

Catalogue no. 12-001-X  
ISSN 1492-0921

## Survey Methodology

# A generalization of inverse probability weighting

by Alain Théberge

Release date: June 21, 2022



Statistics  
Canada

Statistique  
Canada

Canada

---

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, [www.statcan.gc.ca](http://www.statcan.gc.ca).

You can also contact us by

**Email at** [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Telephone**, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on [www.statcan.gc.ca](http://www.statcan.gc.ca) under "Contact us" > "[Standards of service to the public.](#)"

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2022

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

*Cette publication est aussi disponible en français.*

---

# A generalization of inverse probability weighting

Alain Th  berge<sup>1</sup>

## Abstract

In finite population estimation, the inverse probability or Horvitz-Thompson estimator is a basic tool. Even when auxiliary information is available to model the variable of interest, it is still used to estimate the model error. Here, the inverse probability estimator is generalized by introducing a positive definite matrix. The usual inverse probability estimator is a special case of the generalized estimator, where the positive definite matrix is the identity matrix. Since calibration estimation seeks weights that are close to the inverse probability weights, it too can be generalized by seeking weights that are close to those of the generalized inverse probability estimator. Calibration is known to be optimal, in the sense that it asymptotically attains the Godambe-Joshi lower bound. That lower bound has been derived under a model where no correlation is present. This too, can be generalized to allow for correlation. With the correct choice of the positive definite matrix that generalizes the calibration estimators, this generalized lower bound can be asymptotically attained. There is often no closed-form formula for the generalized estimators. However, simple explicit examples are given here to illustrate how the generalized estimators take advantage of the correlation. This simplicity is achieved here, by assuming a correlation of one between some population units. Those simple estimators can still be useful, even if the correlation is smaller than one. Simulation results are used to compare the generalized estimators to the ordinary estimators.

**Key Words:** Calibration estimator; Godambe-Joshi lower bound; Horvitz-Thompson estimator; Moore-Penrose inverse; Vaccination rate.

## 1. Introduction

The usual inverse probability estimator of the total for a population of  $N$  units is

$$\hat{\theta}_{\text{IP}} = \sum_{i=1}^N \frac{\delta_i y_i}{\pi_i}, \quad (1.1)$$

where  $y_i$  is the variable of interest for unit  $i$ ,  $\delta_i$  is 1 or 0 depending on whether  $i$  is in the sample  $s$  or not, and  $\pi_i > 0$  is the probability that  $i$  is in  $s$ . Note that the expectation of  $\delta_i$  is  $\pi_i$ , this makes  $\hat{\theta}_{\text{IP}}$  unbiased for  $\theta = \sum_{i=1}^N y_i$ . It is also known as the Horvitz-Thompson estimator, presented in Horvitz and Thompson (1952). In this paper, estimators that can draw some strength from units not in  $s$  will be presented.

Here is an example of such an estimator for a population of  $N$  units that is partitioned into  $N_p = N/2$  pairs  $\{2i-1, 2i\}$  ( $i=1, 2, \dots, N_p$ ),

$$\hat{\theta}_{\text{LIM}} = \sum_{i=1}^{N_p} \frac{2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - (y_{2i-1} + y_{2i})\delta_{2i-1}\delta_{2i} + (y_{2i-1} - y_{2i})\delta_{2i-1}\delta_{2i} \pi_{\text{diff } i}}{\pi_{2i-1} + \pi_{2i} - \pi_{2i-1 \ 2i}}, \quad (1.2)$$

where  $\pi_{2i-1 \ 2i} = E(\delta_{2i-1}\delta_{2i})$  is the probability that both units  $2i-1$  and  $2i$  are in  $s$ , and  $\pi_{\text{diff } i} = (\pi_{2i} - \pi_{2i-1}) / \pi_{2i-1 \ 2i}$ . It can be verified that  $\hat{\theta}_{\text{LIM}}$  is also unbiased.

1. Alain Th  berge, Ottawa, Ontario, Canada, K4C 1E2. E-mail: alain.theberge1@gmail.com.

It should be noted that the denominators in (1.2) correspond to the probability that at least one unit of the pair is in the sample. Thus, this estimator is reminiscent of inverse probability weighting, except it is based on pairs, instead of individual units. The numerators in (1.2) correspond to a value assigned to each pair with at least one sampled unit, and each observed pair is given a weight equal to the inverse of the probability of being observed. From the observation of only one unit of a pair, the estimator (1.2) assigns a value to the pair, and if the units of a pair are strongly correlated, this may be an efficient way to utilize this correlation. The estimator is a special case of a more general one that applies to more general populations, not only those with units grouped in pairs. Because it yields examples that give some insight into the general estimator, and because those examples can be given an explicit form that is simple to interpret and understand, Section 6 and Section 7 will also be about the case where the population, or a domain, is partitioned into pairs. The generalized inverse probability estimator is presented in Section 2; it depends on a parameter  $\Sigma$ , a positive definite  $N \times N$  matrix. In Section 3, the new estimator is applied to the problem of calibration. The choice of the parameter  $\Sigma$  is discussed in Section 4. In Section 5, it is seen that, with the right choice for  $\Sigma$ , the generalized calibration estimator is optimal, in the sense that it asymptotically attains a generalization of the Godambe-Joshi lower bound. Simple examples are given in Section 6, and the results of a simulation are presented in Section 7. Section 8 summarizes the paper.

## 2. The generalized inverse probability estimator

Estimators in this paper utilize a positive definite matrix  $\Sigma \in \mathbb{R}^{N \times N}$ . A matrix formulation of the estimators will therefore be useful. For a vector of interest  $\mathbf{y}' = (y_1, y_2, \dots, y_N)$  and  $\mathbf{1}_{N \times 1}$  a vector of ones, the inverse probability estimator of the total  $\theta = \sum_{i=1}^N y_i = \mathbf{y}' \mathbf{1}_{N \times 1}$  can be written

$$\begin{aligned} \hat{\theta}_{\text{IP}} &= \sum_{i=1}^N \frac{\delta_i y_i}{\pi_i} \\ &= \mathbf{y}' \Delta_s (E(\Delta_s))^{-1} \mathbf{1}_{N \times 1}, \end{aligned} \quad (2.1)$$

where  $\pi_i = E(\delta_i)$  is assumed greater than 0 for  $i = 1, 2, \dots, N$ , and  $\Delta_s$  is the  $N \times N$  diagonal matrix of the  $\delta_i$ .

