A Generalization of the Horvitz-Thompson Estimator

Alain Théberge[1]
[1]alain.theberge1@gmail.com

Abstract
In finite population estimation, the 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. The estimator works well when the variance matrix of the vector of interest, or the vector of residuals if a model is used, is diagonal; when there is no correlation. Here, the Horvitz-Thompson estimator is generalized. The generalized estimator will be useful in the presence of correlation. Since calibration estimation seeks weights that are close to the Horvitz-Thompson weights, it too can be generalized by seeking weights that are close to those of the generalized Horvitz-Thompson estimator. Calibration is known to be optimal, in the sense that it asymptotically attains the Godambe-Joshi lower bound. That lower bound has also been derived under the assumption that no correlation is present. This too, can be generalized to allow for correlation. Generalized calibration asymptotically attains the generalized lower bound.

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 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 Horvitz-Thompson estimator.

Key Words: calibration estimator, Godambe-Joshi lower bound, Horvitz-Thompson estimator

1. Introduction

For a simple random sample of a population of N units which is grouped into $N_p = N/2$ pairs, there is a simple unbiased estimator of the total, $\sum_{k=1}^{N} y_k$, that is an alternative to the Horvitz-Thompson estimator presented in Horvitz and Thompson (1952). It uses no auxiliary information, and it is not model based.

The new estimator is

$$\sum_{i=1}^{N_p} \frac{\left(2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - \left(y_{2i-1} + y_{2i}\right)\delta_{2i-1}\delta_{2i}\right)}{P\left(\text{observe } i\text{-th pair}\right)}, \tag{1}$$

where N_p is the number of pairs on the frame, $P\left(\text{observe } i\text{-th pair}\right)$ is the probability that at least one unit of the pair is in the sample s, and $\delta_k = 1$ if $k \in s$, $\delta_k = 0$ otherwise.

The new estimator is reminiscent of the Horvitz-Thompson estimator, except it works with pairs, instead of individual units. It assigns a value to each pair with a sampled unit. The value associated to the *i*-th pair is twice the value of the sampled unit if only one unit is sampled, and it is the sum of the two values if both units are sampled. This approach is possible because the probability of sampling one unit is equal to the probability of sampling the other unit of the pair. The estimator is a special case of a more general one that applies to more complex sampling plans. Because it yields examples that are simple to interpret and understand, Section 6 and Section 7 will also be about the case where the population, or a domain, is grouped into pairs. The generalized Horvitz-Thompson estimator is presented in the next section, it depends on a parameter $\mathbf{\Sigma}$, a positive definite $N \times N$ matrix. The choice of that parameter is discussed in Section 3. In Section 4, the new estimator is applied to the problem of calibration. In Section 5, we see that the resulting 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.

2. The Generalized Horvitz-Thompson Estimator

For a vector of interest $\mathbf{y}'=(y_1, y_2, \ldots, y_N)$, the Horvitz-Thompson estimator of the total $\theta=\sum_{k=1}^{N} y_k$ can be written

$$\begin{aligned}\hat{\theta}_{HT} &= \sum_{k=1}^{N} \frac{\delta_k y_k}{\pi_k} \\ &= \mathbf{y}'\mathbf{\Delta}_s \left(E(\mathbf{\Delta}_s)\right)^{-1} \mathbf{1}_{N\times1} , \qquad (2)\end{aligned}$$

where $\pi_k = E(\delta_k)$ is assumed greater than 0 for $k=1,2,\ldots,N$, and $\mathbf{\Delta}_s$ is the $N \times N$ diagonal matrix of the δ_k.

The generalization of the Horvitz-Thompson estimator relies on the Moore-Penrose inverse of a matrix $\mathbf{M}$, denoted $\mathbf{M}^{\dagger}$. The unique Moore-Penrose inverse always exists, and it is equal to the ordinary inverse if the latter exists. In particular, $\mathbf{\Delta}_s^{\dagger} = \mathbf{\Delta}_s$. More properties of the Moore-Penrose inverse can be found in Ben-Israel and Greville (2002).

