REPAIR OF TWO-PHASE CALIBRATION METHODOLOGY IN SURVEY SAMPLING

Sarjinder Singh[1]

ABSTRACT

In the present investigation, a question has been answered raised by the Deville and Särndal (1992) calibration approach to the eminent survey statisticians working at the U.S. Bureau of Census, Statistics Canada, Australian Bureau of Statistics, and their consultants from different universities across the world. This particular paper has been designed especially to repair the defective two-phase calibration methodology of Hidiroglou and Särndal (1995, 1998), and hence that of Singh (2000). The proposed methodology is based on the Golden Jubilee Year–2003, celebrated by Singh (2003, 2004a, 2004b), of the linear regression estimator owed to Hansen, Hurwitz and Madow (1953) for its outstanding performance. In this paper, it has been shown that chain regression estimator is unique among its class of estimators. There is no need of simulation study, because the repaired theoretical results are crystal clear. Based on Singh (2003, 2004a, 2004b) work, some modifications in the statistical package General Estimation System (GES), developed by the Statistics Canada, have been strongly recommended.

KEY WORDS: Calibration; Estimation of total and variance, Two-phase sampling , Use of auxiliary information.

RÉSUMÉ

Une question soulevée par l'approche par calibration de Deville et Särndal (1992) aux éminents statisticiens de sondage travaillant au Bureau du recensement des États-Unis, à Statistique Canada, au Bureau australien des statistiques et de leurs consultants, a été résolue par notre recherche. Cet article est particulièrement conçu pour réparer le travail défectueux de Hidiroglou et Särndal (1995, 1998) sur la méthodologie du calage à deux phases et par conséquent celui de Singh (2000).La méthodologie proposée est basée sur la célébration du jubilé d'or en 2003 de l'estimateur par régression linéaire dû à Hansen, Hurwitz et Madow (1953) par Singh (2003, 2004a, 2004b) pour sa performance extraordinaire. Cet article démontrait que l'estimateur de régression en chaîne est unique dans sa classe d'estimateurs. Il n'est pas nécessaire de faire d'étude de simulation puisque les résultats sont limpides. En se basant sur les travaux de Singh (2003, 2004a, 2004b), des modifications au Système généralisé d'estimation développé par Statistique Canada ont été fortement recommandées.

MOTS CLÉS : Calage; échantillonnage en deux phases; utilisation d'information auxiliaire; estimation du total et de la variance.

1. INTRODUCTION

As highlighted by Hidiroglou and Särndal (1995, 1998) that the use of two-phase sampling is a powerful and cost-effective technique and hence has eminent role is survey sampling. Let the population be represented by $\Omega = \{1, 2, ..., i, ..., N\}$. A first-phase probability sample s_1 $(s_1 \subset \Omega)$ is drawn from the population Ω using a sampling design that generates the selection probabilities π_{1i}. Given that the first-phase sample s_1 has been drawn, the second-phase sample s_2 $(s_2 \subset s_1 \subset \Omega)$ is selected from s_1, with a sampling design with the selection probabilities $\pi_{2i} = \pi_{i|s_1}$. Evidently, the first-phase and second-phase sampling weights are defined as $d_{1i} = 1/\pi_{1i}$ and $d_{2i} = 1/\pi_{2i}$, respectively. The overall sampling weight for a selected i^{th} unit in the second-phase sample s_2 will be $d_i^* = d_{1i} d_{2i}$. Then usual unbiased estimator of population total, Y, in two-phase sampling is given by $\hat{Y}_{\mathrm{u}} = \sum_{i \in s_2} d_i^* y_i$. Hidiroglou and Särndal (1995, 1998) and Singh (2000) have considered another estimator of population total, Y, as $\hat{Y}_{\mathrm{HS}} = \sum_{i \in s_2} \tilde{d}_i^* y_i$, where $\tilde{d}_i^*$ are the ultimate calibrated weights and can be obtained in several ways. It is interesting to note that in multi-phase sampling, we can obtain calibrated weights at each phase separately. For example in the present situation, we are considering two-phase sampling, hence two-types of calibrated weights are possible. (a) First-phase calibrated weights (b) Second-phase calibrated weights. A combination of (a) and (b) will lead to ultimate calibrated weights in two-phase sampling. We will discuss these methods one by one to avoid any kind of confusion.