The generalization of the inverse probability estimator relies on the Moore-Penrose inverse of a matrix  $\mathbf{M}$ , denoted  $\mathbf{M}^\dagger$ . The Moore-Penrose inverse is unique and always exists; it is equal to the ordinary inverse if the latter exists. A precise definition and properties of the Moore-Penrose inverse can be found in Ben-Israel and Greville (2002). In particular, it can be verified that  $\Delta_s^\dagger = \Delta_s$ . Since it is also true that  $\Delta_s^2 = \Delta_s$ , if  $\mathbf{I} \in \mathbb{R}^{N \times N}$  is the identity matrix, the inverse probability estimator can be written

$$\begin{aligned} \hat{\theta}_{\text{IP}} &= \mathbf{y}' \Delta_s (E(\Delta_s))^{-1} \mathbf{1}_{N \times 1} \\ &= \mathbf{y}' (\Delta_s \mathbf{I} \Delta_s)^\dagger (E(\Delta_s \mathbf{I} \Delta_s)^\dagger)^{-1} \mathbf{1}_{N \times 1}. \end{aligned} \quad (2.2)$$

If in (2.2), the identity matrix is replaced by any  $N \times N$  positive definite matrix  $\Sigma$ , one obtains the generalized inverse probability estimator or the generalized Horvitz-Thompson estimator,

$$\hat{\theta}_{\text{GIP}}(\Sigma) = \mathbf{y}' (\Lambda_s \Sigma \Lambda_s)^\dagger \left( E (\Lambda_s \Sigma \Lambda_s)^\dagger \right)^{-1} \mathbf{1}_{N \times 1}. \tag{2.3}$$

In the phrase “inverse probability”, the matrix  $E (\Lambda_s \Sigma \Lambda_s)^\dagger$  is now the “probability” and  $\left( E (\Lambda_s \Sigma \Lambda_s)^\dagger \right)^{-1}$  is the new “inverse probability”. The ordinary inverse probability estimator is simply a special case of  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , which can be obtained by choosing  $\Sigma = \mathbf{I}$ . As will be seen in c) below, one now has a family of unbiased estimators,  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , parameterized by  $\Sigma$ .

### 2.1 Notes on the generalized inverse probability estimator

- a) Although the vector  $\mathbf{y}$  appears in the estimator, only the sampled units affect the estimator’s value. This is because  $(\Lambda_s \Sigma \Lambda_s)^\dagger = \Lambda_s (\Lambda_s \Sigma \Lambda_s)^\dagger$ , thus (2.3) could have been written

$$\hat{\theta}_{\text{GIP}}(\Sigma) = (\Lambda_s \mathbf{y})' (\Lambda_s \Sigma \Lambda_s)^\dagger \left( E (\Lambda_s \Sigma \Lambda_s)^\dagger \right)^{-1} \mathbf{1}_{N \times 1}. \tag{2.4}$$

The proof of this and of many other results stated here may be found in Théberge (2017). The  $N \times 1$  vector  $\mathbf{w}_{s, \text{GIP}}(\Sigma) = (\Lambda_s \Sigma \Lambda_s)^\dagger \left( E (\Lambda_s \Sigma \Lambda_s)^\dagger \right)^{-1} \mathbf{1}_{N \times 1}$  gives the weights of  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , and all the units not in sample have a weight of zero.

- b) The matrix  $E (\Lambda_s \Sigma \Lambda_s)^\dagger$  is invertible under the assumptions that  $\pi_i = E(\delta_i)$  is greater than zero for  $i = 1, 2, \dots, N$  and that  $\Sigma$  is positive definite. Thus, (2.3) is well defined.
- c) By taking the expectation of (2.3), one immediately sees that  $\hat{\theta}_{\text{GIP}}(\Sigma)$  is unbiased for estimating  $\theta = \mathbf{y}' \mathbf{1}_{N \times 1}$ . This is true for any positive definite  $\Sigma$ . A poor choice of  $\Sigma$  may mean an estimator with a high variance, but it does not cause a bias.
- d) Often, there is no closed-form formula for  $E (\Lambda_s \Sigma \Lambda_s)^\dagger$ , but for single stage sampling plans at least, it can be easily approximated. One simply takes the average of a large number of values of  $(\Lambda_s \Sigma \Lambda_s)^\dagger$ , each computed for a different sample obtained with the sampling plan. The computation does not require the knowledge of any of the variables of interest. It is a “desk exercise” in the sense that it does not require contacting the units. It can even be carried out before the actual sample is selected.
- e) It is well known that for a total estimator utilizing a regression vector  $\beta$ ,  $\hat{T}(\beta)$ , is asymptotically equivalent in terms of bias and variance to the estimator  $\hat{T}(\hat{\beta}_s)$  where  $\hat{\beta}_s$  is an estimator that converges in probability to  $\beta$ . Similarly,  $\hat{\theta}_{\text{GIP}}(\hat{\Sigma}_s)$  has the same asymptotic bias and variance as  $\hat{\theta}_{\text{GIP}}(\Sigma)$  if the positive definite matrix  $\hat{\Sigma}_s$  converges in probability to the positive definite matrix  $\Sigma$ . In essence, if the sample size is sufficiently large, the error introduced by estimating  $\Sigma$  by  $\hat{\Sigma}_s$  is negligible compared to the error in  $\hat{\theta}_{\text{GIP}}(\Sigma)$  due to the sampling of units. All asymptotic results in this paper assume that the sampling plan is non informative (see, for example Cassel, Särndal and Wretman, 1977).