For any $N \times N$ positive diagonal matrix $\mathbf{\Sigma}$ one can express the Horvitz-Thompson estimator as

$$\begin{aligned}\hat{\theta}_{HT} &= \mathbf{y}'\mathbf{\Delta}_s \left(E(\mathbf{\Delta}_s)\right)^{-1} \mathbf{1}_{N\times1} \\ &= \mathbf{y}'(\mathbf{\Delta}_s)^{\dagger} \left(E(\mathbf{\Delta}_s)^{\dagger}\right)^{-1} \mathbf{1}_{N\times1} \\ &= \mathbf{y}'(\mathbf{\Delta}_s\mathbf{\Delta}_s)^{\dagger} \left(E(\mathbf{\Delta}_s\mathbf{\Delta}_s)^{\dagger}\right)^{-1} \mathbf{1}_{N\times1} \\ &= \mathbf{y}'(\mathbf{\Delta}_s\mathbf{\Delta}_s)^{\dagger} \mathbf{\Sigma}^{\dagger}\mathbf{\Sigma}\left(E(\mathbf{\Delta}_s\mathbf{\Delta}_s)^{\dagger}\right)^{-1} \mathbf{1}_{N\times1} \\ &= \mathbf{y}'(\mathbf{\Delta}_s\mathbf{\Sigma}\mathbf{\Delta}_s)^{\dagger} \left(E(\mathbf{\Delta}_s\mathbf{\Sigma}\mathbf{\Delta}_s)^{\dagger}\right)^{-1} \mathbf{1}_{N\times1} . \qquad (3)\end{aligned}$$

If in (3), we replace the positive diagonal matrix $\mathbf{\Sigma}$ by any $N \times N$ positive definite matrix $\mathbf{\Sigma}$, we obtain the generalized Horvitz-Thompson estimator

$$\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right) = \mathbf{y}'\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger} \left(E\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}\right)^{-1} \mathbf{1}_{N\times 1} . \tag{4}$$

The vector $\mathbf{w}_{s\,GHT}\left(\mathbf{\Sigma}\right) = \left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger} \left(E\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}\right)^{-1} \mathbf{1}_{N\times 1}$ gives the weights of $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$. As shown in (3), $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right) = \hat{\theta}_{HT}$ when $\mathbf{\Sigma}$ is diagonal. It is immediately seen from (4) that the generalized Horvitz-Thompson estimator is unbiased. In fact, it is unbiased regardless of the choice made for the positive definite matrix $\mathbf{\Sigma}$. The matrix $E\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}$ is invertible under the assumptions that $\mathbf{\Sigma}$ is positive definite, and $\pi_k = E\left(\delta_k\right)$ is greater than 0 for $k = 1, 2, \ldots, N$. This ensures that $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ is well defined. The proofs of many results given here can be found in Théberge (2017). Because $\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger} = \mathbf{\Delta}_s \left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}$, it is seen by substitution in (4) that the generalized estimator only depends on the observed values of $\mathbf{y}$, as any estimator should.

Often, there is no closed-form formula for $E\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}$, but it can be easily approximated. One simply takes the average of a large number of values of $\left(\mathbf{\Delta}_s \mathbf{\Sigma} \mathbf{\Delta}_s\right)^{\dagger}$, each computed for a different sample obtained with the same sampling plan. The computation does not require the knowledge of any of the variables of interest.

