Comparison of Traditional and Model Assisted Estimators in Inverse Random Sampling with Replacement

Sureeporn Sungsuwan[1], Prachoom Suwattee [1]

[1] School of Applied Statistics, National Institute of Development Administration (NIDA)
118 Sareethai, Bangkapi, Bangkok, Thailand 10240

Abstract. In this paper, we proposed two model assisted estimators of the population total and the total in a given set in inverse random sampling with replacement. The precision of the proposed estimators are compared with the estimators given by Greco and Naddeo (2007). The simulation results show that the precision of the two proposed estimators and the Greco-Naddeo estimators are not much different at low correlation between the study (*Y*) and the auxiliary (*X*) variables. With high correlation between the variables, the two proposed estimators are more precise than the Greco-Naddeo estimators.

1 Introduction

Inverse sampling is a method of sampling which requires continued drawing of units until certain specified conditions dependent on the results of those drawings have been fulfilled. The population may be divided into 2 disjoint subgroups, a group satisfying some condition, denoted by C and another group not satisfying the condition, denoted by $\overline{C}$. We do not know which set a unit belongs to until the unit is sampled and observed. For this situation, we may use inverse sampling design which requires continue drawing until a fixed number of units in the set C are obtained in the sample.

Recently, many papers on inverse sampling appeared in various statistical journals. Christman and Lan (2001) considered inverse simple random sampling with and without replacement. They gave an unbiased estimator of the population total and its variance for each case of sampling. Salehi and Seber (2001, 2004) also considered inverse simple random sampling without replacement. They gave an unbiased estimator of the population total and its variance based on the Murthy's estimator. Greco and Naddeo (2007) considered inverse sampling with replacement when the population units were drawn with unequal probabilities. They derived an unbiased estimator of the population total, its variance and an unbiased variance estimator under the design. For inverse simple random sampling, they also gave an estimator of the population total and its variance which was equivalent to the expression given by Christman and Lan (2001).

In some situations, estimators of certain parameters can be derived from information on auxiliary variable. Särndal, et. al (1992) proposed model assisted estimators to improve its precision. In this study, we used model assisted approach to improve the traditional estimators under the inverse random sampling with replacement.

2 Traditional Estimators

Let $U=\{u_1,u_2,...,u_N\}$ be a population of N units. For simplicity, we denote the i^{th} unit by its label i, so the set of N population units is written as $U=\{1,...,i,...,N\}$ with study values $\{y_1,y_2,...,y_N\}$, respectively. Divide U into 2 disjoint subsets, C and $\overline{C}$ according to the $y-$values. Let C be the set of M units satisfying a condition and $\overline{C}$ the set of $N-M$ units not satisfying the condition. It is assumed that a unit satisfies the condition if it has the value of a variable y greater than or equal to a constant c. We can write $C=\{i_1,i_2,...,i_M\}$ and $\overline{C}=\{i_{M+1},i_{M+2},...,i_N\}$ where $U=C\cup\overline{C}$ and $C\cap\overline{C}=\varnothing$.

Consider inverse simple random sampling with replacement from a population of size N when the sampling continues until a prespecified number m of units from the set C are obtained in the sample of size n. This sample can be divided into 2 disjoint subsets, the first is of m units from C denoted by s_C and the other of $n-m$ units from $\overline{C}$ denoted by $s_{\overline{C}}$ and $s=s_C\cup s_{\overline{C}}$ where $s_C\cap s_{\overline{C}}=\varnothing$. In this case, n is a random variable with negative binomial distribution (Lan, 1999).

Let y_i be the value of the study variable y from unit i, $i\in U$, and let $T_y=\sum_{i\in U}y_i$ be the total of this study variable. Greco and Naddeo (2007) gave an unbiased estimator of T_y as

$$\hat{T}_{y,GN}=N[\hat{P}\,\bar{y}_C+(1-\hat{P})\,\bar{y}_{\overline{C}}] \tag{1}$$

where $\hat{P}=(m-1)/(n-1)$ is an unbiased estimator of the proportion of units in the set C, $\bar{y}_C=m^{-1}\sum_{i\in s_C}y_i$, $\bar{y}_{\overline{C}}=(n-m)^{-1}\sum_{i\in s_{\overline{C}}}y_i$.