[1] Sarjinder Singh (sarjinder@yahoo.com), Department of Statistics, St. Cloud State University, St. Cloud, MN 56301—4498, USA:

2. FIRST-PHASE CALIBRATION METHODOLOGY OF HIDIROGLOU AND SARNDAL

Let x_{1i}, and x_{2i} be the two auxiliary variables associated with the study variable y_i. Let $x_{1i}, i \subset \Omega$ and $x_{2i}, i \in s_1$ be known. Obviously, an unbiased estimator of population total X_2 from the first-phase sample information is given by $\hat{X}_2 = \sum_{i \in s_1} d_{1i} x_{2i}$. A calibrated estimator of population total X_2 is given by $\hat{X}_2^* = \sum_{i \in s_1} \tilde{d}_{1i} x_{2i}$, where $\tilde{d}_{1i}$ are the calibrated weights obtained from the first-phase sample information. Following Hidiroglou and Sarndal (1995, 1998), Singh (2000) considered a defective first-phase calibrated methodology to choose weights $\tilde{d}_{1i}$ such that the chi square distance defined as $D_1 = \sum_{i \in s_1} \left\{ \left(\tilde{d}_{1i} - d_{1i} \right)^2 / \left(d_{1i} q_{1i} \right) \right\}$ is minimum subject to the first-phase calibration constraint defined as $\sum_{i \in s_1} \tilde{d}_{1i} x_{1i} = X_1$. The resultant calibrated estimator of the population total X_2 becomes $\hat{X}_2^* = \sum_{i \in s_1} d_{1i} x_{2i} + \hat{\beta}_{1(\mathrm{ds})} \left(X_1 - \sum_{i \in s_1} d_{1i} x_{1i} \right)$ where $\hat{\beta}_{1(\mathrm{ds})} = \left(\sum_{i \in s_1} d_{1i} q_{1i} x_{1i} x_{2i} \right) \Big/ \left(\sum_{i \in s_1} d_{1i} q_{1i} x_{1i}^2 \right)$. Note that there is no choice of weights q_{1i} that the estimator $\hat{X}_2^*$ reduces to the traditional linear regression estimator. In the next paragraph, we suggest a new calibration methodology which leads to a traditional linear regression type estimator and repairs the first-phase methodology of Hidiroglou and Särndal (1995, 1998) and hence that of Singh (2000).

2.1 Repaired First-Phase Calibration Methodology

A repaired calibrated estimator of population total X_2 is given by $\hat{X}_2^{\oplus} = \sum_{i \in s_1} \tilde{d}_{1i}^{\oplus} x_{2i}$, where $\tilde{d}_{1i}^{\oplus}$ are called the repaired calibrated weights obtained from the first-phase sample information. We choose the first-phase calibrated weights $\tilde{d}_{1i}^{\oplus}$ such that the chi-square distance $D_1^{\oplus} = \sum_{i \in s_1} \left\{ \left(\tilde{d}_{1i}^{\oplus} - d_{1i} \right)^2 / \left(q_{1i}^{\oplus} d_{1i} \right) \right\}$ is minimum subject to the two-calibration constraints:

$$\sum_{i \in s_1} \tilde{d}_{1i}^{\oplus} = \sum_{i \in s_1} d_{1i} \text{ and } \sum_{i \in s_1} \tilde{d}_{1i}^{\oplus} x_{1i} = X_1 . \tag{2.1.1}$$

