Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(2) 2010 273-288

Variance Estimation of a Generalized Regression Predictor

Raghunath Arnab[1*], D.K. Shangodoyin[1] and Sarjinder Singh[2]
[1]*University of Botswana, Botswana*
[2]*Texas A&M University-Kingsville, USA*

Received 18 January 2010; Revised 10 March 2010; Accepted 01 April 2010

SUMMARY

The generalized regression predictor (greg) is used for the estimation of a finite population total when the study variable is well related to the auxiliary variable. Särndal (1982), proposed a few estimators for variance of the greg. In this paper, we have derived the lower bounds of variances of the estimators of the variances of greg belonging to certain classes of estimators under a superpopulation model. The proposed optimal variance estimators attaining the lower bound cannot be used in practice since they involve unknown parameters. Hence, some alternative variance estimators are proposed. Simulation studies reveal that the proposed alternative estimators are more efficient than existing alternatives proposed by Särndal (1982).

Keywords: Generalized regression predictor, Superpopulation Model, Variance estimation.

1. INTRODUCTION

Consider a finite population $U = \{1, ..., i, .., N\}$ of N identifiable units. Let $y_i(x_i)$ be the value of the study (auxiliary) variable of the i^{th} unit of the population. The values of y_i's are unknown before survey but the values of x_i's are assumed to be known and positive. Here we consider the problem of estimation of the finite population total $Y = \sum_{i \in U} y_i$ using a sample s selected with probability $p(s)$ by a fixed effective size (n) sampling design p. The inclusion probabilities of units i and the pair of units $i \neq j$ are denoted respectively by π_i and π_{ij} and assumed to be positive. Särndal (1982) and Särndal *et al*. (1989) recommended the use of the generalized regression predictor (greg) for estimation of the finite population total Y. It is well known that the greg predictor is asymptotically design unbiased (ADU) for Y, irrespective of the validity of any model. Several authors including Särndal (1982), Kott (1990), Särndal (1992) and Zou (1999) proposed variance estimators for the greg to facilitate the estimation of confidence interval for the population total Y. Although the variance estimators proposed by Särndal (1982) are ADU, under large samples, but little is known about their efficiencies. In this paper we have proposed a few alternative variance estimators and compare their performances with existing estimators under the following superpopulation model;

Model M: $\quad y_i = \beta x_i + \epsilon_i,\ i \in U \qquad (1.1)$

where β is an unknown constant, ϵ_i's are error components, independently distributed with $E_m(\epsilon_i) = 0$ and $V_m(\epsilon_i) = \sigma_i^2 = \sigma^2 x_i^g$, where σ^2 (> 0) is unknown and an approximate value of g is known from the past experience but the actual value is unknown. Here E_m, V_m denote respectively expectation and variance with respect to the superpopulation model M.

The generalized regression predictor (greg) for Y proposed by Särndal (1982) is given by

$$t_g = \sum_{i \in U} \frac{y_i}{\pi_i} I_{si} + \hat{\beta}_Q \left(X - \sum_{i \in U} \frac{x_i}{\pi_i} I_{si} \right) \qquad (1.2)$$

[*]*Corresponding author* : Raghunath Arnab
E-mail address : arnabr@mopipi.ub.bw

where $I_{si} = 1$ if $i \in s$ and 0 if $i \notin s$,

$$\hat{\beta}_Q = \frac{\sum_{i \in s} x_i y_i/(c_i\pi_i)}{\sum_{i \in s} x_i^2/(c_i\pi_i)} = \frac{\sum_{i \in U} Q_i x_i y_i I_{si}}{\sum_{i \in U} Q_i^2 x_i^2 I_{si}} \text{ and } c_i\ (>0)\text{'s}$$

are suitably chosen constants, $Q_i = 1/(c_i\pi_i)$ and $X = \sum_{i \in U} x_i$.

Let $E_p(V_p)$ denotes expectation (variance) with respect to the sampling design p. The mean square error and the variance of t_g are each approximated by Särndal (1982) as

$$V_{SR} = \frac{1}{2}\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j}\right)^2 \tag{1.3}$$

where

$$\Delta_{ij} = \pi_i\pi_j - \pi_{ij},\ E_i = y_i - \beta_Q x_i \text{ and}$$

$$E_p(\hat{\beta}_Q) \cong \frac{E_p \sum_{i \in s} Q_i x_i y_i}{E_p \sum_{i \in s} Q_i x_i^2} = \frac{\sum_{i \in U} Q_i x_i y_i \pi_i}{\sum_{i \in U} Q_i x_i^2 \pi_i}$$

$$= \frac{\sum_{i \in U} x_i y_i / c_i}{\sum_{i \in U} x_i^2 / c_i} = \beta_Q \tag{1.4}$$

Särndal (1982) proposed two alternative estimators for V_{SR} as follows;

$$\hat{v}_{SR}(1) = \frac{1}{2}\sum_{i \neq}\sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{\hat{E}_i}{\pi_i} - \frac{\hat{E}_j}{\pi_j}\right)^2 \text{ and}$$

$$\hat{v}_{SR}(2) = \frac{1}{2}\sum_{i \neq}\sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{g_{si}\hat{E}_i}{\pi_i} - \frac{g_{sj}\hat{E}_j}{\pi_j}\right)^2$$

where $\hat{E}_i = y_i - \hat{\beta}_Q x_i$ and

$$g_{si} = 1 + \left(X - \sum_{i \in s} \frac{x_i}{\pi_i}\right) \cdot \frac{Q_i \pi_i x_i}{\sum_{i \in s} Q_i x_i^2}$$

Remark 1. For an simple random sampling without replacement (SRSWOR) design $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$. If we choose $Q_i = 1/x_i$, then β_Q reduces to $R = Y/X$ and V_{SR} becomes equal to $V^* = \frac{(1-f)}{n(N-1)}\sum_{i=1}^{N}(y_i - Rx_i)^2$. The expression of V^* was derived by Cochran (1977). In this situation $\hat{v}_{SR}(1)$ and $\hat{v}_{SR}(2)$ becomes equal to

$$v_1 = \frac{(1-f)}{n(n-1)}\sum_{i \in s}(y_i - rx_i)^2 \text{ and}$$

$$v_2 = \frac{(1-f)}{n}\left(\frac{\bar{X}}{\bar{x}}\right)^2\sum_{i \in s}(y_i - rx_i)^2 \text{ respectively.}$$

In our present paper, we propose lower bounds of unbiased estimators of the variances of greg predictor belonging to certain classes of estimators under the superpopulation model (1.1). The optimal estimators, attaining the lower bounds, are also derived. The proposed optimal estimators cannot be used in practice since they involve unknown model parameters. So, we have proposed a few alternative calibrated estimators following Singh *et al*. (1988). Efficiencies of the proposed estimators are compared with existing alternatives. Simulation studies reveal that the estimator $\hat{v}_{SR}(2)$ has a lower efficiency than the Yates-Grundy analogue estimator $\hat{v}_{SR}(1)$ for the model (1.1) with a lower value of ($g < 1.5$) but for the higher value of g, $\hat{v}_{SR}(2)$ is better than $\hat{v}_{SR}(1)$. All the proposed calibrated estimators including the optimum one are found almost equally efficient and they perform much better than $\hat{v}_{SR}(1)$ and $\hat{v}_{SR}(2)$ in most of the situations around the true value of g. The proposed estimators fare better than the existing alternatives under the model $M^*: y_i = \alpha + \beta x_i + \epsilon_i$ where α is a constant, β and ϵ_i's are as in model (1.1).

2. OPTIMUM VARIANCE ESTIMATION

Under the superpopulation model (1.1) β_Q can be approximately written as

$$\beta_Q \cong \frac{E_m E_p \sum_{i\hat{I}U} Q_i x_i y_i I_{si}}{E_m E_p \sum_{i\hat{I}U} Q_i x_i^2 I_{si}} = \beta$$

Then for large sample size n, V_{SR} can be approximated as

$$V = \frac{1}{2}\sum_{i\neq}\sum_{j\in U}\left(\frac{e_i}{\pi_i}-\frac{e_j}{\pi_j}\right)^2 \Delta_{ij}$$

$$= \sum_{i\in U} a_i e_i^2 + \sum_{i\neq}\sum_{j\in U} a_{ij} e_i e_j \quad (2.1)$$

where

$$e_i = y_i - \beta x_i,\ a_i = \left(\frac{1}{\pi_i}-1\right) \quad \text{and}$$

$$a_{ij} = \left(\frac{\pi_{ij}}{\pi_i \pi_j}-1\right)$$

The empirical investigations in section 4 show that both the expressions V and V_{SR} serve as a very good approximation for $M(t_g)$, the mean square error of t_g.