3. The Choice of the Positive Definite Matrix $\mathbf{\Sigma}$

Different choices of $\mathbf{\Sigma}$ will generally lead to different generalized Horvitz-Thompson estimators. A matrix $\mathbf{\Sigma}$ is an appropriate choice to use for $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$, if a model ξ with $V_{\xi}\left(\mathbf{y}\right) = \mathbf{\Sigma}$ is an appropriate model for $\mathbf{y}$. The estimator remains unbiased, even if the matrix $\mathbf{\Sigma}$ used in $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ is different from the variance matrix under the model, but there are advantages for the two to be as close as possible. The ordinary Horvitz-Thompson estimator uses (4) with a diagonal $\mathbf{\Sigma}$. It is often used, and it is always unbiased, even though a more appropriate model for $\mathbf{y}$ would have $V_{\xi}\left(\mathbf{y}\right)$ non-diagonal. One could use $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ with a positive definite matrix $\mathbf{\Sigma}$ which is closer to $V_{\xi}\left(\mathbf{y}\right)$ than a diagonal matrix would be. As for the variance of $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$, it may often be higher than that of the ordinary Horvitz-Thompson estimator, even with the choice of $\mathbf{\Sigma} = V_{\xi}\left(\mathbf{y}\right)$. The advantage of the generalization of the Horvitz-Thompson estimator comes from its use in a generalization of calibration, as discussed in the next section.

The use of a block-diagonal matrix simplifies the computation of inverses needed in (4). 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 to the rating of another worker with the same employer; the race or religion of a couple is often the same. 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 $\mathbf{\Sigma}$ corresponding to persons from a same household should be identical. Such a matrix $\mathbf{\Sigma}$ is not positive definite, but it is the limit of a sequence of positive definite matrices, and the limit of the corresponding generalized Horvitz-Thompson could be computed. The example given in the introduction is based on this idea.

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 Horvitz-Thompson estimator from the correlations will be reduced. On the other hand, the strength borrowed from the correlations by the ordinary Horvitz-Thompson estimator is nil.

If an estimator $\hat{\mathbf{\Sigma}}$ converges to $\mathbf{\Sigma}$ in probability, then the bias and variance of $\hat{\theta}_{GHT}\left(\hat{\mathbf{\Sigma}}\right)$ are asymptotically the same as those of $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$. In practice, even if the general form of $\mathbf{\Sigma}$ depends on $N\left(N-1\right)/2$ covariances, the number of parameters in $\mathbf{\Sigma}$ should be small compared to the sample size. Using the Horvitz-Thompson estimator means assuming all covariances are zero. When using the generalized Horvitz-Thompson estimator, there is nothing wrong with assuming that those covariances depend on one, or a few, parameters, and that those parameters are considered fixed, rather than estimated from the sample. That is to say, consider $\hat{\mathbf{\Sigma}}=\mathbf{\Sigma}$.

4. 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 Horvitz-Thompson estimator of the known population size will have a variance greater than zero. The sum of the weights of the generalized Horvitz-Thompson estimator is often a worse estimator of the population size; it will often vary, even when the sample size is fixed.

An estimator whose estimates of the known population size vary cannot be seen as very reliable. Indeed, for many choices of $\mathbf{\Sigma}$, the variance of the generalized Horvitz-Thompson estimator will often be worse than that of the ordinary Horvitz-Thompson estimator. Even if the matrix $\mathbf{\Sigma}$ used for $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ is equal to $V_{\xi}\left(\mathbf{y}\right)$, the sum of the weights of $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$, noted $S\left(\mathbf{\Sigma}\right)$, will generally be a worse estimator of the population size than the sum of the weights of $\hat{\theta}_{HT}$.

To fix the problem that the ordinary Horvitz-Thompson 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 improvement can be made to the generalized estimator: $\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}\right)=\left(N/S\left(\mathbf{\Sigma}\right)\right)\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$. Although $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ is often more variable than $\hat{\theta}_{HT}$, with an appropriate choice for $\mathbf{\Sigma}$, $\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}\right)$ will generally be preferable to $\hat{\theta}_{CAL}$. Before giving more details about the optimality of $\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}\right)$, the definition of $\hat{\theta}_{GCAL}(\boldsymbol{\Sigma})$ 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, the following problem is addressed:

