

Estimation when the Covariance Structure of the Variable of Interest is Positive Definite

Alain Théberge¹

Generalized regression (GREG) estimation uses a model that assumes that the values of the variable of interest are not correlated. An extension of the GREG estimator to the case where the vector of interest has a positive definite covariance structure is presented in this article. This extension can be translated to the calibration estimators. The key to this extension lies in a generalization of the Horvitz-Thompson estimator which, in some sense, also assumes that the values of the variable of interest are not correlated. The Godambe-Joshi lower bound is another result which assumes a model with no correlation. This is also generalized to a vector of interest with a positive definite covariance structure, and it is shown that the generalized calibration estimator asymptotically attains this generalized lower bound. Properties of the new estimators are given, and they are compared with the Horvitz-Thompson estimator and the usual calibration estimator. The new estimators are applied to the Canadian Reverse Record Check survey and to the problem of variance estimation.

Key words: Asymptotic setup; calibration estimators; Godambe-Joshi lower bound; Horvitz-Thompson estimator; Moore-Penrose inverse.

1. Introduction

Let s be a sample drawn from a population of size N according to a sampling plan p , let $\mathbf{y} = (y_1, y_2, \dots, y_N)'$ be a vector of interest, and let $\mathbf{c} = (c_1, c_2, \dots, c_N)'$ be a vector of known constants. The parameter to estimate is $\theta = \mathbf{y}'\mathbf{c}$. A commonly used estimator is that of Horvitz and Thompson (1952). This estimator can be written $\hat{\theta}_{HT} = \mathbf{y}'\mathbf{W}_s\mathbf{HT}\mathbf{c}$, where $\mathbf{W}_s\mathbf{HT} = \mathbf{\Delta}_s(E(\mathbf{\Delta}_s))^{-1}$ with $\mathbf{\Delta}_s \in \mathbb{R}^{N \times N}$ the diagonal matrix of the δ_k , $k = 1, 2, \dots, N$, with δ_k equal to 1 if unit $k \in s$ and 0 otherwise (it is assumed that $E(\delta_k) = \pi_k > 0$, $k = 1, 2, \dots, N$). The weight matrix $\mathbf{W}_s\mathbf{HT}$ is diagonal. Even in the absence of auxiliary data, other useful estimators exist. An estimator of the form $\mathbf{y}'\mathbf{W}_s\mathbf{c}$, where \mathbf{W}_s is not necessarily diagonal, will be proposed. An unbiased estimator is wanted, thus $E(\mathbf{W}_s) = \mathbf{I}_N$ will be required, where \mathbf{I}_N is the identity matrix of order N . Not requiring the weight matrix \mathbf{W}_s to be diagonal could prove useful if the variance matrix of \mathbf{y} is not diagonal. For example, from the frame, it could be known that units 1 and 2 are twins, that $y_1 = y_2 = y_{twin}$, without knowing the value y_{twin} . Noting $\pi_{kl} = E(\delta_k\delta_l)$, an alternative to the Horvitz-Thompson estimator $\sum_{k=1}^N \frac{\delta_k y_k}{\pi_k}$ for the population total, is the unbiased estimator $\frac{2y_{twin}(\delta_1 + \delta_2 - \delta_1\delta_2)}{\pi_1 + \pi_2 - \pi_{12}} + \sum_{k=3}^N \frac{\delta_k y_k}{\pi_k}$. That is, if either unit 1 or unit 2 are sampled, the

¹ 1221 du Gouverneur, Ottawa, Ontario K4C 1E2, Canada. Email: alain.theberge1@gmail.com

Acknowledgments: The author wishes to thank the Associate Editor and four referees for their valuable comments and suggestions that helped to improve the quality of the article significantly.

value $2y_{rwin}$ is given a weight equal to the inverse of the probability of selecting either of the two units. The number of twins in the sample being random, this will add to the variance of this estimator. However, if both this new estimator and the Horvitz-Thompson estimator are calibrated so that the sum of their weights equals the population size, then the calibrated new estimator is superior to the similarly calibrated Horvitz-Thompson estimator, because it makes use of the information that units 1 and 2 are twins by acknowledging that observing one of the two units is equivalent to observing both. This article will suggest estimators that can improve on the Horvitz-Thompson estimator if some of the y s are simply correlated, without necessarily being equal. For example, because of the increased risk of transmission, the incidence of the flu in two individuals from the same household are two correlated events. Depending on the variable of interest, other examples may occur for workers clustered by establishment.

In the next section, what is meant by “the variance matrix of \mathbf{y} ” is made more precise through the asymptotic setup. With the help of the Moore-Penrose inverse, a generalization of the Horvitz-Thompson estimator, $\hat{\theta}_{GHT}$, is presented in Section 3. The generalized Horvitz-Thompson estimator will depend on an estimate of the variance matrix of \mathbf{y} and will reduce to the usual Horvitz-Thompson estimator when that variance matrix estimate is diagonal. [Godambe and Joshi \(1965\)](#) gave a lower bound applicable to unbiased estimators under the assumption that the variance matrix of \mathbf{y} is diagonal. In Section 4, their result is also generalized to a positive definite variance matrix. In Section 5, the calibration problem, as stated in [Deville and Särndal \(1992\)](#) and in [Théberge \(1999\)](#), is generalized; the desired weights should be close to those of $\hat{\theta}_{GHT}$ rather than those of $\hat{\theta}_{HT}$. The solution to that problem will lead to generalized calibration estimators. Generalized calibration estimators are shown to be optimal in the sense that they asymptotically attain the generalized Godambe-Joshi lower bound. In Section 6, the problem of computing the weights of the generalized estimators is examined with an example where the variance matrix of \mathbf{y} is block diagonal. Modified versions of the generalized estimators are described in Section 7. The new estimators are compared to that of Horvitz-Thompson and to the calibration estimator in Section 8. Applications to the Canadian Reverse Record Check Survey and to the problem of variance estimation are given in Sections 9 and 10 respectively. Finally, concluding remarks are found in Section 11.

2. Asymptotic Setup

In order to discuss the large sample properties of an estimator $\hat{\theta}$, an asymptotic setup will be needed. Such setups have been described by [Brewer \(1979\)](#) and by [Isaki and Fuller \(1982\)](#). The setup shall serve two main purposes: (1) to establish a link between the setup and the variance matrix of the variable of interest, (2) to establish three results that will be useful for deriving the asymptotic properties of $\hat{\theta}_{GHT}$ and calibrated estimators. The setup described here is one that serves those purposes.

Given an auxiliary information matrix $\mathbf{X} \in \mathbb{R}^{N \times q}$ assumed to be of full rank, a sequence of increasingly large populations and samples is generated with the help of an N -dimensional distribution ξ of mean $\mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\beta} \in \mathbb{R}^q$, and variance $\boldsymbol{\Sigma}$, with $\boldsymbol{\Sigma}$ positive definite, and the sampling plan p . The sequence starts with the original population and the original sample. The t th population, of size tN , is obtained by adding N units to the $(t - 1)$ th

population. With respect to the auxiliary information, those added units are identical to the original population. The vector of interest of the added units is generated with the distribution ξ . From the added units, a sample of units is selected using the plan p , and together with the units of the $(t - 1)$ th sample, they will form the t th sample of expected size $n_t = mn$, where n is the expected size of the original sample.

More precisely, if $\mathbf{1}_{a \times b} \in \mathbb{R}^{a \times b}$ is a matrix of ones and $\mathbf{I}_a \in \mathbb{R}^{a \times a}$ is the identity matrix of dimension a , then define $\mathbf{X}_t \in \mathbb{R}^{tN \times q}$ equal to $\mathbf{1}_{t \times 1} \otimes \mathbf{X}$, the auxiliary information matrix of the t th population. Set $\mathbf{c}_t = t^{-1}(\mathbf{1}_{t \times 1} \otimes \mathbf{c})$ for estimating a mean, that is, if $\mathbf{c} = N^{-1}\mathbf{1}_{N \times 1}$, but set $\mathbf{c}_t = (\mathbf{1}_{t \times 1} \otimes \mathbf{c})$ for estimating a total, that is, if $\mathbf{c} = \mathbf{1}_{N \times 1}$. More generally, set $\mathbf{c}_t = t^{\gamma-1}(\mathbf{1}_{t \times 1} \otimes \mathbf{c})$ if $\theta = \mathbf{y}'\mathbf{c} = O_p(N^\gamma)$. Define $\mathbf{y}_t = (\mathbf{y}'_{[1]} \ \mathbf{y}'_{[2]} \ \dots \ \mathbf{y}'_{[t]})'$ and $\theta_t = \mathbf{y}'_t \mathbf{c}_t$, where the subscript $[i]$ is used to denote the N units that belong to population i , but not to

population $(i - 1)$, $\mathbf{\Delta}_t = \begin{pmatrix} \mathbf{\Delta}_{[1]} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{\Delta}_{[t]} \end{pmatrix}$ is the diagonal matrix of the δ_k , where δ_k is

equal to 1 if unit k is sampled and 0 if not (to ease the notation, in this section, the subscript s denoting the sample will be omitted from $\mathbf{\Delta}$). With this setup, $E_\xi(\mathbf{y}_t) = \mathbf{X}_t \boldsymbol{\beta}$ and $V_\xi(\mathbf{y}_t) = \mathbf{I}_t \otimes \boldsymbol{\Sigma}$.

Before presenting the asymptotic results of this section, for any matrix \mathbf{F} , let \mathbf{F}^\dagger denote the Moore-Penrose inverse, and if \mathbf{F} is positive definite, then define $\mathbf{Q}_\mathbf{F} = (E_p((\mathbf{\Delta}\mathbf{F}\mathbf{\Delta})^\dagger))^{-1}$, where $\mathbf{\Delta} = \mathbf{\Delta}_{[1]}$. It will be shown in the following section that $\mathbf{Q}_\mathbf{F}$ is well defined if and only if $\pi_k > 0$, $k = 1, 2, \dots, N$.

Let $\mathbf{T} \in \mathbb{R}^{q \times q}$ and $\mathbf{U} \in \mathbb{R}^{N \times N}$ be symmetric positive definite matrices, $\mathbf{U}_t = \mathbf{I}_t \otimes \mathbf{U}$ and

$$\hat{\boldsymbol{\beta}}_t = \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \mathbf{X}'_t (\mathbf{\Delta}_t \mathbf{U}_t \mathbf{\Delta}_t)^\dagger \mathbf{X}_t \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}'_t (\mathbf{\Delta}_t \mathbf{U}_t \mathbf{\Delta}_t)^\dagger \mathbf{y}_t. \tag{1}$$

With this asymptotic setup, the following three results hold.

RESULT 1. If the sampling plan is noninformative (see, for example [Cassel et al. 1977](#)), then $\hat{\boldsymbol{\beta}}_t \rightarrow \boldsymbol{\beta}$ in probability.

The next two results apply to a positive definite estimator, $\hat{\boldsymbol{\Sigma}}$, of $V_\xi(\mathbf{y}) = \boldsymbol{\Sigma}$. One must first

define a block diagonal matrix $\hat{\boldsymbol{\Sigma}}_t = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{[1]} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{\boldsymbol{\Sigma}}_{[t]} \end{pmatrix}$ where $\hat{\boldsymbol{\Sigma}}_{[i]}$ is the estimator of $\boldsymbol{\Sigma}$

based on the sample represented by $\mathbf{\Delta}_{[i]}$ and then define $\mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} = \left(E_p \left(\left(\mathbf{\Delta}_t \hat{\boldsymbol{\Sigma}}_t \mathbf{\Delta}_t \right)^\dagger \right) \right)^{-1}$.

RESULT 2. For a positive definite estimator $\hat{\boldsymbol{\Sigma}}$, the difference $\mathbf{X}'_t (\mathbf{\Delta}_t \hat{\boldsymbol{\Sigma}}_t \mathbf{\Delta}_t)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t \mathbf{c}_t$ is $O_p(t^{\gamma-1/2})$.

RESULT 3. For a positive definite estimator $\hat{\boldsymbol{\Sigma}}$ such that $\hat{\boldsymbol{\Sigma}} \rightarrow \boldsymbol{\Sigma}$ in probability, the difference $\mathbf{X}'_t (\mathbf{\Delta}_t \hat{\boldsymbol{\Sigma}}_t \mathbf{\Delta}_t)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t (\mathbf{\Delta}_t \boldsymbol{\Sigma}_t \mathbf{\Delta}_t)^\dagger \mathbf{Q}_{\boldsymbol{\Sigma}_t} \mathbf{c}_t$ is $o_p(t^{\gamma-1/2})$.