The variance of the estimator in (1) is

$$V(\hat{T}_{y,GN}) = N^2 \left[(\overline{Y}_C - \overline{Y}_{\overline{C}})^2 V_n(\hat{P}) + \frac{\sigma^2_{y_C}}{m} E_n(\hat{P}^2) + \frac{\sigma^2_{y_{\overline{C}}}}{m-1} E_n[\hat{P}(1-\hat{P})] \right] \quad (2)$$

where $(\overline{Y}_C, \sigma^2_{y_C})$ and $(\overline{Y}_{\overline{C}}, \sigma^2_{y_{\overline{C}}})$ are the means and the variances of the study variable in the set C and $\overline{C}$, respectively. $E_n(\cdot)$, $V_n(\cdot)$ are expectation and variance with respect to the distribution of n and $V(\cdot)$ is variance with respect to the sampling design. They also gave an unbiased estimator of $V(\hat{T}_{y,GN})$ as

$$\hat{V}(\hat{T}_{y,GN}) = N^2 \left[(\overline{y}_C - \overline{y}_{\overline{C}})^2 \frac{\hat{P}(1-\hat{P})}{n-2} + \frac{s^2_{y_C}}{m} \hat{P}_2 + \frac{s^2_{y_{\overline{C}}}}{m-1} [\hat{P} - \frac{m-1}{m-2} \hat{P}_2] \right] \quad (3)$$

where $\hat{P}_2 = [(m-1)(m-2)] / [(n-1)(n-2)]$ and $s^2_{y_C}$, $s^2_{y_{\overline{C}}}$ are the unbiased sample variances of the study variable in the sets C and $\overline{C}$.

For given *n*, the selection procedure under inverse sampling is the same as the selection procedure under stratified sampling with 2 strata where m and $n-m$ are selected from the first and the second stratum. The sample results in the two groups are independent (Greco and Naddeo, 2007). Let $T_{y_C} = \sum_{i \in C} y_i$ be the total of the study variable *y* in the set C. An unbiased estimator of T_{y_C} is given by the first part of expression in (1) with variance

$$V(\hat{T}_{y_{C,GN}}) = N^2 \left[\overline{Y}_C^2 V_n(\hat{P}) + \frac{\sigma^2_{y_C}}{m} E_n(\hat{P}^2) \right] \quad (4)$$

An unbiased estimator of $V(\hat{T}_{y_{C,GN}})$ is

$$\hat{V}(\hat{T}_{y_{C,GN}}) = N^2 \left[\overline{y}_C^2 \frac{\hat{P}(1-\hat{P})}{n-2} + \frac{s^2_{y_C}}{m} \hat{P}_2 \right] \quad (5)$$

3 Model Assisted Estimators of the Population Totals

In the situation that we use the auxiliary value to improve the precision of an estimator. Suppose that $(x_i, y_i), i \in s$ is observed where y_i is the value of study variable of unit i and x_i is auxiliary value. The set $\{(x_i, y_i), i \in s_C\}$ and $\{(x_i, y_i), i \in s_{\overline{C}}\}$ associated with m units satisfy the condition and $n-m$ units not satisfy the condition, respectively. From a finite population of size N, assume that $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$. Applying an estimator of β_1 of Särndal, et. al (1992), we can estimate β_1 using the sample data, that is

$$b_1 = \frac{\left(\frac{\hat{M}}{m}\right)\sum_{i\in s_C}(x_i - \overline{x}_C)(y_i - \overline{y}_C) + \left(\frac{N-\hat{M}}{n-m}\right)\sum_{i\in s_{\overline{C}}}(x_i - \overline{x}_{\overline{C}})(y_i - \overline{y}_{\overline{C}})}{\left(\frac{\hat{M}}{m}\right)\sum_{i\in s_C}(x_i - \overline{x}_C)^2 + \left(\frac{N-\hat{M}}{n-m}\right)\sum_{i\in s_{\overline{C}}}(x_i - \overline{x}_{\overline{C}})^2} \tag{6}$$

where $\hat{M} = N\hat{P}$ is an unbiased estimator of M, $\overline{x}_C = m^{-1}\sum_{i\in s_C} x_i$, $\overline{x}_{\overline{C}} = (n-m)^{-1}\sum_{i\in s_{\overline{C}}} x_i$. A model assisted estimator of the population total is given by

$$\tilde{T}_y = \hat{M}[\overline{y}_C + b_1(\overline{X} - \overline{x}_C)] + (N-\hat{M})[\overline{y}_{\overline{C}} + b_1(\overline{X} - \overline{x}_{\overline{C}})] \tag{7}$$

where $\overline{X} = N^{-1}\sum_{i\in U} x_i$ is the population mean of auxiliary variable x_i, $i \in U$.