Calibration Problem: Among the weight vectors $\mathbf{w}_s \in \mathbb{R}^N$ in the range of $\boldsymbol{\Delta}_s$, i.e. non-sampled units should have a weight of 0, which minimize $\|\mathbf{X}'\mathbf{w}_s - \mathbf{X}'\mathbf{1}_{N\times1}\|_{\mathbf{T}}$, i.e. which "best" satisfy the calibration equations, seek one that minimizes $\|\mathbf{w}_s - \mathbf{w}_{s\,GHT}(\boldsymbol{\Sigma})\|_{\mathbf{U}}$, i.e. as close as possible to the weights of $\hat{\theta}_{GHT}(\boldsymbol{\Sigma})$, 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, $\mathbf{X}'\mathbf{w}_s = \mathbf{X}'\mathbf{1}_{N\times1}$, 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 $\mathbf{w}_{s\,GHT}(\boldsymbol{\Sigma})$. 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.

The solution to the calibration problem yields

$$\hat{\theta}_{GCAL}(\boldsymbol{\Sigma}) = \mathbf{y}'\mathbf{w}_{s\,GCAL}(\boldsymbol{\Sigma})$$

$$= \hat{\mathbf{y}}'\mathbf{c} + (\mathbf{y} - \hat{\mathbf{y}})'\,\mathbf{w}_{s\,GHT}(\boldsymbol{\Sigma}), \qquad (5)$$

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

$$\hat{\boldsymbol{\beta}} = \mathbf{T}^{1/2}\left(\mathbf{T}^{1/2}\mathbf{X}'(\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^{\dagger}\,\mathbf{X}\mathbf{T}^{1/2}\right)^{\dagger}\mathbf{T}^{1/2}\mathbf{X}'(\boldsymbol{\Delta}_s\mathbf{U}\boldsymbol{\Delta}_s)^{\dagger}\,\mathbf{y}\,. \qquad (6)$$

The estimator $\hat{\theta}_{GCAL}(\boldsymbol{\Sigma})$ is asymptotically unbiased.

It can be seen from the form of (5), that $\hat{\theta}_{GCAL}(\boldsymbol{\Sigma})$ is also a regression estimator that uses a model ξ such that $E_{\xi}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $V_{\xi}(\mathbf{y}) = \boldsymbol{\Sigma}$, a positive definite matrix. When viewed as a regression estimator, it is important to realize that $\hat{\theta}_{GCAL}(\boldsymbol{\Sigma})$ is asymptotically unbiased, regardless of both parts of the model; the parameter $\boldsymbol{\beta}$, and the parameter $\boldsymbol{\Sigma}$.

5. The Generalized Godambe-Joshi Lower Bound

For any unbiased estimator $\hat{\theta}$ of the population total θ, Godambe and Joshi (1965) have given a lower bound for the value of $E_\xi V_p\left(\hat{\theta}\right)$ under the assumption that the variance matrix $V_\xi\left(\mathbf{y}\right)$ was diagonal. That result can be generalized:

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

Just as the calibration estimator asymptotically attains the Godambe-Joshi lower bound, the generalized calibration estimator, with $\mathbf{\Sigma} = V_\xi\left(\mathbf{y}\right)$, asymptotically attains the generalized Godambe-Joshi lower bound.

The fact that $\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}\right)$ asymptotically attains the generalized Godambe-Joshi lower bound shows that the generalized Horvitz-Thompson estimator performs well when applied to residuals, as it does in (5), even though it is not recommended in general. Similarly, the ordinary Horvitz-Thompson estimator can run into problems 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 <u>linear</u> unbiased estimators. In fact, an example with $V_\xi\left(\mathbf{y}\right)$ not diagonal, of a non-linear unbiased estimator which does better than the lower bound is given in Théberge (2017).

6. Example

There are cases simple enough for $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\right)$ to be given explicitly. Say $\mathbf{\Sigma}\left(\rho\right)$ is a block-diagonal matrix where each of $N_p = N/2$ blocks equals $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, with $-1 < \rho < 1$.

