A New Model Assisted Chi-square Distance Function for Calibration of Design Weights

Matthew Stearns and **Sarjinder Singh**
Department of Statistics
St. Cloud State University
St. Cloud, MN 56301-4498, USA
E-mail: sarjinder@yahoo.com

ABSTRACT

In this paper, we have proposed a new model assisted chi-square distance function to improve the estimator of the general parameter of interest considered by Rao (1994) and Singh (2001). It has been noted that the single model assisted calibration constraint studied by Farrell and Singh (2002, 2005), Arnab and Singh (2003), and Wu (2003) is not helpful to calibrate the Sen (1953) and Yates and Grundy (1953) estimator of the variance of the traditional linear regression estimator under the optimal design of Godambe and Joshi (1965). Three estimators of the proposed linear regression type estimator of the general parameter of interest are introduced and compared. New two-dimensional linear regression models are introduced which are found to be very useful, unlike a simulation based on a couple of thousands of random samples, in comparing the estimators of variance. The use of knowledge of the model parameters in assisting the estimators of variance has been found to be beneficial. The most attractive feature is that it has been shown theoretically that the proposed method of calibration remains always more efficient than the GREG estimator. No doubt all the statistical packages like GES, SUDDAN, STATA, CALMAR etc. could be improved without any hesitation and the general public could get trustworthy estimates from different organizations using these statistical packages after making the necessary amendments with the methodology developed here. It looks that one day this paper will rule the world of calibration methodology.

Keywords: Model assisted calibration; Linear regression estimator; GREG; Estimation of total and variance

1. Introduction

The proper use of auxiliary information in survey sampling has an eminent role in improving the precision of the estimates of the parameters of interest of the study variable. Cochran (1940) considered the problem of estimation of the population mean $\bar{Y}$ of the study variable y by using known information on the population mean $\bar{X}$ of the auxiliary variable x, and introduced a ratio estimator. It remains more efficient than the sample mean estimator if the correlation between the two variables remains positive and high. Murthy (1964) studied the product estimator and it remains more efficient than the sample mean estimator if the correlation between the two variables is negative and high. Under simple random and without replacement (SRSWOR) sampling Hansen, Hurwitz and Madow (1953) proposed a linear regression estimator of the population mean $\bar{Y}$ as:

$$\bar{y}_{\text{lr}} = \bar{y} + \hat{\beta}_{\text{ols}}(\bar{X} - \bar{x}) \tag{1.1}$$

where $\hat{\beta}_{ols} = s_{xy} \big/ s_x^2$. The variance of $\bar{y}_{\text{lr}}$, to the first order of approximation, is given by:

$$V(\bar{y}_{\text{lr}}) = (\frac{1-f}{n})[S_y^2 + \beta_{ols}^2 S_x^2 - 2\beta_{ols} S_{xy}] \tag{1.2}$$

where $\beta_{ols} = S_{xy} \big/ S_x^2$, $(n-1)s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, is unbiased for $(N-1)S_{xy} = \sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$ and $f = n/N$. The estimator $\bar{y}_{\text{lr}}$ has been found to be more efficient than ratio, product, and sample mean estimator for non-zero correlation between the study and auxiliary variable. Deville and Sarndal (1992) introduced a Generalized Regression (GREG) estimator as:

$$\bar{y}_{\text{ds}} = \bar{y} + \hat{\beta}_{\text{ds}}(\bar{X} - \bar{x}) \tag{1.3}$$

where, under SRSWOR sampling, $\hat{\beta}_{\text{ds}} = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2$ is an estimator of $\beta_{\text{ds}} = \sum_{i=1}^{N} X_i Y_i \Big/ \sum_{i=1}^{N} X_i^2$. Obviously, the variance of $\bar{y}_{\text{ds}}$, to the first order of approximation, is given by:

$$V(\bar{y}_{\text{ds}}) = (\frac{1-f}{n})[S_y^2 + \beta_{\text{ds}}^2 S_x^2 - 2\beta_{\text{ds}} S_{xy}] \tag{1.4}$$

Now $V(\bar{y}_{\text{ds}})$ will be less than $V(\bar{y}_{\text{lr}})$ if and only if ($\beta_{ds} < \beta_{ols}$ and $\beta_{ds} > \beta_{ols}$) but both inequalities will never hold. Thus we can say that the linear regression estimator remains better than the GREG estimator. The GREG became very famous during the last 15 years among public servants in the government institutions like the US Bureau of Census, Statistics Canada, and the Australian Bureau of Statistics; as well as private organizations like WESTAT, RAND etc. because it came through a calibration approach. The concept of linear weighting of sample survey data can be found in Bethlehem and Keller (1987). Consider a population $\Omega = \{1, 2, .., i, .., N\}$, from which a sample s $(s \subset \Omega)$ is drawn with any probability proportional to size and without replacement (PPSWOR) sampling design $p(\cdot)$. The inclusion probabilities $\pi_i = P(i \in s)$ and $\pi_{ij} \in P(i \in s, j \in s)$ are assumed to be positive and known. Deville and Särndal (1992) used calibration on the known population total, X, to modify the basic sampling design weights, $d_i = 1/\pi_i$, that appear in the Horvitz and Thompson (1952) estimator:

$$\hat{Y}_{\text{HT}} = \sum_{i \in s} d_i y_i \tag{1.5}$$

An estimator:

$$\hat{Y}_{\text{G}} = \sum_{i \in s} w_i y_i \tag{1.6}$$

was proposed by Deville and Särndal (1992), with weights w_i as close as possible in an average sense to the d_i for a given measurement and subject to the calibration constraint:

$$\sum_{i \in s} w_i x_i = X \tag{1.7}$$

Minimization of the chi square distance function:

$$D = \sum_{i \in s} \left(w_i - d_i\right)^2 \big/ (d_i q_i) \tag{1.8}$$

between the new weights w_i and design weights d_i leads to the GREG of population total, Y, given by:

$$\hat{Y}_{\text{G}} = \sum_{i \in s} d_i y_i + \hat{\beta}_{\text{ds}}(X - \sum_{i \in s} d_i x_i) \tag{1.9}$$

where $\hat{\beta}_{\text{ds}} = \sum_{i \in s} d_i q_i x_i y_i \Big/ \sum_{i \in s} d_i q_i x_i^2$ and q_i are suitably chosen constants. As pointed out by Singh (2003), $\hat{\beta}_{ds}$ in (1.9) is not an ordinary least square estimator, and hence GREG can never be as efficient as the linear regression estimator $\bar{y}_{lr}$ unless the regression line passes through the origin. If $q_i = 1/x_i$, the GREG reduces to the ratio estimator. Singh, Horn, and Yu (1998) pointed out that there is no choice of q_i such that the GREG reduces to product estimator. Following Särndal, Swensson, and Wretman (1989), Deville and Särndal (1992), Särndal (1996), Rao (1997), and several others, the GREG can be written as:

$$\hat{Y}_{\text{G}} = \sum_{i \in s} d_i e_i + \hat{\beta}_{\text{ds}} X \tag{1.10}$$