Let C_{nhq} be the class of non-homogeneous quadratic unbiased estimator of V of the form

$$v_{nhq} = c_s + \sum_{i\in U} c_{si} I_{si} e_i^2 + \sum_{i\neq}\sum_{j\in U} c_{sij} I_{sij} e_i e_j \quad (2.2)$$

$$= c_s + v_{hq} \quad (2.3)$$

where $I_{sij} = I_{si} I_{sj}$

$$v_{hq} = \sum_{i\in U} c_{si} I_{si} e_i^2 + \sum_{i\neq}\sum_{j\in U} c_{sij} I_{sij} e_i e_j$$

c_s, c_{si} and c_{sij} ($=c_{sji}$) are constants free from y_i's and satisfy the unbiasedness condition

$$E_p(c_s) = \sum_s c_s p(s) = 0),$$

$$E_p(c_{si} I_{si}) = \sum_s c_{si} I_{si} p(s) = a_i \text{ for } i \in U$$

$$\text{and}\quad E_p(c_{sij} I_{sij}) = \sum_s c_{sij} I_{sij} p(s) = a_{ij} \text{ for } i \neq j \in U \quad (2.4)$$

Theorem 1: Under model M

$$E_m V_p(v_{nhq}) \geq \sum_{i\in U} \delta_i a_i^2 \left(\frac{1}{\pi_i}-1\right) + 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2 \sigma_j^2 a_{ij}^2 \left(\frac{1}{\pi_{ij}}-1\right)$$

$$= E_m V_p(v_0) \text{ for } v_{nhq} \in C_{nhq}$$

where

$$\delta_i = V_m(e_i^2) = \mu_4(i) - \sigma_i^4,\ \mu_4(i) = E_m(e_i^4)$$

and

$$v_0 = \sum_{i\in U} \frac{a_i}{\pi_i} e_i^2 I_{si} + \sum_{i\neq}\sum_{j\in U} \frac{a_{ij}}{\pi_{ij}} e_i e_j I_{si} I_{sj} + \sum_{i\in U} a_i \sigma_i^2 - \sum_{i\in U} \frac{a_i \sigma_i^2}{\pi_i} I_{si}$$

Proof. Following Cassel *et al*. (1977), we get

$$E_m V_p(v_{nhq}) = E_m E_p[v_{nhq} - V]^2$$
$$= E_p V_m(v_{nhq}) + E_p[E_m(v_{nhq} - V)]^2 - V_m(V) \quad (2.5)$$

$$V_m(v_{nhq}) = V_m(v_{hq})$$
$$= V_m\left(\sum_{i\in U} c_{si} I_{si} e_i^2\right) + V_m\left(\sum_{i\neq}\sum_{j\in U} c_{sij} I_{sij} e_i e_j\right)$$
$$+ 2\ C_m\left(\sum_{i\in U} c_{si} e_i^2 I_{si}, \sum_{i\neq}\sum_{j\in U} c_{sij} e_i e_j I_s\right) \quad (2.6)$$

where C_m denotes covariance with respect to the model M.

Since e_i's are distributed independently, we get

$$V_m\left(\sum_{i\in U} c_{si} I_{si} e_i^2\right) = \sum_{i\in U} \{V_m(e_i^2)\}(c_{si}^2 I_{si}) + \sum_{i\neq}\sum_{j\in U} \{C_m(e_i^2, e_j^2)\}(c_{si} I_{si}, c_{sj} I_{sj})$$

$$= \sum_{i\in U} \delta_i (c_{si}^2 I_{si}) \quad (2.7)$$

$$V_m(\sum_{i\neq}\sum_{j\in U} c_{sij} I_{sij} e_i e_j)$$

$$= 2\sum_{i\neq}\sum_{j\in U} \left\{V_m(e_i e_j)\right\}(c_{sij} I_{sij})^2$$

$$+\sum_{i\neq}\sum_{j\neq}\sum_{k\neq}\sum_{l\in U} C_m(e_i e_j, e_k e_l)(c_{sij} I_{sij}\cdot c_{skl} I_{skl})$$

$$+ 4\sum_{i\neq}\sum_{j\neq}\sum_{k\in U} C_m(e_i e_j, e_i e_k)(c_{sij} I_{sij}\cdot c_{sik} I_{sil})$$

$$= 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2\sigma_j^2 (c_{sij} I_{sij})^2 \quad (2.8)$$

$$C_m(\sum_{i\in U} c_{si} I_{si} e_i^2, \sum_{i\neq}\sum_{j\in U} c_{sij} I_{sij} e_i e_j)]$$

$$= 2\sum_{i\neq}\sum_{j\in U} C_m(e_i^2, e_i e_j)(c_{si} I_{si}\cdot c_{sij} I_{sij})$$

$$+ \sum_{i\neq}\sum_{j\neq}\sum_{k\in U} C_m\{e_i, e_j e_k\}(c_{si} I_{si}\cdot c_{sjk} I_{sjk}) = 0 \quad (2.9)$$

Equations (2.7), (2.8) and (2.9) yield

$$E_p V_m(v_{nhq}) = E_p V_m(v_{hq})$$

$$= E_p[\sum_{i\in U} \delta_i (c_{si}^2 I_{si}) + 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2\sigma_j^2 (c_{sij} I_{sij})^2]$$

$$\geq \sum_{i\in U} \delta_i \frac{a_i^2}{\pi_i} + 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2\sigma_j^2 \frac{a_{ij}^2}{\pi_{ij}} \quad (2.10)$$

(since

$$E_p(c_{si}^2 I_{si}) = \sum_s (c_{si}^2 I_{si}) p(s) \geq \frac{\{\sum_s (c_{si} I_{si}) p(s)\}^2}{\sum_s I_{si} p(s)}$$

$$= \frac{a_i^2}{\pi_i} \text{ and}$$

$$E_p(c_{sij} I_{sij})^2 = \sum_s (c_{sij} I_{sij})^2 p(s) \geq \frac{\{\sum_s (c_{sij} I_{sij}) p(s)\}^2}{\sum_s I_{sij} p(s)}$$

$$= \frac{a_{ij}^2}{\pi_{ij}} \quad \text{(using (2.4))}$$

Equality in (2.10) holds if $c_{si} = \dfrac{a_i}{\pi_i}$ and

$$c_{sij} = \frac{a_{ij}}{\pi_{ij}}$$

$$[E_m(v_{nhq} - V)]^2 = [c_s + E_m(\sum_{i\in U} c_{si} I_{si} e_i^2$$

$$+\sum_{i\neq}\sum_{j\in s} c_{sij} I_{sij} e_i e_j - \sum_{i\in U} a_i E_i^2 - \sum_{i\neq}\sum_{j\in s} a_{ij} e_i e_j)]^2$$

$$= [c_s + \sum_{i\in U} c_{si} I_{si}\sigma_i^2 - \sum_{i\in U} a_i\sigma_i^2]^2 \quad (2.11)$$

From (2.10) and (2.11), we note that $E_m V_p(v_{nhq})$ is minimized when

$$c_{si} = \frac{a_i}{\pi_i},\ c_{sij} = \frac{a_{ij}}{\pi_{ij}} \text{ and } c_s = \sum_{i\in U} a_i\sigma_i^2 - \sum_{i\in U} \frac{a_i\sigma_i^2}{\pi_i} I_{si}$$

and the minimum value of $E_m V_p(v_{nhq})$ is

$$\sum_{i\in U} \delta_i \frac{a_i^2}{\pi_i} + 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2\sigma_j^2 \frac{a_{ij}^2}{\pi_{ij}} - V_m(V)$$

$$= \sum_{i\in U} \delta_i a_i^2 (\frac{1}{\pi_i} - 1) + 2\sum_{i\neq}\sum_{j\in U} \sigma_i^2\sigma_j^2 a_{ij}^2 (\frac{1}{\pi_{ij}} - 1)$$

Q.E.D.