Theorem 1. A model assisted estimator in equation (7) is biased.

Proof. $E(\tilde{T}_y) = E_n E[\tilde{T}_y \mid n]$

$= E_n\{E[(\hat{M}\,\overline{y}_C + (N-\hat{M})\overline{y}_{\overline{C}}) \mid n] - E[b_1(\hat{M}\,\overline{x}_C + (N-\hat{M})\overline{x}_{\overline{C}} - T_x) \mid n]\}$

$= E_n[\hat{M}\overline{Y}_C + (N-\hat{M})\overline{Y}_{\overline{C}}] - E_n[E[b_1(\hat{M}\overline{x}_C + (N-\hat{M})\overline{x}_{\overline{C}}) \mid n] - E(b_1 T_x) \mid n]$

$= M\,\overline{Y}_C + (N-M)\overline{Y}_{\overline{C}} - \{E[b_1(\hat{M}\,\overline{x}_C + (N-\hat{M})\overline{x}_{\overline{C}})] - E_n E(b_1 \mid n) T_x]\}$

$= T_y - \{E[b_1(\hat{M}\,\overline{x}_C + (N-\hat{M})\overline{x}_{\overline{C}})] - E(b_1) E_n[E(\hat{M}\,\overline{x}_C + (N-\hat{M})\overline{x}_{\overline{C}}) \mid n]\}$

$= T_y - Cov(b_1, \hat{T}_x)$ where $T_x = \sum_{i \in U} x_i$ and $\hat{T}_x = \hat{M}\, \bar{x}_C + (N - \hat{M})\, \bar{x}_{\bar{C}}$.
Bias of $\tilde{T}_y$ is $B(\tilde{T}_y) = E(\tilde{T}_y) - T_y = -Cov(b_1, \hat{T}_x)$.
where $E(\cdot)$ is expectation with respect to the sampling design.

Since the properties of a model assisted estimator with respect to the sampling design usually cannot be studied exactly, because of the complex form (Särndal, et. al, 1992 : 235). Thus an approximated technique is used to study the properties.

Theorem 2. An estimator in equation (7) is approximately

$$\tilde{T}_y^* = M\, \bar{y}_C + (N-M)\bar{y}_{\bar{C}} - B_1[(M\, \bar{x}_C + (N-M)\bar{x}_{\bar{C}}) - N\bar{X}] + [(\bar{Y}_C - B_1\bar{X}_C) - (\bar{Y}_{\bar{C}} - B_1\bar{X}_{\bar{C}})](\hat{M} - M) \quad (8)$$

with $MSE(\tilde{T}_y) \approx V(\tilde{T}_y^*)$

$$= \frac{M^2}{m}(\sigma_{E_C}^2 - \sigma_{E_{\bar{C}}}^2) + \sigma_{E_{\bar{C}}}^2 \frac{(m+1)}{m^2} MN + (\bar{E}_C - \bar{E}_{\bar{C}})^2 V_n(\hat{M}). \quad (9)$$

Proof. For given *n*, from Taylor linearization technique,

$$\tilde{T}_y = \hat{M}[\bar{y}_C + b_1(\bar{X} - \bar{x}_C)] + (N - \hat{M})[\bar{y}_{\bar{C}} + b_1(\bar{X} - \bar{x}_{\bar{C}})]$$

$$= \hat{M}\, \bar{y}_C + (N - \hat{M})\bar{y}_{\bar{C}} + b_1[N\bar{X} - (\hat{M}\, \bar{x}_C + (N - \hat{M})\, \bar{x}_{\bar{C}})]$$

$$= h(\bar{y}_C, \bar{y}_{\bar{C}}, b_1, \bar{x}_C, \bar{x}_{\bar{C}}, \hat{M})$$