Note that the first condition in (2.1.1) is a requirement of the chi square test owed to Prof. Karl Pearson (1900, 1922)/Sir R.A. Fisher (1915, 1922) and is ignored by all the followers of Deville and Sarndal (1992). The choice of $q_{1i}^{\oplus}$ decides the form of the estimator. The repaired first-phase Lagrange function is then defined as

$$L_1 = \sum_{i \in s_1} \frac{\left(\tilde{d}_{1i}^{\oplus} - d_{1i} \right)^2}{q_{1i}^{\oplus} d_{1i}} - 2\lambda_{11}^{\oplus} \left[\sum_{i \in s_1} \tilde{d}_{1i}^{\oplus} x_{1i} - X_1 \right] - 2\lambda_{12}^{\oplus} \left[\sum_{i \in s_1} \tilde{d}_{1i}^{\oplus} - \sum_{i \in s_1} d_{1i} \right] \tag{2.1.2}$$

where $\lambda_{11}^{\oplus}$ and $\lambda_{12}^{\oplus}$ denote the first-phase Lagrange multipliers. On setting $\partial L_1 / \partial \tilde{d}_{1i}^{\oplus} = 0$, we get

$$\tilde{d}_{1i}^{\oplus} = d_{1i} + \lambda_{11}^{\oplus} q_{1i}^{\oplus} d_{1i} x_{1i} + \lambda_{12}^{\oplus} d_{1i} q_{1i}^{\oplus} \tag{2.1.3}$$

On substituting (2.1.3) in both equations of (2.1.1) , we have

$$\lambda_{11}^{\oplus} \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i} \right) + \lambda_{12}^{\oplus} \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} \right) = 0 \quad \text{and} \quad \lambda_{11}^{\oplus} \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i}^2 \right) + \lambda_{12}^{\oplus} \left(\sum_{i \in s_1} d_{1i} q_{1i} x_{1i} \right) = \left(X_1 - \sum_{i \in s_1} d_{1i} x_{1i} \right) . \tag{2.1.7}$$

On solving the system of equations in (2.1.7) for $\lambda_{11}^{\oplus}$ and $\lambda_{12}^{\oplus}$, and substituting back in (2.1.3), we have the repaired first-phase calibrated estimator of X_2 is given by

$$\hat{X}_2^{\oplus} = \sum_{i \in s_2} d_i x_{2i} + \hat{\beta}_{1(\mathrm{ols})} \left[X_1 - \sum_{i \in s_1} d_{1i} x_{1i} \right] \tag{2.1.8}$$

where $\hat{\beta}_{1(\mathrm{ols})} = \left\{ \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i} x_{2i} \right) \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} \right) - \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{2i} \right) \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i} \right) \right\} \Big/ \left\{ \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} \right) \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i}^2 \right) - \left(\sum_{i \in s_1} d_{1i} q_{1i}^{\oplus} x_{1i} \right)^2 \right\}$. (2.1.9)

Clearly the estimator (2.1.8) is a traditional linear regression estimator due to Hansen, Hurwitz and Madow (1953), and is a repaired version of the first-phase calibrated estimator of Hidiroglou and Särndal (1995, 1998) and hence that of Singh (2000). Note that there is no choice of $q_{1i}^{\oplus}$ which reduces the estimator (2.1.8) into ratio or product method of estimation, and leads to the following result, "The traditional first-phase linear regression estimator of the population total X_2 is unique among its class of estimators".

3. SECOND-PHASE CALIBRATION METHODOLOGY OF HIDIROGLOU AND SARNDAL

Obviously, an unbiased estimator of population total, Y, in two-phase sampling is defined as: $\hat{Y}_u = \sum_{i\in s_2} d_{1i}d_{2i}y_i$ .Following Hidiroglou and Särndal (1995, 1998), Singh (2000) considered another defective estimator of population total, Y , defined as $\hat{Y}_c = \sum_{i\in s_2} \tilde{d}_i^* y_i$, where $\tilde{d}_i^*$ are called the second-phase calibrated weights. The second-phase calibrated weights $\tilde{d}_i^*$ are chosen such that the chi-square distance function defined as: $D_2 = \sum_{i\in s_2} \left\{ \left(\tilde{d}_i^* - d_{1i}d_{2i}\right)^2 / \left(d_{1i}d_{2i}q_{2i}\right) \right\}$ is minimum subject to the calibration constraint defined as $\sum_{i\in s_2} \tilde{d}_i^* x_{2i} = \hat{X}_2^*$ where $\hat{X}_2^*$ is obtained after fixing first-phase calibration. The choice of q_{2i} makes different forms of estimators in two-phase sampling. The resultant second-phase calibrated estimator was given by