Since $n_t = tn$ and $N_t = tN$, what, for example, is $O_p(t^{\gamma-1/2})$, is also $O_p\left(n_t^{\gamma-1/2}\right)$ and $O_p\left(N_t^{\gamma-1/2}\right)$. The proofs of these three results can be found in Appendix A. This asymptotic setup incorporates the superpopulation model ξ ; a separate superpopulation model is not needed. This avoids possible inconsistencies between the asymptotic setup's model and that of a superpopulation.

3. A Generalization of the Horvitz-Thompson Estimator

In this section and the next section, an auxiliary data matrix is not needed, or at least, it need not be known. Only a positive definite estimate, $\hat{\Sigma}$, of the variance matrix implied by the setup, $V_{\xi}(\mathbf{y}) = \Sigma$, will be needed. In the absence of auxiliary data, the Horvitz-Thompson estimator, $\hat{\theta}_{HT} = \mathbf{y}'\Delta_s(E_p(\Delta_s))^{-1}\mathbf{c}$, is an estimator that is often used. In order to generalize the Horvitz-Thompson estimator, it will first be proven that $E_p\left(\left(\Delta_s\hat{\Sigma}\Delta_s\right)^\dagger\right)$ is nonsingular if and only if $\pi_k > 0, k = 1, 2, \dots, N$, where \mathbf{F}^\dagger denotes the Moore-Penrose inverse of the matrix \mathbf{F} .

LEMMA 1. *If $\mathbf{F}_i \in \mathbb{R}^{N \times N} i = 1, 2, \dots, K$ are symmetric positive semi-definite matrices, then the null space of $\mathbf{F} = \sum_{i=1}^K \mathbf{F}_i$, noted $\mathcal{N}(\mathbf{F})$, equals $\cap_{i=1}^K \mathcal{N}(\mathbf{F}_i)$.*

The proof of Lemma 1 is given in Appendix A. From Ben-Israel and Greville (2002, Exercise 2.38), $\mathcal{N}\left(\left(\Delta_s\hat{\Sigma}\Delta_s\right)^\dagger\right) = \mathcal{N}\left(\Delta_s\hat{\Sigma}\Delta_s\right) = \mathcal{N}(\Delta_s)$. Applying Lemma 1, the matrix sum $E_p\left(\left(\Delta_s\hat{\Sigma}\Delta_s\right)^\dagger\right)$ is invertible if and only if $\cap_{\text{all samples } s} \mathcal{N}(\Delta_s) = \mathbf{0}$, that is, $\pi_k > 0, k = 1, 2, \dots, N$.

Note $\mathbf{Q}_{\hat{\Sigma}} = \left(E_p\left(\left(\Delta_s\hat{\Sigma}\Delta_s\right)^\dagger\right)\right)^{-1}$ and define

$$\begin{aligned} \hat{\theta}_{GHT} &= \mathbf{y}'\left(\Delta_s\hat{\Sigma}\Delta_s\right)^\dagger \mathbf{Q}_{\hat{\Sigma}}\mathbf{c} \\ &= \mathbf{y}'\mathbf{W}_{sGHT}\mathbf{c}. \end{aligned} \tag{2}$$

It is readily seen that regardless of the choice of $\hat{\Sigma}$, $\hat{\theta}_{GHT}$ is unbiased for estimating $\theta = \mathbf{y}'\mathbf{c}$. Also, although $\hat{\theta}_{GHT}$ depends on $\hat{\Sigma}$, it does not depend on auxiliary data. As required of an estimator, the rows of \mathbf{W}_{sGHT} corresponding to nonsampled units are all $\mathbf{0}$, that is, $\mathbf{W}_{sGHT} = \Delta_s\mathbf{W}_{sGHT}$. This follows from the following lemma, also proven in Appendix A, and the fact that Δ_s is an orthogonal projection, that is, Δ_s is symmetric and $\Delta_s^2 = \Delta_s$.

LEMMA 2. *Let $\mathbf{P} \in \mathbb{R}^{N \times N}$ be an orthogonal projection;*

- (a) *If $\mathbf{F} \in \mathbb{R}^{q \times N}$, then $(\mathbf{FP})^\dagger = \mathbf{P}(\mathbf{FP})^\dagger$.*
- (b) *If $\mathbf{F} \in \mathbb{R}^{N \times q}$, then $(\mathbf{PF})^\dagger = (\mathbf{PF})^\dagger\mathbf{P}$.*
- (c) *If $\mathbf{F} \in \mathbb{R}^{N \times N}$, then $(\mathbf{PFP})^\dagger = \mathbf{P}(\mathbf{PFP})^\dagger\mathbf{P}$.*

Thus, if $\mathbf{W}_{[s]GHT} \in \mathbb{R}^{n \times N}$ and $\mathbf{y}_{[s]} \in \mathbb{R}^n$ are the submatrices of $\mathbf{W}_s GHT$ and \mathbf{y} respectively, with rows corresponding to the sampled units, then

$$\hat{\theta}_{GHT} = \mathbf{y}'_{[s]} \mathbf{W}_{[s]GHT} \mathbf{c}. \tag{3}$$

It will be shown in Section 4 that among linear unbiased estimators $\hat{\theta}$, $\hat{\theta}_{GHT}$ minimizes $E_p V_{\xi}(\hat{\theta})$.

If $\hat{\Sigma}$ is a diagonal matrix, then $\hat{\theta}_{GHT}$ reduces to the Horvitz-Thompson estimator. Because, for diagonal matrices $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^{N \times N}$, $(\mathbf{F}_1 \mathbf{F}_2)^\dagger = \mathbf{F}_1^\dagger \mathbf{F}_2^\dagger$, diagonal matrices permute, $\Delta_s^\dagger = \Delta_s$, $\Delta_s^2 = \Delta_s$, and because $\hat{\Sigma}^\dagger = \hat{\Sigma}^{-1}$, it follows that

$$\begin{aligned} \hat{\theta}_{GHT} &= \mathbf{y}' (\Delta_s \hat{\Sigma} \Delta_s)^\dagger \left(E_p (\Delta_s \hat{\Sigma} \Delta_s)^\dagger \right)^{-1} \mathbf{c} \\ &= \mathbf{y}' \Delta_s \hat{\Sigma}^{-1} \hat{\Sigma} (E_p (\Delta_s))^{-1} \mathbf{c} \\ &= \mathbf{y}' \Delta_s (E_p (\Delta_s))^{-1} \mathbf{c} \\ &= \hat{\theta}_{HT}. \end{aligned} \tag{4}$$

Somewhat more generally, if for every possible sample s , $\Delta_s \hat{\Sigma} \Delta_s$ is diagonal, then $\hat{\theta}_{GHT}$ will also reduce to the Horvitz-Thompson estimator.

Note that the Horvitz-Thompson estimator, which uses a diagonal $\hat{\Sigma}$, is unbiased, even if a more appropriate estimate of Σ would have a more complex structure; $\hat{\theta}_{GHT}$ is just as forgiving.

When $\hat{\Sigma}$ is diagonal, so is $\mathbf{W}_s GHT$. The weights on the diagonal, the Horvitz-Thompson weights, are often referred to as the design weights. If $\hat{\Sigma}$ is not diagonal, then the weight matrix $\mathbf{W}_s GHT$ and the vector $\mathbf{W}_s GHT \mathbf{c}$ depend on both the sampling design and on $\hat{\Sigma}$. It may not be appropriate to refer to $\mathbf{W}_s GHT$ or $\mathbf{W}_s GHT \mathbf{c}$ as design weights.

The following is a simple consequence of Result 3 and will be needed before discussing the variance of $\hat{\theta}_{GHT}$.

RESULT 4. If $\hat{\Sigma}$ is a positive definite estimator and $\hat{\Sigma} \rightarrow \Sigma$ in probability, then $\hat{\theta}_{GHT} = \mathbf{y}' (\Delta_s \hat{\Sigma} \Delta_s)^\dagger \mathbf{Q}_{\hat{\Sigma}} \mathbf{c}$ and $\theta_{GHT} = \mathbf{y}' (\Delta_s \Sigma \Delta_s)^\dagger \mathbf{Q}_{\Sigma} \mathbf{c}$ are asymptotically equivalent.

Under the conditions of the preceding result, one has

$$\begin{aligned} V_p(\hat{\theta}_{GHT}) &\doteq V_p(\mathbf{y}' (\Delta_s \hat{\Sigma} \Delta_s)^\dagger \mathbf{Q}_{\hat{\Sigma}} \mathbf{c}) \\ &= V_p(\text{vec}(\mathbf{y}' (\Delta_s \hat{\Sigma} \Delta_s)^\dagger \mathbf{Q}_{\hat{\Sigma}} \mathbf{c})) \\ &= (\mathbf{Q}_{\hat{\Sigma}} \mathbf{c} \otimes \mathbf{y})' V_p(\text{vec}((\Delta_s \hat{\Sigma} \Delta_s)^\dagger)) (\mathbf{Q}_{\hat{\Sigma}} \mathbf{c} \otimes \mathbf{y}) \\ &= \|\mathbf{Q}_{\hat{\Sigma}} \mathbf{c} \otimes \mathbf{y}\|_{V_p(\text{vec}((\Delta_s \hat{\Sigma} \Delta_s)^\dagger))}^2, \end{aligned} \tag{5}$$

where $\text{vec}(\mathbf{F})$ denotes the vector obtained by stacking the successive columns of the matrix \mathbf{F} .

In practice, Σ is unknown. The statistician will simply assume that a matrix Σ of a certain structure reflects the correlations among the population units. The matrix may, or

may not, depend on certain parameters that need to be estimated. For example, in the concluding section, Σ is a block diagonal matrix where each block equals $\mathbf{1}_{2 \times 2}$; there are no parameters to estimate. In Section 9, the correlation between persons of a same household is estimated; however, in that particular example, the estimated correlation is not used directly; a compromise that works for two important variables of interest is chosen. The computation of $\hat{\theta}_{GHT}$ also requires the computation of $\mathbf{Q}_{\hat{\Sigma}}$. Although it can be difficult to find a closed form expression for $\mathbf{Q}_{\hat{\Sigma}} = \left(E_p \left(\left(\Delta_s \hat{\Sigma} \Delta_s \right)^\dagger \right) \right)^{-1}$, its value can be approximated by repeatedly sampling the population using the sampling plan p , computing the average over the samples of $(\Delta_s \hat{\Sigma} \Delta_s)^\dagger$, and inverting that average. If $\hat{\Sigma}$ varies with the sample s , then it would not be possible to compute $(\Delta_s \hat{\Sigma} \Delta_s)^\dagger$ for all the samples. The alternative is to fix $\hat{\Sigma}$ to the estimate obtained for the sample effectively drawn, then $\hat{\theta}_{GHT}$ will be biased, but still asymptotically unbiased. In the case of a two-stage sampling plan, the Horvitz-Thompson weights would likely be applied to the primary sampling units and the methods of this article, including the method just described to compute $\mathbf{Q}_{\hat{\Sigma}}$, would apply to the secondary sampling units. For that purpose, the population consists of the secondary sampling units that belong to the primary sampling units selected in the first stage. For that “population”, Σ would typically be block diagonal with each block corresponding to a selected primary sampling unit.