Thus $\tilde{T}_y$ is nonlinear function of the estimators. From Taylor linearization technique, we approximate this function by a linear function. The partial derivatives with respect to the estimators are needed and we evaluate these partial derivatives at the expected value point. The estimator in (7) becomes

$$\tilde{T}_y \approx M\, \bar{y}_C + (N-M)\bar{y}_{\bar{C}} - B_1[(M\, \bar{x}_C + (N-M)\, \bar{x}_{\bar{C}}) - N\bar{X}]$$

$$+ [(\bar{Y}_C - B_1\bar{X}_C) - (\bar{Y}_{\bar{C}} - B_1\bar{X}_{\bar{C}})](\hat{M} - M)$$

$$= \tilde{T}_y^*.$$

Consider $E(\tilde{T}_y^*) = E_n[E(\tilde{T}_y^* \mid n)]$

$$= E_n\{M\overline{Y}_C + (N-M)\overline{Y}_{\overline{C}} - B_1[(M\overline{X}_C + (N-M)\overline{X}_{\overline{C}}) - N\overline{X}]$$

$$+[(\overline{Y}_C - B_1\overline{X}_C) - (\overline{Y}_{\overline{C}} - B_1\overline{X}_{\overline{C}})](\hat{M} - M)\}$$

$$= T_y.$$

Thus, $\tilde{T}_y^*$ is an unbiased estimator of T_y. To find the variance of $\tilde{T}_y^*$, consider

$$\tilde{T}_y^* = M\,\bar{y}_C + (N-M)\bar{y}_{\overline{C}} - B_1[(M\,\bar{x}_C + (N-M)\bar{x}_{\overline{C}}) - N\overline{X}]$$

$$+[(\overline{Y}_C - B_1\overline{X}_C) - (\overline{Y}_{\overline{C}} - B_1\overline{X}_{\overline{C}})](\hat{M} - M)$$

$$= NB_1\overline{X} + (\frac{M}{m})\sum\nolimits_{i\in s_C} E_i + (\frac{N-M}{n-m})\sum\nolimits_{i\in s_{\overline{C}}} E_i + [\overline{E}_C - \overline{E}_{\overline{C}}](\hat{M} - M) \qquad (10)$$

where $E_i = y_i - B_1 x_i$, $\overline{E}_C = M^{-1}\sum\nolimits_{i\in C} E_i$, $\overline{E}_{\overline{C}} = (N-M)^{-1}\sum\nolimits_{i\in\overline{C}} E_i$.

$$V(\tilde{T}_y^*) = E_n[V(\tilde{T}_y^* \mid n)] + V_n[E(\tilde{T}_y^* \mid n)]$$

$$= M^2\frac{\sigma_{E_C}^2}{m} + (N-M)^2\sigma_{E_{\overline{C}}}^2 E_n\left(\frac{1}{n-m}\right) + (\overline{E}_C - \overline{E}_{\overline{C}})^2 V_n(\hat{M}). \qquad (11)$$

Substituting $E_n\left(\frac{1}{n-m}\right) \approx \frac{mM(N-M) + MN}{m^2(N-M)^2}$, equation (11) becomes

$$V(\tilde{T}_y^*) = \frac{M^2}{m}(\sigma_{E_C}^2 - \sigma_{E_{\overline{C}}}^2) + MN\frac{(m+1)}{m^2}\sigma_{E_{\overline{C}}}^2 + (\overline{E}_C - \overline{E}_{\overline{C}})^2 V_n(\hat{M}) \qquad (12)$$

where $\sigma_{E_C}^2 = M^{-1}\sum\nolimits_{i\in C}(E_i - \overline{E}_C)^2$, $\sigma_{E_{\overline{C}}}^2 = (N-M)^{-1}\sum\nolimits_{i\in\overline{C}}(E_i - \overline{E}_{\overline{C}})^2$.

Thus, an estimator of $V(\tilde{T}_y^*)$ in (12) is in the form

$$\hat{V}(\tilde{T}_y^*) = (s_{e_C}^2 - s_{e_{\overline{C}}}^2)\frac{N\hat{M}}{m}\left(\frac{m-2}{n-2}\right) + s_{e_{\overline{C}}}^2\frac{m+1}{m^2}\hat{M}N + \left[(\bar{e}_C - \bar{e}_{\overline{C}})^2 - \frac{s_{e_C}^2}{m} - \frac{s_{e_{\overline{C}}}^2}{n-m} \right] \frac{\hat{M}(N-\hat{M})}{n-2} \quad (13)$$

where $e_i = y_i - b_1 x_i$, $\bar{e}_C = m^{-1}\sum_{i\in s_C} e_i$, $\bar{e}_{\overline{C}} = (n-m)^{-1}\sum_{i\in s_{\overline{C}}} e_i$,

$s_{e_C}^2 = (m-1)^{-1}\sum_{i\in s_C}(e_i - \bar{e}_C)^2$, $s_{e_{\overline{C}}}^2 = (n-m-1)^{-1}\sum_{i\in s_{\overline{C}}}(e_i - \bar{e}_{\overline{C}})^2$.