Such a block-diagonal matrix is appropriate if the population can be grouped into pairs where, within a pair, the variable of interest is correlated. Then, (4) reduces to

$$\hat{\theta}_{GHT}\left(\mathbf{\Sigma}\left(\rho\right)\right) = \sum_{i=1}^{N_p} \frac{a_{2i-1}y_{2i-1} + a_{2i}y_{2i}}{\left(\pi_{2i-1}\pi_{2i}\left(1-\rho^2\right) + \left(\pi_{2i-1} + \pi_{2i} - \pi_{2i-1\,2i}\right)\pi_{2i-1\,2i}\rho^2\right)}, \tag{7}$$

where

$$a_{2i-1} = \delta_{2i-1}\left[\pi_{2i}\left(1-\rho^2\right)+\pi_{2i-1\,2i}\rho\left(1+\rho\right)\right]+\delta_{2i-1}\delta_{2i}\left[\rho^2\pi_{2i}-\rho\pi_{2i-1}-\rho^2\pi_{2i-1\,2i}\right]$$

$$a_{2i} = \delta_{2i}\left[\pi_{2i-1}\left(1-\rho^2\right)+\pi_{2i-1\,2i}\rho\left(1+\rho\right)\right]+\delta_{2i-1}\delta_{2i}\left[\rho^2\pi_{2i-1}-\rho\pi_{2i}-\rho^2\pi_{2i-1\,2i}\right] \quad (8)$$

$$\pi_{2i-1\,2i} = E\left(\delta_{2i-1}\delta_{2i}\right) \quad .$$

Once again, this generalized Horvitz-Thompson estimator is unbiased, regardless of any assumptions made about the variance-covariance matrix of $\mathbf{y}$. It is seen that, as expected when $\mathbf{\Sigma}$ is diagonal, the estimator reduces to the Horvitz-Thompson estimator when $\rho = 0$. The case of $\rho \rightarrow 1$ is a special case of the following limit problem.

Let $\mathbf{\Sigma}(\varepsilon)$ be a sequence of positive definite matrices indexed by ε, such that $\lim \mathbf{\Sigma}(\varepsilon) = \mathbf{\Sigma}_1$, the block-diagonal matrix with each block equal to $\mathbf{1}_{2\times 2}$. Such a block-diagonal matrix is appropriate if the population can be grouped into pairs where the variable of interest is perfectly correlated (e.g. household income of two-person households). Because $\mathbf{\Sigma}_1$ is only positive semi-definite and not positive definite, we cannot define the generalized Horvitz-Thompson estimator with $\mathbf{\Sigma}_1$. However it can be defined if we replace the correlation of 1 with a correlation of 0.999. The limit, for an infinite number of nines after the decimal point is of interest. More generally, it can be shown that $\lim \hat{\theta}_{GHT}\left(\mathbf{\Sigma}(\varepsilon)\right)$, noted $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}_1\right)$, can be written

$$\hat{\theta}_{GHT}\left(\mathbf{\Sigma}_1\right) = \sum_{i=1}^{N_p} \frac{2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - \left(y_{2i-1}+y_{2i}\right)\delta_{2i-1}\delta_{2i} + \left(y_{2i-1}-y_{2i}\right)\delta_{2i-1}\delta_{2i}\pi_{diff\ i}}{P\left(\text{observe } i\text{-th pair}\right)}, \quad (9)$$

where $P\left(\text{observe } i\text{-th pair}\right) = \pi_{2i-1} + \pi_{2i} - \pi_{2i-1\ 2i}$, $N_p = N/2$ is the number of pairs in the population, and $\pi_{diff\ i} = \left(\pi_{2i} - \pi_{2i-1}\right)/\pi_{2i-1\ 2i}$. In particular, $\hat{\theta}_{GHT}\left(\mathbf{\Sigma}_1\right) = \lim_{\rho\rightarrow 1}\hat{\theta}_{GHT}\left(\mathbf{\Sigma}(\rho)\right)$. It is unbiased, for any sample design with known probabilities of inclusion, even if $V_\xi\left(\mathbf{y}\right) \neq \mathbf{\Sigma}_1$. It can be calibrated so that the sum of the weights of the estimator given in (9) is equal to *N*. The resulting estimator is