Remark 2.1. Consider a situation where y_i's are independently and identically distributed with common mean μ, common variance σ^2 and common fourth order moment $\mu_4 = E_m(y_i - \mu)^4$; $x_i = Q_i = 1$ and simple random sampling design (SRSWOR) is used. In this situation

$$\beta = \mu,\ a_i = \frac{N-n}{n},\ a_{ij} = \frac{N-n}{n(N-1)}$$

$$\beta_Q = \frac{Y}{N} = \bar{Y},\ \hat{\beta}_Q = \frac{\sum_{i\in s} y_i}{n} = \bar{y}_s$$

$$t_g = N s_y^2, V = N.\frac{N-n}{n} S_y^2$$

$$v_0 = N.\frac{N-n}{n} s_y^2$$

where $(N-1)S_y^2 = \sum_{i \in U}(y_i - \bar{Y})^2$ and

$$(n-1)s_y^2 = \sum_{i \in s}(y_i - \bar{y}_s)^2$$

Using the Theorem 1, we get

$$E_m V_p(v_{nhq}) \geq \sum_{i \in U} \delta_i a_i^2 (\frac{1}{\pi_i} - 1)$$

$$+ 2\sum_{i \neq} \sum_{j \in U} \sigma_i^2 \sigma_j^2 a_{ij}^2 (\frac{1}{\pi_{ij}} - 1)$$

$$= (N \cdot \frac{N-n}{n})^2 [\frac{\mu_4 - \sigma^4}{Nn} + 2\sigma^4 (\frac{1}{n(n-1)}$$

$$- \frac{1}{N(N-1)})] \quad (2.12)$$

$$= E_m V_p(v_0) = (N \, . \, \frac{N-n}{N})^2 \, E_m V_p \, (s_y^2)$$

The expression in the third bracket of (2.12) was derived by Sengupta (1988) as a lower bound of variances of unbiased estimators of S_y^2.

Remark 2.2. The optimal estimator v_0 given in Theorem1 is not usable in practice since it involves several unknown parameters. So, we obtain an approximate optimal estimator by replacing the parameters using their suitable estimates as follows.

$$\hat{v}_0 = \sum_{i \in U} \frac{a_i}{\pi_i} \hat{e}_i^2 I_{si} + \sum_{i \neq} \sum_{j \in U} \frac{a_{ij}}{\pi_{ij}} \hat{e}_i \hat{e}_j I_{sij}$$

$$+ \sum_{i \in U} a_i \hat{\sigma}_i^2 - \sum_{i \in U} \frac{a_i \hat{\sigma}_i^2}{\pi_i} I_{si}$$

where $\hat{\sigma}_i^2$ is an unbiased estimator of σ_i^2 and $\hat{e}_i = y_i - \hat{\beta} x_i$.

Let us consider C_q, a class of quadratic unbiased estimators of V consisting of estimators of the form

$$v_q = c_s + \sum_{i \in U} c_{si} I_{si} e_i^2 + \sum_{i \neq} \sum_{j \in U} c_{sij} I_{sij} e_i e_j + \sum_{i \in U} d_{si} e_i I_{si}$$

$$= v_{nhq} + \sum_{i \in U} d_{si} e_i I_{si}$$

where v_{nhq} is given in (2.3) and d_{si}'s are constants satisfying the unbiasedness condition $E_p(d_{si}) = 0$. Clearly $C_q \supset C_{nhq}$. Here, we have the following theorem relating to the lower bound of the unbiased estimators of V belonging to the class C_q.

Theorem 2. Under model M with

$$\mu_3(i) = E_m(e_i^3) = 0 \; \forall \quad i \in U$$

$$E_m V_p(v_q) \geq \sum_{i \in U} \delta_i a_i^2 (\frac{1}{\pi_i} - 1)$$

$$+ 2\sum_{i \neq} \sum_{j \in U} \sigma_i^2 \sigma_j^2 a_{ij}^2 (\frac{1}{\pi_{ij}} - 1)$$

$$= E_m V_p(v_0) \text{ for } v_q \in Cq$$

where v_0 is as given in the Theorem1.

Proof. Letting $v_q = a_s + \sum_{i \in U} d_{si} e_i I_{si} + v_{hq}$

$$v_{hq} = \sum_{i \in U} c_{si} I_{si} e_i^2 + \sum_{i \neq} \sum_{j \in U} c_{sij} I_{sij} e_i e_j$$

and using (2.5), we get

$$E_m V_p(v_q) = E_p V_m(v_q) + E_p[E_m(v_q - V)]^2 - V_m(V) \quad (2.13)$$

and

$$E_p V_m(v_q) = E_p[V_m(v_{hq}) + V_m(\sum_{i \in s} d_{si} e_i)$$

$$+ 2C_m(v_{hq}, \sum_{i \in s} d_{si} e_i)]$$

$$= E_p[V_m(v_{hq}) + V_m(\sum_{i \in s} d_{si} e_i)]$$

since $C_m(v_{hq}, \sum_{i \in s} d_{si} e_i) = 0$ as $\mu_3(i) = 0 \; \forall \; i \in U$

Further using (2.10), we get

$$E_p V_m(v_q) \geq \sum_{i \in U} \delta_i \frac{a_i^2}{\pi_i} + 2\sum_{i \neq} \sum_{j \in U} \sigma_i^2 \sigma_j^2 \frac{a_{ij}^2}{\pi_{ij}} \quad (2.14)$$

Equality in (2.14) holds if $c_{si} = \dfrac{a_i}{\pi_i}$

$$c_{sij} = \frac{a_{ij}}{\pi_{ij}} \text{ and } d_{si} = 0$$

$$[E_m(v_q - V)]^2 = [c_s + \sum_{i\in U} c_{si} I_{si} \sigma_i^2 - \sum_{i\in U} a_i \sigma_i^2]^2$$
$$= 0$$

when
$$c_s = \sum_{i\in U} a_i \sigma_i^2 - \sum_{i\in U} c_{si} I_{si} \sigma_i^2 \quad (2.15)$$

Finally putting (2.14) and (2.15) in (2.13) we can verify the Theorem 2.

Let C_u be the class of unbiased estimator of V, then using the Theorem 1 and result of Liu (1974), we have the following theorem.

Theorem 3. Let y_i's be independent and identically distributed random variables with a common mean μ, common variance σ^2 and common fourth order central moment $\mu_4 = E_m(y_i - \mu)^4$. Then for any fixed effective size n sampling design with $\pi_{ij} > 0$

$$E_m V_p(v) \geq (\mu_4 - \sigma^4) \sum_{i\in U} a_i^2 (\frac{1}{\pi_i} - 1)$$
$$+ 2\sigma^4 \sum_{i\neq}\sum_{j\in U} a_{ij}^2 (\frac{1}{\pi_{ij}} - 1)$$
$$= E_m V_p(v_0) \;\; \forall\, v \in C_u$$

3. CALIBRATED VARIANCE ESTIMATORS

In this section we will consider calibrated estimators for V, the approximate mean square error of t_g under the model M with $\sigma_i^2 = \sigma^2 x_i^g$.