A model assisted estimator of the population total in the set C is given by

$$\tilde{T}_{y_C} = \hat{M}\,\bar{y}_C + b_1[N\overline{X} - (\hat{M}\bar{x}_C + (N-\hat{M})\bar{x}_{\overline{C}})] \quad (14)$$

Applying theorem 2, the mean squared error of the total in equation (14) is approximately

$$MSE(\tilde{T}_{y_C}) \approx \frac{M^2}{m}(\sigma_{E_C}^2 - B_1^2\sigma_{x_{\overline{C}}}^2) + B_1^2\sigma_{x_{\overline{C}}}^2\frac{(m+1)}{m^2}MN + [\overline{Y}_C - B_1(\overline{X}_C - \overline{X}_{\overline{C}})]^2 V_n(\hat{M}) \quad (15)$$

where $\sigma_{x_{\overline{C}}}^2 = (N-M)^{-1}\sum_{i\in\overline{C}}(x_i - \overline{X}_{\overline{C}})^2$. An estimator of $MSE(\tilde{T}_{y_C})$ is

$$M\hat{S}E(\tilde{T}_{y_C}) = (s_{e_C}^2 - b_1^2 s_{x_{\overline{C}}}^2)\frac{N\hat{M}}{m}\left(\frac{m-2}{n-2}\right) + b_1^2 s_{x_{\overline{C}}}^2\frac{(m+1)}{m^2}\hat{M}N + \left\{ [(\bar{y}_C - b_1(\bar{x}_C - \bar{x}_{\overline{C}})]^2 - \frac{s_{y_C}^2}{m} - b_1^2\frac{s_{x_C}^2}{m} - b_1^2\frac{s_{x_{\overline{C}}}^2}{n-m} + 2b_1\,\mathrm{cov}(\bar{x}_C, \bar{y}_C) \right\} \frac{\hat{M}(N-\hat{M})}{n-2} \quad (16)$$

where $s_{x_C}^2 = (m-1)^{-1}\sum_{i \in s_C}(x_i - \bar{x}_C)^2$, $s_{x_{\bar{C}}}^2 = (n-m-1)^{-1}\sum_{i \in s_{\bar{C}}}(x_i - \bar{x}_{\bar{C}})^2$, $\text{cov}(\bar{x}_C, \bar{y}_C) = [m(m-1)]^{-1}\sum_{i \in s_C}(x_i - \bar{x}_C)(y_i - \bar{y}_C)$.

4 Comparison of Estimators

Since the precision of model assisted estimators and the estimators obtained from Greco and Naddeo (2007) cannot be compared directly from the expressions of the variances and mean squared errors, the comparison is carried out by simulation. The simulation is based on repeated sampling from a generated finite population of size 10,000 from the model $Y_i = 10 + 0.3X_i + \varepsilon_i$ where $\varepsilon_i \sim N(0,1)$ and X has normal distribution. The population values of (x_i, y_i) are generated with linear relationship between the auxiliary X and the study variable Y with the correlation (ρ) 0.1, 0.3, 0.5, 0.7 and 0.9. The population proportion of units in the set C in this simulation is set to be 0.01, 0.05, and 0.2. The number m of units satisfying the condition in the sample is equals to 5, 7 and 17. For each situation, the L samples are drawn. For a sample l we calculate the total estimates, $\hat{\theta}_l$, $l = 1,...,L$ and also the average of the total estimates, $\bar{\hat{\theta}} = L^{-1}\sum_{l=1}^{L}\hat{\theta}_l$. The variance estimate is obtained from $\hat{V}(\hat{\theta}) = (L-1)^{-1}\sum_{l=1}^{L}(\hat{\theta}_l - \bar{\hat{\theta}})^2$. The mean squared error estimate is given by $M\hat{S}E(\hat{\theta}) = \hat{V}(\hat{\theta}) + [Bi\hat{a}s(\hat{\theta})]^2$ where $Bi\hat{a}s(\hat{\theta}) = \bar{\hat{\theta}} - \theta$. In this study, we let $L = 1{,}000$ for each situation. We compute the mean squared errors of the two model assisted estimators and compare with the variance estimates of the unbiased estimators given by Greco and Naddeo (2007). Further more, the average sample size and relative efficiency is also obtained. The results are shown in Table 1 and Table 2.