$$\hat{Y}_c = \sum_{i\in s_2} d_{1i}d_{2i}y_i + \hat{\beta}_{2(ds)}\left[\hat{X}_2^* - \sum_{i=1}^{n} d_{1i}d_{2i}x_{2i}\right] \text{ where } \hat{\beta}_{2(ds)} = \sum_{i\in s_2} d_{1i}d_{2i}q_{2i}x_{2i}y_i \Big/ \sum_{i\in s_2} d_{1i}d_{2i}q_{2i}x_{2i}^2 . \quad (3.1)$$

Case 1. If $q_{1i} = x_{1i}^{-1}$, and $q_{2i} = x_{2i}^{-1}$, then the calibrated estimator $\hat{Y}_c$ at (3.1) reduces to a chain ratio type estimator of population total, Y , defined as

$$\hat{Y}_{\text{Ratio}} = \sum_{i\in s_2} d_{1i}d_{2i}\left(\sum_{i\in s_1} d_{1i}x_{2i} \Big/ \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}\right)\left(X_1 \Big/ \sum_{i\in s_1} d_{1i}x_{1i}\right) \quad (3.2)$$

Note that the estimator (3.2) is a well known chain ratio type estimator studied by Chand (1975), Kiregyera (1980), Sahoo and Swain (1983, 1986), and Srivastava, Khare and Srivastava (1990).

Case 2. If $q_{1i} = 1$ and $q_{2i} = 1$, then the resultant calibrated estimator (3.1) becomes

$$\hat{Y}_{\text{Reg}} = \sum_{i\in s_2} d_{1i}d_{2i}y_i + \hat{\beta}_{2(ds)}\left[\sum_{i\in s_1} d_{1i}x_{2i} - \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}\right] + \hat{\beta}_{2(ds)}\hat{\beta}_{1(ds)}\left[X_1 - \sum_{i\in s_1} d_{1i}x_{1i}\right] \quad (3.7)$$

where, $\hat{\beta}_{2(ds)} = \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}y_i \Big/ \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}^2$ and $\hat{\beta}_{1(ds)} = \sum_{i\in s_1} d_{1i}x_{1i}x_{2i} \Big/ \sum_{i\in s_1} d_{1i}x_{1i}^2$. Note that $\hat{\beta}_{1(ds)}$ and $\hat{\beta}_{2(ds)}$ are not the ordinary least square estimators of the regression coefficients, and denies the claim of Singh (2000) that $\hat{Y}_{\text{Reg}}$ is a chain regression type estimator studied by Singh and Singh (1991), which is also a member of the general class of chain type estimators proposed by Ahmed (1997), Prasad, Singh and Singh (1996), Singh and Upadhyaya (1995), Singh, Singh and Singh (1994), Srivastava, Khare and Srivastava (1990) and Upadhyaya, Khushwaha and Singh (1990).

Caution!: *Note that the logic of construction of different forms of estimators in the General Estimation System (GES) by the Statistics Canada with different choice of weights looks misleading, although it forms different mathematical formulae. Further note that the traditional linear regression estimator automatically reduces to ratio estimator if the regression line passes through the origin, and thus the second requirement of choice of weights to form ratio type estimator using the GES becomes ridiculous.*