4. A Generalization of the Godambe and Joshi Lower Bound

Although it wasn't in the context of an asymptotic setup and although it was assumed that $V_\xi(\mathbf{y}) = \Sigma$ was diagonal, for any unbiased estimator $\hat{\theta}$ of θ , [Godambe and Joshi \(1965\)](#) have given a lower bound for the value of $E_\xi V_p(\hat{\theta})$. The derivation of that lower bound used the following identity:

$$E_\xi V_p(\hat{\theta}) = E_p V_\xi(\hat{\theta}) + E_p [E_\xi(\hat{\theta} - \theta)]^2 - V_\xi(\theta). \quad (6)$$

Also, for any linear unbiased estimator $\hat{\theta}$ of θ

$$\begin{aligned} E_p V_\xi(\hat{\theta}) &= E_p V_\xi(\theta_{GHT}^* + (\hat{\theta} - \theta_{GHT}^*)) \\ &= E_p V_\xi(\theta_{GHT}^*) + E_p V_\xi(\hat{\theta} - \theta_{GHT}^*) + 2 E_p \text{Cov}_\xi(\theta_{GHT}^*, (\hat{\theta} - \theta_{GHT}^*)) \\ &\geq E_p V_\xi(\theta_{GHT}^*) + 2 E_p \text{Cov}_\xi(\theta_{GHT}^*, (\hat{\theta} - \theta_{GHT}^*)) \\ &= E_p V_\xi(\theta_{GHT}^*), \end{aligned} \quad (7)$$

because for any linear unbiased estimator $\hat{\theta}$ of θ , $E_p \text{Cov}_\xi(\theta_{GHT}^*, \hat{\theta}) = 0$. To show this, let $\hat{\theta}$ be written $\mathbf{y}' \Delta_s \boldsymbol{\lambda}_{0s} + \kappa_s$, with $\boldsymbol{\lambda}_{0s} \in \mathbb{R}^N$ and κ_s independent of \mathbf{y} , may depend on the sample s . Setting $\mathbf{y} = \mathbf{0}$ yields $E_p(\kappa_s) = 0$. The following derivation uses Lemma 2

as well as [Ben-Israel and Greville \(2002, Exercise 2.21\)](#):

$$\begin{aligned}
 E_p \text{Cov}_\xi \left(\theta_{GHT}^*, \hat{\theta} \right) &= E_p \text{Cov}_\xi \left(\mathbf{y}' (\Delta_s \Sigma \Delta_s)^\dagger \mathbf{Q}_\Sigma \mathbf{c}, \mathbf{y}' \Delta_s \boldsymbol{\lambda}_{0s} \right) \\
 &= E_p \left(\mathbf{c}' \mathbf{Q}_\Sigma (\Delta_s \Sigma \Delta_s)^\dagger \text{Cov}_\xi(\mathbf{y}, \mathbf{y}) \Delta_s \boldsymbol{\lambda}_{0s} \right) \\
 &= E_p \left(\mathbf{c}' \mathbf{Q}_\Sigma \Delta_s (\Delta_s \Sigma \Delta_s)^\dagger \Delta_s \Sigma \Delta_s \boldsymbol{\lambda}_{0s} \right) \\
 &= E_p \left(\mathbf{c}' \mathbf{Q}_\Sigma \Delta_s \boldsymbol{\lambda}_{0s} + \kappa_s \right) \\
 &= 0,
 \end{aligned} \tag{8}$$

because $\mathbf{c}' \mathbf{Q}_\Sigma \Delta_s \boldsymbol{\lambda}_{0s} + \kappa_s$ is the unbiased estimator $\hat{\theta}$ with $\mathbf{y} = \mathbf{Q}_\Sigma \mathbf{c}$ as the vector of interest.

The inequality (7) is what makes θ_{GHT}^* , and $\hat{\theta}_{GHT}$, special. Taken together with (6) it shows that for any linear unbiased estimator, $\hat{\theta}$,

$$E_\xi V_p(\hat{\theta}) \geq E_p V_\xi \left(\theta_{GHT}^* \right) - V_\xi(\theta). \tag{9}$$

Knowing that $V_\xi(\theta) = \mathbf{c}' \Sigma \mathbf{c} = \|\mathbf{c}\|_\Sigma^2$ and that

$$\begin{aligned}
 E_p V_\xi \left(\theta_{GHT}^* \right) &= E_p \left(\mathbf{c}' \mathbf{Q}_\Sigma (\Delta_s \Sigma \Delta_s)^\dagger \Sigma (\Delta_s \Sigma \Delta_s)^\dagger \mathbf{Q}_\Sigma \mathbf{c} \right) \\
 &= \mathbf{c}' \mathbf{Q}_\Sigma E_p \left((\Delta_s \Sigma \Delta_s)^\dagger \Delta_s \Sigma \Delta_s (\Delta_s \Sigma \Delta_s)^\dagger \right) \mathbf{Q}_\Sigma \mathbf{c} \\
 &= \mathbf{c}' \mathbf{Q}_\Sigma E_p (\Delta_s \Sigma \Delta_s)^\dagger \mathbf{Q}_\Sigma \mathbf{c} \\
 &= \mathbf{c}' \mathbf{Q}_\Sigma \mathbf{c} \\
 &= \|\mathbf{c}\|_{\mathbf{Q}_\Sigma}^2,
 \end{aligned} \tag{10}$$

allows the following generalization to a positive definite matrix, Σ , of a lower bound given in [Godambe and Joshi \(1965\)](#).

RESULT 5. For any linear unbiased estimator, $\hat{\theta}$ of $\theta = \mathbf{y}' \mathbf{c}$, if $V_\xi(\mathbf{y}) = \Sigma$ is positive definite, then

$$E_\xi V_p(\hat{\theta}) \geq \|\mathbf{c}\|_{\mathbf{Q}_\Sigma - \Sigma}^2. \tag{11}$$

If Σ is a diagonal matrix equal to $\text{diag}(\sigma_k^2)_{k=1,2,\dots,N}$, then (11) reduces to $E_\xi V_p(\hat{\theta}) \geq \|\mathbf{c}\|_{(E_p(\Delta_s)^{-1} - \mathbf{I}_N) \Sigma}^2 = \sum_k \left(\frac{1}{\pi_k} - 1 \right) \sigma_k^2 c_k^2$, which is the Godambe and Joshi lower bound, usually given for a total, that is, for $\mathbf{c} = \mathbf{1}_{N \times 1}$. If $\mathbf{c} = \mathbf{1}_{N \times 1}$, then the generalized lower bound equals the sum of all entries in the matrix $\mathbf{Q}_\Sigma - \Sigma$. It should be noted that [Godambe and Joshi \(1965\)](#) had proven that, if Σ is diagonal, the lower bound holds for the class of all unbiased estimators, not only for the class of linear unbiased estimators. What if Σ is allowed not to be diagonal? Does the generalized Godambe-Joshi lower bound apply to all unbiased estimators? The answer is no, and a nonlinear counter-example is given in Appendix B.

When the variance matrix Σ is diagonal, it is known that the calibration estimator introduced by [Deville and Särndal \(1992\)](#), or the equivalent generalized regression estimator, asymptotically attains the Godambe and Joshi lower bound (see, for example [Särndal et al. 1992](#)). In the next section, the calibration estimator will be generalized to the case of a positive definite variance matrix Σ . It will then be shown that this generalized calibration estimator asymptotically attains the lower bound given in (11).

5. A Generalization of the Calibration Estimator

Define $\mathbf{1}'_{N \times 1} \mathbf{W}_s \text{GHT} \mathbf{c}$ to be the effective sample weight for estimating $\theta = \mathbf{y}'\mathbf{c}$. The variance of the effective sample weight will often be larger if $\hat{\Sigma}$ is nondiagonal, and this variance will negatively affect the estimator $\hat{\theta}_{\text{GHT}}$. This source of variance can be eliminated with the use of a weight vector \mathbf{w}_s that satisfies the calibration equation $\mathbf{1}'_{N \times 1} \mathbf{w}_s = \mathbf{1}'_{N \times 1} \mathbf{c}$. In this section, to estimate $\theta = \mathbf{y}'\mathbf{c}$, an estimator $\hat{\theta}_{\text{GCAL}} = \mathbf{y}'\mathbf{w}_s \text{GCAL}$ will be derived through calibration using an auxiliary variable matrix \mathbf{X} assumed to be of full rank. More precisely, noting $\mathbf{w}_s \text{GHT} = \mathbf{W}_s \text{GHT} \mathbf{c}$, the following problem is addressed:

Calibration Problem: Among the weight vectors \mathbf{w}_s in the range of $\Delta_s, \mathcal{R}(\Delta_s)$, (nonsampled units should have a weight of 0) which minimize $\|\mathbf{X}'\mathbf{w}_s - \mathbf{X}'\mathbf{c}\|_{\mathbf{T}}$, that is, which best satisfy the calibration equations, seek one that minimizes $\|\mathbf{w}_s - \mathbf{w}_s \text{GHT}\|_{\mathbf{U}}$, that is, a weight vector as close as possible to the generalized Horvitz-Thompson weights, where $\mathbf{T} \in \mathbb{R}^{q \times q}$ and $\mathbf{U} \in \mathbb{R}^{N \times N}$ are positive semi-definite matrices.

Weights, \mathbf{w}_s , that satisfy the calibration equations, $\mathbf{X}'\mathbf{w}_s = \mathbf{X}'\mathbf{c}$, do not always exist, especially if the number of equations, q , is high relative to the sample size. To prepare for this eventuality, the matrix \mathbf{T} is at the statistician’s disposal for specifying the relative importance of the q calibration equations.

This formulation of the calibration problem generalizes that of [Théberge \(1999\)](#), where \mathbf{T} and \mathbf{U} were diagonal matrices and the Horvitz-Thompson weights were used instead of the generalized Horvitz-Thompson weights.

Setting $\mathbf{v} = \mathbf{w}_s - \mathbf{w}_s \text{GHT}$, a minimum-norm least-squares solution is sought. A helpful theorem is given in [Rao and Mitra \(1971\)](#).

THEOREM 1. *Let $\mathbf{T} \in \mathbb{R}^{q \times q}$ and $\mathbf{U} \in \mathbb{R}^{N \times N}$ be symmetric positive semi-definite matrices, also let $\mathbf{A} \in \mathbb{R}^{q \times N}$ and $\mathbf{b} \in \mathbb{R}^q$. There is a matrix $\mathbf{G} \in \mathbb{R}^{N \times q}$ such that $\mathbf{G}\mathbf{b}$ minimizes $\|\mathbf{v}\|_{\mathbf{U}}$ among the vectors $\mathbf{v} \in \mathbb{R}^N$ which minimize $\|\mathbf{A}\mathbf{v} - \mathbf{b}\|_{\mathbf{T}}$, if and only if*

$$\mathbf{T}\mathbf{A}\mathbf{G} = \mathbf{T}\mathbf{A} \quad \mathbf{U}\mathbf{G}\mathbf{A} = \mathbf{U}\mathbf{G} \quad \mathbf{T}\mathbf{A}\mathbf{G}' = (\mathbf{T}\mathbf{A}\mathbf{G}')' \quad \mathbf{U}\mathbf{G}\mathbf{A}' = (\mathbf{U}\mathbf{G}\mathbf{A}')'. \tag{12}$$

Choices for \mathbf{G} are

$$\begin{aligned} \mathbf{G} &= (\mathbf{I}_N - (\mathbf{P}_{\mathcal{N}(\mathbf{A}'\mathbf{T}\mathbf{A})} \mathbf{U} \mathbf{P}_{\mathcal{N}(\mathbf{A}'\mathbf{T}\mathbf{A})})^\dagger \mathbf{U}) (\mathbf{A}'\mathbf{T}\mathbf{A})^\dagger \mathbf{A}'\mathbf{T}, \\ &= \mathbf{U}^{-1} \mathbf{A}'\mathbf{T}^{1/2} (\mathbf{T}^{1/2} \mathbf{A}\mathbf{U}^{-1} \mathbf{A}'\mathbf{T}^{1/2})^\dagger \mathbf{T}^{1/2} \quad \text{if } \mathbf{U} \text{ is positive definite,} \end{aligned} \tag{13}$$

where $\mathbf{T}^{1/2}$ is the symmetric positive semi-definite square root of \mathbf{T} and \mathbf{P}_S is the orthogonal projection on S , a subspace of \mathbb{R}^N .

The first part of the theorem is proven in [Rao and Mitra \(1971\)](#), where other choices for \mathbf{G} are given. It is shown in Appendix C that (13) does satisfy (12). To compute $\mathbf{P}_{\mathcal{N}(\mathbf{A}'\mathbf{T}\mathbf{A})}$, the identity $\mathbf{P}_{\mathcal{N}(\mathbf{F})} = \mathbf{I} - \mathbf{F}^\dagger \mathbf{F}$ can be used.

The first choice of \mathbf{G} given in (13) is derived from Ben-Israel and Greville (2002, Corollary 8.2) which is itself a consequence of the generalized Gauss-Markov theorem, see Zyskind and Martin (1969) and Albert (1973). The second choice of \mathbf{G} given in (13) is derived from Théberge (1999).

If \mathbf{U} is positive definite, there are two other possible forms for \mathbf{G} , namely

$$\mathbf{G} = \mathbf{U}^{-1/2}(\mathbf{T}^{1/2}\mathbf{A}\mathbf{U}^{-1/2})^\dagger\mathbf{T}^{1/2} \quad (14)$$

and

$$\mathbf{G} = \mathbf{U}^{-1/2}(\mathbf{U}^{-1/2}\mathbf{A}'\mathbf{T}\mathbf{A}\mathbf{U}^{-1/2})^\dagger\mathbf{U}^{-1/2}\mathbf{A}'\mathbf{T}, \quad (15)$$

where $\mathbf{U}^{1/2}$ is the symmetric positive semi-definite square root of \mathbf{U} . Applying the identity $\mathbf{F}^\dagger = \mathbf{F}'(\mathbf{F}\mathbf{F}')^\dagger$ to the Moore-Penrose inverse on the right-hand side of (14) yields the second part of (13); applying the identity $\mathbf{F}^\dagger = (\mathbf{F}'\mathbf{F})^\dagger\mathbf{F}'$ to the Moore-Penrose inverse on the right-hand side of (14) yields (15). Both of those identities are found in Ben-Israel and Greville (2002, Exercise 1.18).