$$\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}_1\right) = \sum_{i=1}^{N_p} \frac{2y_{2i-1}\delta_{2i-1} + 2y_{2i}\delta_{2i} - \left(y_{2i-1}+y_{2i}\right)\delta_{2i-1}\delta_{2i} + \left(y_{2i-1}-y_{2i}\right)\delta_{2i-1}\delta_{2i}\pi_{diff\ i}}{\nu_p / N_p}, \quad (10)$$

where $\nu_p = \sum_{i=1}^{N_p} (\delta_{2i-1} + \delta_{2i} - \delta_{2i-1}\delta_{2i})$ is the number of pairs in the sample. If, for every pair $i = 1, 2, ..., N_p$, the two units have the same probability of inclusion, that is $\pi_{2i-1} = \pi_{2i}$, then $\pi_{diff\ i} = 0$ and the simplification will result in the example given in Section 1. If the variable of interest of both units of the *i*-th pair take the same value, $y_{twin\ i}$, then the estimators (9) and (10) will reduce to

$$\hat{\theta}_{GHT}\left(\mathbf{\Sigma}_1\right) = \sum_{i=1}^{N_p} \frac{2\, y_{twin\,i}\left(\delta_{2i-1} + \delta_{2i} - \delta_{2i-1}\delta_{2i}\right)}{P\left(\text{observe } i\text{-th pair}\right)} \quad (11)$$

and

$$\hat{\theta}_{GCAL}\left(\mathbf{\Sigma}_1\right) = \sum_{i=1}^{N_p} \frac{2\, y_{twin\,i}\left(\delta_{2i-1} + \delta_{2i} - \delta_{2i-1}\delta_{2i}\right)}{\nu_p / N_p} \quad (12)$$

respectively. We see from (11) and (12) that the generalized estimators are reminiscent of their ordinary cousins, except we are working with pairs, instead of individual units. The generalized calibration estimator (10) is optimized for $\mathbf{\Sigma}_1$, but it can still have a lower variance than both, the Horvitz-Thompson estimator and the ordinary calibration estimator, if the correlation between the units of a pair is strong (e.g. race, religion or education level of a couple).

There are modified versions of the generalized Horvitz-Thompson estimator and of the generalized calibration estimator. If $\pi_{diff\ i}=0$, the modified calibration estimator optimized for $\mathbf{\Sigma}_1$ becomes:

$$\hat{\theta}_{MGCAL}\left(\mathbf{\Sigma}_1\right)=2\sum_{i=1}^{N_p}\frac{y_{2i-1}\delta_{2i-1}+y_{2i}\delta_{2i}-\left(y_{2i-1}+y_{2i}\right)\delta_{2i-1}\delta_{2i}}{\tilde{v}_p / N_p}, \tag{13}$$

where $\tilde{v}_p=\sum_{i=1}^{N_p}(\delta_{2i-1}+\delta_{2i}-2\delta_{2i-1}\delta_{2i})$ is the number of pairs with exactly one unit sampled.

The modified versions have the advantage of having a closed form; there is no need to compute the expectation of $\left(\mathbf{\Delta}_s\mathbf{\Sigma}\mathbf{\Delta}_s\right)^{\dagger}$. For a positive definite matrix $\mathbf{\Sigma}$, they are defined as

$$\hat{\theta}_{MGHT}\left(\mathbf{\Sigma}\right)=\mathbf{y}'\mathbf{\Delta}_s\mathbf{\Sigma}^{-1}\mathbf{\Delta}_s\left(\mathbf{\Sigma}^{-1}\circ\mathbf{\Pi}\right)^{-1}\mathbf{1}_{N\times 1} \tag{14}$$

and

$$\hat{\theta}_{MGCAL}\left(\mathbf{\Sigma}\right)=\hat{\mathbf{y}}'\mathbf{c}+\left(\mathbf{y}-\hat{\mathbf{y}}\right)'\mathbf{w}_{s\,MGHT}\left(\mathbf{\Sigma}\right), \tag{15}$$

