54788.

Rodriguez-Veiga, P., Tansey, K. and Balzter, H., 2014. *Global Biomass Information System: Mapping above Ground Biomass, Uncertainty, and Forest Area using Multi-Platform Earth Observation Datasets*. European Centre of Excellence in Earth Observation Research Training Deliverable D2.

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

Stage, A.R., 1973. *Prognosis model for stand development* (Vol. 137). Intermountain Forest & Range Experiment Station, Forest Service, US Department of Agriculture.

Sullivan, A.D. and Clutter, J.L., 1972. A simultaneous growth and yield model for loblolly pine. *Forest Science*, *18*(1), pp.76-86.

ACKNOWLEDGEMENTS

I value and appreciate comments and suggestions by Steen S. Magnussen and Stephen Stehman, both of whom reviewed the Abstract and Introduction (pages 2-9) to this Technical Report. Any errors or omissions remain my sole responsibility.

DISCLAIMER

APPENDIX: MULTINOMIAL EXAMPLE

Hansen *et al.* (1953) and Särndal *et al.* (1992) assume arbitrary coefficients in the difference estimator are independent from design-based (*e.g.*, HT) population estimates for the study variables and the auxiliary residuals. However, computation of the minimum-variance coefficients for GMDe in Section 6 (page 16) require those same π-estimates. Hypothetically, different realizations of the sample would result in different π-estimates, and therefore, different minimum-variance coefficients. While GMDe population estimates remain unbiased, omission of that propagated random error results in biased GMDE population estimates, the degree to which is not well understood.

One possible mitigation for variance underestimation is to shrink the minimum-variance GMDe coefficients towards zero to the extent that no hypothetical realization of the random sample would yield minimum-variance coefficients that were closer to zero. This is Alternative #2 (page 42).

Assume for the following that there is one study variable and one auxiliary variables, which is a special case of Section 7.2.1.5 (page 33).

Equations (53) and (55) express the minimum-variance GMDe estimate for study variable *k* and its variance as a function of the correlation between study variable *k* and auxiliary variable 1 in the special case of normalized π-estimates ($\hat{v}_{i*}^{-} = \hat{v}_{1*}^{-} = 1$) in Section 5 (page 14)

$$\left(\hat{t}_{k*\text{opt}}^{+}\right)_1 = \hat{t}_{k*}^{-} - \hat{\rho}_{k1}^{-} r_{1*}^{-} \tag{53}$$

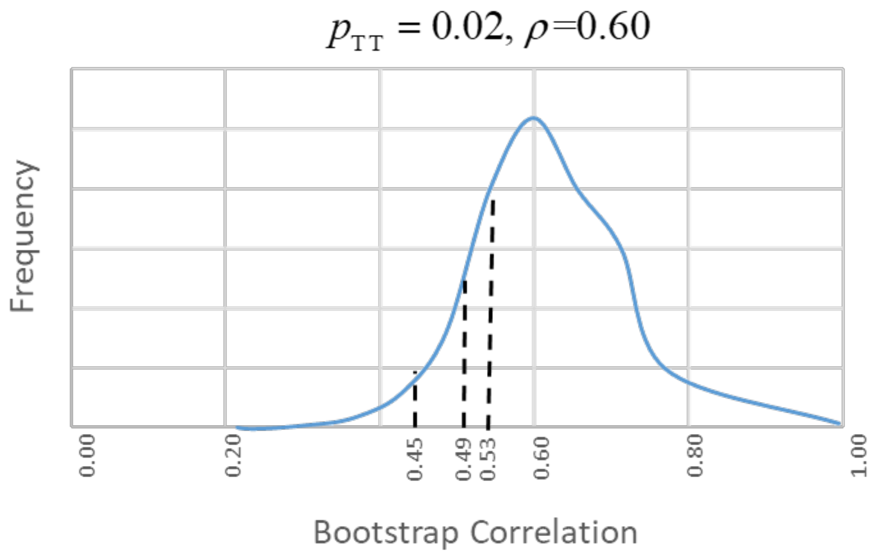
$$\hat{v}_{k*\text{opt}}^{+} = 1 - \left(\hat{\rho}_{k1}^{-}\right)^2 \tag{55}$$

Hypothetically, different realizations of the π-estimates for $\hat{\rho}_{k1}^{-}$ would produce different GMDe population estimates and variances for study variable *k*.