Now noting

$$M(t_g) = V = \sum_{i\in U} a_i e_i^2 + \sum_{i\neq}\sum_{j\in U} a_{ij} e_i e_j$$
$$= \frac{1}{2} \sum_{i\neq}\sum_{j\in U} \Delta_{ij} (\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j})^2$$

we get the Horvitz-Thomson (1952) and Yates-Grundy's (1953) type design unbiased estimators of V respectively as follows;

$$v_{HT} = \sum_{i\in U} a_i \frac{e_i^2}{\pi_i} I_{si} + \sum_{i\neq}\sum_{j\in U} a_{ij} \frac{e_i e_j}{\pi_{ij}} I_{sij} \quad (3.1)$$

and

$$v_{YG} = \frac{1}{2} \sum_{i\neq}\sum_{j\in U} \Delta_{ij} (\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j})^2 I_{sij} \quad (3.2)$$

3.1. Calibration of v_{HT}

Following Deville and Särndal (1992), we propose an adjustment of v_{HT} as follows:

$$v_{HT}(C) = \sum_{i\in U} a_i w_i e_i^2 I_{si} + \sum_{i\neq}\sum_{j\in U} a_{ij} w_{ij} e_i e_j I_{sij} \quad (3.3)$$

where w_i and w_{ij} are calibrated weights determined by minimizing distance function

$$D = \sum_{i\in U} \frac{\pi_i (w_i - \frac{1}{\pi_i})^2 I_{si}}{q_i} + \sum_{i\neq}\sum_{j\in U} \pi_{ij} \frac{(w_{ij} - \frac{1}{\pi_{ij}})^2 I_{sij}}{q_{ij}} \quad (3.4)$$

(q_i and q_{ij}'s are suitably chosen weights) subject to the calibration constraints

$$E_m[v_{HT}(c)] = E_m(V)$$

i.e.
$$\sum_{i\in U} a_i w_i x_i^g I_{si} = \sum_{i\in U} a_i x_i^g \quad (3.5)$$

Minimization of (3.4) subject to (3.5) yields

$$w_i = \frac{1}{\pi_i} + \lambda q_i x_i^g \cdot \frac{a_i}{\pi_i}, \;\; w_{ij} = \frac{1}{\pi_{ij}}$$

and

$$v_{HT}(c) = v_{HT} + B_{HT}(\sum_{i\in U} a_i x_i^g - \sum_{i\in U} a_i x_i^g \frac{1}{\pi_i} I_{si}) \quad (3.6)$$

where
$$\lambda = \frac{\sum_{i\in U} a_i x_i^g - \sum_{i\in U} a_i x_i^g \frac{1}{\pi_i} I_{si})}{\sum_{i\in U} \frac{1}{\pi_i} a_i^2 q_i x_i^{2g} I_{si}}$$
and

$$B_{HT} = \frac{\sum_{i\in U} \frac{1}{\pi_i} a_i^2 q_i x_i^g e_i^2 I_{si}}{\sum_{i\in U} \frac{1}{\pi_i} a_i^2 q_i x_i^{2g} I_{si}}$$

Remark 3.1. The estimator B_{HT} is a model unbiased estimator of σ^2 since $E_m(B_{HT}) = \sigma^2$. The estimator $v_{HT}(c)$ reduces to the optimal estimator v_0 when B_{HT} equals to σ^2. The estimator $v_{HT}(c)$ given in (3.6) is not usable in practice in general since it involves unknown e_i's. Hence replacing e_i by its estimate $\hat{e}_i$ in $v_{HT}(c)$, we get

$$\hat{v}_{HT}(c) = \hat{v}_{HT} + \hat{B}_{HT}\left(\sum_{i\in U} a_i x_i^g - \sum_{i\in U} a_i x_i^g \frac{1}{\pi_i} I_{si}\right)$$

where $\hat{B}_{HT} = \dfrac{\sum_{i\in U} \frac{1}{\pi_i} a_i^2 q_i x_i^g \hat{e}_i^2 I_{si}}{\sum_{i\in U} \frac{1}{\pi_i} a_i^2 q_i x_i^{2g} I_{si}}$ and

$$\hat{v}_{HT} = \sum_{i\in U} a_i \frac{\hat{e}_i^2}{\pi_i} I_{si} + \sum_{i\neq}\sum_{j\in U} a_{ij} \frac{\hat{e}_i \hat{e}_j}{\pi_{ij}} I_{sij}$$

3.1.1 Particular Cases

For the choices of $q_i = \dfrac{1}{a_i x_i^g}$ and $q_i = \dfrac{1}{a_i}$, $\hat{v}_{HT}(c)$ reduces to $\hat{v}_{HT}(c1)$ and $\hat{v}_{HT}(c2)$ respectively as follows;

$$\hat{v}_{HT}(c1) = \hat{v}_{HT} + \hat{B}_{HT}(1)\left(\sum_{i\in U} a_i x_i^g - \sum_{i\in U} a_i x_i^g \frac{1}{\pi_i} I_{si}\right)$$

where

$$\hat{B}_{HT}(1) = \frac{\sum_{i\in U} \frac{1}{\pi_i} a_i \hat{e}_i^2 I_{si}}{\sum_{i\in U} \frac{1}{\pi_i} a_i x_i^g I_{si}}$$

$$= \frac{\sum_{i\in U} \frac{1}{\pi_i} a_i \hat{e}_i^2 I_{si}}{\sum_{i\in U} \frac{1}{\pi_i} a_i x_i^g I_{si}} \cdot \sum_{i\in U} a_i x_i^g$$

$$+ \sum_{i\neq}\sum_{j\in U} a_{ij} \frac{\hat{e}_i \hat{e}_j}{\pi_{ij}} I_{sij} \tag{3.7}$$

$$\hat{v}_{HT}(c2) = \hat{v}_{HT} + \hat{B}_{HT}(2)\left(\sum_{i\in U} a_i x_i^g - \sum_{i\in U} a_i x_i^g \frac{1}{\pi_i} I_{si}\right) \tag{3.8}$$

where $\hat{B}_{HT}(2) = \dfrac{\sum_{i\in U} \frac{1}{\pi_i} a_i x_i^g \hat{e}_i^2 I_{si}}{\sum_{i\in U} \frac{1}{\pi_i} a_i x_i^{2g} I_{si}}$

3.2 Calibration of v_{YG}

The proposed calibration of v_{YG} is

$$v_{YG}(C) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U} \Delta_{ij} w_{ij} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2 I_{sij} \tag{3.9}$$

Here w_{ij} are calibrated weights determined by minimizing distance function

$$D^* = \sum_{i\neq}\sum_{j\in s} \pi_{ij} \frac{\left(w_{ij} - \frac{1}{\pi_{ij}}\right)^2}{q_{ij}} \tag{3.10}$$

subject to the constraints

$$E_m[v_{YG}(c)] = E_m(V)$$

i.e. $$\sum_{i\neq}\sum_{j\in U} \Delta_{ij} w_{ij} \left(\frac{x_i^g}{\pi_i^2} + \frac{x_j^g}{\pi_j^2}\right) I_{sij}$$

$$= \sum_{i\neq}\sum_{j\in U} \Delta_{ij} \left(\frac{x_i^g}{\pi_i^2} + \frac{x_j^g}{\pi_j^2}\right)$$

(q_{ij}'s are suitably chosen weights)
Minimization (3.10) leads

$$w_{ij} = \frac{1}{\pi_{ij}} + \mu q_{ij} \left(\frac{x_i^g}{\pi_i^2} + \frac{x_j^g}{\pi_j^2}\right) \cdot \frac{\Delta_{ij}}{\pi_{ij}}$$

$$v_{YG}(c) = v_{YG} + B_{YG}\left\{\sum_{i\neq}\sum_{j\in U} \Delta_{ij} \left(\frac{x_i^g}{\pi_i^2} + \frac{x_j^g}{\pi_j^2}\right) - \sum_{i\neq}\sum_{j\in U} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{x_i^g}{\pi_i^2} + \frac{x_j^g}{\pi_j^2}\right) I_{sij}\right\} \tag{3.11}$$

where

$$\mu = \frac{\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) - \sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}\}}{\sum_{i \neq}\sum_{j \in U} \frac{q_{ij}}{\pi_{ij}} \Delta_{ij}^2 \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)^2 I_{sij}}$$

and

$$B_{YG} = \frac{\sum_{i \neq}\sum_{j \in U} \frac{q_{ij}}{\pi_{ij}} \Delta_{ij}^2 \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)\left(\frac{e_i}{\pi_i}-\frac{e_j}{\pi_j}\right)^2 I_{sij}}{\sum_{i \neq}\sum_{j \in U} \frac{q_{ij}}{\pi_{ij}} \Delta_{ij}^2 \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)^2 I_{sij}}$$

Remark 3.2. It can be checked that B_{YG} is a model unbiased estimator of σ^2 as $E_m(B_{YG}) = \sigma^2$. Here, the estimator $v_{YG}(c)$ is also not usable in practice since it involves unknown e_i's. Hence replacing e_i by its estimate $\hat{e}_i$ in $v_{YG}(c)$, we get;

$$\hat{v}_{YG}(c) = \hat{v}_{YG} + \hat{B}_{YG} \cdot \left[\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) - \sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}\right]$$

where $$\hat{B}_{YG} = \frac{\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}^2}{\pi_{ij}} q_{ij} \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)\left(\frac{\hat{e}_i}{\pi_i}-\frac{\hat{e}_i}{\pi_i}\right)^2 I_{sij}}{\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}^2}{\pi_{ij}} q_{ij} \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)^2 I_{sij}}$$

and $$\hat{v}_{YG} = \frac{1}{2}\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{\hat{e}_i}{\pi_i}-\frac{\hat{e}_j}{\pi_j}\right)^2 I_{sij}$$