- f) When  $\Sigma = \mathbf{I}$ , then  $\hat{\theta}_{\text{GIP}}(\Sigma)$  reduces to the ordinary inverse probability estimator,  $\hat{\theta}_{\text{IP}}$ , as given in (2.1). This is the justification for referring to  $\hat{\theta}_{\text{GIP}}(\Sigma)$  as the generalized inverse probability estimator or the generalized Horvitz-Thompson estimator. It will be seen later, why this particular unbiased extension of the ordinary Horvitz-Thompson estimator is of interest.
- g) An arbitrary symmetric positive definite matrix  $\Sigma$  may contain up to  $N(N+1)/2$  distinct parameters. It is not feasible to specify so many values. If the sample  $s$  is utilized to estimate those parameters, the task of estimating  $N(N+1)/2$  parameters from  $n < N$  observations is clearly impossible. A simpler choice must be used. The simplest choice utilizes  $\Sigma = \mathbf{I}$ , as seen in f). There are other choices that have a reasonable number of parameters. One example is given in Section 6.
- h) For estimating a domain total  $\mathbf{y}'\mathbf{c}$  where  $\mathbf{c} = (c_1, \dots, c_i, \dots, c_N)'$  is a vector of known constants with  $c_i = 1$  or 0 depending on whether unit  $i$  is in the domain or not, it suffices to replace (2.3), which is for estimating the population total, with  $\mathbf{y}'(\Delta_s \Sigma \Delta_s)^\dagger (E(\Delta_s \Sigma \Delta_s)^\dagger)^{-1} \mathbf{c}$ . The weight vector  $(\Delta_s \Sigma \Delta_s)^\dagger (E(\Delta_s \Sigma \Delta_s)^\dagger)^{-1} \mathbf{c}$  varies with each domain described by  $\mathbf{c}$ ; however the weight matrix,  $(\Delta_s \Sigma \Delta_s)^\dagger (E(\Delta_s \Sigma \Delta_s)^\dagger)^{-1}$ , does not depend on the domain. There are  $N - n$  rows of this matrix that are nil. Even though there are potentially  $nN$  elements of the weight matrix that are non zero, post-multiplication by  $\mathbf{c}$  will give the weight vector for any domain described by  $\mathbf{c}$ .
- i) One possibility for the matrix  $\Sigma$  is one where all the diagonal elements are the same, and all the off-diagonal elements are the same. In this way, all the units are the same with respect to  $\Sigma$ . However, if all units are the same with respect to the sampling plan, for example simple random sampling or Bernoulli sampling, and if all units are the same with respect to the parameter estimated, for example a total or an average for all units, then by symmetry, every sampled unit will have the same weight. Since both  $\hat{\theta}_{\text{IP}}$  and  $\hat{\theta}_{\text{GIP}}(\Sigma)$  are unbiased, both estimators will have the same weights. Nonetheless, for domain parameters, because some units are in the domain and some not, the symmetry argument no longer holds and the value of the off-diagonal elements of  $\Sigma$  may make a difference in  $\hat{\theta}_{\text{GIP}}(\Sigma)$ .
- j) By setting  $\mathbf{y} = \mathbf{1}_{N \times 1}$  in  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , the estimator simply becomes the sum of all the weights of the sampled units and the parameter estimated becomes  $\mathbf{1}_{1 \times N} \mathbf{1}_{N \times 1} = N$ , the known total number of units. However, the sum of the weights does not necessarily equal  $N$ . This does not bode well for the variance of  $\hat{\theta}_{\text{GIP}}(\Sigma)$ . To fix this, calibration can be used. Calibration was introduced by Deville and Särndal (1992). At its simplest, it would consist of scaling the inverse probability weights, generalized or not, by a common factor so that the resulting final weights do add up to  $N$ . Even for the ordinary inverse probability estimator, for some sampling plans, the sum of the design weights does not necessarily equal  $N$ , and here too, the solution lies in calibration. The subject of calibration is examined in the next section.

### 3. The generalized calibration estimator

The sum of the weights of an estimator is an estimate of the known population size,  $N$ . When the sampling plan is such that the sample size is not fixed, the ordinary inverse probability estimator of the population size will have a variance greater than zero. The sum of the weights of  $\hat{\theta}_{GIP}(\Sigma)$ , noted  $S(\Sigma)$ , is often a worse estimator of the population size than the sum of the weights of  $\hat{\theta}_{IP}$ ; it will often vary, even when the sample size is fixed. An estimator whose estimates of the population size vary, cannot be seen as very reliable.

To fix the problem that the ordinary inverse probability estimator experiences when the sample size is variable, calibration can be used. The weights of  $\hat{\theta}_{CAL}$  are calibrated so that their sum equals the population size,  $N$ . A similar fix can be made to the generalized estimator:  $\hat{\theta}_{GCAL}(\Sigma) = (N / S(\Sigma)) \hat{\theta}_{GIP}(\Sigma)$ . The definition of  $\hat{\theta}_{GCAL}(\Sigma)$  will be expanded to include the possibility of more calibration equations involving more auxiliary variables. The use of calibration equations was presented in Deville and Särndal (1992).

With an auxiliary variable matrix  $\mathbf{X} \in \mathbb{R}^{N \times q}$  assumed to be of full rank and noting  $\|\mathbf{v}\|_{\mathbf{M}} = (\mathbf{v}'\mathbf{M}\mathbf{v})^{1/2}$  the weighted Euclidean norm of the vector  $\mathbf{v}$ , the following problem is addressed:

**Calibration Problem:** Among the weight vectors  $\mathbf{w}_s \in \mathbb{R}^N$  in the range of  $\Delta_s$ , i.e., non-sampled units should have a weight of 0, which minimize  $\|\mathbf{X}'\mathbf{w}_s - \mathbf{X}'\mathbf{1}_{N \times 1}\|_{\mathbf{T}}$ , i.e., which “best” satisfy the  $q$  calibration equations, seek one that minimizes  $\|\mathbf{w}_s - \mathbf{w}_{s,GIP}(\Sigma)\|_{\mathbf{U}}$ , i.e., as close as possible to the weights of  $\hat{\theta}_{GIP}(\Sigma)$ , where  $\mathbf{T} \in \mathbb{R}^{q \times q}$  and  $\mathbf{U} \in \mathbb{R}^{N \times N}$  are positive definite matrices.

Weights,  $\mathbf{w}_s$ , that satisfy the calibration equations,  $\mathbf{X}'\mathbf{w}_s = \mathbf{X}'\mathbf{1}_{N \times 1}$ , 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. The matrix  $\mathbf{U}$  specifies the relative importance given to each unit when measuring the distance from  $\mathbf{w}_{s,GIP}(\Sigma)$ . This formulation of the calibration problem generalizes that of Théberge (1999), where  $\mathbf{T}$  and  $\mathbf{U}$  were diagonal matrices, and the inverse probability, or Horvitz-Thompson, weights were used instead of the generalized inverse probability weights.

The solution to the calibration problem yields

$$\begin{aligned} \hat{\theta}_{GCAL}(\Sigma) &= \mathbf{y}'\mathbf{w}_{s,GCAL}(\Sigma) \\ &= \hat{\mathbf{y}}'\mathbf{1}_{N \times 1} + (\mathbf{y} - \hat{\mathbf{y}})'\mathbf{w}_{s,GIP}(\Sigma), \end{aligned} \tag{3.1}$$

where  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$  with

$$\hat{\boldsymbol{\beta}} = \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{y}. \tag{3.2}$$

The estimator  $\hat{\theta}_{GCAL}(\Sigma)$  is asymptotically unbiased. Also, if  $\hat{\Sigma}_s \rightarrow \Sigma$  in probability, then the bias and variance of  $\hat{\theta}_{GCAL}(\hat{\Sigma}_s)$  are asymptotically the same as those of  $\hat{\theta}_{GCAL}(\Sigma)$ . The rate at which  $\hat{\Sigma}_s \rightarrow \Sigma$  will depend on the estimator  $\hat{\Sigma}_s$  and on the number of parameters in  $\Sigma$ .