The Sen (1953) and Yates and Grundy (1953) form of the first estimator of variance of the GREG considered by Deville and Sarndal (1992) is given by:

$$\hat{V}_{ds(1)} = \frac{1}{2} \sum_{i \neq j \in s} \sum D_{ij} (d_i e_i - d_j e_j)^2 \tag{1.11}$$

where $D_{ij} = d_{ij} \Theta_{ij}$, $\Theta_{ij} = (\pi_i \pi_j - \pi_{ij})$, $d_{ij} = \pi_{ij}^{-1}$, $i \neq j$, and $e_i = y_i - \hat{\beta}_{\text{ds}} x_i$. The second estimator studied by Deville and Sarndal (1992) can easily be derived from (1.11) by replacing $(d_i e_i - d_j e_j)^2$ with $(w_i e_i - w_j e_j)^2$. Singh, Horn, and Yu (1998) were the first to apply a higher order calibration approach to estimate the variance of the GREG estimator. Farrell and Singh (2005) considered the superpopulation model:

$$M: y_i = \beta x_i + \sqrt{v(x_i)}\, e_i \tag{1.12}$$

where $y_1, y_2, ..., y_N$ are independently distributed, $E_m(e_i) = 0$, and $V_m(e_i) = \sigma^2$. The parameters β and σ^2 are unknown, but the form of the function $v(x_i)$ is assumed to be known. The variance of the Horvitz and Thompson (1952) estimator $\hat{Y}_{\text{HT}}$ in the Sen (1953) and Yates and Grundy (1953) form is:

$$V_{\text{syg}}(\hat{Y}_{\text{HT}}) = \frac{1}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{ij} \Delta_{y,i,j}^2 \tag{1.13}$$

where $\Delta_{y,i,j} = (d_i y_i - d_j y_j)$. An unbiased estimator for $V_{\text{syg}}(\hat{Y}_{\text{HT}})$ is:

$$\hat{V}_{\text{syg}}(\hat{Y}_{\text{HT}}) = \frac{1}{2} \sum_{i \neq j \in s} \sum d_{ij} \Theta_{ij} \Delta_{y,i,j}^2 \tag{1.14}$$

where $d_{ij} = 1/\pi_{ij}$ are the true design weights as considered by Fuller (1970) and Sitter and Wu (2002). Farrell and Singh (2005) considered a new estimator of variance as:

$$\hat{V}_{\text{FS}}\left(\hat{Y}_{\text{HT}}\right) = \frac{1}{2} \sum_{i \neq j \in s} \sum w_{ij}^{\circ} \Theta_{ij} \Delta_{y,i,j}^2 \tag{1.15}$$

with a new set of weights w_{ij}° such that the two-dimensional chi square distance function:

$$D = \frac{1}{2} \sum_{i \neq j \in s} \sum (w_{ij}^{\circ} - d_{ij})^2 \big/ (d_{ij} q_{ij})^2 \tag{1.16}$$

is minimum subject to the calibration constraints given by:

$$\frac{1}{2} E_m \sum_{i \neq j \in s} \sum w_{ij}^{\circ} \Theta_{ij} \Delta_{y,i,j}^2 = E_m V_{\text{syg}}(\hat{Y}_{\text{HT}}) \tag{1.17}$$

Under model (1.12) the equation (1.17) leads to a new set of higher order calibration constraints:

$$\frac{1}{2} \sum_{i \neq j \in s} \sum w_{ij}^{\circ} \Theta_{ij} \Delta_{x,i,j}^2 = V_{\text{syg}}(\hat{X}_{\text{HT}}), \tag{1.18}$$

and

$$\frac{1}{2} \sum_{i \neq j \in s} \sum w_{ij}^{\circ} \Theta_{ij} v_{i,j} = \frac{1}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{ij} v_{i,j} \tag{1.19}$$

where $V_{\text{syg}}(\hat{X}_{\text{HT}}) = \frac{1}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{ij} \Delta_{x,i,j}^2$, $\Delta_{x,i,j} = (d_i x_i - d_j x_j)$ and $v_{i,j} = d_i^2 v(x_i) + d_j^2 v(x_j)$. Equation (1.18) is similar to that suggested by Singh, Horn, and Yu (1998), also studied by Sitter and Wu (2002), and the equation (1.19) is due to Farrell and Singh (2002) and is dependent on $v(x_i)$. Farrell and Singh (2005) considered the need to study the effect of both constraints on the resultant estimators, because it is easy to see that the Wu and Sitter (2001) and the Sitter and Wu (2002) estimators are special cases of (1.18) and (1.19). Wu (2003) also mentioned the need to study the effect of both of these constraints. Minimization of (1.16) subject to the calibration constraints in (1.18) and (1.19) leads to a new, model assisted and design based, calibrated estimator of the variance of the GREG and is given by:

$$\begin{aligned} \hat{V}_{\text{FS}} = {} & \hat{V}_{\text{ds(1)}} + \hat{\beta}_1 (V_{\text{syg}}(\hat{X}_{\text{HT}}) - \hat{V}_{\text{syg}}(\hat{X}_{\text{HT}})) \\ & + \hat{\beta}_2 \left(\frac{1}{2} \sum_{i \neq j \in \Omega} \sum \Theta_{ij} v_{i,j} - \frac{1}{2} \sum_{i \neq j \in s} \sum d_{ij} \Theta_{ij} v_{i,j}\right) \end{aligned} \tag{1.20}$$

where

$\hat{\beta}_1 = (P^* B - Q^* C) \big/ (AB - C^2)$ and $\hat{\beta}_2 = (Q^* A - P^* C) \big/ (AB - C^2)$,

$P^* = \sum_{i \neq j \in s} \sum d_{ij} q_{ij} \Theta_{ij}^2 \Delta_{x,i,j}^2 (d_i \hat{e}_i - d_j \hat{e}_j)^2$, $B = \sum_{i \neq j \in s} \sum d_{ij} q_{ij} \Theta_{ij}^2 v_{i,j}^2$,

$Q^{*}=\sum_{i\neq j\in s}\sum d_{ij}q_{ij}\Theta_{ij}^{2}v_{i,j}(d_i\hat{e}_i-d_j\hat{e}_j)^2$, $A=\sum_{i\neq j\in s}\sum d_{ij}q_{ij}\Theta_{ij}^{2}\Delta_{x,i,j}^{4}$, $C=\sum_{i\neq j\in s}\sum d_{ij}q_{ij}\Theta_{ij}^{2}v_{i,j}\Delta_{x,i,j}^{2}$. In the very recent work of Farrell and Singh (2005), note carefully that $\hat{\beta}_1$ and $\hat{\beta}_2$ in (1.20) are also not in the least square form of estimators and need to be corrected. Note that if $\pi_i\propto\sqrt{v(x_i)}$, (1.19) reduces to:

$$\sum_{i\neq j\in s}\sum w_{ij}^{\circ}\Theta_{ij}=\sum_{i\neq j\in\Omega}\sum\Theta_{ij} \tag{1.21}$$

No doubt the Farrell and Singh (2005) calibration approach is also appropriate under the Godambe and Joshi (1965) condition for the lower bound of the variance. The reason may be that the GREG is not the best estimator like the linear regression estimator due to Hansen, Hurwitz and Madow (1953). Note that for SRSWOR sampling where Θ_{ij} is constant for all i,j , equation (1.21) reduces to:

$$\sum_{i\neq j\in s}\sum w_{ij}^{\circ}=N(N-1) \tag{1.22}$$

which is the same calibration constraint studied by Sitter and Wu (2002).