3.2.1 Particular Cases

For the respective choices of $q_{ij} = [\Delta_{ij}(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2})]^{-1}$ and $q_{ij} = \frac{1}{\Delta_{ij}}$, $\hat{v}_{YG}(c)$ reduces to $\hat{v}_{YG}(c1)$ and $\hat{v}_{YG}(c2)$ as follows;

$$\hat{v}_{YG}(c1) = \hat{v}_{YG} + \hat{B}_{YG}(1) \cdot \left[\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) - \sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}\right]$$

$$= \hat{v}_{YG} \cdot \frac{\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)}{\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}} \qquad (3.12)$$

where $\hat{B}_{YG}(1) = \hat{v}_{YG}\left[\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}\right]^{-1}$

$$\hat{v}_{YG}(c2) = \hat{v}_{YG} + \hat{B}_{YG}(2) \cdot \left[\sum_{i \neq}\sum_{j \in U} \Delta_{ij}\left(\frac{x_i^g}{\pi_i^2}\right) - \sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}}\left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right) I_{sij}\right]$$

where $$\hat{B}_{YG}(2) = \frac{\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)\left(\frac{\hat{e}_i}{\pi_i}-\frac{\hat{e}_j}{\pi_j}\right)^2 I_{sij}}{\sum_{i \neq}\sum_{j \in U} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{x_i^g}{\pi_i^2}+\frac{x_j^g}{\pi_j^2}\right)^2 I_{sij}}$$

4. SIMULATION STUDIES

First of all, we conduct a simulation study to test how the formulae V and V_{SR} serve as an approximation for $M(t_g)$, the mean square of t_g. For the present simulation studies, we have used "Mathematica Software" thoroughly. Using this package, we generate 20 populations each of sizes $N = 100$ so that the values of y_i's and x_i's satisfy the model $y_i = \beta x_i + \epsilon_i$ for $i = 1,\ldots,100$. We first select $\tilde{\epsilon}_i$'s (i = 1, ..., 100) as a random sample from $N(0, 1)$ so that $\epsilon_i = \sigma x_i^{g/2}\tilde{\epsilon}_i$ follows normal distribution with mean zero and variance $\sigma^2 x_i^g$. We then select two sets of x_i's as samples, each of sizes 100 from gamma populations with parameters $\gamma = 10$ and 15 respectively. From each of the given set (i) $\beta = 4$, $\sigma = 1$, $\gamma = 10$ and (ii) $\beta = 2$, $\sigma = 1$, $\gamma = 15$, 10 values of g are chosen, so that we generate 20 populations each of sizes 100 satisfying the model (1.1). From each of the populations, R = 20,000 independent samples each of sizes n = 25 are selected using Midzuno-Sen (M-S, 1952, 1953) sampling scheme. The first two order inclusion probabilities for M-S sampling schemes are respectively

$$\pi_i = \frac{N-n}{N-1} p_i + \frac{n-1}{N-1} \text{ and}$$

$$\pi_{ij} = \frac{(n-1)(N-n)}{(N-1)(N-2)} \cdot (p_i + p_j) + \frac{(n-1)(n-2)}{(N-1)(N-2)}$$

where $p_i = x_i / X$

From each of these sample s_r, r = 1, ..., R, the values of t_g are obtained as $t_g(s_r) = \sum_{i \in s_r} \frac{y_i}{\pi_i} + \hat{\beta}_Q (X - \sum_{i \in s_r} \frac{x_i}{\pi_i})$ using $Q_i = 1/(\pi_i x_i)$ and 1. The mean square error of t_g, is computed as $M(t_g) = \frac{1}{R}\sum_{r=1}^{R}[t_g(s_r) - Y]^2$. The relative percentage accuracies of the proposed formulae V and that V_{SR} for approximating the mean square error of t_g are computed as $AC(1) = | V_{SR}/M(t_g) - 1| \times 100$ and $AC(2) = | V/M(t_g) - 1 | \times 100$. The values of $AC(1)$ and $AC(2)$ are given in the following Table 1. From the Table 1, it is evident that both formulae serve good approximations for the mean square error of t_g.

To compare the relative efficiencies of the proposed alternative calibrated estimators $v(3) = \hat{v}_{YG}(c1)$, $v(4) = \hat{v}_{YG}(c2)$, $v(5) = \hat{v}_{HT}(c1)$, $v(6) = \hat{v}_{HT}(c2)$ and the optimum estimator $v(0) = v_0$ with the existing alternatives $v(1) = \hat{v}_{SR}(1)$ and $v(2) = \hat{v}_{SR}(2)$, we generate 8 populations each of size N =100,

Table 1. Relative percentage accuracy V_{SR} and V

	$\beta = 4, \sigma = 1, \gamma = 15$				$\beta = 2, \sigma = 1, \gamma = 10$			
	$Q_i = 1/\pi_i x_i$		$Q_i = 1$		$Q_i = 1/\pi_i x_i$		$Q_i = i$	
g	AC1	AC2	AC1	AC2	AC1	AC2	AC1	AC2
0	.041	.137	.142	.004	.694	.160	.783	.129
.2	.003	.093	.205	.080	.674	.093	.772	.176
.5	.061	.020	.305	.196	.640	.006	.750	.249
.8	.133	.060	.441	.318	.602	.105	.721	.326
1.0	.185	.117	.484	.401	.573	.171	.697	.378
1.2	.239	.177	.560	.485	.542	.237	.669	.433
1.5	.327	.271	.677	.614	.491	.336	.620	.518
1.8	.420	.372	.797	.746	.433	.436	.559	.608
2.0	.486	.442	.879	.835	.390	.503	.513	.670
2.5	.659	.628	1.088	1.060	.260	.674	.376	.836

corresponding to different values of β, σ, γ and g. As stated earlier we initially select a random sample $\tilde{\epsilon}_i$'s (i = 1,..,100) of size 100 from N(0,1). Then we select 2 sets of random samples x_i's each of sizes 100 from Gamma population with parameter (γ) 5, and 15, respectively. For each of the two chosen combinations, $\beta = 5$, $\sigma = 3$, $\gamma = 5$ and $\beta = 1$, $\sigma = 1$, $\gamma = 15$ we generate 4 sets of values of y_i's for four different values of g as .5, 1.0, 1.5 and 2.0 using the relation $y_i = \beta x_i + \sigma x_i^{g/2} \tilde{\epsilon}_i$, i = 1, ..., 100. From each of the 12 sets of populations, we select R = 2,000 independent samples using Midzuno-Sen (1952, 53) sampling procedure as described earlier. The relative percentage efficiencies (R_j's) are computed for 24 different values of g using the following the formula

$$R_j = \frac{\bar{v}(1)}{\bar{v}(j)} \times 100$$

where

$$\bar{v}(j) = \frac{1}{R}\sum_{r=1}^{R}[v_r(j) - M(t_g)]^2 \text{ for } j = 0,1,2,3,4,5,6; \text{ and}$$

$v_r(j)$ is the value of $v(j)$ obtained from the sample r (=1, ..., R). The relative efficiencies of the variance estimators are presented in Table 2. From the Table 2, it is evident that all the proposed calibrated variance