The difference between  $\hat{\theta}_{\text{GCAL}}(\Sigma)$  and the ordinary calibration estimator,  $\hat{\theta}_{\text{GCAL}}(\mathbf{I}_{N \times N})$ , is simply the use of generalized inverse probability weights to estimate the sum of the residues, rather than the usual inverse probability weights. This was to be expected given that in one case we are, in the calibration problem, seeking weights that minimize  $\|\mathbf{w}_s - \mathbf{w}_{s\text{GIP}}(\Sigma)\|_{\mathbf{U}}$ , instead of weights that minimize  $\|\mathbf{w}_s - \mathbf{w}_{s\text{IP}}\|_{\mathbf{U}}$ , where  $\mathbf{w}_{s\text{IP}} = \mathbf{w}_{s\text{GIP}}(\mathbf{I}_{N \times N})$  are the usual inverse probability weights.

The following result is proven in the Appendix: for any  $\alpha \in \mathbb{R}^N$ , if  $\Delta_s \alpha$  is in the range of  $\Delta_s \mathbf{X}$ , then the weighted sum of residuals,  $(\mathbf{y} - \hat{\mathbf{y}})' (\Delta_s \mathbf{U} \Delta_s)^\dagger \alpha$ , is zero. A vector  $\mathbf{v}$  is said to be in the range of a matrix  $\mathbf{F}$  if there exists a vector  $\lambda$  such that  $\mathbf{v} = \mathbf{F}\lambda$ . In particular, if the matrix  $\mathbf{U}$  is diagonal and written  $\mathbf{U} = \mathbf{A}^{-1} \mathbf{D}$ , where  $\mathbf{A} = (E(\Delta_s))^{-1}$  is the diagonal matrix of the ordinary inverse probability weights and  $\mathbf{D} \in \mathbb{R}^{N \times N}$  is an arbitrary positive diagonal matrix, then with  $\alpha = \mathbf{D}\mathbf{c}$  the result gives that  $(\mathbf{y} - \hat{\mathbf{y}})' (\Delta_s \mathbf{A}^{-1} \mathbf{D} \Delta_s)^\dagger \mathbf{D}\mathbf{c} = (\mathbf{y} - \hat{\mathbf{y}})' \mathbf{A} \Delta_s \mathbf{c}$  is zero if  $\Delta_s \mathbf{D}\mathbf{c}$  is in the range of  $\Delta_s \mathbf{X}$ . This is similar to result 6.5.1 of Särndal, Swensson and Wretman (1992), for example, where  $\mathbf{c}$  is a vector of ones and the diagonal elements of  $\mathbf{D}$  are variances.

It can be seen from the form of (3.1), that  $\hat{\theta}_{\text{GCAL}}(\Sigma)$  is also a regression estimator that uses a model  $\xi$  such that  $E_\xi(\mathbf{y}) = \mathbf{X}\beta$ . Despite the notation used in (3.2), calibration estimators do not use models, instead there are calibration equations. When viewed as a regression estimator, it is important to realize that  $\hat{\theta}_{\text{GCAL}}(\Sigma)$  is asymptotically design unbiased, regardless of the choice of the model parameter  $\beta$ , and regardless of the choice of the positive definite matrix  $\Sigma$ .

## 4. The choice of the positive definite matrix $\Sigma$

Different choices for  $\Sigma$  will generally lead to different generalized inverse probability estimators and different generalized calibration estimators. The advantage of the generalization of the inverse probability estimator comes from its use in a generalization of calibration, as seen in Section 3, and the optimality of generalized calibration, as discussed in Section 5. It will be seen that a matrix  $\Sigma$  is an appropriate choice to use for  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , if a model  $\xi$  with  $V_\xi(\mathbf{y}) = \Sigma$  is an appropriate model for  $\mathbf{y}$ . Even if the assumption that  $V_\xi(\mathbf{y}) = \Sigma$  is wrong, the estimator  $\hat{\theta}_{\text{GIP}}(\Sigma)$  remains design unbiased and the estimator  $\hat{\theta}_{\text{GCAL}}(\Sigma)$  remains asymptotically design unbiased. The generalized calibration estimators with  $\Sigma = V_\xi(\mathbf{y})$  can be said to be model assisted as opposed to model based or model dependent (see Särndal et al., 1992, Section 6.7). The ordinary calibration estimators,  $\hat{\theta}_{\text{CAL}}$  use (3.1) with  $\Sigma = \mathbf{I}$ . A model that fits the population perfectly is not necessary, but hopefully a better model than one with  $V_\xi(\mathbf{y}) = \mathbf{I}$  can be utilized. In fact, if  $\Sigma$  is any positive diagonal matrix, then  $\hat{\theta}_{\text{GIP}}(\Sigma)$  will result in the ordinary inverse probability estimator, and the generalized calibration estimator will result in the ordinary calibration estimator. Often, a more appropriate model for  $\mathbf{y}$  would have  $V_\xi(\mathbf{y})$  non-diagonal. As for the variance of  $\hat{\theta}_{\text{GIP}}(\Sigma)$ , it may be higher than that of the ordinary inverse probability estimator, even if  $V_\xi(\mathbf{y}) = \Sigma$ . It is the calibration of  $\hat{\theta}_{\text{GIP}}(\Sigma)$  that yields, as will be seen in Section 5, an optimal estimator.

The use of a block-diagonal matrix simplifies the computation of inverses needed in (2.3). Blocks may correspond to persons of a household, students of a class, workers of an establishment, dwellings of a

block, etc. It is often natural for units belonging to the same block to have a correlated variable of interest. For example, how one worker rates their employer is likely correlated with the rating of another worker of the same employer; the race or religion of a couple is often the same. In such cases, a multistage sampling plan would often be used, but it will be assumed here that a single stage plan is used. This could be because a single stage sampling plan was more suitable for other variables of interest of the same survey, or because some unit level characteristics are so important, that it is desirable to stratify at the population level so that the sample can be targeted at certain strata. For example, it may be important to stratify persons by age, but households can't be stratified by age.

In the simulation presented in this paper, the vaccination status of individuals in two-person households is made to be correlated. An extreme case presents itself if the blocks are persons of a same household and the variable of interest is household income. In such a case the correlation is perfect, and lines of  $\Sigma$  corresponding to persons from a same household should be identical. Such a matrix  $\Sigma$  is not positive definite, but it is the limit of a sequence of positive definite matrices, and the limit of the corresponding sequence of generalized inverse probability estimators could be computed. The example (1.2) given in the introduction is based on this idea.