In the present investigation, three new estimators of the variance of the new proposed linear regression estimator of the general parameter of interest are studied under different situations. A set of new notations is introduced so that the final results can be put into compact form.

2. General parameter of interest

We consider a general parameter of interest:

$$H_y=\sum_{j\in\Omega}h(y_j) \text{ and } \overline{H}_y=N^{-1}H_y \tag{2.1}$$

for a specified function h. The choice of $h(y)=y$ gives the population total $H_y=Y$ and the population mean $\overline{H}=\overline{Y}$, while the choice $h(y)=\Delta(t_y-y)$ with $\Delta(a_y)=1$ when $a_y\geq 0$ and $\Delta(a_y)=0$ otherwise give the distribution function:

$$\overline{H}_y=F(t_y)=N^{-1}\sum_{j\in\Omega}\Delta(t_y-y_j) \tag{2.2}$$

for each t_y. Rao (1994) suggested a general class of estimators of H_y given by:

$$\hat{H}_y=\sum_{i\in s}d_i(s)h(y_i) \tag{2.3}$$

where the basic weights $d_i(s)$ can depend both on s and $i(i\in s)$ and satisfy the design unbiasedness condition. The choice $h(y)=y$ in (2.3) gives the Godambe (1955) class of estimators of total. If $d_i(s)=d_i$ (or equivalently $d_i(s)=1$ and $h(y_i)=d_iy_i$) and then (2.3) reduces to the Horvitz and Thompson (1952) estimator of population total. If $d_i(s)=w_i$ and $h(y_i)=I(y_i\leq t)$, then (2.3) reduces to the estimator $\hat{F}(t)$ suggested by Silva and Skinner (1995). Rao (1994) suggested an estimator to estimate the variance of the estimator $\hat{H}_y$ as:

$$\hat{V}(\hat{H}_y)=\sum_{i<j\in s}\sum D_{ij}(s)w_iw_j(z_i-z_j)^2 \tag{2.4}$$

where $z_i=h(y_i)/w_i$ and weights $D_{ij}(s)$ can depend both on s and $(i,j)\in s$, and satisfy the unbiasedness condition. It is remarkable that (2.4) depends on the conditional theory that $\hat{H}_y$ equals H_y where $h(y_i)\propto w_i$. The Sen (1953) and Yates and Grundy (1953) estimator of the variance of the Horvitz and Thompson (1952) estimator is a special case of (2.4) with $w_i=\pi_i$ and $D_{ij}(s)=d_{ij}d_id_j\Theta_{ij}$ for any fixed sample size design. Rao (1994) considered the following estimator of the general parameter of interest as:

$$\hat{H}_{\text{Rao}}=\sum_{i\in s}w_i(s)h(y_i) \tag{2.5}$$

where $w_i(s)$ are the calibrated weights such that the chi square distance function:

$$D=\frac{1}{2}\sum_{i\in s}\frac{\{w_i(s)-d_i(s)\}^2}{d_i(s)q_i} \tag{2.6}$$

is minimum subject to the linear constraint:

$$\sum_{i\in s}w_i(s)h(x_i)=H_x \tag{2.7}$$

where $H_x=\sum_{j\in\Omega}h(x_j)$ and q_i are suitably chosen real weights. Obviously, Rao (1994) leads to a GREG of the general parameter of interest H_y as:

$$\hat{H}_{\text{Rao}}=\hat{H}_y+\hat{\beta}_{\text{rao}}(H_x-\hat{H}_x) \tag{2.8}$$

where:

$$\hat{\beta}_{\text{rao}}=\frac{\sum_{i\in s}d_i(s)q_ih(x_i)h(y_i)}{\sum_{i\in s}d_i(s)q_i\{h(x_i)\}^2}. \tag{2.9}$$

Note that $\hat{\beta}_{\text{rao}}$ in (2.9) is not the same as has been used in the traditional linear regression estimator due to Hansen, Hurwitz and Madow (1953).

3. Proposed new chi-square distance function

In this section, we consider a new estimator of the general parameter H_y as:

$$\hat{H}_{\text{ms}}=\sum_{i\in s}\omega_i^{\Diamond}(s)h(y_i) \tag{3.1}$$

Note that under the model:

$$\text{M}:h(y_i)=\alpha+\beta\,h(x_i)+e_i^{\Diamond} \tag{3.2}$$

where α is an intercept, β is a slope, $E_M(e_i^{\Diamond})=0$, $E_M(e_i^{\Diamond 2})=\sigma^2\sqrt{v(h(x_i))}$, and $v(h(x_i))$ is known, (*see* Royall 1970a, 1970b, 1970c, 1971, Hajek 1981, Bellhouse 1984, Pokropp 2002). In (3.1) we consider $\omega_i^{\Diamond}$ as the calibrated *diamond* weights such that the model assisted chi-square distance function defined as:

$$D = \frac{1}{2}\sum_{i\in s}\frac{\sqrt{v(h(x_i))}(\omega_i^{\Diamond}(s) - d_i(s))^2}{d_i(s)q_i^{\Diamond}} \quad (3.3)$$

is minimum subject to the two constraints, given by:

$$\sum_{i\in s}\omega_i^{\Diamond}(s) = \sum_{i\in s}d_i(s) \text{ and } \sum_{i\in s}\omega_i^{\Diamond}(s)h(x_i) = H_x . \quad (3.4)$$