This appendix explores one possible method to mitigate the incorrect assumption that the minimum-variance GMDe coefficients are independent from the prior π-estimates. What if the arbitrary GMDe coefficient was some smaller value than the minimum-variance coefficient? The smaller the arbitrary coefficient, *i.e.*, correlation $\hat{\rho}_{k1}^{-}$ in Equation (55), the less likely a hypothetical realization of the sample survey would have a population estimate for correlation $\hat{\rho}_{k1}^{-}$ that is even smaller. This would reduce the dependence of the realized GMDe estimates on the realized sample.

Figure 4 is an example. Assume the realized estimate of the correlation is $\hat{\rho}_{k1}^{-} = 0.60$. Use the bootstrap resampling method to approximate the hypothetical sampling distribution for correlation $\hat{\rho}_{k1}^{-}$. The following simulation assumes simple random sampling with equal inclusion probabilities; sample size n=1000; 100 bootstrap resamplings, and prevalence of 0.02 for population elements with a nonzero value in the population (*i.e.*, expectation is that 2 of every 100 population elements have categorical study variable equal to 1).

Figure 4 Bootstrap distribution (smoothed) of π-estimates for the correlation ρ between categorical study variable y and categorical auxiliary variable x. In



this example, p_{TT}=0.02 population elements have the categorical trait, and the expected correlation between the study variable and auxiliary variables is ρ=0.60. Five percent of the bootstrap samples estimated correlation ρ<0.45, 10-percent estimated ρ<0.49; and 25-percent estimated ρ<0.53.

Five percent of the bootstrap samples have an estimated correlation ρ<0.45. If ρ=0.45 were used instead of $\hat{\rho}_{k1}^{-} = 0.60$ in the arbitrary coefficient for GMDe estimate from Equation (55), then it would be relatively unlikely that GMDe from a different hypothetical realization of the sample would have an even smaller estimate. However, this would reduce the absolute magnitude of GMDe coefficients, which would reduce the statistical efficiency relative to GMDe with the minimum-variance coefficients; but this choice of arbitrary coefficient would reduce dependence of GMDe coefficients on the realized sample. In other words, bias in GMDe variance estimates would be reduced by sacrificing GMDe variance reduction.

This appendix explores this mitigation tactic with an example in which:

- Simple random sample of population with equal inclusion probabilities;
- Single categorical study variable;
- Single categorical auxiliary residual;
- The association between the study variable and auxiliary variable is characterized by a 2κ2 contingency table;

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

- Correlation between the study variable and auxiliary variable varies from moderate $\hat{\rho}_{k1}^{-} = 0.60$ to high $\hat{\rho}_{k1}^{-} = 0.90$.
- Prevalence of the study variable varies from rare $p_{\bullet\mathrm{T}} = p_{\mathrm{T}\bullet} = 0.02$ to more common $p_{\bullet\mathrm{T}} = p_{\mathrm{T}\bullet} = 0.25$.