Table 2. Efficiencies of the proposed variance estimators

	(g = .5, β = 5, σ = 3, γ = 5) α = 0						(g = .5, β = 5, σ = 3, γ = 5) α = 2						(g = .5, β = 5, σ = 3, γ = 5) α = 5					
g	R_0	R_2	R_3	R_4	R_5	R_6	R_0	R_2	R_3	R_4	R_5	R_6	R_0	R_2	R_3	R_4	R_5	R_6
0	99	69	98	98	99	99	100	64	100	100	100	100	100	62	100	100	100	100
.1	100	69	100	100	100	100	100	64	100	100	100	100	99	62	99	99	99	99
.2	100	69	100	100	100	100	100	64	100	100	100	100	98	62	98	98	98	98
.3	101	69	101	101	101	101	99	64	99	99	99	99	97	62	97	97	97	97
.4	101	69	101	101	101	101	99	64	99	99	99	99	96	62	96	96	96	96
.5	101	69	101	101	101	101	98	64	98	98	98	98	95	62	95	95	95	95
.6	101	69	100	100	101	101	97	97	64	97	97	94	97	62	93	93	94	94
.7	100	69	100	100	100	100	96	64	95	96	96	96	93	62	92	92	92	93
.8	100	69	99	99	99	100	95	64	94	94	95	95	91	62	90	90	91	91
.9	99	69	97	98	98	99	94	64	92	93	93	94	90	62	89	89	89	90
1.0	98	69	95	96	97	98	93	64	90	91	92	93	89	62	87	87	88	89
1.1	96	69	93	94	95	96	91	64	88	89	90	91	88	62	85	86	86	88
1.2	95	69	91	93	93	95	90	64	86	87	88	90	87	62	83	84	85	87
1.3	94	69	89	91	91	94	88	64	83	85	85	88	85	62	81	82	83	85
1.4	92	69	86	88	88	92	87	64	81	83	83	87	84	62	79	81	81	84
1.5	91	69	83	86	85	91	85	64	78	81	80	85	83	62	76	79	79	83
1.6	89	69	79	84	82	89	84	64	75	78	77	84	81	62	74	77	76	81
1.7	87	69	76	81	79	87	82	64	71	76	74	82	80	62	71	75	74	80
1.8	86	69	72	78	76	86	80	64	68	74	71	80	78	62	69	73	71	78
1.9	84	69	69	76	72	84	78	64	65	71	68	78	77	62	66	71	69	77
2.0	82	69	65	73	69	82	76	64	61	69	65	76	75	62	63	69	66	75
2.2	78	69	57	68	61	78	73	64	55	64	58	73	72	62	57	64	60	72
2.5	72	69	46	59	50	72	66	64	45	56	48	66	66	62	48	57	51	66
3.0	61	69	30	46	34	61	56	64	30	44	33	56	56	62	34	46	37	56
	(g =.5, β = 1, σ = 1, γ = 15) α = 0						(g = .5, β = 1, σ = 1, γ = 15) α = 2						(g = .5, β = 1, σ = 1, γ = 15) α = 5					
0	100	83	100	100	100	100	99	94	99	99	99	99	99	106	99	99	99	99
.1	100	83	100	100	100	100	100	94	100	100	100	100	99	106	99	99	99	99
.2	100	83	100	100	100	100	100	94	100	100	100	100	100	106	100	100	100	100
.3	100	83	100	100	100	100	101	94	101	101	101	101	101	106	101	101	101	101
.4	100	83	100	100	100	100	101	94	101	101	101	101	102	106	102	102	102	102
.5	100	83	99	99	100	100	101	94	101	101	101	101	102	106	102	102	102	102
.6	99	83	99	99	99	99	102	94	102	102	102	102	103	106	103	103	103	103
.7	99	83	99	99	99	99	102	94	102	102	102	102	104	106	104	104	104	104
.8	98	83	98	98	98	98	102	94	102	102	102	102	104	106	105	105	104	104

.9	98	83	97	97	97	98	102	94	102	102	102	102	105	106	105	105	105	105
1.0	97	83	96	96	97	97	102	94	102	102	102	102	106	106	106	106	106	106
1.1	96	83	95	95	96	96	102	94	101	102	102	102	106	106	106	107	106	106
1.2	95	83	94	94	95	95	102	94	101	101	101	102	107	106	107	107	107	107
1.3	94	83	93	93	94	94	101	94	100	101	101	101	107	106	107	108	107	107
1.4	93	83	91	92	92	93	101	94	100	100	100	101	108	106	108	108	108	108
1.5	92	83	90	91	91	92	101	94	99	100	100	101	108	106	108	108	108	108
1.6	91	83	88	89	90	91	100	94	98	99	99	100	109	106	109	109	109	109
1.7	90	83	87	88	88	90	100	94	97	98	98	100	109	106	109	109	109	109
1.8	89	83	85	86	87	89	99	94	96	97	97	99	109	106	109	109	109	109
1.9	88	83	83	85	85	88	98	94	95	96	96	98	110	106	109	110	110	110
2.0	86	83	81	83	83	86	98	94	93	95	95	98	110	106	110	110	110	110
2.2	84	83	77	80	79	84	96	94	90	93	92	96	110	106	110	110	110	110
2.5	80	83	70	74	73	80	94	94	85	89	87	94	111	106	109	109	109	111
3.0	73	83	58	65	61	73	89	94	73	81	77	89	111	106	104	108	106	111
(g = 1, β = 5, σ = 3, γ = 5) α = 0							(g = 1, β = 5, σ = 3, γ = 5) α = 2						(γ =1, β = 5, σ = 3, γ = 5) α =5					
0	97	93	97	97	97	97	98	85	97	97	98	98	99	76	99	99	99	99
.1	99	93	99	99	99	99	99	85	99	99	100	99	100	76	99	99	100	100
.2	101	93	100	101	101	101	101	85	101	101	101	101	100	76	100	100	100	100
.3	103	93	103	103	103	103	102	85	102	102	102	102	101	76	101	101	101	101
.4	105	93	105	105	105	105	104	85	104	104	103	104	101	76	101	101	101	101
.5	106	93	107	107	106	106	104	85	105	105	104	104	101	76	101	101	101	101
.6	108	93	108	108	107	108	105	85	105	105	105	105	101	76	101	101	101	101
.7	109	93	109	109	108	109	105	85	106	106	105	105	100	76	100	100	100	100
.8	109	93	109	110	109	109	105	85	106	106	106	105	100	76	100	99	100	100
.9	110	93	110	110	110	110	105	85	105	105	105	105	99	76	99	99	99	99
1.0	110	93	110	110	110	110	105	85	105	105	105	105	98	76	98	97	98	98
1.1	110	93	109	110	110	110	104	85	104	104	104	104	97	76	96	96	97	97
1.2	109	93	109	109	109	109	103	85	102	102	103	103	95	76	94	94	96	95
1.3	109	93	108	108	108	109	102	85	101	101	102	102	94	76	93	93	94	94
1.4	108	93	106	107	107	108	100	85	99	99	100	100	92	76	90	91	92	92
1.5	107	93	104	105	106	107	99	85	96	97	98	99	90	76	88	88	90	90
1.6	105	93	102	103	104	105	97	85	94	95	96	97	88	76	85	86	87	88
1.7	104	93	99	101	101	104	95	85	91	92	93	95	86	76	82	84	85	86
1.8	102	93	96	98	98	102	93	85	87	90	90	93	84	76	79	81	82	84
1.9	100	93	92	96	95	100	91	85	84	87	87	91	82	76	76	78	79	82
2.0	98	93	88	93	92	98	89	85	80	84	83	89	82	76	72	76	76	80