The function $h(x)$ has similar properties as $h(y)$. One most important and **new** point to note here is that the choice of *diamond* weights $q_i^{\Diamond}$, $i \in s$ in (3.3) is not unique but their cross products $q_i^{\Diamond} q_j^{\Diamond}$, $i \neq j \in s$ are unique, depending upon the choice of sampling design and heteroscedastic nature of the model used, and are obtained by solving $n(n-1)/2$ non-linear equations given by:

$$q_i^{\Diamond} q_j^{\Diamond} = D_{ij}\{d_i(s)d_j(s)\}^{-1}\sqrt{v[h(x_i)]v[h(x_j)]} \quad (3.5)$$

for $i \neq j \in s$. Now the minimization of the chi-square distance function (3.3) subject to both constraints in (3.4) leads to a new unique estimator of H_y as:

$$\hat{H}_{ms} = \sum_{i\in s}\omega_i^{\Diamond}(s)h(y_i) = \hat{H}_y + \hat{\beta}_{\text{ols}}(H_x - \hat{H}_x) \quad (3.6)$$

where:

$$\hat{\beta}_{\text{ols}} = \frac{\left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}h(x_i)h(y_i)}{\sqrt{\upsilon[h(x_i)]}}\right)\left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}}{\sqrt{\upsilon[h(x_i)]}}\right) - \left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}h(y_i)}{\sqrt{\upsilon[h(x_i)]}}\right)\left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}h(x_i)}{\sqrt{\upsilon[h(x_i)]}}\right)}{\left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}}{\sqrt{\upsilon[h(x_i)]}}\right)\left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}[h(x_i)]^2}{\sqrt{\upsilon[h(x_i)]}}\right) - \left(\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}h(x_i)}{\sqrt{\upsilon[h(x_i)]}}\right)^2} \quad (3.7)$$

or equivalently:

$$\hat{\beta}_{\text{ols}} = \frac{\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]v[h(x_j)]}}\left[h(x_i)h(y_i) - h(x_i)h(y_j)\right]}{\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]v[h(x_j)]}}\left[\{h(x_i)\}^2 - h(x_i)h(x_j)\right]}$$

$$= \frac{\frac{1}{2}\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]v[h(x_j)]}}\left[h(x_i)h(y_i) - h(x_i)h(y_j) - h(x_j)h(y_i) + h(x_j)h(y_j)\right]}{\frac{1}{2}\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]\,v[h(x_j)]}}\left[\{h(x_i)\}^2 - h(x_i)h(x_j) - h(x_i)h(x_j) + \{h(x_j)\}^2\right]}$$

$$= \frac{\frac{1}{2}\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]v[h(x_j)]}}\left[h(x_i) - h(x_j)\right]\left[h(y_i) - h(y_j)\right]}{\frac{1}{2}\sum_{i\neq j\in s}\frac{d_i(s)q_i^{\Diamond}d_j(s)q_j^{\Diamond}}{\sqrt{v[h(x_i)]v[h(x_j)]}}\left[h(x_i) - h(x_j)\right]^2} \quad (3.8)$$

which is clearly the Sen (1953) and Yates and Grundy (1953) form of the estimator of regression coefficient.

What a cute result!

Note that although there are several choices and methods of choosing diamonds weights in (3.8) (or (3.7)), we would suggest a best choice that beats the world of calibration methodology, and makes the proposed estimator work well under any PPSWOR sampling design. A list of 50 PPSWOR sampling designs can be had from Brewer and Hanif (1983).

If $d_i(s) = 1$, $h(x_i) = d_i x_i$, $h(y_i) = d_i y_i$ then the estimator $\hat{H}_{\text{ms}}$ becomes the unique traditional linear regression estimator of population total under any PPSWOR sampling and is given by:

$$\hat{Y}_{\text{ms}} = \hat{Y}_{\text{HT}} + \hat{\beta}_{\text{ols}}(X - \hat{X}_{\text{HT}}) \quad (3.9)$$

We suggest to make a unique choice in pairs of the *diamond* weights $q_i^{\Diamond}$ and $q_j^{\Diamond}$ such that:

$$q_i^{\Diamond} q_j^{\Diamond} = D_{ij}\sqrt{v[h(x_i)]v[h(x_j)]}, \quad i \neq j \in s \quad (3.10)$$

then $\hat{\beta}_{\text{ols}}$ in (3.8) (or (3.7)) becomes:

$$\hat{\beta}_{\text{ols}} = \frac{\sum_{i\neq j\in s}D_{ij}[d_i^2 x_i y_i - d_i x_i d_j x_j]}{\sum_{i\neq j\in s}D_{ij}[d_i^2 x_i^2 - d_i x_i d_j x_j]}$$

$$= \frac{\frac{1}{2}\sum_{i\neq j\in s}D_{ij}[d_i^2 x_i y_i - d_i x_i d_j x_j - d_j x_j d_i y_i + d_j^2 x_j y_j]}{\frac{1}{2}\sum_{i\neq j\in s}D_{ij}[d_i^2 x_i^2 - d_i x_i d_j x_j - d_j x_j d_i x_i + d_j^2 x_j^2]}$$

$$= \frac{\frac{1}{2}\sum_{i\neq j\in s}D_{ij}(d_i x_i - d_j x_j)(d_i y_i - d_j y_j)}{\frac{1}{2}\sum_{i\neq j\in s}D_{ij}(d_i x_i - d_j x_j)^2} = \frac{\hat{\text{cov}}(\hat{Y}_{\text{HT}}, \hat{X}_{\text{HT}})}{\hat{v}(\hat{X}_{\text{HT}})} \quad (3.11)$$

which is the estimator of the regression coefficient β_{ols} under any PPSWOR sampling. Now we have the following theorem:

Theorem 3.1. The estimator $\hat{Y}_{\text{ms}}$ is always more efficient than the GREG $\hat{Y}_{\text{G}}$ under any PPSWOR sampling.

Proof. Under any PPSWOR sampling, to the first order of approximation, we have:

$$V(\hat{Y}_{\text{ms}}) \approx V(\hat{Y}_{\text{HT}}) + \beta_{\text{ols}}^2 V(\hat{X}_{\text{HT}}) - 2\beta_{\text{ols}}\text{Cov}(\hat{Y}_{\text{HT}}, \hat{X}_{\text{HT}}) \quad (3.12)$$

and

$$V(\hat{Y}_{\text{G}}) \approx V(\hat{Y}_{\text{HT}}) + \beta_{\text{ds}}^2 V(\hat{X}_{\text{HT}}) - 2\beta_{ds}\text{Cov}(\hat{Y}_{\text{HT}}, \hat{X}_{\text{HT}}) \quad (3.13)$$

From (3.12) and (3.13), the proposed estimator $\hat{Y}_{\text{ms}}$ will be more efficient than the GREG due to Deville and Sarndal (1992) $\hat{Y}_{\text{G}}$ if:

$$V(\hat{Y}_{\text{ms}}) < V(\hat{Y}_{\text{G}})$$

or if:

$$(\beta_{\text{ols}} - \beta_{\text{ds}})[(\beta_{\text{ols}} + \beta_{\text{ds}})V(\hat{X}_{\text{HT}}) - 2\text{Cov}(\hat{X}_{\text{HT}}, \hat{Y}_{\text{HT}})] < 0 \quad (3.14)$$

Note that:

$$\beta_{\text{ols}} = \text{Cov}(\hat{Y}_{\text{HT}}, \hat{X}_{\text{HT}}) / \text{V}(\hat{X}_{\text{HT}}),$$

thus on dividing both sides of (3.14) by $V(\hat{X}_{\text{HT}})$, we have:

$$(\beta_{\text{ols}} - \beta_{\text{ds}})^2 > 0$$

which is always true, and thus proves the theorem.