9.1 Categorical Study Variable and Categorical Auxiliary Variable

Let binary study variable y_κ=1 if categorical variable Y_κ=True for population element κ; y_κ=0 otherwise. Let binary auxiliary variable x_κ=1 if categorical variable X_κ=True for population element κ; x_κ=0 otherwise. The population has a total of *N* elements, and a simple random sample of *n* elements is used to estimate population parameters. Define those population parameters with the contingency table of joint proportions (probabilities) as

	X=True	*X*=False	
Y= True	p_{TT}	p_{TF}	$p_{\mathrm{T}\bullet} = p_{\mathrm{TT}} + p_{\mathrm{TF}}$
Y= False	p_{FT}	$p_{\mathrm{FF}} = 1-\left(p_{\mathrm{TT}} + p_{\mathrm{FT}} + p_{\mathrm{TF}}\right)$	$p_{\mathrm{F}\bullet} = p_{\mathrm{FT}} + p_{\mathrm{FF}}$
	$p_{\bullet\mathrm{T}} = p_{\mathrm{TT}} + p_{\mathrm{FT}}$	$p_{\bullet\mathrm{F}} = p_{\mathrm{TF}} + p_{\mathrm{FF}}$	1

where p_{TT} is the proportion of population elements for which study variable *Y*=True and auxiliary variable *X*=True; p_{FF} is the proportion of population elements for which study variable *Y*=False and auxiliary variable *X*=False; p_{TF} is the proportion of population elements for which study variable *Y*= True and auxiliary variable *X*=False; and p_{FT} is the proportion of population elements for which study variable *Y*=False and auxiliary variable *X*= True. The classifiers for the study variables and auxiliary variables agree in the first two cases, and they disagree in the latter two cases (*i.e.*, misclassification).

To simplify this hypothetical example, assume the margins of the contingency table agree, *i.e.*, $p_{\bullet\mathrm{T}} = p_{\mathrm{T}\bullet}$; and the proportion of population elements in which the study variable *X*=True is known $p_{\mathrm{T}\bullet}$

	X=True	*X*=False	
Y= True	p_{TT}	$p_{F\mathrm{T}} = \left(p_{\mathrm{T}\bullet} - p_{\mathrm{TT}}\right)$	$p_{\mathrm{T}\bullet}$
Y=False	$p_{\mathrm{FT}} = p_{F\mathrm{T}} = \left(p_{\bullet\mathrm{T}} - p_{\mathrm{TT}}\right)$	$p_{\mathrm{FF}} = 1 - p_{\mathrm{TT}} - 2\left(p_{\mathrm{T}\bullet} - p_{\mathrm{TT}}\right)$	$p_{\mathrm{F}\bullet} = \left(1 - p_{\mathrm{T}\bullet}\right)$
	$p_{\bullet\mathrm{T}} = p_{\mathrm{T}\bullet}$	$p_{\bullet\mathrm{F}} = \left(1 - p_{\mathrm{T}\bullet}\right)$	1

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Given these assumptions, two parameters $p_{\mathrm{T}\bullet}$ and p_{TT} sufficiently define the hypothetical contingency table

	X=True	*X*=False	
Y= True	p_{TT}	$\left(p_{\mathrm{T}\bullet}-p_{\mathrm{TT}}\right)$	$p_{\mathrm{T}\bullet}$
Y=False	$\left(p_{\bullet\mathrm{T}}-p_{\mathrm{TT}}\right)$	$p_{\mathrm{FF}}=1-p_{\mathrm{TT}}-2\left(p_{\mathrm{T}\bullet}-p_{\mathrm{TT}}\right)$	$\left(1-p_{\mathrm{T}\bullet}\right)$
	$p_{\mathrm{T}\bullet}$	$\left(1-p_{\mathrm{T}\bullet}\right)$	1

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

Table 2. Examples used in bootstrap simulations in Section 8.4 (page 53): three different prevalences of the categorical variables in the population $p_{T\bullet} = p_{T\bullet} = \{0.25, 0.10, 0.02\}$ from Sections 13.1 and 13.2; and four different correlations $\rho = \{0.60, 0.70, 0.80.0.90\}$ between π-estimates of the study variable and the auxiliary variable from Section 13.3

Correlation between HT population estimates for study and auxiliary variables $\rho = 0.60$