2.2	93	93	80	87	84	93	84	85	72	78	76	84	75	76	65	70	69	75
2.5	86	93	67	77	71	86	76	85	59	68	64	76	68	76	54	61	58	68
3.0	72	93	45	60	49	72	63	85	40	52	44	63	56	76	37	47	41	56
	($g = 1, \beta = 1, \sigma = 1, \gamma = 15$) $\alpha = 0$						($g = 1, \beta = 1, \sigma = 1, \gamma = 15$) $\alpha = 2$						($g = 1, \beta = 1, \sigma = 1, \gamma = 15$) $\alpha = 5$					
0	99	93	99	99	99	99	99	98	99	99	99	99	99	106	98	98	99	99
.1	100	93	100	100	100	100	100	98	100	100	100	100	99	106	99	99	100	99
.2	100	93	100	100	100	100	100	98	100	100	100	100	100	106	100	100	100	100
.3	101	93	101	101	101	101	101	98	101	101	101	101	101	106	101	101	101	101
.4	101	93	101	101	101	101	102	98	102	102	102	102	102	106	102	102	102	102
.5	102	93	102	102	102	102	103	98	103	103	102	103	103	106	103	103	103	103
.6	102	93	102	102	102	102	103	98	103	103	103	103	104	106	104	104	104	104
.7	102	93	102	102	102	102	104	98	104	104	103	104	105	106	105	105	105	105
.8	102	93	102	102	102	102	104	98	104	104	104	104	106	106	105	106	105	106
.9	102	93	102	102	102	102	104	98	104	104	104	104	106	106	106	106	106	106
1.0	102	93	102	102	102	102	104	98	104	104	104	104	107	106	107	107	106	107
1.1	101	93	101	101	101	101	104	98	104	104	104	104	107	106	107	107	107	107
1.2	101	93	101	101	101	101	104	98	104	104	104	104	108	106	107	108	107	108
1.3	101	93	100	100	100	101	104	98	104	104	104	104	108	106	108	108	108	108
1.4	100	93	99	99	100	100	104	98	103	103	103	104	108	106	108	108	108	108
1.5	99	93	98	98	99	99	103	98	103	103	103	103	109	106	108	108	108	109
1.6	98	93	97	97	98	98	103	98	102	102	102	103	109	106	108	108	108	109
1.7	97	93	96	96	97	97	102	98	101	101	102	102	109	106	108	108	108	109
1.8	96	93	94	95	96	96	102	98	100	100	101	102	109	106	108	108	108	109
1.9	95	93	93	93	94	95	101	98	99	99	100	101	109	106	107	108	108	109
2.0	94	93	91	92	93	94	100	98	97	98	99	100	109	106	107	108	107	109
2.2	92	93	87	89	89	92	98	98	94	96	96	98	108	106	105	106	106	108
2.5	88	93	81	83	83	88	95	98	88	91	91	95	107	106	102	104	103	107
3.0	80	93	68	74	72	80	89	98	76	82	80	89	104	106	93	98	95	104
	($g = 1.5, \beta = 5, \sigma = 3, \gamma = 5$) $\alpha = 0$						($g = 1.5, \beta = 5, \sigma = 3, \gamma = 5$) $\alpha = 2$						($g = 1.5, \beta = 5, \sigma = 3, \gamma = 5$) $\alpha = 5$					
0	97	119	96	96	97	97	97	112	96	96	97	97	97	101	97	97	97	97
.1	99	119	99	99	99	99	99	112	99	99	99	99	99	101	99	99	99	99
.2	102	119	101	101	102	102	102	112	101	101	102	102	101	101	101	101	101	101
.3	104	119	104	104	104	104	104	112	104	104	104	104	103	101	103	103	103	103
.4	107	119	107	107	106	107	106	112	106	106	106	106	105	101	105	105	105	105
.5	109	119	109	109	109	109	109	112	108	109	108	109	107	101	107	107	106	107

.6	112	119	111	112	111	112	111	112	110	111	110	111	108	101	108	108	108	108
.7	114	119	114	114	113	114	112	112	112	113	112	112	109	101	109	109	109	109
.8	116	119	116	116	115	116	114	112	114	114	113	114	110	101	110	110	110	110
.9	118	119	117	118	117	118	115	112	115	116	115	115	110	101	111	111	111	110
1.0	120	119	119	120	118	120	116	112	117	117	116	116	111	101	111	111	111	111
1.1	121	119	121	121	120	121	117	112	117	118	117	117	110	101	111	111	111	110
1.2	122	119	122	122	121	122	118	112	118	118	118	118	110	101	111	111	111	110
1.3	123	119	123	123	122	123	118	112	118	118	118	118	109	101	111	110	111	109
1.4	123	119	123	124	123	123	117	112	118	118	118	117	108	101	110	109	110	108
1.5	123	119	123	124	123	123	117	112	118	117	118	117	107	101	108	107	109	107
1.6	123	119	123	123	123	123	116	112	117	116	117	116	106	101	107	106	108	106
1.7	123	119	122	123	123	123	115	112	115	115	116	115	104	101	105	104	106	104
1.8	122	119	121	121	122	122	113	112	113	113	115	113	102	101	102	101	104	102
1.9	121	119	119	120	120	121	112	112	111	111	113	112	100	101	99	99	101	100
2.0	119	119	117	118	119	119	110	112	108	108	110	110	97	101	96	96	99	97
2.2	115	119	110	113	113	115	105	112	101	102	104	105	92	101	89	89	92	92
2.5	107	119	97	103	102	107	96	112	88	92	92	96	83	101	76	79	80	83
3.0	91	119	71	82	77	91	80	112	63	72	68	80	68	101	54	61	58	68
(g = 1.5, β = 1, σ = 1, γ = 15) α = 0							(g = 1.5, β = 1, σ = 1, γ = 15) α = 2						(g = 1.5, β = 1, σ = 1, γ = 15) α = 5					
0	98	104	98	98	98	98	98	111	98	98	98	98	98	107	98	98	98	98
.1	100	104	99	99	100	100	99	111	99	99	99	99	100	107	99	99	100	100
.2	101	104	101	101	101	101	101	111	101	101	101	101	101	107	101	101	101	101
.3	102	104	102	102	102	102	102	111	102	102	102	102	102	107	102	102	102	102
.4	103	104	103	103	103	103	103	111	103	103	103	103	103	107	103	103	103	103
.5	104	104	104	104	104	104	104	111	104	104	104	104	104	107	104	104	104	104
.6	105	104	105	105	104	105	105	111	105	106	105	105	105	107	105	105	105	105
.7	105	104	105	105	105	105	107	111	106	107	106	107	106	107	106	106	106	106
.8	106	104	106	106	106	106	108	111	107	108	107	108	107	107	107	107	106	107
.9	107	104	107	107	106	107	108	111	108	109	108	108	107	107	108	108	107	107
1.0	107	104	107	107	107	107	109	111	109	109	109	109	108	107	108	108	108	108
1.1	107	104	107	107	107	107	110	111	110	110	109	110	109	107	109	109	108	109
1.2	107	104	108	108	107	107	111	111	111	111	110	111	109	107	109	109	109	109
1.3	108	104	108	108	108	108	111	111	111	111	111	111	109	107	109	109	109	109
1.4	107	104	108	108	108	107	112	111	111	112	111	112	109	107	109	109	109	109
1.5	107	104	107	107	107	107	112	111	112	112	111	112	109	107	109	109	109	109
1.6	107	104	107	107	107	107	112	111	112	112	112	112	109	107	109	109	109	109

1.7	106	104	106	106	107	106	112	111	112	112	112	112	109	107	109	109	109	109
1.8	106	104	105	105	106	106	112	111	112	112	112	112	109	107	108	108	108	109
1.9	105	104	005	105	105	105	112	111	111	112	111	1112	108	107	107	108	108	108
2.0	104	104	103	103	104	104	112	111	111	111	111	112	107	107	106	107	107	107
2.2	102	104	101	101	102	102	111	111	109	110	110	111	106	107	104	104	105	106
2.5	99	104	95	96	97	99	109	111	105	107	107	109	103	107	99	100	101	103
3.0	91	104	83	86	86	91	104	111	95	99	98	104	96	107	87	91	90	96
(g = 2, β = 5, σ = 3, γ = 5) α = 0							(g = 2, β =5, σ = 3, γ =5) α = 2						(g = 2, β = 5, σ = 3, γ = 5) α = 5					
0	97	137	96	96	97	97	97	134	96	96	97	97	97	127	96	96	97	97
.1	99	137	99	99	99	99	99	134	99	99	99	99	99	127	99	99	99	99
.2	102	137	101	101	102	102	102	134	101	101	102	102	102	127	101	101	102	102
.3	104	137	104	104	104	104	104	134	104	104	104	104	104	127	104	104	104	104
.4	107	137	107	107	107	107	107	134	107	107	107	107	107	127	106	107	106	107
.5	110	137	109	110	109	110	110	134	109	110	109	110	110	127	109	109	109	110
.6	113	137	112	113	112	113	113	134	112	113	111	113	112	127	111	112	111	112
.7	116	137	115	116	114	116	116	134	115	115	114	116	115	127	114	114	113	115
.8	119	137	118	119	117	119	119	134	117	118	116	119	117	127	116	117	115	117
.9	123	137	120	122	119	123	122	134	120	121	119	122	119	127	118	119	117	119
1.0	125	137	123	125	122	125	124	134	122	124	121	124	122	127	120	122	119	122
1.1	128	137	125	128	124	128	127	134	125	126	123	127	123	127	122	123	121	123
1.2	131	137	128	131	126	131	129	134	127	129	125	129	125	127	124	125	123	125
1.3	133	137	130	133	129	133	131	134	129	131	127	131	126	127	126	127	124	126
1.4	136	137	133	136	131	136	133	134	131	133	129	133	127	127	127	128	126	127
1.5	138	137	135	138	133	138	134	134	133	134	131	134	127	127	128	128	127	127
1.6	139	137	137	140	135	139	135	134	134	136	133	135	128	127	129	129	128	128
1.7	140	137	139	141	137	140	136	134	135	137	134	136	127	127	129	129	128	127
1.8	141	137	140	142	138	141	136	134	136	137	135	136	127	127	129	128	128	127
1.9	142	137	141	143	140	142	136	134	136	137	135	136	126	127	128	127	128	126
2.0	142	137	141	143	140	142	135	134	136	136	136	135	124	127	127	125	127	124
2.2	141	137	140	142	141	141	133	134	134	133	134	133	120	127	123	121	124	120
2.5	136	137	134	136	136	136	126	134	126	126	128	126	112	127	113	112	116	112
3.0	120	137	111	116	117	120	110	134	101	105	107	110	95	127	88	91	93	95