The theorem must be modified to take into account that \mathbf{v} is constrained to $\mathcal{R}(\Delta_s)$, a subspace S of \mathbb{R}^N . This is done by applying the method of Ben-Israel and Greville (2002, sec. 2.9) and minimizing $\|\mathbf{v}\|_{\mathbf{U}} = \|\mathbf{P}_S\mathbf{z}\|_{\mathbf{U}} = \|\mathbf{z}\|_{\mathbf{P}_S\mathbf{U}\mathbf{P}_S}$ among the vectors which minimize $\|\mathbf{A}\mathbf{P}_S\mathbf{z} - \mathbf{b}\|_{\mathbf{T}}$. Using the preceding theorem to find the optimal \mathbf{z} , gives the constrained analog:

THEOREM 2. Let $\mathbf{T} \in \mathbb{R}^{q \times q}$ and $\mathbf{U} \in \mathbb{R}^{N \times N}$ be symmetric positive semi-definite matrices, $\mathbf{A} \in \mathbb{R}^{q \times N}$, $\mathbf{b} \in \mathbb{R}^q$ and \mathbf{P}_S be the orthogonal projection on S a subspace of \mathbb{R}^N . There is a matrix $\mathbf{G} \in \mathbb{R}^{N \times q}$ such that $\mathbf{G}\mathbf{b}$ minimizes $\|\mathbf{v}\|_{\mathbf{U}}$ among the vectors $\mathbf{v} \in S$ which minimize $\|\mathbf{A}\mathbf{v} - \mathbf{b}\|_{\mathbf{T}}$, if and only if

$$\begin{aligned} \mathbf{T}\mathbf{A}\mathbf{P}_S\mathbf{G}\mathbf{A}\mathbf{P}_S &= \mathbf{T}\mathbf{A}\mathbf{P}_S & \mathbf{P}_S\mathbf{U}\mathbf{P}_S\mathbf{G}\mathbf{A}\mathbf{P}_S\mathbf{G} &= \mathbf{P}_S\mathbf{U}\mathbf{P}_S\mathbf{G} \\ \mathbf{T}\mathbf{A}\mathbf{P}_S\mathbf{G} &= (\mathbf{T}\mathbf{A}\mathbf{P}_S\mathbf{G})' & \mathbf{P}_S\mathbf{U}\mathbf{P}_S\mathbf{G}\mathbf{A}\mathbf{P}_S &= (\mathbf{P}_S\mathbf{U}\mathbf{P}_S\mathbf{G}\mathbf{A}\mathbf{P}_S)'. \end{aligned} \quad (16)$$

Choices for \mathbf{G} are

$$\begin{aligned} \mathbf{G} &= (\mathbf{I}_N - (\mathbf{P}_{\mathcal{N}(\mathbf{P}_S\mathbf{A}'\mathbf{T}\mathbf{A}\mathbf{P}_S)}\mathbf{P}_S\mathbf{U}\mathbf{P}_S\mathbf{P}_{\mathcal{N}(\mathbf{P}_S\mathbf{A}'\mathbf{T}\mathbf{A}\mathbf{P}_S)})^\dagger\mathbf{P}_S\mathbf{U}\mathbf{P}_S)(\mathbf{P}_S\mathbf{A}'\mathbf{T}\mathbf{A}\mathbf{P}_S)^\dagger\mathbf{A}'\mathbf{T}, \\ &= (\mathbf{P}_S\mathbf{U}\mathbf{P}_S)^\dagger\mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{A}(\mathbf{P}_S\mathbf{U}\mathbf{P}_S)^\dagger\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \quad \text{if } \mathbf{U} \text{ is positive definite.} \end{aligned} \quad (17)$$

Applying this to our calibration problem stated at the beginning of this section means setting $\mathbf{P}_S = \Delta_s$, $\mathbf{b} = \mathbf{X}'\mathbf{c} - \mathbf{X}'\mathbf{w}_s\text{GHT}$ and $\mathbf{A} = \mathbf{X}'$. This gives the generalized calibration weight vector

$$\mathbf{w}_s\text{GCAL} = \mathbf{w}_s\text{GHT} + \mathbf{G}(\mathbf{X}'\mathbf{c} - \mathbf{X}'\mathbf{w}_s\text{GHT}), \quad (18)$$

and the calibration estimator becomes

$$\begin{aligned} \hat{\theta}_{\text{GCAL}} &= \mathbf{y}'\mathbf{w}_s\text{GCAL} \\ &= \hat{\mathbf{y}}'\mathbf{c} + (\mathbf{y} - \hat{\mathbf{y}})'\mathbf{w}_s\text{GHT}, \end{aligned} \quad (19)$$

where $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = \mathbf{G}'\mathbf{y}$ and \mathbf{G} is given by (17) with $\mathbf{P}_S = \Delta_s$ and $\mathbf{A} = \mathbf{X}'$.

It should be noted that the weight vector \mathbf{w}_{sGCAL} can be written in the form $\mathbf{W}_{sGCAL}\mathbf{c}$ with $\mathbf{W}_{sGCAL} \in \mathbb{R}^{N \times N}$. For (18) to hold true for any vector \mathbf{c} , one must have $\mathbf{W}_{sGCAL} = \mathbf{W}_{sGHT} + \mathbf{G}(\mathbf{X}' - \mathbf{X}'\mathbf{W}_{sGHT})$.

In the remainder of this section, it will be assumed that \mathbf{U} is positive definite and the second choice for \mathbf{G} given in (17) will be used. Then, one can write

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{X}\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{X}'(\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^\dagger\mathbf{X}\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2}\mathbf{X}'(\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^\dagger\mathbf{y}\end{aligned}\quad (20)$$

and

$$\begin{aligned}\mathbf{w}_{sGCAL} &= \mathbf{w}_{sGHT} + (\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^\dagger\mathbf{X}\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{X}'(\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^\dagger\mathbf{X}\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \\ &\quad \times (\mathbf{X}'\mathbf{c} - \mathbf{X}'\mathbf{w}_{sGHT}).\end{aligned}\quad (21)$$

If one notes $\mathbf{w}_{[s]GCAL} \in \mathbb{R}^n$, $\mathbf{w}_{[s]GHT} \in \mathbb{R}^n$, $\mathbf{U}_{[ss]} \in \mathbb{R}^{n \times n}$ and $\mathbf{X}_{[s]} \in \mathbb{R}^{n \times q}$, the subvectors and submatrices with lines corresponding to the sampled units, then using Lemma 2,

$$\begin{aligned}\mathbf{w}_{[s]GCAL} &= \mathbf{w}_{[s]GHT} + \mathbf{U}_{[ss]}^{-1}\mathbf{X}_{[s]}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{X}'_{[s]}\mathbf{U}_{[ss]}^{-1}\mathbf{X}_{[s]}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \\ &\quad \times (\mathbf{X}'\mathbf{c} - \mathbf{X}'_{[s]}\mathbf{w}_{[s]GHT}),\end{aligned}\quad (22)$$

because the weights of nonsampled units are zero. Thus, $\hat{\boldsymbol{\theta}}_{GCAL} = \mathbf{y}'_{[s]}\mathbf{w}_{[s]GCAL}$. This shows that for computing $\hat{\boldsymbol{\theta}}_{GCAL}$ the population parameter $\mathbf{X}'\mathbf{c}$ must be known, but the individual rows of \mathbf{X} need only be known for those corresponding to sampled units. It is seen that the weights given by (22) could be interpreted as those from a GREG estimator, see [Cassel et al. \(1977\)](#), except that the Horvitz-Thompson weights are replaced with those of the generalized Horvitz-Thompson estimator, a matrix \mathbf{T} has been introduced in case $\mathbf{X}_{[s]}$ is not of full rank and the matrix \mathbf{U} would be set equal to $\mathbf{Q}_{\hat{\boldsymbol{\Sigma}}}$ in a GREG estimator. For $\hat{\boldsymbol{\Sigma}}$ diagonal, $\mathbf{Q}_{\hat{\boldsymbol{\Sigma}}}$ reduces to $\hat{\boldsymbol{\Sigma}}(E_p(\boldsymbol{\Delta}_s))^{-1}$. Equation (20) with $\mathbf{U} = \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}}$ is thus a generalization of the value of $\hat{\mathbf{y}}$ for a GREG estimator when $\hat{\boldsymbol{\Sigma}}$, the estimated variance matrix of \mathbf{y} under the model, is not necessarily diagonal.

If $\mathbf{U}^{1/2}$ is the unique positive definite square root of \mathbf{U} , then defining $\mathbf{Z}_{[s]} = \mathbf{U}_{[ss]}^{-1/2}\mathbf{X}_{[s]}$ yields

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{GCAL} &= \mathbf{y}'_{[s]}\mathbf{w}_{[s]GCAL} \\ &= \mathbf{y}'_{[s]}\mathbf{w}_{[s]GHT} + \left(\mathbf{U}_{[ss]}^{-1/2}\mathbf{y}_{[s]}\right)' \mathbf{Z}_{[s]}'\mathbf{T}^{1/2} \left(\mathbf{T}^{1/2}\mathbf{Z}'_{[s]}\mathbf{Z}_{[s]}'\mathbf{T}^{1/2}\right)^\dagger \\ &\quad \times \mathbf{T}^{1/2}(\mathbf{X}'\mathbf{c} - \mathbf{X}'_{[s]}\mathbf{w}_{[s]GHT}).\end{aligned}\quad (23)$$

If \mathbf{T} is also positive definite and if $\mathbf{X}_{[s]}$ is of full rank, then (23) simplifies to

$$\hat{\boldsymbol{\theta}}_{GCAL} = \mathbf{y}'_{[s]}\mathbf{w}_{[s]GHT} + \left(\mathbf{U}_{[ss]}^{-1/2}\mathbf{y}_{[s]}\right)' \mathbf{Z}_{[s]}'(\mathbf{Z}'_{[s]}\mathbf{Z}_{[s]})^{-1}(\mathbf{X}'\mathbf{c} - \mathbf{X}'_{[s]}\mathbf{w}_{[s]GHT}). \quad (24)$$

In this form, a parallel can be drawn with the use of “instrumental variables”, as for example, in [Estevao and Särndal \(2003\)](#).

Replacing $\hat{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}$ in $\hat{\theta}_{GCAL}$ and noting $\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta}$ gives the random variable $\theta_{GCAL}^* = (\mathbf{y}^*)' \mathbf{c} + (\mathbf{y} - \mathbf{y}^*)' \mathbf{w}_{s\,GHT}$. The bias of θ_{GCAL}^* is zero. Also, $\hat{\theta}_{GCAL}$ and θ_{GCAL}^* are asymptotically equivalent. Indeed,

$$\begin{aligned} t^{1/2-\gamma}(\hat{\theta}_{GCALt} - \theta_{GCALt}^*) &= t^{1/2-\gamma}(\mathbf{y}_t^* - \hat{\mathbf{y}}_t)' \left((\Delta_t \hat{\boldsymbol{\Sigma}}_t \Delta_t)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{c}_t \right) \\ &= t^{1/2-\gamma}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_t)' \left[\mathbf{X}'_t (\Delta_t \hat{\boldsymbol{\Sigma}}_t \Delta_t)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t \mathbf{c}_t \right] \end{aligned} \tag{25}$$

tends to 0 in probability since, from the results of Section 2, $\mathbf{X}'_t (\Delta_t \hat{\boldsymbol{\Sigma}}_t \Delta_t)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t \mathbf{c}_t$ is $O_p(t^{\gamma-1/2})$ and $\hat{\boldsymbol{\beta}}_t \rightarrow \boldsymbol{\beta}$ in probability. This leads to the following result.

RESULT 6. For any positive definite matrix \mathbf{U} , if the calibration equations can be satisfied, $E_\xi V_p(\theta_{GCAL}^*)$ attains the lower bound given in (11).

To prove this, use (6) with $\hat{\theta} = \theta_{GCAL}^*$ while noting that $E_\xi(\theta_{GCAL}^* - \theta) = 0$, that $V_\xi(\theta) = \|\mathbf{c}\|_{\boldsymbol{\Sigma}}^2$, and that $E_p V_\xi(\theta_{GCAL}^*) = E_p V_\xi(\hat{\theta}_{GHT}) \doteq E_p V_\xi(\theta_{GHT}^*) = \|\mathbf{c}\|_{\mathbf{Q}_\Sigma}^2$. It should be noted that generally, $\hat{\theta}_{GHT}$ does not attain the lower bound; calibration is required.