	$p_{T\bullet} = p_{T\bullet} = 0.25$			$p_{T\bullet} = p_{T\bullet} = 0.10$			$p_{T\bullet} = p_{T\bullet} = 0.02$	
.	x=True	x=False	.	x=True	x=False	.	x=True	x=False
y=True	66%	11%	y=True	62%	4%	y=True	61%	1%
Y=False	34%	89%	y=False	38%	96%	y=False	39%	99%
.	100%	100%	.	100%	100%	.	100%	100%

Correlation between HT population estimates for study and auxiliary variables $\rho = 0.70$

	$p_{T\bullet} = p_{T\bullet} = 0.25$			$p_{T\bullet} = p_{T\bullet} = 0.10$			$p_{T\bullet} = p_{T\bullet} = 0.02$	
.	x=True	x=False	.	x=True	x=False		x=True	x=False
y=True	73%	9%	y=True	71%	3%	y=True	70%	1%
Y=False	27%	91%	y=False	29%	97%	y=False	30%	99%
.	100%	100%	.	100%	100%		100%	100%

Correlation between HT population estimates for study and auxiliary variables $\rho = 0.80$

	$p_{T\bullet} = p_{T\bullet} = 0.25$			$p_{T\bullet} = p_{T\bullet} = 0.10$			$p_{T\bullet} = p_{T\bullet} = 0.02$	
.	x=True	x=False		x=True	x=False		x=True	x=False
y=True	82%	6%	y=True	81%	2%	y=True	80%	0%
y=False	18%	94%	y=False	19%	98%	y=False	20%	100%
.	100%	100%		100%	100%	.	100%	100%

Correlation between HT population estimates for study and auxiliary variables $\rho = 0.90$

	$p_{T\bullet} = p_{T\bullet} = 0.25$			$p_{T\bullet} = p_{T\bullet} = 0.10$			$p_{T\bullet} = p_{T\bullet} = 0.02$	
.	x=True	x=False		x=True	x=False	0	x=True	x=False
y=True	90%	3%	y=True	90%	1%	y=True	90%	0%
y=False	10%	97%	y=False	10%	99%	y=False	10%	100%
.	100%	100%		100%	100%	.	100%	100%

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

9.2 THE 4×1 VECTOR TRANSFORMATION FOR 2×2 CONTINGENCY TABLE

Next, transform the 2×2 contingency table into a 4×1 vector of population proportions

$$\mathbf{p} = \begin{bmatrix} p_{\mathrm{TT}} \\ p_{\mathrm{FT}} \\ p_{\mathrm{FT}} \\ p_{\mathrm{FF}} \end{bmatrix} \quad \text{where} \begin{cases} p_{\mathrm{TF}} = p_{\mathrm{FT}} = \left(p_{\mathrm{T}\bullet} - p_{\mathrm{TT}}\right) \\ p_{\mathrm{FF}} = 1 - p_{\mathrm{TT}} - 2\left(p_{\mathrm{T}\bullet} - p_{\mathrm{TT}}\right) \end{cases} \tag{68}$$

The multinomial distribution provides the covariance matrix for vector p in Equation (68)

$$\mathbf{V_p} = n\begin{bmatrix} p_{\mathrm{TT}} & -p_{\mathrm{TT}}p_{\mathrm{FT}} & -p_{\mathrm{TT}}p_{\mathrm{FT}} & -p_{\mathrm{TT}}p_{\mathrm{FF}} \\ -p_{\mathrm{TT}}p_{\mathrm{FT}} & p_{\mathrm{FT}}\left(1-p_{\mathrm{FT}}\right) & -p_{\mathrm{FT}}p_{\mathrm{FT}} & -p_{\mathrm{FT}}p_{\mathrm{FF}} \\ -p_{\mathrm{TT}}p_{\mathrm{FT}} & -p_{\mathrm{FT}}p_{\mathrm{FT}} & p_{\mathrm{FT}}\left(1-p_{\mathrm{FT}}\right) & -p_{\mathrm{FT}}p_{\mathrm{FF}} \\ -p_{\mathrm{TT}}p_{\mathrm{FF}} & -p_{\mathrm{FT}}p_{\mathrm{FF}} & -p_{\mathrm{FT}}p_{\mathrm{FF}} & p_{\mathrm{FF}}\left(1-p_{\mathrm{FF}}\right) \end{bmatrix} \tag{69}$$