	$(g = 2, \beta = 1, \sigma = 1, \gamma = 15)$ $\alpha = 0$						$(g = 2, \beta = 1, \sigma = 1, \gamma = 15)$ $\alpha = 2$						$(g = 2, \beta = 1, \sigma = 1, \gamma = 15)$ $\alpha = 5$					
0	98	116	98	98	98	98	98	118	98	98	98	98	98	119	98	98	98	98
.1	99	116	99	99	99	99	99	118	99	99	99	99	99	119	99	99	99	99
.2	101	116	101	101	101	101	101	118	101	101	101	101	101	119	101	101	101	101
.3	102	116	102	102	102	102	102	118	102	102	102	102	102	119	102	102	102	102
.4	104	116	104	104	104	104	104	118	104	104	104	104	104	119	104	104	104	104
.5	105	116	105	105	105	105	105	118	105	105	105	105	106	119	105	106	105	106
.6	107	116	107	107	106	107	107	118	107	107	106	107	107	119	107	107	107	107
.7	108	116	108	108	108	108	108	118	108	108	108	108	109	119	108	109	108	109
.8	109	116	109	110	109	109	110	118	110	110	109	110	110	119	110	110	109	110
.9	111	116	111	111	110	111	111	118	111	111	110	111	111	119	111	112	111	111
1.0	112	116	112	112	111	112	112	118	112	112	111	112	113	119	112	113	112	113
1.1	113	116	113	113	112	113	113	118	113	114	113	113	114	119	114	114	113	114
1.2	114	116	114	114	113	114	114	118	114	115	114	114	115	119	115	115	114	115
1.3	115	116	115	115	114	115	115	118	115	116	114	115	116	119	116	116	115	116
1.4	115	116	115	116	115	115	116	118	116	116	115	116	117	119	117	117	116	117
1.5	116	116	116	116	115	116	117	118	117	117	116	117	118	119	118	118	117	118
1.6	116	116	116	116	116	116	117	118	117	117	117	117	118	119	118	119	118	118
1.7	116	116	117	117	116	116	117	118	118	118	117	117	119	119	119	119	118	119
1.8	116	116	117	117	116	116	118	118	118	118	117	118	119	119	119	120	119	119
1.9	116	116	117	117	117	116	118	118	118	118	118	118	119	119	119	120	119	119
2.0	116	116	116	116	116	116	117	118	118	118	118	117	119	119	119	120	119	119
2.2	115	116	115	115	116	115	117	118	117	117	117	117	119	119	119	119	119	119
2.5	112	116	112	111	113	112	114	118	114	114	115	114	117	119	116	117	117	117
3.0	105	116	101	102	104	105	108	118	104	105	106	108	112	119	108	109	110	112

estimators $v(3)$, $v(4)$, $v(5)$, $v(6)$ including the optimum estimator $v(0)$ are almost equally efficient for all the populations. These estimators are robust around the neighborhood of the true value of g. It is found that all the proposed estimators are more efficient than $v(1)$ and $v(2)$ for all the population in the neighborhood of the true values of g (=.5, 1, 1.5 and 2). The estimator $v(2) = \hat{v}_{SR}(2)$ found to be much inefficient than $v(1) = \hat{v}_{SR}(1)$ for all the populations with lower values of $g(<1.5)$. However for higher values of g (≥ 1.5), $v(2)$ perform reasonably well. It is also noted that none of calibrated estimators shown negative values (details are not given here). In practice, we never sure whether the model M in (1.1) is appropriate in a given situation. So, we examine performances of the variance estimators if the alternative model $M^*: y_i = \alpha + \beta x_i + \epsilon_i$ holds in practice. Here α is an unknown constant and β, γ and ϵ_i's are as in (1.1). To compare the performances of the proposed variance estimators, we generate additional 16 populations with $\alpha = 2$ and 5 keeping values of β, γ, σ and g unchanged. The simulation studies reveal that all the variance estimators are highly sensitive with the values of α. Here also we find that all the variance estimators including the optimum one are almost equally efficient and perform much better than the existing variance estimators $\hat{v}(1)$ and $\hat{v}(2)$ proposed by Särndal (1982). So, the proposed estimators are robust not only with respect to g but also with α. So, in conclusion, we say that if no definite values of g or α are known, one should use any of the proposed calibrated estimators. But if g is known to be small (<1.5), $v(2)$ should not be used at any cost.

5. CONCLUSION

The lower bounds of the estimators of the design variances of a generalized regression predictor (greg) under certain superpopulation models have been derived. The optimum variance estimators attaining the lower bounds depend on unknown model parameters and hence can not be used in practice. Several calibrated variance estimators for the variance of the greg have been proposed. Simulation studies reveal that all the proposed calibrated estimators including the optimum one are approximately equally efficient and robust with respect to the model parameter g and intercept α. All the proposed estimators are more efficient that existing alternatives proposed by Särndal (1982) around the true value of g. In presence of the super- population model, one may need to estimate the model variance rather than the design variance. The problem of estimation of the model variance of the greg is not considered here and it might be subject of future research.

ACKNOWLEDGEMENT

The authors are grateful to the referee for his comments that lead to improvements of the revised version of this paper.

REFERENCES

Cassel, C.M., Särndal, C.E. and Wretman, J.W. (1976). Some results of generalized difference and generalized regression estimation for a finite population. *Biometrika,* **63**, 615-620.

Cochran,W.G. (1977). *Sampling Techniques*. 3rd edition, Wiley, New York.

Horvitz, D.G. and Thompson, D.J. (1952). A generalisation of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc*., **47**, 663-685.

Kott, P.S. (1990). Estimating the conditional variances of a design consistent regression estimator. *J. Statist. Plann. Inf.*, **24**, 287-296.

Liu,T.P. (1974). Bayes estimation for variance of a finite population. *Metrika*, **21**,127-132.

Murthy, M.N. (1977). *Sampling Theory and Methods*. Statistical Publishing Society.

Midzuno, J. (1952). On the estimating system with probabilities proportional to sum of sizes. *Ann. Ins. Statist. Math*., **3**, 99-107.

Särndal, C.E. (1982). Implications of survey designs for generalized regression estimators of linear functions. *J. Statist. Plann. Inf.*, **7**, 155-170.

Särndal, C.E., Swensson, B.E. and Wretman, J.H. (1989). The weighed residual technique for estimating the variance of the general regression estimator for the finite population total. *Biometrika*, **76**, 527-537.

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

Sen, A.R. (1953). On the estimator of variance in sampling with varying probabilities. *J. Ind. Soc. Agril. Statist*., **5**, 119-127.

Sengupta, S. (1988). Optimality of design unbiased strategy for estimating finite population variance. *Sankhyâ*, **B,** 50.

Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of the regression estimator: Higher level calibration approach. *Survey Methodology*, **24**,41-50.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc*., **B15**, 253-261.

Zou, G. (1999). Variance estimation for unequal probability sampling. *Metrika*, **50**, 71-82.