If $d_i(s)=1$, $h(x_i)=d_i h^{\circ}(x_i)$, $h(y_i)=d_i h^{\circ}(y_i)$, where $h^{\circ}(x_i)$ and $h^{\circ}(y_i)$ are any functions similar to $h(x_i)$ and $h(y_i)$, or they could be the same for simplicity, the estimator $\hat{H}_{\text{ms}}$ becomes:

$$\hat{H}_{ms} = \hat{H}_y + \hat{\beta}_{ols}(H_x - \hat{H}_x) \quad (3.15)$$

we suggest to make a unique choice in pairs of the *diamond* weights $q_i^{\Diamond}$ and $q_j^{\Diamond}$ such that:

$$q_i^{\Diamond} q_j^{\Diamond} = D_{ij}\sqrt{v[h(x_i)]v[h(x_j)]} \quad (3.16)$$

then $\hat{\beta}_{\text{ols}}$ in (3.8) (or (3.7)) becomes:

$$\hat{\beta}_{\text{ols}} = \frac{\sum\limits_{i\neq j\in s} D_{ij}[d_i^2 h(x_i)h(y_i) - d_i h(x_i) d_j h(x_j)]}{\sum\limits_{i\neq j\in s} D_{ij}[d_i^2\{h(x_i)\}^2 - d_i h(x_i) d_j h(x_j)]} = \frac{\hat{\text{cov}}(\hat{H}_{\text{y}}, \hat{H}_{\text{x}})}{\hat{v}(\hat{H}_{\text{x}})} \quad (3.17)$$

which is an estimator of the regression coefficient β_{ols} for any PPSWOR design. Now we have the following theorem:

Theorem 3.2. The estimator $\hat{H}_{\text{ms}}$ is always more efficient than the estimator $\hat{H}_{\text{Rao}}$.

Proof. Under any PPSWOR sampling, to the first order of approximation, we have:

$$V(\hat{H}_{\text{ms}}) \approx V(\hat{H}_y) + \beta_{\text{ols}}^2 V(\hat{H}_x) - 2\beta_{\text{ols}}\text{Cov}(\hat{H}_y, \hat{H}_x) \quad (3.18)$$

and

$$V(\hat{H}_{\text{Rao}}) \approx V(\hat{H}_y) + \beta_{\text{rao}}^2 V(\hat{H}_x) - 2\beta_{\text{rao}}\text{Cov}(\hat{H}_y, \hat{H}_x) \quad (3.19)$$

From (3.18) and (3.19), the proposed estimator $\hat{H}_{\text{ms}}$ will be more efficient than $\hat{H}_{\text{Rao}}$ if:

$$(\beta_{\text{ols}} - \beta_{\text{rao}})^2 > 0$$

which is always true, and thus proves the theorem.

Theorem 3.3. The choice of *diamond* weights $q_i^{\Diamond}$ in (3.5) is not unique, but the final resultant estimator is unique.

Proof. For simplicity, let $n=2$, and $v[h(x_i)]=1$, then:

$$q_1^{\Diamond} q_2^{\Diamond} = D_{12} \quad (3.20)$$

where D_{12} is known. Thus for any real $q_1^{\Diamond}$ there exists a real $q_2^{\Diamond}$ such that (3.20) is satisfied.

Let $n=3$ then there exists $q_i^{\Diamond}$, $i=1,2,3$ such that the following three equations are satisfied:

$$q_1^{\Diamond} q_2^{\Diamond} = D_{12} \quad (3.21)$$

$$q_1^{\Diamond} q_3^{\Diamond} = D_{13} \quad (3.22)$$

$$q_2^{\Diamond} q_3^{\Diamond} = D_{23} \quad (3.23)$$

Thus for $n=3$ the choice of $q_i^{\Diamond}$, $i=1,2,3$ may be unique. But note that for $n=4$, there will be six equations and four unknowns. Hence the theorem.

From the above theorems, the Berger, Tirari and Tille (2003) claim, based on some simulation results, that their estimator remains slightly better than the Montanari (1987) estimator becomes doubtful. The Anderson and Thornburn (2005) simulation results also need to be reinvestigated. All the chapters related to GREG in Sarndal and Lundstrom (2005) can be improved. This paper confirms the recommendation of Singh (2003, 2004) that all the statistical packages, like GES (***a prestige product of Statistics Canada***), SUDDAN, STATA, and CALMAR etc., used by private and government organizations need to be modified so that, in the future, the general public should get trustworthy estimates from the organizations who are using these packages.

4. Verification by weighted least square method

Consider the minimization of the weighed error sum of squares (WSSE) given by:

$$\text{WSSE} = \sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}\hat{e}_i^{\Diamond 2}}{\sqrt{\upsilon[h(x_i)]}} = \min\sum_{i\in s}\frac{d_i(s)q_i^{\Diamond}\left[h(y_i) - \hat{\alpha}_{\text{ols}} - \hat{\beta}_{\text{ols}}\, h(x_i)\right]^2}{\sqrt{\upsilon[h(x_i)]}}$$

we will have the same $\hat{\beta}_{ols}$ as in Section 3.

5. Estimation of variance

In this section, we will first introduce the variance of the estimator of the general parameter and then we will discuss the usual estimator and three new estimators of the variance.

5.1. Variance of the proposed estimator

The proposed estimator $\hat{H}_{ms}$ can be written as:

$$\hat{H}_{ms} = \sum_{i\in s} d_i h(y_i) + \hat{\beta}_{ols}[H_x - \sum_{i\in s} d_i h(x_i)]$$

$$= \sum_{i\in s} d_i \hat{e}_i^{\Diamond} + \hat{\alpha}_{ols}\sum_{i\in s} d_i + \hat{\beta}_{ols} H_x \quad (5.1.1)$$

where $\hat{e}_i^{\Diamond} = h(y_i) - \hat{\alpha}_{ols} - \hat{\beta}_{ols} h(x_i)$ (5.1.2)

Assuming that $\hat{\alpha}_{ols}\sum_{i\in s} d_i + \hat{\beta}_{ols} H_x$ is approximately constant, then the variance of $\hat{H}_{ms}$ can be approximated as:

$$V(\hat{H}_{ms}) \doteq \frac{1}{2}\sum_{i\neq j\in\Omega}\sum \Theta_{ij}(d_i e_i^{\Diamond} - d_j e_j^{\Diamond})^2 \quad (5.1.3)$$

where $e_i^{\Diamond} = h(y_i) - \alpha_{\text{ols}} - \beta_{\text{ols}}\, h(x_i)$.