Define study variable y as the proportion of population elements in which Y_κ= True, and define auxiliary variable x as the proportion of population elements in which X_κ= True. GMDe uses population estimate for auxiliary variable $\hat{x}$ to improve the population estimate for study variable $\hat{y}$.

$$\begin{aligned} \begin{bmatrix} y \\ x \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_{\mathrm{TT}} \\ p_{\mathrm{TF}} \\ p_{\mathrm{FT}} \\ p_{\mathrm{FF}} \end{bmatrix} \\ &= \begin{bmatrix} p_{\mathrm{TT}} + \left(p_{\mathrm{T}\bullet} - p_{\mathrm{TT}}\right) \\ p_{\mathrm{TT}} + \left(p_{\bullet\mathrm{T}} - p_{\mathrm{TT}}\right) \end{bmatrix} \\ &= \begin{bmatrix} p_{\mathrm{T}\bullet} \\ p_{\bullet\mathrm{T}} \end{bmatrix} \end{aligned} \tag{70}$$

The 2×2 covariance matrix for the linear transformation in Equation (70) uses the 4×4 covariance matrix is

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

$$\mathbf{V_p} = n\left(\begin{bmatrix}1 & 1 & 0 & 0\\ 1 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix} p_{\mathrm{TT}} & -p_{\mathrm{TT}}p_{\mathrm{TF}} & -p_{\mathrm{TT}}p_{\mathrm{FT}} & -p_{\mathrm{TT}}p_{\mathrm{FF}} \\ -p_{\mathrm{TF}}p_{\mathrm{TT}} & p_{\mathrm{TF}}\left(1-p_{\mathrm{TF}}\right) & -p_{\mathrm{TF}}p_{\mathrm{FT}} & -p_{\mathrm{TF}}p_{\mathrm{FF}} \\ -p_{\mathrm{FT}}p_{\mathrm{TT}} & -p_{\mathrm{FT}}p_{\mathrm{TF}} & p_{\mathrm{FT}}\left(1-p_{\mathrm{FT}}\right) & -p_{\mathrm{FT}}p_{\mathrm{FF}} \\ -p_{\mathrm{FF}}p_{\mathrm{TT}} & -p_{\mathrm{FF}}p_{\mathrm{TF}} & -p_{\mathrm{FF}}p_{\mathrm{FT}} & p_{\mathrm{FF}}\left(1-p_{\mathrm{FF}}\right)\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}\right)$$

$$= n\begin{bmatrix} p_{\mathrm{TT}} - 2p_{\mathrm{TF}}p_{\mathrm{TT}} - p_{\mathrm{TF}}^2 + p_{\mathrm{TF}} & p_{\mathrm{TT}} - p_{\mathrm{TF}}p_{\mathrm{TT}} - p_{\mathrm{FT}}p_{\mathrm{TT}} - p_{\mathrm{TF}}p_{\mathrm{FT}} \\ p_{\mathrm{TT}} - p_{\mathrm{TF}}p_{\mathrm{TT}} - p_{\mathrm{FT}}p_{\mathrm{TT}} - p_{\mathrm{TF}}p_{\mathrm{FT}} & p_{\mathrm{TT}} - 2p_{\mathrm{FT}}p_{\mathrm{TT}} - p_{\mathrm{FT}}^2 + p_{\mathrm{FT}}\end{bmatrix}$$

where $p_{\mathrm{FT}} = p_{\mathrm{TF}} = \left(p_{\mathrm{T\bullet}} - p_{\mathrm{TT}}\right)$ (71)