For the generalized calibration estimator to asymptotically attain the lower bound, it is important for the generalized Horvitz-Thompson weights, $\mathbf{w}_{s\,GHT}$, to be calculated with a matrix $\hat{\boldsymbol{\Sigma}}$ that satisfies the conditions of Result 3. Also, the same auxiliary variables as appear in the model are to be used for the calibration, so that $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}$. Note that $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}$ whatever the choice of the positive definite matrix \mathbf{U} , and the choice has no impact on whether or not the generalized calibration estimator asymptotically attains the lower bound. The case of a variance matrix $\boldsymbol{\Sigma}$ which is not diagonal has been examined before, see, for example, [Montanari and Ranalli \(2002\)](#). The focus is usually on the choice of the estimator $\hat{\boldsymbol{\beta}}$, or more precisely on the choice of the matrix \mathbf{U} . Result 6 puts the importance of \mathbf{U} in perspective.

It was seen in Section 3 that if $\hat{\boldsymbol{\Sigma}}$ is diagonal, then $\mathbf{w}_{s\,GHT}$ reduces to the usual Horvitz-Thompson weights. If the matrices \mathbf{U} and \mathbf{T} are also chosen to be diagonal, then the generalized calibration estimator reduces to the usual calibration estimator as given in [Théberge \(1999\)](#).

Assuming that $\hat{\boldsymbol{\Sigma}}$ satisfies the conditions of Result 3, the variance of the generalized calibration estimator is

$$\begin{aligned} V_p(\hat{\theta}_{GCAL}) &\doteq V_p(\theta_{GCAL}^*) \\ &= V_p((\mathbf{y} - \mathbf{y}^*)' \mathbf{w}_{s\,GHT}) \\ &\doteq \|\mathbf{Q}_\Sigma \mathbf{c} \otimes (\mathbf{y} - \mathbf{y}^*)\|_{V_p(\text{vec}((\Delta_t \boldsymbol{\Sigma} \Delta_t)^\dagger))}^2 \end{aligned} \tag{26}$$

Theorem 1 may also be used to find an optimal vector $\boldsymbol{\beta}$; one which minimizes the variance. Among the vectors $\boldsymbol{\beta}$ which minimize

$$\begin{aligned} V_p(\hat{\theta}_{GCAL}) &\doteq \|\mathbf{Q}_{\Sigma}\mathbf{c} \otimes (\mathbf{y} - \mathbf{y}^*)\|_{V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))}^2 \\ &= \|(\mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{X})\boldsymbol{\beta} - \mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{y}\|_{V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))}^2, \end{aligned} \quad (27)$$

the one that minimizes $\|\boldsymbol{\beta}\|_{\mathbf{U}}^2$ can be found by applying Theorem 1 with $\mathbf{A} = \mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{X}$, $\mathbf{b} = \mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{y}$, $\mathbf{T} = V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))$ and by using \mathbf{G} given by (15). This gives

$$\begin{aligned} \boldsymbol{\beta}_{opt} &= \mathbf{U}^{-1/2}[\mathbf{U}^{-1/2}(\mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{X})' V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))(\mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{X})\mathbf{U}^{-1/2}]^\dagger \mathbf{U}^{-1/2} \\ &\quad \times (\mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{X})' V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))(\mathbf{Q}_{\Sigma}\mathbf{c} \otimes \mathbf{y}). \end{aligned} \quad (28)$$

If Σ is diagonal, then $V_p(\hat{\theta}_{GCAL}) \doteq \|\text{diag}(\mathbf{c})(\mathbf{y} - \mathbf{y}^*)\|_{\mathbf{A}\mathbf{\Pi}\mathbf{A} - \mathbf{1}_{N \times N}}^2 = \|\text{diag}(\mathbf{c})\mathbf{X}\boldsymbol{\beta} - \text{diag}(\mathbf{c})\mathbf{y}\|_{\mathbf{A}\mathbf{\Pi}\mathbf{A} - \mathbf{1}_{N \times N}}^2$, where $\text{diag}(\mathbf{v})$ denotes the diagonal matrix formed from the vector \mathbf{v} , $\mathbf{\Pi}$ is the matrix of the second-order inclusion probabilities and $\mathbf{A} = (E(\Delta_s))^{-1}$. Applying Theorem 1, again with \mathbf{G} given by (15), to find the optimal $\boldsymbol{\beta}$ will give

$$\begin{aligned} \boldsymbol{\beta}_{opt} &= \mathbf{U}^{-1/2}[\mathbf{U}^{-1/2}\mathbf{X}'\text{diag}(\mathbf{c})(\mathbf{A}\mathbf{\Pi}\mathbf{A} - \mathbf{1}_{N \times N})\text{diag}(\mathbf{c})\mathbf{X}\mathbf{U}^{-1/2}]^\dagger \mathbf{U}^{-1/2} \\ &\quad \times \mathbf{X}'\text{diag}(\mathbf{c})(\mathbf{A}\mathbf{\Pi}\mathbf{A} - \mathbf{1}_{N \times N})\text{diag}(\mathbf{c})\mathbf{y}, \end{aligned} \quad (29)$$

a result similar to that found in Montanari (1998), if one sets $\mathbf{U} = \mathbf{I}$ and $\mathbf{c} = \mathbf{1}_{N \times 1}$.

The variance $V_p(\hat{\theta}_{GCAL})$ is an unbiased estimator, under ξ , of $E_\xi V_p(\hat{\theta}_{GCAL}) \doteq \|\mathbf{c}\|_{\mathbf{Q}_{\Sigma} - \Sigma}^2$. It is then possible for $V_p(\hat{\theta}_{GCAL})$ to be smaller than the lower bound $E_\xi V_p(\hat{\theta}_{GCAL}) \doteq \|\mathbf{c}\|_{\mathbf{Q}_{\Sigma} - \Sigma}^2$. Also, for any other linear unbiased estimator, $\hat{\theta}$, the variance $V_p(\hat{\theta})$ is an unbiased estimator, under ξ , of $E_\xi V_p(\hat{\theta}) \geq \|\mathbf{c}\|_{\mathbf{Q}_{\Sigma} - \Sigma}^2$. The fact that $V_p(\hat{\theta}_{GCAL})$ is an unbiased estimator of a parameter not greater than the parameter estimated by $V_p(\hat{\theta})$ is not a guarantee that $V_p(\hat{\theta}_{GCAL})$ is not greater than $V_p(\hat{\theta})$, but it is a point in favor of $\hat{\theta}_{GCAL}$.

For a population parameter, $\boldsymbol{\Omega}$, one could use the asymptotic setup to define $\lim_{U \rightarrow U_\infty} \boldsymbol{\Omega} = E_\xi(\boldsymbol{\Omega})$. For example, with $\mathbf{U} \in \mathbb{R}^{N \times N}$ a symmetric positive definite matrix, a population regression parameter $\mathbf{B} = (\mathbf{X}'\mathbf{Q}_U^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_U^{-1}\mathbf{y}$ can be defined; by definition, one has $\lim_{U \rightarrow U_\infty} \mathbf{B} = E_\xi(\mathbf{B})$, which in this case is $\boldsymbol{\beta}$. In that sense, $\lim_{U \rightarrow U_\infty} (V_p(\hat{\theta}_{GCAL})) \leq \lim_{U \rightarrow U_\infty} (V_p(\hat{\theta}))$ for any linear unbiased estimator $\hat{\theta}$.

6. Example

The computation of $\hat{\theta}_{GHT}$ and of $\hat{\theta}_{GCAL}$ requires the computation of $\mathbf{Q}_{\hat{\Sigma}}$. An iterative method of computation was described at the end of Section 3. In this section, an example is examined where it is possible to obtain a closed form expression for $\mathbf{Q}_{\hat{\Sigma}}$. It will be assumed that $\hat{\Sigma}$ is a block diagonal matrix. Units of a same block could be persons of a same household, workers of a same establishment, children of a same school, or similar

groupings. Let us say that there are F blocks, then

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\Sigma}_F \end{pmatrix}. \tag{30}$$

In this example, it will be assumed that units belonging to the same block have the same variance and the same covariance. Note that multiplying $\hat{\Sigma}_f$ $f = 1, \dots, F$ by a scalar leaves \mathbf{W}_s GHT unchanged, even if the scalar varies with f . More precisely, if N_f is the size of block f , it will be assumed that $\hat{\Sigma}_f = \mathbf{I}_{N_f} + \rho_f(\mathbf{1}_{N_f \times N_f} - \mathbf{I}_{N_f})$ $f = 1, \dots, F$ with $\frac{-1}{N_f-1} < \rho_f < 1$.

With $\hat{\Sigma}$ of this form, it is possible to find a closed form expression for the block diagonal matrix

$$\mathbf{Q}_{\hat{\Sigma}} = \begin{pmatrix} \mathbf{Q}_{\hat{\Sigma}_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{\hat{\Sigma}_F} \end{pmatrix}. \tag{31}$$

For block f , $\mathbf{Q}_{\hat{\Sigma}_f}^{-1} = E_p \left(\left(\Delta_{s_f} \hat{\Sigma}_f \Delta_{s_f} \right)^\dagger \right)$, where Δ_{s_f} is the $N_f \times N_f$ submatrix of Δ_s that corresponds to block f . Conditioning on the number of units of the block that are sampled, S_f , one obtains

$$\mathbf{Q}_{\hat{\Sigma}_f}^{-1} = \sum_{n_f=1}^{N_f} P(S_f = n_f) E_p \left(\left(\Delta_{s_f} \hat{\Sigma}_f \Delta_{s_f} \right)^\dagger | S_f = n_f \right). \tag{32}$$

The probabilities, $P(S_f = n_f)$, can be expressed as a function of the inclusion probabilities. If one writes

$$P(S_f = n_f) = \sum_{i=n_f}^{N_f} \left(k(i) \sum_{\text{block } f} \pi^{[i]} \right) \quad 1 \leq n_f \leq N_f \tag{33}$$

with $\sum_{\text{block } f} \pi^{[i]}$ being the sum of all probabilities of inclusion of order i , where all i units are in block f , then the application of the inclusion-exclusion principle will yield the recurrence relation

$$k(i) = - \sum_{j=1}^{i-n_f} \binom{i}{j} k(i-j) \quad n_f + 1 \leq i \leq N_f \tag{34}$$

with $k(n_f) = 1$. An example with $N_f = 5$ is given in Appendix D.

Note that $E_p \left(\left(\Delta_{s_f} \hat{\Sigma}_f \Delta_{s_f} \right)^\dagger | S_f = 1 \right) = E_p(\Delta_{s_f} | S_f = 1)$, which is the diagonal matrix of the first order conditional (on $S_f = 1$) inclusion probabilities. Finally, for $n_f \geq 2$,

$E_p \left(\left(\mathbf{\Delta}_{s_f} \hat{\mathbf{\Sigma}}_f \mathbf{\Delta}_{s_f} \right)^\dagger \mid S_f = n_f \right) = E_p(\mathbf{\Delta}_{s_f} \mathbf{M}_f \mathbf{\Delta}_{s_f} \mid S_f = n_f) = \mathbf{\Pi}_{n_f} \circ \mathbf{M}_f$, where the diagonal elements of $\mathbf{M}_f \in \mathbb{R}^{N_f \times N_f}$ are equal to $\frac{-(n_f-2)\rho_f-1}{(n_f-1)\rho_f^2-(n_f-2)\rho_f-1}$, whereas the offdiagonal elements are equal to $\frac{\rho_f}{(n_f-1)\rho_f^2-(n_f-2)\rho_f-1}$. $\mathbf{\Pi}_{n_f}$ is the matrix of second order conditional (on $S_f = n_f$) probabilities of inclusion, and \circ denotes the Hadamard product, that is, element-wise multiplication. The value of the elements of \mathbf{M}_f come from inverting an $n_f \times n_f$ submatrix of $\hat{\mathbf{\Sigma}}_f$.