If  $\Sigma$  is block-diagonal with blocks  $\Sigma_1, \Sigma_2, \dots, \Sigma_B$ , then because both the Moore-Penrose inverse and the ordinary inverse of a block-diagonal matrix is the block-diagonal matrix of inverses, the estimator  $\hat{\theta}_{GIP}(\Sigma)$  can be decomposed into

$$\begin{aligned} \hat{\theta}_{GIP}(\Sigma) &= \sum_{b=1}^B \hat{\theta}_{GIP_b}(\Sigma_b) \\ &= \sum_{b=1}^B \mathbf{y}'_b (\Delta_{s_b} \Sigma_b \Delta_{s_b})^\dagger \left( E(\Delta_{s_b} \Sigma_b \Delta_{s_b})^\dagger \right)^{-1} \mathbf{1}_{N_b \times 1}, \end{aligned} \tag{4.1}$$

where  $N_b$  is the size of block  $b$ ,  $\mathbf{y}_b$  and  $\Delta_{s_b}$  are the sub-vector and sub-matrix respectively, which correspond to block  $b$ .

If the population is partitioned into blocks of correlated units, the variable defining the blocks must be on the frame. But that variable need not be perfect. For example, a unit's household may only be known at the time of the survey, but using an outdated household variable available on the frame will still be useful, while not introducing any bias. It simply means that the strength borrowed by the generalized inverse probability estimator from the correlations will be reduced. On the other hand, the strength borrowed from the correlations by the ordinary inverse probability estimator is nil.

If a positive definite estimator  $\hat{\Sigma}_s$  converges to a positive definite  $\Sigma$  in probability, then the bias and variance of  $\hat{\theta}_{GIP}(\hat{\Sigma}_s)$  are asymptotically the same as those of  $\hat{\theta}_{GIP}(\Sigma)$ . In practice, even if the general form of  $\Sigma$  depends on  $N(N-1)/2$  covariances, the number of parameters in  $\Sigma$  should be small compared to the sample size. Using the inverse probability estimator means assuming all covariances are zero. When using the generalized inverse probability estimator, one could assume that those covariances depend on a few parameters, and that those parameters are considered fixed, rather than estimated from the sample. In the examples of Section 6,  $\Sigma$  depends on only one parameter,  $\rho$ , and its value is assumed to be 1.

## 5. The generalized Godambe-Joshi lower bound

For any unbiased estimator  $\hat{\theta}$  of the population total  $\theta$ , if  $V_p(\hat{\theta})$  is the variance of  $\hat{\theta}$  under the sampling plan, Godambe and Joshi (1965) have given a lower bound for the value of  $E_{\xi}V_p(\hat{\theta})$  under the assumption that the variance matrix  $V_{\xi}(\mathbf{y})$  was diagonal. That lower bound is the sum of the elements of the diagonal matrix  $\left((E(\Delta_s))^{-1} - \mathbf{I}\right)V_{\xi}(\mathbf{y})$ . That result is generalized in the following paragraph.

For any linear unbiased total estimator,  $\hat{\theta}$ , if  $V_{\xi}(\mathbf{y})$  is positive definite, then  $E_{\xi}V_p(\hat{\theta})$  is not lower than the sum of the elements of the matrix  $\left(E(\Delta_s V_{\xi}(\mathbf{y}) \Delta_s^{\dagger})^{-1} - V_{\xi}(\mathbf{y})\right)$ . It is easily verified that the usual Godambe-Joshi lower bound is obtained if  $V_{\xi}(\mathbf{y})$  is diagonal.

Just as the calibration estimator asymptotically attains the Godambe-Joshi lower bound, the generalized calibration estimator with  $\Sigma = V_{\xi}(\mathbf{y})$ , asymptotically attains the generalized Godambe-Joshi lower bound, regardless of the value of the matrices  $\mathbf{X}$ ,  $\mathbf{T}$  and  $\mathbf{U}$ . The link between the value of those three matrices and the value of  $V_{\xi}(\mathbf{y})$  is not examined in this paper, but the calibration problem stated in Section 3 does clarify the role of each of those matrices. The derivation of the generalized lower bound and the proof of the optimality of the generalized calibration estimator are given in Théberge (2017).

The fact that  $\hat{\theta}_{\text{GCAL}}(V_{\xi}(\mathbf{y}))$  asymptotically attains the generalized Godambe-Joshi lower bound shows that the generalized inverse probability estimator performs well when applied to residuals, as it does in (3.1), even though it is not recommended in general. Similarly, the ordinary inverse probability estimator can be inefficient if the sample size is random, but will perform well if applied to residuals.

It should be noted that, contrary to the ordinary Godambe-Joshi lower bound, the generalized lower bound applies only to *linear* unbiased estimators. In fact, an example with  $V_{\xi}(\mathbf{y})$  not diagonal, of a non-linear unbiased estimator which does better than the lower bound is given in Théberge (2017).

## 6. Examples

There are cases simple enough for  $\hat{\theta}_{\text{GIP}}(\Sigma)$  to be given explicitly. Say  $\Sigma(\rho)$  is a block-diagonal matrix where each of  $N_p = N/2$  blocks equals  $\sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ , with  $-1 < \rho < 1$ . Such a block-diagonal matrix corresponds to a model of a population which can be partitioned into pairs  $\{2i-1, 2i\}$  ( $i=1, 2, \dots, N_p$ ) where, within a pair, the variable of interest is correlated. Then, (2.3) reduces to

$$\hat{\theta}_{\text{GIP}}(\Sigma(\rho)) = \sum_{i=1}^{N_p} \frac{a_{2i-1}y_{2i-1} + a_{2i}y_{2i}}{\left(\pi_{2i-1}\pi_{2i}(1-\rho^2) + (\pi_{2i-1} + \pi_{2i} - \pi_{2i-12i})\pi_{2i-12i}\rho^2\right)}, \quad (6.1)$$

where

$$\begin{aligned} a_{2i-1} &= \delta_{2i-1} \left[ \pi_{2i}(1-\rho^2) + \pi_{2i-12i}\rho(1+\rho) \right] + \delta_{2i-1}\delta_{2i} \left[ \rho^2\pi_{2i} - \rho\pi_{2i-1} - \rho^2\pi_{2i-12i} \right] \\ a_{2i} &= \delta_{2i} \left[ \pi_{2i-1}(1-\rho^2) + \pi_{2i-12i}\rho(1+\rho) \right] + \delta_{2i-1}\delta_{2i} \left[ \rho^2\pi_{2i-1} - \rho\pi_{2i} - \rho^2\pi_{2i-12i} \right]. \end{aligned} \quad (6.2)$$