5.2. Usual estimator of the variance

By using the method of moments (MOM), a usual estimator of variance of $\hat{H}_{ms}$ in the Sen (1953) and Yates and Grundy (1953) form is given by:

$$\hat{v}_0(\hat{H}_{ms}) = \frac{1}{2}\sum_{i\neq j\in\Omega}\sum D_{ij}(s)\Phi_{ij}^2 \quad (5.2.1)$$

where $\Phi_{ij}^2 = w_i w_j (d_i \hat{e}_i^{\Diamond} - d_j \hat{e}_j^{\Diamond})^2$. A slightly better estimator of variance can always be obtained from (5.2.1) by using:

$$\Phi_{ij}^2 = w_i w_j (\omega_i^{\Diamond}\hat{e}_i^{\Diamond} - \omega_j^{\Diamond}\hat{e}_j^{\Diamond})^2 \quad (5.2.2)$$

5.3. Estimator of variance when the variance of the estimator of any parameter of the auxiliary variable is known

We consider the first estimator of the variance as:

$$\hat{\upsilon}_{m(1)}(\hat{H}_{ms}) = \hat{v}_0(\hat{H}_{ms}) + \hat{\beta}_1[V(\hat{H}_x) - \hat{\upsilon}(\hat{H}_x)] \quad (5.3.1)$$

where $\hat{\beta}_1$ is given by:

$$\text{Minimize}: \sum_{i\neq j\in s}\sum D_{ij}(s)Q_{ij}^{(1)}[\Phi_{ij}^2 - \hat{\alpha}_1 - \hat{\beta}_1\Gamma_{ij}^2]^2 \quad (5.3.2)$$

with $\Gamma_{ij} = d_i h(x_i) - d_j h(x_j)$.

Alternatively minimize:

$$D = \frac{1}{2}\sum_{i\neq j\in s}\sum(\omega_{ij}^{(1)} - D_{ij}(s))^2 \Big/ (D_{ij}(s)Q_{ij}^{(1)}) \quad (5.3.3)$$

subject to:

$$\sum_{i\neq j\in s}\sum \omega_{ij}^{(1)} = \sum_{i\neq j\in s}\sum D_{ij}(s) \quad (5.3.4)$$

and

$$\frac{1}{2}\sum_{i\neq j\in s}\sum \omega_{ij}^{(1)}[d_i h(x_i) - d_j h(x_j)]^2 = V(\hat{H}_x) \quad (5.3.5)$$

Then the calibrated estimator:

$$\hat{\upsilon}_{m(1)}(\hat{H}_{ms}) = \frac{1}{2}\sum_{i\neq j\in s}\sum \omega_{ij}^{(1)}\Phi_{ij}^2 \quad (5.3.6)$$

reduces to (5.3.1).

5.4. Estimator of variance when model parameters are known

We consider the second estimator of the variance as:

$$\hat{v}_{m(2)}(\hat{H}_{ms}) = \hat{\upsilon}_0(\hat{H}_{ms}) + \hat{\beta}_2(V_m - \hat{v}_m) \quad (5.4.1)$$

where

$V_m = \frac{1}{2}\sum_{i\neq j\in\Omega}\sum \Theta_{ij}\delta_{ij}$, $\hat{v}_m = \frac{1}{2}\sum_{i\neq j\in s}\sum D_{ij}(s)\delta_{ij}$,

$\delta_{ij} = d_i^2\sqrt{v[h(x_i)]} + d_j^2\sqrt{v[h(x_j)]}$, and $\hat{\beta}_2$ is given by:

$$\text{Minimize}: \sum_{i\neq j\in s}\sum D_{ij}(s)Q_{ij}^{(2)}[\Phi_{ij}^2 - \hat{\alpha}_2 - \hat{\beta}_2\delta_{ij}]^2 \quad (5.4.2)$$

Alternatively minimize:

$$D = \frac{1}{2}\sum_{i\neq j\in s}\sum(\omega_{ij}^{(2)} - D_{ij}(s))^2 \Big/ (D_{ij}(s)Q_{ij}^{(2)}) \quad (5.4.3)$$

subject to:

$$\sum_{i\neq j\in s}\sum \omega_{ij}^{(2)} = \sum_{i\neq j\in s}\sum D_{ij}(s) \quad (5.4.4)$$

and

$$\sum_{i\neq j\in s}\sum w_{ij}^{(2)}\delta_{ij} = \sum_{i\neq j\in s}\sum \Theta_{ij}\delta_{ij} \quad (5.4.5)$$

Then the calibrated estimator:

$$\hat{v}_{m(2)}(\hat{H}_{ms}) = \frac{1}{2}\sum_{i\neq j\in s}\sum w_{ij}^{(2)}\Phi_{ij}^2 \quad (5.4.6)$$

leads to the estimator (5.4.1).

5.4.1. Use of optimal design weights

Note that the estimator $\hat{\upsilon}_{m(2)}(\hat{H}_{ms})$ is a corrected version of the estimator recently studied by Farrell and Singh (2002, 2005), Arnab and Singh (2003), and Wu (2003). Further note that if $\pi_i \propto [v\{h(x_i)\}]^{-1/4}$ and for any fixed sample size design for which: $\sum_{j\neq i\in\Omega}\pi_{ij} = (n-1)\pi_i$, $\sum_{j\neq i\in\Omega}\pi_j = (n-1)$ and $\sum_{i\in\Omega}\pi_i = n$, then $\sum_{i\neq j\in\Omega}\sum \Theta_{ij} = 0$ and (5.4.5) becomes:

$$\sum_{i\neq j\in s}\sum \omega_{ij}^{(2)} = 0 \quad (5.4.7)$$

Thus the Lagrange function becomes:

$$L = \frac{1}{2}\sum_{i\neq j\in s}\sum(\omega_{ij}^{(2)} - D_{ij}(s))^2 \Big/ (D_{ij}(s)Q_{ij}^{(2)}) - \lambda\left[0 - \sum_{i\neq j\in s}\sum D_{ij}(s)\right]$$

On setting $dL \big/ d\omega_{ij}^{(2)} = 0$, we have:

$$\omega_{ij}^{(2)} = D_{ij}(s) \quad (5.4.8)$$

that is, there is no change in the design weights if we consider the use of optimal design weights due to Godambe and Joshi (1965). Recall that under such circumstances the estimator of the variance of GREG reduces to zero as reported by Farrell and Singh (2002).