7. A Modification to the Generalized Estimators

Note that $(\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})$ is positive definite. Indeed, for any non-zero vector $\mathbf{z} \in \mathbb{R}^N$, $E_p(V_\xi(\mathbf{z}' \mathbf{\Delta}_s \mathbf{\Sigma}^{-1} \mathbf{y})) = \mathbf{z}' (\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi}) \mathbf{z} > 0$. If $\mathbf{\Sigma}$ is known, instead of $\hat{\theta}_{GHT} = \theta_{GHT}^* = \mathbf{y}' (\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s)^\dagger (E_p((\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s)^\dagger))^{-1} \mathbf{c}$, one could use

$$\begin{aligned} \hat{\theta}_{MGHT} &= \mathbf{y}' \mathbf{\Delta}_s \mathbf{\Sigma}^{-1} \mathbf{\Delta}_s (E_p(\mathbf{\Delta}_s \mathbf{\Sigma}^{-1} \mathbf{\Delta}_s))^{-1} \mathbf{c} \\ &= \mathbf{y}' \mathbf{\Delta}_s \mathbf{\Sigma}^{-1} \mathbf{\Delta}_s (\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{c}. \end{aligned} \tag{35}$$

If $\mathbf{\Sigma}$ must be estimated, it can be shown that an asymptotically equivalent estimator could be obtained by replacing $\mathbf{\Sigma}$ with $\hat{\mathbf{\Sigma}}$ in (35), if $\hat{\mathbf{\Sigma}}$ satisfies the conditions of Result 3. Contrary to $\hat{\theta}_{GHT}$, which relies on the computation of \mathbf{Q}_Σ , $\hat{\theta}_{MGHT}$ is readily given by a closed-form formula. It is seen that $\hat{\theta}_{MGHT}$ is an estimator; it does not depend on unobserved values of \mathbf{y} . Knowledge of $\mathbf{\Pi}$ is required, thus two-phase sampling for example, may be problematic. If $\mathbf{\Sigma}$ is diagonal, then $\hat{\theta}_{MGHT} = \hat{\theta}_{GHT} = \hat{\theta}_{HT}$. Like $\hat{\theta}_{GHT}$, $\hat{\theta}_{MGHT}$ is unbiased. Also, a closed-form formula can be given for its variance:

$$\begin{aligned} V_p(\hat{\theta}_{MGHT}) &= \|(\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{c} \otimes \mathbf{y}\|_{V_p(\text{vec}(\mathbf{\Delta}_s \mathbf{\Sigma}^{-1} \mathbf{\Delta}_s))}^2 \\ &= \|(\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{c} \otimes \mathbf{y}\|_{V_p(\text{diag}(\text{vec}(\mathbf{\Sigma}^{-1})) \mathbf{\Delta}_s^{(2)} \mathbf{1}_{N^2 \times 1})}^2 \\ &= \|\text{diag}(\text{vec}(\mathbf{\Sigma}^{-1})) (\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{c} \otimes \mathbf{y}\|_{\mathbf{\Pi}^{(4)} - \text{vec}(\mathbf{\Pi})(\text{vec}(\mathbf{\Pi}))'}^2, \end{aligned} \tag{36}$$

where $\mathbf{\Delta}_s^{(2)} = \text{diag}(\text{vec}(\mathbf{\Delta}_s \mathbf{1}_{N \times N} \mathbf{\Delta}_s))$, and $\mathbf{\Pi}^{(4)} = E(\mathbf{\Delta}_s^{(2)} \mathbf{1}_{N^2 \times N^2} \mathbf{\Delta}_s^{(2)})$ is a matrix of fourth-order inclusion probabilities.

Noting $\hat{\theta}_{MGHT} = \mathbf{y}' \mathbf{w}_{sMGHT}$, the calibration problem could now be changed in order to find weights as close as possible to \mathbf{w}_{sMGHT} , instead of \mathbf{w}_{sGHT} . The resulting estimator would be

$$\begin{aligned} \hat{\theta}_{MGCAL} &= \mathbf{y}' \mathbf{w}_{sMGCAL} \\ &= \hat{\mathbf{y}}' \mathbf{c} + (\mathbf{y} - \hat{\mathbf{y}})' \mathbf{w}_{sMGHT}. \end{aligned} \tag{37}$$

The estimator $\hat{\theta}_{MGCAL}$ is asymptotically unbiased and

$$V_p(\hat{\theta}_{MGCAL}) \doteq \|\text{diag}(\text{vec}(\mathbf{\Sigma}^{-1})) (\mathbf{\Sigma}^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{c} \otimes (\mathbf{y} - \mathbf{y}^*)\|_{\mathbf{\Pi}^{(4)} - \text{vec}(\mathbf{\Pi})(\text{vec}(\mathbf{\Pi}))'}^2. \tag{38}$$

Of course, it is not expected that $\hat{\theta}_{MGAL}$ will attain the lower bound given in (11). However, it does not rely on generalized inverses, and does not require the computation of \mathbf{Q}_{Σ} .

8. Estimator Comparison

In this section, six estimators are compared: the Horvitz-Thompson estimator, the calibration estimator, the generalized Horvitz-Thompson estimator, the generalized calibration estimator, and the modified versions of the latter two as described in Section 7. All estimators are, at least asymptotically, unbiased. For comparing their variance, a population of 1,000 units, 200 clusters of five units each, was used. The variable of interest was generated from a normal distribution with mean 10 and variance 2. This was done in such a way that units from the same cluster have a covariance of one, whereas units from different clusters are independent. The parameter to be estimated is the population mean.

The variance, or asymptotic variance, of the generalized estimators and of the modified generalized estimators was computed with a block diagonal matrix Σ , with each 5×5 block having the value 2 on the diagonal and 1 offdiagonal, thus reflecting the distribution used to generate the population. The variances were computed under two sampling plans: simple random sampling, and Poisson sampling with the five units from a block being selected with probability (0.15, 0.15, 0.2, 0.2, 0.3). It was assumed that $\mathbf{X} = \mathbf{1}_{N \times 1}$ for computing the asymptotic variances of the calibration estimator, the generalized calibration estimator, and the modified generalized calibration estimator. That is, the only calibration equation is the one specifying that the sum of the weights should equal the population size – the trivial calibration equation. The value of $V_p(\text{vec}((\Delta_s \Sigma \Delta_s)^\dagger))$, needed to compute (5) and (26), was approximated by computing $(\Delta_s \Sigma \Delta_s)^\dagger$ for 10,000 different samples, all drawn according to the appropriate sampling plan: simple random sampling or Poisson sampling. With the covariance matrix used for generating the population, the lower bound given in (11) equals 0.0070 under simple random sampling, and 0.0075 under Poisson sampling.

Table 1 gives the variances of the estimators, or their asymptotic variances in the case of calibrated estimators, under the sampling plan. The table shows that, for simple random sampling, the generalized Horvitz-Thompson estimator is much less precise than the regular Horvitz-Thompson estimator. The explanation for this was given at the beginning of Section 5. The generalized Horvitz-Thompson estimator was not meant to be optimal; its interest lies in relation (7). In contrast, the generalized calibration estimator outperforms the Horvitz-Thompson estimator. Note that under simple random sampling with $\mathbf{X} = \mathbf{1}_{N \times 1}$, the calibration estimator is equal to the Horvitz-Thompson estimator. The asymptotic variance under the sampling plan, $V_p(\hat{\theta}_{GICAL})$, is very close to the generalized Godambe-Joshi lower bound for $E_{\xi} V_p(\hat{\theta})$, which for this Σ and sampling plan is equal to 0.0070. The performance of the modified generalized Horvitz-Thompson estimator can be significantly different from that of the generalized Horvitz-Thompson estimator. It cannot be seen as a good approximation of the generalized Horvitz-Thompson estimator. Nevertheless, the modified generalized calibration estimator performs better than the Horvitz-Thompson estimator and practically as well as the generalized calibration estimator for both simple random sampling and the Poisson sampling plan.

Table 1. Variances of six estimators of the mean.

Estimator	Simple random sampling	Poisson sampling
Horvitz-Thompson	0.0077	0.4458
Calibration	–	0.0084
Generalized Horvitz-Thompson	0.0477	0.2848
Generalized calibration	0.0069	0.0073
Modified generalized Horvitz-Thompson	0.0361	0.3237
Modified generalized calibration	0.0070	0.0076
Generalized Godambe-Joshi lower bound	0.0070	0.0075

With the Poisson sample being of random size, it is not surprising that the noncalibrated estimators (Horvitz-Thompson, generalized Horvitz-Thompson, and modified generalized Horvitz-Thompson) are performing poorly with this sampling plan. The generalized calibration estimator outperforms the calibration estimator. Its asymptotic variance is comparable to the generalized Godambe-Joshi lower bound for $E_{\xi}V_p(\hat{\theta})$, which for this Σ and sampling plan is equal to 0.0075.

Since the calibration estimator is the generalized calibration estimator computed with a diagonal matrix Σ , the asymptotic variance for the calibration estimator in Table 1 shows what can happen if the generalized version is used with a matrix Σ different from the true variance matrix, $V_{\xi}(\mathbf{y})$, . . . the generalized calibration estimator could become the ordinary calibration estimator.

9. Application to the Canadian Reverse Record Check Survey

The Reverse Record Check (RRC) is a Canadian postcensal undercoverage survey used in conjunction with the Census of Population and a postcensal overcoverage study to arrive at population estimates; see [Statistics Canada \(2015\)](#). In this section, the estimates and the methodology used for the Canadian Territory of Yukon for the 2011 RRC are examined, and the generalized estimates are compared to the current one. A list frame of persons is sampled with stratified random sampling. There is one large take-all stratum, where all units are enumerated by the Census, and take-some strata comprising units that are either enumerated by the Census, missed by the Census, or out of scope for the Census. The main objective of the RRC is to estimate the number of missed persons. Information on the frame is available to group persons by household. Because of the Census methodology, the variable “missed”, which takes the value 1 if the person is missed and 0 if not, is highly correlated for persons belonging to the same household. If the Census enumerated (or missed) someone, it likely enumerated (or missed) the other members of that household. With the current RRC methodology, the Horvitz-Thompson weight of responding units is multiplied by a factor to account for nonresponse and a factor to account for frame undercoverage. The latter factor is such that the estimate of persons enumerated coincides with the equivalent Census number. In this application, the Horvitz-Thompson weights are replaced with the generalized weights. For the generalized calibrated weights, the calibration equation simply ensured that the sum of the calibrated weights equalled the

Table 2. RRC estimates of missed persons.

Estimator	Estimate	Variance estimate
Horvitz-Thompson	5,272	91,727
Generalized Horvitz-Thompson	5,150	82,920
Generalized calibration	5,137	82,505
Modified generalized Horvitz-Thompson	5,194	86,945
Modified generalized calibration	5,173	85,999

number of units in the stratum. The same nonresponse adjustment factors were used, and the frame undercoverage adjustment factors were all computed so that the estimates of persons enumerated coincides with the equivalent Census number: 29,982. The generalized weights were computed assuming that the correlation structure is block diagonal, each block representing the persons of a same household, according to the frame information. All offdiagonal elements of each block are set to 0.95. This is because the estimates of the correlations within households are 0.956 for the variable “missed” and 0.947 for the variable “enumerated”. The correlations are less than one because the Census sometimes partly enumerates or partly misses a household, and because the frame household may differ from the census household.

The estimates obtained along with the corresponding variance estimates are given in Table 2. The Horvitz-Thompson estimates are those currently used by the survey. The variance estimate of the generalized Horvitz-Thompson estimator is lower than that of the Horvitz-Thompson estimator, in spite of having an additional source of variance, as discussed at the beginning of Section 5. This is because, for all estimators, the last step in computing the estimates is a calibration on the number of persons enumerated. The variance estimates of the generalized calibration estimator and the modified generalized calibration estimator, those that would be used in practice, are lower than that of the Horvitz-Thompson estimator. The advantage of the modified generalized estimator is that its computation did not require calculating the matrix \mathbf{Q}_{Σ} , although it was easy to approximate this matrix by repeatedly sampling the frame one million times, and using the method described at the beginning of Section 6. The estimates of missed persons are not significantly different from one another.

10. Variance Estimation

Statisticians are better at estimating totals or weighted totals than they are at estimating variances. Why not write variances in the form of weighted totals? A variance can be written in the form $\theta_{\text{var}} = \mathbf{y}'\mathbf{V}\mathbf{y} = (\mathbf{y} \otimes \mathbf{y})'\text{vec}(\mathbf{V})$, with $\mathbf{V} \in \mathbb{R}^{N \times N}$. Such parameters have been estimated in Sections 3, 5, and 7. There are N^2 units, each corresponding to a pair of units of the original population, with a vector of interest equal to $\mathbf{y} \otimes \mathbf{y}$ and $\mathbf{c} = \text{vec}(\mathbf{V})$. The methods of this article apply here, because in general, $\text{Cov}(y_i y_j, y_k y_l) \neq 0$ for $(i, j) \neq (k, l)$. Whatever the asymptotic setup, $V_{\xi}(\mathbf{y} \otimes \mathbf{y}) = \Sigma_2$ will not be a positive definite matrix. For example, the row of Σ_2 which corresponds to unit (i, j) $i \neq j$ is equal to the row which corresponds to unit (j, i) . In fact, the event $(i, j) \in s$ is identical to the event $(j, i) \in s$. Thus, from the N^2 units, only the $N(N + 1)/2$ with $j \geq i$ need to be kept. The

estimators suggested in this article would require the inversion of a matrix of order $N(N + 1)/2$.