Once again, this generalized inverse probability estimator is unbiased, for any value of  $\rho$ , “correct” or not. It is seen that as expected, when  $\rho=0$ , the estimator reduces to the inverse probability estimator. The

value of  $\rho$  cannot simply be set to one in (6.1), because  $\Sigma(1)$  is not positive definite. However, the limit of (6.1) as  $\rho \rightarrow 1$  results in the estimator (1.2) given in the Introduction,  $\hat{\theta}_{LIM}$ . It can be calibrated so that the sum of the weights is equal to  $N$ . If the probabilities of inclusion do not vary with  $i = 1, 2, \dots, N_p$ , the resulting estimator is

$$\hat{\theta}_{LICAL} = \frac{N_p}{\nu_p} \sum_{i=1}^{N_p} 2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - (y_{2i-1} + y_{2i})\delta_{2i-1}\delta_{2i} + (y_{2i-1} - y_{2i})\delta_{2i-1}\delta_{2i}\pi_{diff\ i}, \quad (6.3)$$

where  $\nu_p = \sum_{i=1}^{N_p} (\delta_{2i-1} + \delta_{2i} - \delta_{2i-1}\delta_{2i})$  is the number of pairs with at least one unit in the sample. It is easy to verify, by setting  $\mathbf{y} = \mathbf{1}_{N \times 1}$  in (6.3), that the sum of the weights of  $\hat{\theta}_{LICAL}$  is equal to  $2N_p = N$ . The generalized calibration estimator (6.3) is optimized for  $\rho \rightarrow 1$ , but it can still have a lower variance than both, the inverse probability estimator and the ordinary calibration estimator, if the correlation between the units of a pair is strong (for example, race, religion or education level of a couple). Since a variable indicating which unit is paired with which, must be on the frame, a calibration at the pair level would be possible. The calibration would ensure that the sum of the weights of the sampled units of a pair would equal 2. However, the low number of observations per calibration group would not ensure the validity of asymptotic results and could result in significant biases.

There are modified versions of the generalized inverse probability estimator and of the generalized calibration estimator. The modified versions have the advantage of having a closed form; there is no need to compute the expectation of  $(\Lambda_s \Sigma \Lambda_s)^\dagger$ . They also do not rely on the Moore-Penrose inverse. For a positive definite matrix  $\Sigma$ , they are defined as

$$\hat{\theta}_{MGIP}(\Sigma) = \mathbf{y}' \Lambda_s \Sigma^{-1} \Lambda_s (\Sigma^{-1} \circ \mathbf{\Pi})^{-1} \mathbf{1}_{N \times 1} \quad (6.4)$$

and

$$\hat{\theta}_{MGICAL}(\Sigma) = \hat{\mathbf{y}}' \mathbf{1}_{N \times 1} + (\mathbf{y} - \hat{\mathbf{y}})' \mathbf{w}_{sMGHT}(\Sigma), \quad (6.5)$$

where  $\mathbf{\Pi} = (\pi_{kl}) = (E(\delta_k \delta_l)) \in \mathbb{R}^{N \times N}$  is the matrix of second order probabilities of inclusion,  $\mathbf{w}_{sMGIP}(\Sigma)$  is the vector of weights of  $\hat{\theta}_{MGIP}(\Sigma)$ , and  $\circ$  denotes the Hadamard product, i.e., element-wise multiplication. With  $\hat{\theta}_{MGIP}(\Sigma)$ , the “probability” part of the phrase “inverse probability” is  $\Sigma^{-1} \circ \mathbf{\Pi}$ . The modified generalized estimators are also unbiased, or at least asymptotically unbiased in the case of  $\hat{\theta}_{MGICAL}(\Sigma)$ . The usual estimators  $\hat{\theta}_{IP}$  and  $\hat{\theta}_{CAL}$  are obtained if  $\Sigma = \mathbf{I}$ .

If  $\rho \rightarrow 1$ , the modified generalized inverse probability estimator,  $\hat{\theta}_{MGIP}(\Sigma(\rho))$ , becomes:

$$\hat{\theta}_{MLIM} = \sum_{i=1}^{N_p} w_{2i-1} y_{2i-1} + w_{2i} y_{2i}, \quad (6.6)$$

where

$$w_{2i-1} = \frac{\delta_{2i-1} (\pi_{2i} + \pi_{2i-1\ 2i}) - \delta_{2i-1} \delta_{2i} (\pi_{2i-1} + \pi_{2i-1\ 2i})}{\pi_{2i-1} \pi_{2i} - \pi_{2i-1\ 2i}^2} \quad (6.7)$$

and

$$w_{2i} = \frac{\delta_{2i}(\pi_{2i-1} + \pi_{2i-1, 2i}) - \delta_{2i-1}\delta_{2i}(\pi_{2i} + \pi_{2i-1, 2i})}{\pi_{2i-1}\pi_{2i} - \pi_{2i-1, 2i}^2}. \quad (6.8)$$

If the sampling plan is such that  $\pi_{2i} = \pi_{2i-1}$  for any  $i=1, 2, \dots, N_p$ , and if both units of that pair are sampled, then the weights of both units will be zero. That some sampled units may not contribute to the estimator, in some circumstances, is an undesirable property of  $\hat{\theta}_{\text{MLIM}}$ .

One characteristic of the estimator  $\hat{\theta}_{\text{LIM}}$  is somewhat surprising. It is constructed in such a way that for each observed pair, that is each pair with at least one unit in the sample, the numerator in (1.2) corresponds to a value for the pair's variable of interest total. The numerator of the  $i^{\text{th}}$  term is 0 if neither unit  $2i-1$  nor unit  $2i$  are observed, it is  $2y_{2i-1}$  if only unit  $2i-1$  of the pair is sampled, it is  $2y_{2i}$  if only unit  $2i$  of the pair is sampled, and it is  $(y_{2i-1} + y_{2i}) + (y_{2i-1} - y_{2i})\pi_{\text{diff } i}$  if both units of the pair are sampled. The unexpected characteristic is that when both units of a pair  $i$  ( $i=1, 2, \dots, N_p$ ) are observed, the estimated value for the pair's total is not the known total  $y_{2i-1} + y_{2i}$ . This is the motivation for yet another estimator and its calibrated version, where the estimate for a pair, while still being unbiased, will agree with the known total when both units of the pair are sampled. The alternative estimators are