9.3 CORRELATION BETWEEN STUDY VARIABLE AND AUXILIARY VARIABLE

The correlation ρ between population estimates $\hat{y}$ and $\hat{x}$ is

$$\rho = \frac{v_{yx}}{\sqrt{v_y v_x}} \quad \text{for } p_{\mathrm{FT}} = p_{\mathrm{TF}}$$
$$= \frac{n\left(p_{\mathrm{TT}} - 2p_{\mathrm{TT}}p_{\mathrm{FT}} - p_{\mathrm{FT}}^2\right)}{\sqrt{n\left(p_{\mathrm{TT}} - 2p_{\mathrm{TT}}p_{\mathrm{FT}} - p_{\mathrm{FT}}^2 + p_{\mathrm{FT}}\right)}\sqrt{n\left(p_{\mathrm{TT}} - 2p_{\mathrm{TT}}p_{\mathrm{FT}} - p_{\mathrm{FT}}^2 + p_{\mathrm{FT}}\right)}}$$
$$= \frac{p_{\mathrm{TT}} - 2p_{\mathrm{TT}}p_{\mathrm{FT}} - p_{\mathrm{FT}}^2}{p_{\mathrm{TT}} - 2p_{\mathrm{TT}}p_{\mathrm{FT}} - p_{\mathrm{FT}}^2 + p_{\mathrm{FT}}}$$
$$= \frac{p_{\mathrm{TT}}^2 - p_{\mathrm{T\bullet}}^2 + p_{\mathrm{TT}}}{p_{\mathrm{TT}}^2 - p_{\mathrm{T\bullet}}^2 + p_{\mathrm{T\bullet}}} \tag{72}$$

The following example uses the proportion of the population $p_{\mathrm{T\bullet}}$ in which study variable Y=True and the correlation between population estimates $\hat{y}$ and $\hat{x}$ in Equation (72). In order to set *a prior* levels of correlation, solve Equation (72) for population proportion p_{TT} given values for $p_{\mathrm{T\bullet}}$ and ρ

$$\left(1-\rho\right)p_{\mathrm{TT}}^2 + p_{\mathrm{TT}} - \left(1-\rho\right)p_{\mathrm{T\bullet}}^2 - \rho p_{\mathrm{T\bullet}} = 0 \tag{73}$$

Note that Equations (72) and (73) do not include the sample size *n* in the simple random sample with equal inclusion probabilities, which simplifies generalization.

9.4 BOOTSTRAP EXPERIMENTS

This example uses a bootstrap simulation for a simulated simple random sample of *n*=1000 with equal inclusion probabilities. There are 12 different scenarios: three different prevalences of the categorical variables in the population $p_{\mathrm{T*}}=p_{\mathrm{*T}}$={0.25,0.10,0.02} in Sections 8.1 and 8.2; and ρ={0.60,0.70,0.80.0.90} from Section 8.3four different correlations between π-estimates of the study variable and the auxiliary variable ρ={0.60,0.70,0.80.0.90} from Section 8.3.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

The assumption is that a suboptimal choice for the arbitrary GMDe coefficient (*i.e.*, shrinking the minimum-variance coefficient towards zero) can lessen the dependence between GMDe coefficients and sampling error in π-estimates. If the suboptimal choice were near the lesser tail of the bootstrap distribution for the correlation between population estimates for the study variable and auxiliary residual, then hypothetical alternative realizations of the π-estimate for that correlation would be less affected by those realizations.

The direct metrics are the bootstrap estimates of correlation ρ_{boot} at three different percentiles (25-percentile, 10-percentile, 5-percentile) of the bootstrap distribution. Figure 4 is an example, in which the prevalence of the categorical variable is 0.02 (2%) of population elements, and the correlation between population estimates for the study variable and the auxiliary residual is ρ=0.60. If GMDe coefficient in Equations (53) and (55) were calculated with correlation ρ=0.49, which corresponds to the 5-precentile of the bootstrap distribution for the correlation, rather that the realized correlation estimate of ρ=0.60, then approximately 95% of hypothetical samples would have a higher correlation (*i.e.*, greater variance reduction). The premise is that this would mitigate underestimation of GMDe variances caused by sampling error in estimating the correlation.