Table 1. Average sample size $(\bar{n})$, variance estimates of the unbiased estimators $(\hat{V}(\hat{T}_{y,GN}))$, mean squared error estimates of model assisted estimator $(M\hat{S}E(\tilde{T}_y))$ and relative efficiencies $(M\hat{S}E(\tilde{T}_y)/\hat{V}(\hat{T}_{y,GN}))$.

ρ	P	m	$\bar{n}$	$\hat{V}(\hat{T}_{y,GN})$ $\times 10^4$	$M\hat{S}E(\tilde{T}_y)$ $\times 10^4$	$\frac{M\hat{S}E(\tilde{T}_y)}{\hat{V}(\hat{T}_{y,GN})}$
.1	.01	5	594	18.67	18.59	0.995
		7	786	13.26	13.20	0.995
		17	1,808	5.71	5.69	0.996
	.05	5	119	104.11	104.85	1.007
		7	161	72.35	70.97	0.981
		17	363	26.06	25.92	0.994
	.2	5	30	371.95	385.62	1.037
		7	40	302.80	316.82	1.046
		17	90	113.97	113.06	0.992
.3	.01	5	579	21.45	20.01	0.933
		7	819	12.82	11.89	0.928
		17	1,790	5.98	5.37	0.898
	.05	5	120	107.71	97.58	0.906
		7	158	82.32	77.04	0.936
		17	358	29.15	26.67	0.915
	.2	5	30	391.98	384.31	0.980
		7	40	264.25	249.31	0.943
		17	90	122.40	116.36	0.951
.5	.01	5	598	20.71	15.66	0.756
		7	794	15.20	11.80	0.776
		17	1,799	5.65	4.38	0.774
	.05	5	120	111.23	83.92	0.755
		7	162	69.35	54.70	0.789
		17	364	30.87	22.64	0.733
	.2	5	30	403.03	354.13	0.879
		7	40	306.88	252.92	0.824
		17	90	125.54	100.54	0.801
.7	.01	5	585	22.21	10.69	0.481
		7	785	15.59	7.91	0.507
		17	1,808	5.98	2.94	0.491
	.05	5	119	101.34	54.14	0.534
		7	160	69.60	38.16	0.548
		17	361	30.71	15.96	0.520
	.2	5	29	453.20	269.52	0.595
		7	40	294.18	180.60	0.614
		17	89	116.34	64.05	0.551

Table 1. Continued

ρ	P	m	$\bar{n}$	$\hat{V}(\hat{T}_{y,GN})$ $\times 10^4$	$M\hat{S}E(\tilde{T}_y)$ $\times 10^4$	$\frac{M\hat{S}E(\tilde{T}_y)}{\hat{V}(\hat{T}_{y,GN})}$
.9	.01	5	616	19.58	3.66	0.187
		7	794	14.06	2.91	0.207
		17	1,790	6.16	1.12	0.182
	.05	5	122	107.39	19.81	0.184
		7	164	77.25	12.93	0.167
		17	356	28.17	5.55	0.197
	.2	5	30	395.53	96.00	0.243
		7	41	274.11	59.93	0.219
		17	90	116.87	25.08	0.215

From Table 1 it is seen that the average sample size increase if we increases the values of *m*. However, if the population proportion increases, smaller average sample sizes are obtained. Considering the variance estimates and the mean squared errors estimates, we see that if the population proportion increases the variance estimates and the mean squared errors also increase. If the correlation between *X* and *Y* is less than 0.5 the variances of the unbiased estimators and the mean squared errors of the model assisted estimators are not much different for any level of proportion and any a number *m* in the samples. If the correlation between *X* and *Y* is greater than or equal 0.5 the model assisted estimates have smaller mean squared errors than the variance estimates of unbiased estimates for any level of proportion and the number *m*.

Table 2. Variance estimates of the unbiased estimator $(\hat{V}(\hat{T}_{y_C,GN}))$, mean squared error estimates of model assisted estimator $(M\hat{S}E(\tilde{T}_{y_C}))$ and relative efficiencies $(M\hat{S}E(\tilde{T}_{y_C})/\hat{V}(\hat{T}_{y_C,GN}))$.