where $\mathbf{\Pi}=\left(\pi_{kl}\right)=\left(E\left(\delta_k\delta_l\right)\right)\in\mathbb{R}^{N\times N}$ is the matrix of second order probabilities of inclusion and $\mathbf{w}_{s\,MGHT}\left(\mathbf{\Sigma}\right)$ is the vector of weights of $\hat{\theta}_{MGHT}\left(\mathbf{\Sigma}\right)$. The modified generalized estimators are also unbiased, or at least asymptotically unbiased in the case of $\hat{\theta}_{MGCAL}\left(\mathbf{\Sigma}\right)$.

7. Simulation Results

For this simulation, a population of 1,000 units grouped into 500 pairs was generated. The model ξ used to generate $\mathbf{y}$ was such that the correlation between the units of a pair was 0.8. More precisely, $V_{\xi}\left(\mathbf{y}\right)$ is a block-diagonal matrix with each block proportional to $\begin{pmatrix}1 & 0.8\\ 0.8 & 1\end{pmatrix}$. The population total was 478. A simple random sample of 200 units was selected 10,000 times. For each sample, four estimators of the total were calculated: Horvitz-Thompson, which in this case equals the ordinary calibration estimator, generalized Horvitz-Thompson, generalized calibration, and generalized modified calibration. The generalized estimators were computed assuming, wrongly, that $V_{\xi}\left(\mathbf{y}\right)=\mathbf{\Sigma}_1$. The simple closed-form formulae of the preceding section could thus be used. Their average and variance over the 10,000 repetitions are given in Table 1. The theoretical variance of each estimator, or asymptotic variance in the case of the calibrated estimators, can be calculated, and is also shown.

Table 1: Simulation results comparing four estimators

Estimator	*Total*	*Variance*	*Theoretical variance*
Horvitz-Thompson	478.19	1002	999
Generalized Horvitz-Thompson	478.26	1023	1022
Generalized Calibration	478.28	930	928
Generalized Modified Calibration	478.40	1119	1117

As expected, the bias of each estimator is negligible. Also to be expected, is that the Monte Carlo variance of the estimators is close to the theoretical values. The generalized calibration estimator, with a variance of 930, performed the best. This is in spite of the fact that it was calculated assuming that the correlation between the units of a pair was one. It should be remembered that the Horvitz-Thompson estimator, with a variance of 1002, is also a generalized Horvitz-Thompson estimator, but it is computed assuming that the correlation between the units of a pair is zero. The variance of the generalized Horvitz-Thompson estimator was 1023. That estimator is not expected to perform well, unless it is applied to residuals. Finally, the generalized modified calibration estimator had the highest variance. There are reasons to believe it could do better with a population grouped into larger groups of correlated units, such as triplets of quadruplets.

The generalized Godambe-Joshi lower bound for the model ξ used to generate $\mathbf{y}$ was 923. This is the variance that could be expected of the generalized calibration estimator, if it had been calculated with a matrix $\mathbf{\Sigma} = V_\xi\left(\mathbf{y}\right)$ based on the correct model ξ, where the correlation between units of a pair is 0.8. If, as in the preceding section, $\mathbf{\Sigma}\left(\varepsilon\right)$ is a family of positive definite matrices such that $\lim \mathbf{\Sigma}\left(\varepsilon\right) = \mathbf{\Sigma}_1$, then the limit of the generalized Godambe-Joshi lower bounds for the models with $\mathbf{\Sigma}\left(\varepsilon\right)$ is 888. This is the variance that could be expected of the generalized calibration estimator, if the correlation between units of a same pair was one. On the other hand, the variance of the Horvitz-Thompson estimator is expected to remain the same, since it does not draw any strength from the correlation.

References

Ben-Israel, A. and Greville, T.N.E. 2002. *Generalized Inverses: Theory and Applications (second ed.)*. New York: Springer-Verlag.

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.

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.