5.5. Estimator of variance when both variance and model parameters are known

We consider the third estimator of the variance as:

$$\hat{\upsilon}_{(3)}(\hat{H}_{ms}) = \hat{v}_0(\hat{H}_{ms}) + \hat{\beta}_1^{(0)}(V(\hat{H}_x) - \hat{\upsilon}(\hat{H}_x)) + \hat{\beta}_2^{(0)}\left(\sum_{i\neq j\in\Omega}\sum \Theta_{ij}\delta_{ij} - \sum_{i\neq j\in s}\sum D_{ij}(s)\delta_{ij}\right) \quad (5.5.1)$$

where $\hat{\beta}_1^{(0)}$ and $\hat{\beta}_2^{(0)}$ are the partial regression coefficients while we consider to:

$$\text{Minimize}: \sum_{i\neq j\in s}\sum D_{ij}(s)Q_{ij}^{(3)}\left[\Phi_{ij}^2 - \hat{\alpha}_1^{(o)} - \hat{\beta}_1^{(o)}\Gamma_{ij}^2 - \hat{\beta}_2^{(o)}\delta_{ij}\right]^2 \quad (5.5.2)$$

Alternatively minimize:

$$D = \frac{1}{2}\sum_{i\neq j\in s}\sum(\omega_{ij}^{(3)} - D_{ij}(s))^2 \Big/ (D_{ij}(s)Q_{ij}^{(3)}) \quad (5.5.3)$$

subject to:

$$\sum_{i\neq j\in s}\sum \omega_{ij}^{(3)} = \sum_{i\neq j\in s}\sum D_{ij}(s) \quad (5.5.4)$$

$$\frac{1}{2}\sum_{i\neq j\in s}\sum \omega_{ij}^{(3)}\Gamma_{ij}^2 = V(\hat{H}_x) \quad (5.5.5)$$

and

$$\sum_{i\neq j\in s}\sum \omega_{ij}^{(3)}\delta_{ij} = \sum_{i\neq j\in\Omega}\sum \Theta_{ij}\delta_{ij} \quad (5.5.6)$$

The calibrated estimator:

$$\hat{\upsilon}_{(3)}(\hat{H}_{ms}) = \frac{1}{2}\sum_{i\neq j\in s}\sum \omega_{ij}^{(3)}\Phi_{ij}^2 \quad (5.5.7)$$

reduces to (5.5.1).

6. Simulation setup

For simplicity, here we consider the problem of estimation of population total under SRSWOR sampling with: $h(y_i) = y_i$, $h(x_i) = x_i$, and $v(x_i) = x_i^g$. The estimator $\hat{H}_{ms}$ reduces to:

$$\hat{Y}_{lr} = N[\bar{y} + \hat{\beta}_{ols}(\bar{X} - \bar{x})] \quad (6.1)$$

Let $\hat{v}_0$, $\hat{v}_1$, $\hat{v}_2$ and $\hat{v}_3$ be the four estimators of variance derived from Section 5 under SRSWOR sampling (for detail refer to Stearns (2005)).

The percent relative efficiency (RE) of the first estimator $\hat{v}_1$ with respect to $\hat{v}_0$ can be written as:

$$\text{RE}(0,1) = \frac{V(\hat{v}_0)}{V(\hat{v}_1)} \times 100\% = (1 - \rho^2_{\Phi^2\Gamma^2})^{-1} \times 100\% \quad (6.2)$$

where $\rho_{\Phi^2\Gamma^2}$ denotes the simple correlation between $\Phi^2_{ij} = (e_i^\Diamond - e_j^\Diamond)^2$ and $\Gamma^2_{ij} = (x_i - x_j)^2$ for $i \neq j = 1,2,3,...,N$.

The RE of $\hat{v}_2$ with respect to $\hat{v}_0$ can be written as:

$$\text{RE}(0,2) = \frac{V(\hat{v}_0)}{V(\hat{v}_2)} \times 100\% = (1 - \rho^2_{\Phi^2\delta})^{-1} \times 100\% \quad (6.3)$$

where $\rho_{\Phi^2\delta}$ denotes the simple correlation between $\Phi^2_{ij} = (e_i^\Diamond - e_j^\Diamond)^2$ and $\delta_{ij} = \sqrt{x_i^g} + \sqrt{x_j^g}$ for $i \neq j = 1,2,3,...,N$.

The RE of $\hat{v}_3$ with respect to $\hat{v}_0$ can be written as:

$$\text{RE}(0,3) = \frac{V(\hat{v}_0)}{V(\hat{v}_3)} \times 100\% = (1 - R^2_{\Phi^2|\Gamma^2,\delta})^{-1} \times 100\% \quad (6.4)$$

where $R^2_{\Phi^2|\Gamma^2,\delta}$ denotes the population coefficient of determination while regressing $\Phi^2_{ij} = (e_i^\Diamond - e_j^\Diamond)^2$ on both $\delta_{ij} = \sqrt{x_i^g} + \sqrt{x_j^g}$ and $\Gamma^2_{ij} = (x_i - x_j)^2$ for $i \neq j = 1,2,3,...,N$.

Note that RE expressions in (6.2), (6.3) and (6.4) are free from the sample size, but depend upon on parameters. We first considered the well-known Horvitz and Thompson (1952) data set. The population consists of $N = 20$ units with $Y_i =$ the number of households on the *i-th* block, and $X_i =$ the eye estimated number of households on the *i-th* block. The results obtained from three estimators are presented in **Figure 6.1**. In this population, the estimator $\hat{v}_3$ seems to be best from the minimum variance point of view. The estimator $\hat{v}_1$ remains as efficient as the usual estimator $\hat{v}_0$, thus there is not much gain and the value of RE remains around 102%. The RE of the second estimator $\hat{v}_2$ varies from 122% to 128% with the median 127%, and the relative efficiency of the third estimator $\hat{v}_3$ varies from 123% to 129% with a median efficiency of 128%. The value of g was changed from 0.0 to 4.8 with a step of 0.2 to study different heteroscedastic situations. The change in the value of g has little effect on the RE of the estimator $\hat{v}_1$, but it has more effect for the estimators $\hat{v}_2$ and $\hat{v}_3$. Note that the two estimators $\hat{v}_2$ and $\hat{v}_3$ are not defined for $g = 0.0$.

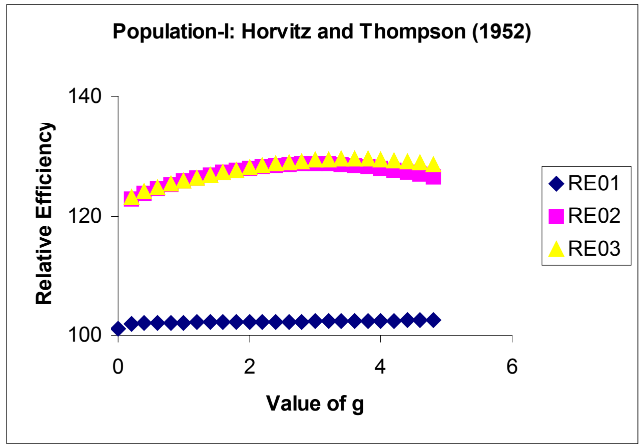

Fig. 6.1. Relative efficiency.

Figure 6.1 shows that the RE of the estimator $\hat{v}_1$ remains substantially lower than the RE of the other two estimators $\hat{v}_2$ and $\hat{v}_3$. Also note that as the value of g increases from 0.0 to 4.8, the RE first increases and then decreases.

Now we consider another large population consisting of $N = 50$ states in the United States. Here y_i represents the number of abortions in the 50 states of the USA during 1992 to 1996 and x_i represents residential population during 2000, and the results so obtained are shown in **Figure 6.2**.