Table 3 provides examples for the cost in statistical efficiency in order to reduce underestimation of GMDe variances. In the example of Figure 4, the variance of GMDe population estimate for the study variable would increase by a factor of 125% relative to the minimum-variance estimate (see Table 3).

Table 3 Increase in GMDe variance to mitigate underestimated GMDe variances.

			Bootstrap Variance Increase		
	ρ	optimum	25 %ile	10 %ile	5 %ile
$p_{\text{T}\bullet} = p_{\bullet\text{T}} = 0.25$	0.60	100%	103%	106%	109%
	0.70		105%	111%	114%
	0.80		106%	113%	117%
	0.90		109%	117%	120%
$p_{\text{T}\bullet} = p_{\bullet\text{T}} = 0.10$	0.60	100%	105%	112%	114%
	0.70		110%	115%	119%
	0.80		111%	118%	123%
	0.90		116%	138%	143%
$p_{\text{T}\bullet} = p_{\bullet\text{T}} = 0.02$	0.60	100%	108%	119%	125%
	0.70		115%	126%	128%
	0.80		122%	140%	155%
	0.90		126%	158%	178%

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx

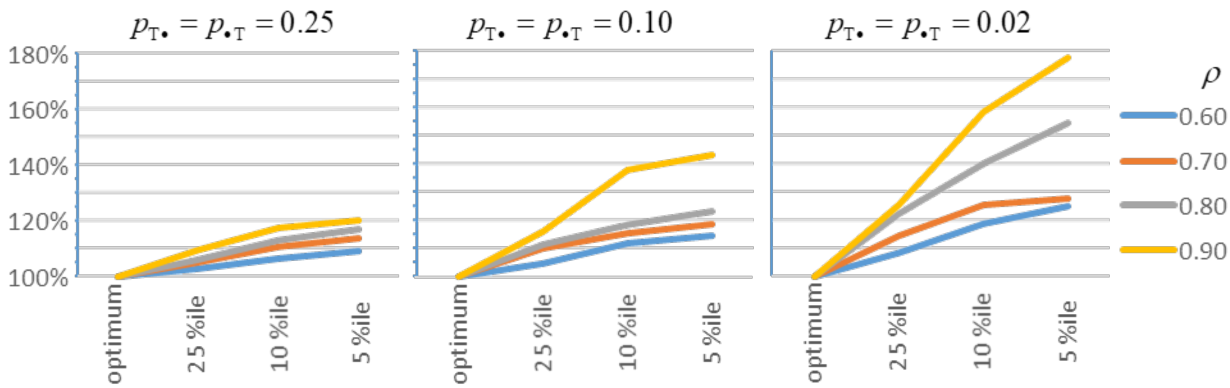


Figure 5 Increase in GMDe variance estimates relative to the minimum-variance estimates with different percentiles (25-percentile, 10-percentile, 5-percentile) of bootstrap distribution for different correlations ρ={0.60,0.70,0.80.0.90} between study variable and auxiliary variable and different prevalences of the categorical variable in the population p_{T*}=p_{*T}={0.25,0.10,0.02}.

Table 3 and Figure 5 offer several insights that mitigate bias in GMDe variances.

- The mitigation cost in statistical efficiency is greater for less prevalent study variables (*e.g.*, p_{T*}=p_{*T}=0.02) and higher correlations (*e.g.*, ρ=0.90) between population estimates for the study variables and the auxiliary residual.
- The cost in statistical efficiency might be too high for rare population attributes *e.g.*, p_{T*}=p_{*T}=0.02), especially with high correlations between population estimates for the study variables and the auxiliary residual.

 DOI: xx.xxxxx/xx.x.x.xxxxx.xxxxx