ρ	P	m	$\hat{V}(\hat{T}_{y_C,GN})$ $\times 10^4$	$M\hat{S}E(\tilde{T}_{y_C})$ $\times 10^4$	$\frac{M\hat{S}E(\tilde{T}_{y_C})}{\hat{V}(\hat{T}_{y_C,GN})}$
.1	.01	5	73.32	73.10	0.997
		7	51.17	51.33	1.003
		17	21.41	21.51	1.005
	.05	5	2,078.71	2,081.28	1.001
		7	1,321.25	1,316.45	0.996
		17	466.72	467.36	1.001
	.2	5	194,105.58	193,979.02	0.999
		7	175,150.20	175,882.51	1.004
		17	55,076.89	55,048.82	0.999

Table 2. Continued

ρ	P	m	$\hat{V}(\hat{T}_{y_C,GN})$ $\times 10^4$	$M\hat{S}E(\tilde{T}_{y_C})$ $\times 10^4$	$\frac{M\hat{S}E(\tilde{T}_{y_C})}{\hat{V}(\hat{T}_{y_C,GN})}$
.3	.01	5	112.45	113.92	1.013
		7	50.59	51.37	1.016
		17	24.09	24.21	1.005
	.05	5	1,989.00	1,955.78	0.983
		7	1,414.49	1,405.65	0.994
		17	519.78	512.01	0.985
	.2	5	19,722.36	19,655.03	0.997
		7	12,938.90	12,843.36	0.993
		17	5,794.40	5,744.83	0.991
.5	.01	5	74.00	73.05	0.987
		7	60.86	59.93	0.985
		17	22.74	22.20	0.976
	.05	5	1,803.18	1,721.48	0.955
		7	1,007.37	977.27	0.970
		17	478.00	456.60	0.955
	.2	5	21,462.76	21,000.72	0.978
		7	16,188.11	15,777.30	0.975
		17	5,697.40	5,542.45	0.973
.7	.01	5	106.80	103.24	0.967
		7	74.95	73.79	0.985
		17	22.14	21.57	0.974
	.05	5	1,947.51	1,791.84	0.920
		7	1,393.06	1,301.17	0.934
		17	513.18	475.73	0.927
	.2	5	22,186.40	20,942.50	0.944
		7	15,616.10	14,795.45	0.947
		17	6,211.88	5,853.69	0.942
.9	.01	5	89.26	84.17	0.943
		7	57.85	55.45	0.959
		17	21.93	20.60	0.939
	.05	5	2,098.27	1,762.31	0.840
		7	1,266.06	1,043.25	0.824
		17	518.34	454.00	0.876
	.2	5	21,376.17	18,806.14	0.880
		7	14,671.47	12,866.34	0.877
		17	5,911.62	5,218.90	0.883

From the results in Table 2, we compare the precision of the total estimates of units in the set C. We see that the variance estimates and the mean squared error estimates increase if the population proportion increase. When $\rho \le 0.7$ the variance estimates of the unbiased estimates and the mean squared error estimates of the model assisted estimates are not quite different for any values of population proportion and the number *m*. At very high correlation between *X* and *Y* $(\rho = 0.9)$ the model assisted estimates have smaller mean squared errors than the unbiased estimates especially when the population proportion is greater than or equal 0.05 but they do not depend on the number *m*.

References

1. Christman, M.C., Lan, F.: Inverse adaptive cluster sampling. Biometrics. 57, 1096-1105 (2001).
2. Greco, L., Naddeo, S.: Inverse sampling with unequal selection probabilities. Communications in Statistics - Theory and Methods. 36, 1039-1048 (2007).
3. Haldane, J. B. S.: On a method of estimating frequencies. Biometrika. 33, 222-225 (1945).
4. Lan, F.: Sequential adaptive sampling designs to estimate abundance in rare populations. Doctorial Dissertation. The American University (1999).
5. Särndal, C.E., Swensson, B., Wretman, J.: Model assisted survey sampling. Springer-Verlag. New York (1992).
6. Salehi, M.M., Seber, G.A.F.: A New proof of Murthy's estimator with applies to sequential sampling. Austral. NZ J. Statist. 43, 281-286 (2001).
7. Salehi, M.M., Seber, G.A.F.: A general inverse sampling scheme and its application to adaptive cluster sampling. Austral. NZ J. Statist. 46, 483-494 (2004).