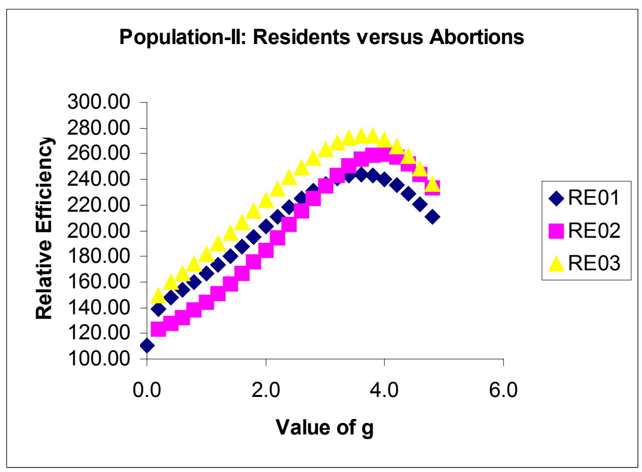

Fig. 6.2. Relative efficiency.

The results based on the large population of 50 units, are much better than those based on the small population. The RE of the estimator $\hat{v}_1$ varies from 110% to 244% with a median RE of 210%. In the same way the RE of the second estimator $\hat{v}_2$ varies from 122% to 259% with a median RE 204%, and that of the third estimator $\hat{v}_3$ varies from 149% to 273% with a median 236%. The REs of the three estimators show a very nice pattern when the value of g changes from 0.0 to 4.8. The RE of the first two estimators $\hat{v}_1$ and $\hat{v}_2$ cross each other, however the RE of the third estimator $\hat{v}_3$ remains a bit higher.

7. Conclusion

The well known statistical packages like GES, SUDDAN, STATA, CALMAR etc. could be improved without any hesitation, and the general public could get trustworthy estimates from different organizations who are using these statistical packages after making the necessary amendments with the methodology developed here.

8. Acknowledgements

The help from a professional English Editor Ms. Melissa Lindsay, Writing Place Center, St. Cloud State University, for editing the entire manuscript has been duly acknowledged. The use of data from the Statistical Abstracts of the United States has also been duly acknowledged.

9. REFERENCES

Anderson, Per Gosta and Thorburn, Daniel (2005). An optimal calibration distance leading to the optimal regression estimator. *Survey Methodology*, 31(1), 95--99.

Arnab, R. and Singh, S. (2003). On estimation of population total using generalized regression predictor. *Proceedings of the Joint Statistical Meeting, California*, 1-- 6 (Available on the CD).

Bellhouse, D.R. (1984). A review of optimal designs in survey sampling. *Canadian J. Statist.*, 12, 53--65.

Berger, Y.G., Tirari, M.E.H. and Tille, Y. (2003). Towards optimal regression estimation in sample surveys. *Australian & New Zealand J. Statistics*, 45, 319--329.

Bethlehem, J.G and Keller, W.J. (1987). Linear weighting of sample survey data. *J. Official Statist.*, 141--153.

Brewer, K.R.W. and Hanif, M. (1983). *Sampling with unequal probabilities.* New York: Springer--Verlag.

Cochran, W.G. (1940). Some properties of estimators based on sampling scheme with varying probabilities. *Australian J. Statist.*, 17, 22-28.

Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc*., 87, 376--382.

Farrell, P.J. and Singh, S. (2002). Re-calibration of higher order calibration weights. *Proc. of the Statistical Society of Canada (Available on the CD).*

Farrell, P.J. and Singh, S. (2005). Model-assisted higher order calibration of estimators of variance. *Australian & New Zealand J. Statist.*, 47, 375--383.

Fuller, W. (1970). Sampling with random stratum boundaries. *J.R. Statist. Soc.*, 32, 208-226.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *J. R. Statist. Soc.,* B, 17, 269--278.

Godambe, V.P. and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations--I. *Ann. Math. Statist.,* 36, 1707--1722.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample survey methods and theory.* New York, John Wiley and Sons, 456--464.

Hajek, J. (1981). *Sampling from a Finite Population*. Marcel Dekker, NY.

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

Montanari, G.E. (1987). Post sampling efficient QR prediction in large sample survey. *Internat. Statist. Rev.*, 55, 191--202.

Murthy, M.N. (1964). Product method of estimation. *Sankhya*, A, 26, 69-74.

Pokropp, F. (2002). Imposed linear structures in conventional sampling theory. *Alligemeines Statist. Archiv,* 86, 333--352.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Official Statistics*, 10(2), 153--165.

Rao, J.N.K. (1997). Developments in sample survey theory: an appraisal. *Canad. J. Statist*., 25, 1--21.

Royall, R.M. (1970a). On finite population sampling theory under certain linear regression models. *Biometrika*,57, 377--387.

Royall, R.M. (1970b). Finite population sampling: on labels in estimation. *Ann. Math*. *Statist.,* 41, 1774--1779.

Royall, R.M. (1970c). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377--387.

Royall, R.M. (1971). Linear regression models in finite population sampling theory. *Foundations of Statist. Infer.* (V.P. Godambe and D.A. Scott, eds). Holt, Rinehart & Winston, Toronto, 259--274.

Särndal, C.E. (1996). Efficient estimators with simple variance in unequal probability sampling. *J. Amer. Statist . Assoc*.,91, 1289--1300.

Särndal, C.E. and Lundstrom, S. (2005). *Estimation in surveys with Nonresponse.* John Wiley & Sons.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76 (3), 527--537.

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

Silva, P.L.D.N. and Skinner, C.J. (1995). Estimating distribution functions with auxiliary information using poststratification. *J. Official Statist.*, 11, 277--294.

Singh, S. (2001). Generalized calibration approach for estimating the variance in survey sampling. *Ann. Ins. Stat. Math.* 53(2), 404--417.

Singh, S. (2003). *Advanced sampling theory with applications: How Michael Selected Amy.* Kluwer Academic Publishers, The Netherlands.

Singh, S. (2004). Golden and Silver Jubilee Year-2003 of the linear regression estimators. *Proceedings of the Joint Statistical Meeting, Toronto, 4382-4389 (Available on the CD)*.

Singh, S., Horn, S. and Yu, F. (1998). Estimation of variance of general regression estimator: Higher level calibration approach. *Survey Methodology,* 24 (1), 41--50.

Sitter, R.R. and Wu, C. (2002). Efficient estimation of quadratic finite population functions. *J. Amer. Statist. Assoc*., 97, 535--543

Stearns, Matthew (2005). *On the estimation of the variance of the general parameter of interest.* Unpublished project submitted to the Department of Statistics, St. Cloud State University, St. Cloud, MN 56301-4498, USA

Wu, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika*, 90, 937--951.

Wu, C. and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.*, 96, 185--193.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. R. Statist. Soc*, 15(B), 253--261.