$$\hat{\theta}_{\text{ALIM}} = \sum_{i=1}^{N_p} \frac{(a_i\delta_{2i-1} + b_i\delta_{2i-1}\delta_{2i})y_{2i-1} + (c_i\delta_{2i} + d_i\delta_{2i-1}\delta_{2i})y_{2i}}{\pi_{2i-1} + \pi_{2i} - \pi_{2i-1, 2i}} \quad (6.9)$$

and

$$\hat{\theta}_{\text{ALCAL}} = \hat{\mathbf{y}}' \mathbf{1}_{N \times 1} + (\mathbf{y} - \hat{\mathbf{y}})' \mathbf{w}_{s\text{ALIM}}(\boldsymbol{\Sigma}), \quad (6.10)$$

where  $\mathbf{w}_{s\text{ALIM}}$  is the vector of weights of  $\hat{\theta}_{\text{ALIM}}$ ,  $a_i + b_i = c_i + d_i = 1$ , motivated by what is wanted when both units of the pair are sampled, and in order to have  $\hat{\theta}_{\text{ALIM}}$  unbiased, one should have  $a_i\pi_{2i-1} + b_i\pi_{2i-1, 2i} = c_i\pi_{2i} + d_i\pi_{2i-1, 2i} = \pi_{2i-1} + \pi_{2i} - \pi_{2i-1, 2i}$ . Therefore, for  $i=1, 2, \dots, N_p$ ,

$$\begin{aligned} a_i &= \frac{\pi_{2i-1} + \pi_{2i} - 2\pi_{2i-1, 2i}}{\pi_{2i-1} - \pi_{2i-1, 2i}} \\ b_i &= \frac{\pi_{2i-1, 2i} - \pi_{2i}}{\pi_{2i-1} - \pi_{2i-1, 2i}} \\ c_i &= \frac{\pi_{2i-1} + \pi_{2i} - 2\pi_{2i-1, 2i}}{\pi_{2i} - \pi_{2i-1, 2i}} \\ d_i &= \frac{\pi_{2i-1, 2i} - \pi_{2i-1}}{\pi_{2i} - \pi_{2i-1, 2i}}. \end{aligned} \quad (6.11)$$

## 7. Simulation results

In this simulation, estimators from the preceding section will be compared to the ordinary inverse probability estimator and the ordinary calibrated estimator. A population of 2,000 individuals grouped into 1,000 two-person households was generated. Persons  $2i$  and  $2i-1$  for  $i=1, 2, \dots, 1,000$  belong to the same household. A variable of interest  $y$  takes the value 1 to represent a vaccinated person, and it takes

the value 0 to represent an unvaccinated person. To simulate how vaccination status can be correlated within household, the method of Lunn and Davies (1998) was used to generate pairs of correlated Bernoulli variables with a probability of 0.7 of a value of 1 and a correlation of 0.8. The actual population generated has 254 households where neither person is vaccinated, 660 households where both are vaccinated, 44 households where only the person with an odd label is vaccinated, and 42 households where only the person with an even label is vaccinated. The total number of persons vaccinated is  $660 \times 2 + 44 + 42 = 1,406$  for a vaccination rate of 0.703. The correlation between persons of the same household is  $(0.66 - 0.704 \times 0.702) / (\sqrt{0.704 \times 0.296} \times \sqrt{0.702 \times 0.298}) = 0.7941$ .

The population was sampled 10,000 times. Each household  $i (i = 1, 2, \dots, N_p)$  was sampled independently; the probability of selecting both units was 0.05, the probability of selecting only unit  $2i - 1$  was 0.10, and the probability of selecting only unit  $2i$  was 0.05. Thus, for each sample, the probabilities of inclusion were  $\pi_{2i-1} = 0.15$ ,  $\pi_{2i} = 0.1$  and  $\pi_{2i-12i} = 0.05$ . This means  $\pi_{\text{diff } i} = (\pi_{2i} - \pi_{2i-1}) / \pi_{2i-12i}$  in (1.2) was chosen to not be zero. This is because when  $\pi_{\text{diff } i}$  is zero,  $\hat{\theta}_{\text{LIM}}$  is a somewhat obvious choice: it is an inverse probability estimator based on pairs where the pair is given a value of  $2y_{2i-1}$  if only unit  $2i - 1$  is sampled, a value of  $2y_{2i}$  if only unit  $2i$  is sampled, and a value of  $y_{2i-1} + y_{2i}$  if both units are sampled. Combined with calibration, it is an obvious competitor to the ordinary calibration estimator. Why not base the estimator on pairs in this way, rather than units, if there is a strong correlation between units of a pair? When  $\pi_{\text{diff } i}$  is zero, it is also true that  $\hat{\theta}_{\text{LIM}} = \hat{\theta}_{\text{ALIM}}$ . It is interesting to find out how  $\hat{\theta}_{\text{LIM}}$  compares when  $\pi_{\text{diff } i}$  is not zero. For each sample, eight estimators of the total were calculated: the inverse probability estimator, the ordinary calibrated estimator, the generalization of the inverse probability estimator and its calibrated version, the modified generalized inverse probability estimator and its calibrated version, and finally the alternative estimator and its calibrated version. For the generalized and modified generalized estimators, including their calibrated versions,  $\Sigma(\rho)$  with  $\rho \rightarrow 1$  was used, as explained in the examples of the preceding section. The simple closed-form formulae of that section can thus be used. For the calibration,  $\mathbf{X} = \mathbf{1}_{N \times 1}$  with  $\mathbf{T} = 1$  and  $\mathbf{U} = \mathbf{I}_{N \times N}$  yields  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \frac{\sum_{k \in s} y_k}{\sum_{k \in s} x_k} \mathbf{1}_{N \times 1}$ . The average total and the variance over the 10,000 repetitions are given in Table 7.1.

**Table 7.1**  
**Simulation results comparing eight estimators**

<b>Estimator and lower bounds</b>	<b>Total</b>	<b>Variance</b>
Inverse probability	1,406.60	13,326
Calibrated inverse probability	1,407.38	3,856
Generalized inverse probability	1,406.41	11,226
Calibrated generalized inverse probability	1,407.08	3,419
Modified generalized inverse probability	1,406.37	16,337
Calibrated modified generalized inverse probability	1406.68	4,932
Alternative	1,406.61	12,447
Calibrated alternative	1,407.12	3,697
Generalized Godambe-Joshi lower bound ( $\rho = 0.8$ )		3,408
Generalized Godambe-Joshi lower bound ( $\rho \rightarrow 1$ )		3,360

All eight estimators are either unbiased or asymptotically unbiased, so as expected, the observed bias of each estimator is negligible, since the real population total is 1,406.