The variance matrix will be diagonal, $Cov(y_i y_j, y_k y_l) = 0$ for $(i, j) \neq (k, l)$, if $V_\xi(\mathbf{y})$ is diagonal and if $E_\xi(\mathbf{y}) = \mathbf{0}$. The last assumption is reasonable if \mathbf{y} is a vector of residues, as would be the case if one is estimating the variance of a calibration estimator (regular, generalized, or modified). With $Cov(y_i y_j, y_k y_l) = 0$ for $(i, j) \neq (k, l)$, the regular calibration estimator will suffice to estimate the variance of a calibrated estimator. A choice of calibration equations must still be made. The findings made in [Théberge \(1999\)](#) remain valid; namely, to use an auxiliary variables matrix in a block diagonal form with two blocks: a trivial model for the cross products terms of $\theta_{\text{var}} = \mathbf{y}'\mathbf{V}\mathbf{y}$, that is, $\mathbf{y}'(\mathbf{V} - (\mathbf{V} \circ \mathbf{I}))\mathbf{y}$ (this will yield a Horvitz-Thompson estimator in the case of fixed-size sampling plans) and a nontrivial model to estimate the squared terms of $\theta_{\text{var}} = \mathbf{y}'\mathbf{V}\mathbf{y}$, that is, $\mathbf{y}'(\mathbf{V} \circ \mathbf{I})\mathbf{y}$. In the examples examined in [Théberge \(1999\)](#), the nontrivial model used for estimating the squared terms of the variance was the one corresponding to a ratio estimator.

Whatever the models are, it is important to use an auxiliary variables matrix in a block diagonal form with two blocks. For example, the intercept used to estimate the cross-product terms has nothing to do with a possible intercept to estimate the squared terms. Therefore there should not be an auxiliary variable taking the value one for all $(i, j) j \geq i$. It is preferable to have an auxiliary variable taking the value one for all $(i, j) j > i$ and zero otherwise, and another auxiliary variable taking the value one for $(i, i) i = 1, 2, \dots, N$ and zero otherwise.

11. Conclusion

An asymptotic setup is necessary to discuss the asymptotic properties of the estimators. The setup used here integrates a superpopulation model. There is no need for a superpopulation model separate from the asymptotic setup. The setup's model does not assume that the units are uncorrelated.

Even the Horvitz-Thompson estimator can be viewed as relying on a model. The generalized Horvitz-Thompson estimator, like the Horvitz-Thompson estimator, is unbiased. Both estimators, but especially the former, can be affected by the variance in the effective sample weight. Even without auxiliary data, it is possible to calibrate the weights so their total equals the population size. If this is done, then the generalized estimator will have a lower asymptotic variance than the ordinary estimator.

The calibration estimator was generalized in two ways: firstly, one is seeking weights close to the generalized Horvitz-Thompson weights; secondly, the matrices \mathbf{T} and \mathbf{U} , used in measuring distances, need no longer be diagonal.

A somewhat easier way to compute the modified generalized calibration estimator was shown to perform practically as well as the generalized calibration estimator in the examples given in this article.

The Godambe-Joshi lower bound can be generalized to the case where the units are correlated. The asymptotic variance of the generalized calibration estimators attains the generalized Godambe-Joshi lower bound, if the model is correct, that is, if $\hat{\Sigma} \rightarrow \Sigma$ in probability and the matrix \mathbf{X} used by the generalized calibration estimator agrees with that of the asymptotic setup. This is regardless of the choice for the matrix \mathbf{U} , used in

measuring the distance between the calibrated weights and the generalized Horvitz-Thompson weights.

By viewing variances as weighted totals, the theory developed here provides a framework for variance estimation. The general case would require the inversion of very large matrices, but there are simplifications to be made if one is estimating the variance of a calibration estimator. Those simplifications will often result in what was called the “hybrid estimator” in [Théberge \(1999\)](#).

Even though there are workarounds, such as dropping variables or using the limit of a positive definite matrix, it would be interesting to generalize the results of this article to the case of Σ positive semi-definite. This strategy of using the methods of this article with a positive definite covariance matrix that differs only infinitesimally from a block diagonal matrix where each block equals $\mathbf{1}_{2 \times 2}$ will allow concluding by revisiting the example in the introduction. The assumption that the correlation between y_{2i-1} and y_{2i} $i = 1, 2, \dots, N/2$ is 1, is weaker than the assumption that the two units are equal, as was done in the introduction. The resulting generalized Horvitz-Thompson estimator of the total is $\sum_{i=1}^{N/2} 2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - (y_{2i-1} + y_{2i})\delta_{2i-1}\delta_{2i} + (y_{2i-1} - y_{2i})\delta_{2i-1}\delta_{2i}(\pi_{2i} - \pi_{2i-1}) / \pi_{2i-1} + \pi_{2i} - \pi_{2i-1}$. Setting $y_1 = y_2$ will yield the term given in the introduction. For example, for N spouses grouped into $N/2$ couples, the variable of interest may be the place of residence (very high correlation), or education level (significant correlation). Using a calibrated version of this estimator to ensure that the sum of the weights equals the population size, will be preferable to using a similarly calibrated version of the Horvitz-Thompson estimator, if the correlation is somewhat close to 1. Using an estimator optimized for a correlation of 1 will often be preferable to using an estimator optimized for a correlation of 0. Note that two samples drawn with the same sampling plan made up of individuals from the same couples will have the same effective sample weight, regardless of how many of the spouses, one or two, are sampled from each observed couple.

Appendix A: Proofs of Results of Section 2 and Lemmas of Section 3

Proof of Result 1:

Note that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_t &= \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \mathbf{X}'_t (\Delta_t \mathbf{U}_t \Delta_t)^\dagger \mathbf{X}_t \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}'_t (\Delta_t \mathbf{U}_t \Delta_t)^\dagger \mathbf{y}_t \\ &= \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \sum_{i=1}^t \mathbf{X}'_i (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger \mathbf{X}_i \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \sum_{i=1}^t \mathbf{X}'_i (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger \mathbf{y}_{[i]} \\ &= \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \mathbf{X}'_t{}^{-1} \sum_{i=1}^t (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger \mathbf{X}_i \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}'_t{}^{-1} \sum_{i=1}^t (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger \mathbf{y}_{[i]}. \end{aligned}$$

The $\Delta_{[i]}$ being independent and identically distributed, from the weak law of large numbers, $t^{-1} \sum_{i=1}^t (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger$ tends in probability to $\mathbf{Q}_U^{-1} = E_p((\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger)$. Since the sampling plan is noninformative, $t^{-1} \sum_{i=1}^t (\Delta_{[i]} \mathbf{U}_{\Delta_{[i]}})^\dagger \mathbf{y}_{[i]}$ tends in probability to $\mathbf{Q}_U^{-1} \mathbf{X} \boldsymbol{\beta}$.

Thus,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_t &\rightarrow \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \mathbf{X}' \mathbf{Q}_U^{-1} \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}' \mathbf{Q}_U^{-1} \mathbf{X} \boldsymbol{\beta} \\ &= \mathbf{T}^{1/2} \left(\mathbf{T}^{1/2} \mathbf{X}' \mathbf{Q}_U^{-1} \mathbf{X} \mathbf{T}^{1/2} \right)^{-1} \mathbf{T}^{1/2} \mathbf{X}' \mathbf{Q}_U^{-1} \mathbf{X} \mathbf{T}^{1/2} \mathbf{T}^{-1/2} \boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

in probability.

Proof of Result 2:

Note that

$$\mathbf{X}'_t \left(\boldsymbol{\Delta}_t \hat{\boldsymbol{\Sigma}}_t \boldsymbol{\Delta}_t \right)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t \mathbf{c}_t = t^\gamma \mathbf{X}' \left(\left(t^{-1} \sum_{i=1}^t \left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \right) \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}} - \mathbf{I}_N \right) \mathbf{c}. \tag{A.39}$$

The expectation, under the plan p , of $\left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger$ is equal to $\mathbf{Q}_{\hat{\boldsymbol{\Sigma}}}^{-1}$. Also, the variance of the vector $\mathbf{X}' \left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}} \mathbf{c}$ is finite. Indeed, there is a finite number of possible values for $\boldsymbol{\Delta}_{[i]}$, and for any $\boldsymbol{\Delta}_{[i]}$, $\left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger$ exists. According to the central limit theorem, $\left(\mathbf{X}' \left(t^{-1} \sum_{i=1}^t \left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \right) \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}} \mathbf{c} - \mathbf{X}' \mathbf{c} \right) t^{1/2}$ converges to a normal distribution with mean $\mathbf{0}_{q \times 1}$ and finite variance. Since $\left(\mathbf{X}' \left(t^{-1} \sum_{i=1}^t \left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \right) \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}} \mathbf{c} - \mathbf{X}' \mathbf{c} \right) t^{1/2}$ is $O_p(1)$, from (A.39) it follows that $\mathbf{X}'_t \left(\boldsymbol{\Delta}_t \hat{\boldsymbol{\Sigma}}_t \boldsymbol{\Delta}_t \right)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}_t} \mathbf{c}_t - \mathbf{X}'_t \mathbf{c}_t$ is $O_p(t^{\gamma-1/2})$.

Proof of Result 3:

The difference can be written as a sum of differences $t^\gamma \mathbf{X}' \left(t^{-1} \sum_{i=1}^t \left(\left(\boldsymbol{\Delta}_{[i]} \hat{\boldsymbol{\Sigma}}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \mathbf{Q}_{\hat{\boldsymbol{\Sigma}}} - \left(\boldsymbol{\Delta}_{[i]} \boldsymbol{\Sigma}_{[i]} \boldsymbol{\Delta}_{[i]} \right)^\dagger \mathbf{Q}_{\boldsymbol{\Sigma}} \right) \right) \mathbf{c}$. The differences under the summation sign tend to $\mathbf{0}$ in probability and the central limit theorem yields the result. Note that the condition of $\hat{\boldsymbol{\Sigma}}_t$ being positive definite is needed to ensure the Moore-Penrose inverse is continuous at $\boldsymbol{\Sigma}_t$ (see Ben-Israel and Greville 2002, 212).

Proof of Lemma 1:

It is obvious that $\mathcal{N}(\mathbf{F}) \supseteq \cap_{i=1}^K \mathcal{N}(\mathbf{F}_i)$. To show that $\mathcal{N}(\mathbf{F}) \subseteq \cap_{i=1}^K \mathcal{N}(\mathbf{F}_i)$, let $\mathbf{v} \in \mathcal{N}(\mathbf{F})$, then $\mathbf{v}' \mathbf{F} \mathbf{v} = \sum_{i=1}^K \mathbf{v}' \mathbf{F}_i \mathbf{v} = 0$. The matrices \mathbf{F}_i being positive semi-definite, one must have $\mathbf{v}' \mathbf{F}_i \mathbf{v} = 0$ $i = 1, 2, \dots, K$. With \mathbf{F}_i also being symmetric, there exists a symmetric positive semi-definite matrix \mathbf{K}_i such that $\mathbf{F}_i = \mathbf{K}_i^2$. Therefore $\mathbf{v}' \mathbf{F}_i \mathbf{v} = 0$ implies $\mathbf{v}' \mathbf{K}'_i \mathbf{K}_i \mathbf{v} = 0$ and one must have $\mathbf{v} \in \mathcal{N}(\mathbf{F}_i)$.

Proof of Lemma 2:

Writing $\mathcal{R}(\mathbf{F})$ for the range of a matrix \mathbf{F} , from [Ben-Israel and Greville \(2002\)](#), Exercise 2.38) it is known that $\mathcal{R}((\mathbf{F}\mathbf{P})^\dagger) = \mathcal{R}(\mathbf{P}\mathbf{F}') \subseteq \mathcal{R}(\mathbf{P})$, which proves a). The proof of b) is obtained by taking the transpose of each side of the identity in a). Finally, from using a) and b) in succession it follows that $(\mathbf{P}\mathbf{F}\mathbf{P})^\dagger = \mathbf{P}(\mathbf{P}\mathbf{F}\mathbf{P})^\dagger = \mathbf{P}(\mathbf{P}\mathbf{F}\mathbf{P})^\dagger\mathbf{P}$.

Appendix B: Counter-example

To estimate a population total ($\mathbf{c} = \mathbf{1}$), an estimator $\hat{\theta} = \hat{\theta}_{GCAL} + \hat{0}$, where $\hat{0}$ is a nonlinear unbiased estimator of 0, will be used. The computation of $\hat{\theta}_{GCAL}$ will use $\hat{\Sigma} = \Sigma$. It will be shown that for the asymptotically equivalent random variable $\theta^* = \theta_{GCAL}^* + \hat{0}$, one has $E_\xi V_p(\theta^*) < E_\xi V_p(\theta_{GCAL}^*) = \text{GGJLB}$, where GGJLB is the generalized Godambe-Joshi lower bound.