The observed variances show that only the four calibrated estimators have reasonable variances. With the sampling plan used for this simulation, only the calibrated estimators can estimate the known population total with zero variance.

The calibrated generalized inverse probability estimator, with a variance of 3,419, performs best. This despite being calculated assuming that the correlation between the units of a pair is one. It should be remembered that the calibrated inverse probability estimator, with a variance of 3,856, is a special case of the calibrated generalized inverse probability estimator, but it is computed assuming that the correlation between the units of a pair is zero. The calibrated alternative estimator, which contrary to the other estimators, has been defined only for a household size of 2, has a variance somewhere in between that of the calibrated versions of the inverse probability and generalized inverse probability estimators. Finally, the calibrated modified generalized estimator had the highest variance of the four calibrated estimators.

The generalized Godambe-Joshi lower bound with the variance matrix,  $V_{\xi}(\mathbf{y})$ , of the model  $\xi$  used to generate  $\mathbf{y}$  is 3,408. This is the asymptotic variance that could be expected of the calibrated generalized estimator, if it had been calculated with a matrix  $\Sigma = V_{\xi}(\mathbf{y})$  based on the correct model  $\xi$ , where the correlation between units of a pair is 0.8. If  $\Sigma(\rho)$  is defined as in the preceding section, and  $\text{GJ}(\Sigma(\rho))$  is the generalized Godambe-Joshi lower bound for the positive definite variance matrix  $V_{\xi}(\mathbf{y}) = \Sigma(\rho)$ , then the limit as  $\rho \rightarrow 1$  of  $\text{GJ}(\Sigma(\rho))$  is 3,360. This is the variance that could be expected of the generalized calibration estimator, if the correlation between units of a same pair was one.

## 8. Summary

The concept of inverse probability estimation can be generalized with a positive definite matrix  $\Sigma$ . There is then a whole family of unbiased estimators parameterized by  $\Sigma$  where one member, with  $\Sigma = \mathbf{I}_{N \times N}$  is the usual inverse probability estimator. The concept of calibration can also be generalized so that weights close to those of the generalized inverse probability estimator are sought. The Godambe and Joshi lower bound of  $E_{\xi} V_p(\hat{\theta})$  can also be generalized to a model  $\xi$  where the variance matrix  $V_{\xi}(\mathbf{y})$  is not necessarily diagonal. The calibrated generalized inverse probability estimator, with  $\Sigma = V_{\xi}(\mathbf{y})$ , asymptotically attains the generalized lower bound for any linear unbiased estimator  $\hat{\theta}$ . The new estimators are model assisted, not model based. They remain unbiased, or at least asymptotically unbiased, even if  $\Sigma \neq V_{\xi}(\mathbf{y})$ .

Examples where the new estimators can be given an explicit form have been presented. Simulations comparing those new estimators with the usual ones have been done. Those simulations show that, while remaining asymptotically unbiased, significant improvements in variance can be obtained in situations where there is significant correlation between some units of the population, as for example there would be,

between persons of a same household with regards to vaccination status. Improvements in variance can still be made, even with  $\Sigma \neq V_{\xi}(\mathbf{y})$ .

## Acknowledgements

I would like to thank the Associate Editor and the referees for their constructive comments and suggestions to improve the paper.

## Appendix

Proof that with  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ ,  $\hat{\boldsymbol{\beta}} = \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{y}$ , where  $\mathbf{T}$  and  $\mathbf{U}$  are positive definite, then for any  $\boldsymbol{\alpha} \in \mathbb{R}^N$ , if  $\Delta_s \boldsymbol{\alpha}$  is in the range of  $\Delta_s \mathbf{X}$ , then the weighted sum of residuals,  $(\mathbf{y} - \hat{\mathbf{y}})' (\Delta_s \mathbf{U} \Delta_s)^\dagger \boldsymbol{\alpha}$ , is zero.

First,

$$\begin{aligned} (\mathbf{y} - \hat{\mathbf{y}})' (\Delta_s \mathbf{U} \Delta_s)^\dagger \boldsymbol{\alpha} &= \mathbf{y}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \left[ \mathbf{I} - \mathbf{X} \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \right] \boldsymbol{\alpha} \\ &= \mathbf{y}' \mathbf{M} \boldsymbol{\alpha}. \end{aligned} \quad (\text{A.1})$$

With  $\Delta_s$  being an orthogonal projection, note that by Lemma 2 of Théberge (2017),  $\mathbf{M} = \mathbf{M} \Delta_s$ , and that by the properties of the Moore-Penrose inverse,  $\mathbf{M} \Delta_s \mathbf{X} \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger = \mathbf{0}$ . For  $\mathbf{T}$  and  $\mathbf{U}$  of full rank, one has that the rank of  $\Delta_s \mathbf{X} \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger$  equals the rank of  $\Delta_s \mathbf{X} \mathbf{T}^{1/2}$ . It then follows that the range of  $\Delta_s \mathbf{X}$ , which equals the range of  $\Delta_s \mathbf{X} \mathbf{T}^{1/2}$ , equals the range of  $\Delta_s \mathbf{X} \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger$  by exercise 1.10 of Ben-Israel and Greville (2002). Therefore, if  $\Delta_s \boldsymbol{\alpha}$  is in the range of  $\Delta_s \mathbf{X}$  which equals the range of  $\Delta_s \mathbf{X} \mathbf{T}^{1/2} \left( \mathbf{T}^{1/2} \mathbf{X}' (\Delta_s \mathbf{U} \Delta_s)^\dagger \mathbf{X} \mathbf{T}^{1/2} \right)^\dagger$ , then we will have  $\mathbf{M} \Delta_s \boldsymbol{\alpha} = \mathbf{M} \boldsymbol{\alpha} = \mathbf{0}$ .

## References

- Ben-Israel, A., and Greville, T.N.E. (2002). *Generalized Inverses: Theory and Applications (Second Ed.)*. New York: Springer-Verlag.
- Cassel, C.-M., Särndal, C.-E. and Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling*. New York: John Wiley & Sons, Inc.
- Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Godambe, V.P., and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations, 1. *Annals of Mathematical Statistics*, 36, 1707-1722.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Lunn, A.D., and Davies, S.J. (1998). A note on generating correlated binary variables. *Biometrika*, 85, 487-490.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Théberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94, 635-644.

Théberge, A. (2017). Estimation when the covariance structure of the variable of interest is positive definite. *Journal of Official Statistics*, 33, 275-299.