Starting with Equation (9)

$$\begin{aligned} E_\xi V_p(\theta^*) &= E_p V_\xi(\theta^*) + E_p(E_\xi(\theta^* - \theta))^2 - V_\xi(\theta) \\ &= E_p V_\xi(\theta_{GCAL}^*) + E_p V_\xi(\hat{0}) + 2E_p \text{Cov}_\xi(\theta_{GCAL}^*, \hat{0}) + E_p(E_\xi(\theta^* - \theta))^2 - V_\xi(\theta) \\ &= E_p V_\xi(\hat{\theta}_{GHT}) - V_\xi(\theta) + E_p V_\xi(\hat{0}) + 2E_p \text{Cov}_\xi(\theta_{GCAL}^*, \hat{0}) + E_p(E_\xi(\hat{0}))^2 \\ &= \text{GGJLB} + E_p V_\xi(\hat{0}) + 2E_p \text{Cov}_\xi(\theta_{GCAL}^*, \hat{0}) + E_p(E_\xi(\hat{0}))^2. \end{aligned}$$

In this example, the population $U = \{1, 2, 3\}$. Under the sampling plan, the samples $s_1 = \{1, 2\}$ and $s_2 = \{2, 3\}$ can each be selected with probability 0.5. The vector $\mathbf{u} \in \mathbb{R}^3$ is composed of independent and identically distributed variables taking the values 1 or -1 each with probability 0.5. Under the model ξ , the vector of interest is $\mathbf{y} = \Sigma^{1/2}\mathbf{u}$,

with $\Sigma^{1/2} = \begin{pmatrix} 2 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Thus, with this model, $E_\xi(\mathbf{y}) = \mathbf{0}$, $V_\xi(\mathbf{y}) =$

$$\Sigma = \begin{pmatrix} 4.25 & 1.5 & 0 \\ 1.5 & 1.25 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{Q}_\Sigma = \begin{pmatrix} 6.7 & 1.5 & 0 \\ 1.5 & 1.25 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

A nonlinear unbiased, under the sampling plan, estimator of 0 is $\hat{0} = -[y_2]$ if s_1 is selected and $\hat{0} = [y_2]$ if s_2 is selected, where $[y_2]$ represents the integer part of y_2 .

Under those conditions, $\text{GGJLB} = \mathbf{c}'(\mathbf{Q}_\Sigma - \Sigma)\mathbf{c} = 3.45$, $E_p V_\xi(\hat{0}) = 0.5$, $E_\xi(\hat{0}) = 0$,

$$\text{Cov}_\xi(\mathbf{y}, [y_2]) = \begin{pmatrix} 1.25 \\ 0.75 \\ 0 \end{pmatrix},$$

$$E_p \text{Cov}_\xi(\theta_{GCAL}^*, \hat{0}) = 0.5 \times \mathbf{c}'\mathbf{Q}_\Sigma((\Delta_{s_2}\Sigma\Delta_{s_2})^\dagger - (\Delta_{s_1}\Sigma\Delta_{s_1})^\dagger)\text{Cov}_\xi(\mathbf{y}, [y_2]) = -0.35 \text{ and}$$

$$\begin{aligned}
E_{\xi}V_p(\theta^*) &= \text{GGJLB} + E_pV_{\xi}(\hat{\theta}) + 2E_p\text{Cov}_{\xi}(\theta_{GCAL}^*, \hat{\theta}) + E_p(E_{\xi}(\hat{\theta}))^2 \\
&= 3.45 + 0.5 - 0.7 + 0 \\
&= 3.25 \\
&< \text{GGJLB}.
\end{aligned}$$

Thus, asymptotically, $E_{\xi}V_p(\hat{\theta}_{GCAL} + \hat{\theta}) < E_{\xi}V_p(\hat{\theta}_{GCAL}) = \text{GGJLB}$. With $N = 3$, the asymptotic properties are not very meaningful. However the example could be expanded to include a large number of strata of size 3, each with the model and sampling plan described above.

Appendix C: Proof that (13) Satisfies (12)

To simplify, $\mathbf{P}_{\mathcal{N}(A'TA)}$ will be denoted by \mathbf{P} . First, using the first part of (13), set $\mathbf{G} = (\mathbf{I}_N - (\mathbf{PUP})^{\dagger}\mathbf{U})(A'TA)^{\dagger}A'T$.

Because for an arbitrary real square matrix \mathbf{M} , $\mathcal{N}(\mathbf{MM}') = \mathcal{N}(\mathbf{M}) = \mathcal{N}(\mathbf{M}')$, it follows that

$$\begin{aligned}
\mathbf{UGA} &= \mathbf{U}(\mathbf{I}_N - (\mathbf{PUP})^{\dagger}\mathbf{U})(A'TA)^{\dagger}A'TA \\
&= \mathbf{U}(\mathbf{I}_N - (\mathbf{PUP})^{\dagger}\mathbf{U})(\mathbf{I}_N - \mathbf{P}) \\
&= \mathbf{U} - \mathbf{U}(\mathbf{PUP})^{\dagger}\mathbf{U} - \mathbf{UP} + \mathbf{UP}(\mathbf{PUP})^{\dagger}(\mathbf{PUP}) \\
&= \mathbf{U} - \mathbf{U}(\mathbf{PUP})^{\dagger}\mathbf{U} - \mathbf{UP} + \mathbf{UP}(\mathbf{I}_N - \mathbf{P}_{\mathcal{N}(\mathbf{PUP})}) \\
&= \mathbf{U} - \mathbf{U}(\mathbf{PUP})^{\dagger}\mathbf{U} - \mathbf{UPP}_{\mathcal{N}(\mathbf{PUP})} \\
&= \mathbf{U} - \mathbf{U}(\mathbf{PUP})^{\dagger}\mathbf{U},
\end{aligned} \tag{C.1}$$

which is symmetrical. Also symmetrical, is

$$\begin{aligned}
\mathbf{TAG} &= \mathbf{TA}(\mathbf{I}_N - (\mathbf{PUP})^{\dagger}\mathbf{U})(A'TA)^{\dagger}A'T \\
&= (\mathbf{TA} - \mathbf{TAP}(\mathbf{PUP})^{\dagger}\mathbf{U})(A'TA)^{\dagger}A'T \\
&= \mathbf{TA}(A'TA)^{\dagger}A'T,
\end{aligned} \tag{C.2}$$

because $\mathbf{TAP} = \mathbf{0}$. From (C.2),

$$\begin{aligned}
\mathbf{TAGA} &= \mathbf{TA}(A'TA)^{\dagger}A'TA \\
&= \mathbf{TA}(\mathbf{I}_N - \mathbf{P}) \\
&= \mathbf{TA}.
\end{aligned} \tag{C.3}$$

Finally, from (C.1),

$$\begin{aligned}
 \mathbf{UGAG} &= (\mathbf{U} - \mathbf{U}(\mathbf{PUP})^\dagger)\mathbf{U}\mathbf{G} \\
 &= \mathbf{UG} - \mathbf{U}(\mathbf{PUP})^\dagger\mathbf{U}(\mathbf{I}_N - (\mathbf{PUP})^\dagger\mathbf{U})(\mathbf{A}'\mathbf{TA})^\dagger\mathbf{A}'\mathbf{T} \\
 &= \mathbf{UG} - [\mathbf{U}(\mathbf{PUP})^\dagger\mathbf{U} - \mathbf{U}(\mathbf{PUP})^\dagger\mathbf{PUP}(\mathbf{PUP})^\dagger\mathbf{U}](\mathbf{A}'\mathbf{TA})^\dagger\mathbf{A}'\mathbf{T} \\
 &= \mathbf{UG}.
 \end{aligned} \tag{C.4}$$

If $\mathbf{G} = \mathbf{U}^{-1}\mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2}$ with \mathbf{U} positive definite, then

$$\mathbf{UGA} = \mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2}\mathbf{A} \tag{C.5}$$

is a symmetrical matrix. Also,

$$\begin{aligned}
 \mathbf{TAG} &= \mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \\
 &= \mathbf{T}^{1/2}\mathbf{P}_{\mathcal{R}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})}\mathbf{T}^{1/2}
 \end{aligned} \tag{C.6}$$

is symmetrical, since an orthogonal projection matrix is symmetrical. From the properties of the Moore-Penrose inverse,

$$\begin{aligned}
 \mathbf{UGAG} &= \mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \\
 &= \mathbf{A}'\mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2} \\
 &= \mathbf{UG}.
 \end{aligned} \tag{C.7}$$

Finally, because for an arbitrary real matrix \mathbf{M} , $\mathcal{R}(\mathbf{MM}') = \mathcal{R}(\mathbf{M})$ and because \mathbf{U} is positive definite, it follows that

$$\begin{aligned}
 \mathbf{TAGA} &= \mathbf{T}^{1/2}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})^\dagger\mathbf{T}^{1/2}\mathbf{A} \\
 &= \mathbf{T}^{1/2}\mathbf{P}_{\mathcal{R}(\mathbf{T}^{1/2}\mathbf{AU}^{-1}\mathbf{A}'\mathbf{T}^{1/2})}\mathbf{T}^{1/2}\mathbf{A} \\
 &= \mathbf{T}^{1/2}\mathbf{P}_{\mathcal{R}(\mathbf{T}^{1/2}\mathbf{A})}\mathbf{T}^{1/2}\mathbf{A} \\
 &= \mathbf{TA}.
 \end{aligned} \tag{C.8}$$

Appendix D: Example of Computing $P(S_f = n_f)$ with a Block of Size 5

$$P(S_f = 5) = \pi^{[5]}$$

$$P(S_f = 4) = \sum_{\text{block}} \pi^{[4]} - 5\pi^{[5]}$$

$$P(S_f = 3) = \sum_{\text{block}} \pi^{[3]} - 4 \sum_{\text{block}} \pi^{[4]} + 10\pi^{[5]}$$

$$P(S_f = 2) = \sum_{\text{block}} \pi^{[2]} - 3 \sum_{\text{block}} \pi^{[3]} + 6 \sum_{\text{block}} \pi^{[4]} - 10 \pi^{[5]}$$

$$P(S_f = 1) = \sum_{\text{block}} \pi^{[1]} - 2 \sum_{\text{block}} \pi^{[2]} + 3 \sum_{\text{block}} \pi^{[3]} - 4 \sum_{\text{block}} \pi^{[4]} + 5 \pi^{[5]}$$

$$P(S_f = 0) = 1 - \sum_{n_f=1}^5 P(S_f = n_f)$$

12. References

- Albert, A. 1973. "The Gauss–Markov Theorem for Regression Models with Possibly Singular Covariances." *SIAM Journal on Applied Mathematics* 24: 182–187.
- Ben-Israel, A. and T.N.E. Greville. 2002. *Generalized Inverses: Theory and Applications (second ed.)*. New York: Springer-Verlag.
- Brewer, K.R.W. 1979. "A Class of Robust Sampling Designs for Large Scale Surveys." *Journal of the American Statistical Association* 74: 911–915.
- Cassel, C.-M., C.E. Särndal, and J.H. Wretman. 1977. *Foundations of Inference in Survey Sampling*. New York: John Wiley & Sons.
- Deville, J.-C. and C.E. Särndal. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87: 376–382.
- Estevao, V.M. and C.E. Särndal. 2003. *A New Perspective on Calibration Estimators*. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 1346–1356.
- Godambe, V.P. and V.M. Joshi. 1965. "Admissibility and Bayes Estimation in Sampling Finite Populations, 1." *Annals of Mathematical Statistics* 36: 1707–1722.
- Horvitz, D.G. and D.J. Thompson. 1952. "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Statistical Association* 47: 663–685.
- Isaki, C.T. and W.A. Fuller. 1982. "Survey Design Under the Regression Superpopulation Model." *Journal of the American Statistical Association* 77: 89–96.
- Montanari, G.E. 1998. "On Regression Estimation of Finite Population Means." *Survey Methodology* 24: 69–77.
- Montanari, G.E. and M.G. Ranalli. 2002. "Asymptotically Efficient Generalised Regression Estimators." *Journal of Official Statistics* 18: 577–590.
- Rao, C.R. and S.K. Mitra. 1971. *Generalized Inverse of Matrices and its Applications*. New York: John Wiley & Sons.
- Särndal, C.E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Statistics Canada. 2015. *Census Technical Report: Coverage*. Statistics Canada Catalogue no. 98-303-X, Ottawa, Ontario. Available at: <https://www12.statcan.gc.ca/census-recensement/2011/ref/guides/98-303-x/index-eng.cfm> (accessed January 2017).

Théberge, A. 1999. “Extensions of Calibration Estimators in Survey Sampling.” *Journal of the American Statistical Association* 94: 635–644.

Zyskind, G. and F.B. Martin. 1969. “On Best Linear Estimation and a General Gauss–Markov Theorem in Linear Models with Arbitrary Non-negative Covariance Structure.” *SIAM Journal on Applied Mathematics* 17: 1190–1202.

Received July 2015

Revised April 2016

Accepted July 2016