3.1. Repaired Second-Phase Calibration Methodology

At "Sarjinder's own workshop of statistical tools", we also repair the second-phase calibrated estimator as $\hat{Y}_c^{\oplus} = \sum_{i\in s_2} \tilde{d}_i^{\oplus} y_i$ where $\tilde{d}_i^{\oplus}$ are called the second-phase calibrated weights. The second-phase calibrated weights $\tilde{d}_i^{\oplus}$ are chosen such that the chi-square distance function $D_2^{\oplus} = \sum_{i\in s_2} \left\{ \left(\tilde{d}_i^{\oplus} - d_{1i}d_{2i}\right)^2 / \left(d_{1i}d_{2i}q_{2i}^{\oplus}\right) \right\}$ is minimum subject to the two-calibration constraints:

$$\sum_{i\in s_2} \tilde{d}_i^{\oplus} = \sum_{i\in s_2} d_{1i}d_{2i} \text{ and } \sum_{i\in s_2} \tilde{d}_i^{\oplus} x_{2i} = \hat{X}_2^{\oplus} \quad (3.1.1)$$

where $\hat{X}_2^{\oplus}$ is same as defined earlier. The repaired second-phase calibrated estimator of the population total Y is given by

$$\hat{Y}_c^{\oplus} = \sum_{i\in s_2} d_{1i}d_{2i}y_i + \hat{\beta}_{2(\text{ols})}\left[\hat{X}_2^{\oplus} - \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}\right] \quad (3.1.2)$$

where

$$\hat{\beta}_{2(\text{ols})} = \left\{\left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}x_{2i}y_i\right)\left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}\right) - \left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}y_i\right)\left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}y_i\right)\right\} \Big/ \left\{\left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}\right)\left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}x_{2i}^2\right) - \left(\sum_{i\in s_2} d_{1i}d_{2i}q_{2i}^{\oplus}x_{2i}\right)^2\right\} \quad (3.1.3)$$

Note that (3.1.2) can be written as

$$\hat{Y}_{\mathrm{Re}\,g}^{\oplus} = \sum_{i\in s_2} d_{1i}d_{2i}y_i + \hat{\beta}_{2(ols)}\left[\sum_{i\in s_1} d_{1i}x_{2i} - \sum_{i\in s_2} d_{1i}d_{2i}x_{2i}\right] + \hat{\beta}_{2(ols)}\hat{\beta}_{1(ols)}\left[X_1 - \sum_{i\in s_1} d_{1i}x_{1i}\right] \tag{3.1.4}$$

where $\hat{\beta}_{1(ols)}$ and $\hat{\beta}_{2(ols)}$ are the ordinary least square estimators, and hence the estimator (3.1.2) (or (3.1.4)) can be claimed as a traditional chain regression type estimator in survey sampling as studied by Tikkiwal (1960). In the next section, we repair the higher order methodology owed to Singh (2000) and hence low level methodology due to Hidiroglou and Särndal (1995, 1998).

4. VARIANCE OF THE REPAIRED CALIBRATED ESTIMATOR

Defining V_1 and V_2 as the variance over all possible first-phase samples and for all second-phase samples from a given first phase sample, then we have

$$V\left(\hat{Y}_c^{\oplus}\right) = E_1V_2\left(\hat{Y}_c^{\oplus}\right) + V_1E_2\left(\hat{Y}_c^{\oplus}\right) = \frac{1}{2}\sum_{i\neq j\in\Omega}\sum \pi_{1ij}\left(\pi_{2i}\pi_{2j} - \pi_{2ij}\right)\left(\frac{e_{2i}^*}{\pi_{1i}\pi_{2i}} - \frac{e_{2j}^*}{\pi_{1i}\pi_{2i}}\right)^2 + \frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_{1i}\pi_{1j} - \pi_{1ij}\right)\left(\frac{e_{1i}^*}{\pi_{1i}} - \frac{e_{1j}^*}{\pi_{1j}}\right)^2 \tag{4.1}$$

where $e_{2i}^* = y_i - \alpha_2 - \beta_2x_{2i}$ are $e_{1i}^* = y_i - \alpha_2 - \beta_2\left\{x_{2i} - \alpha_1 - \beta_1x_{1i}\right\}$ are different error terms than reported by Särndal, Swensson and Wretman (1992).

4.1 Repaired Low Level Calibrated Estimators of Variance

Following SHY methodology of Singh, Horn and Yu (1998), a low level calibrated estimator of variance of $\hat{Y}_c^{\oplus}$ is

$$\hat{V}_{LL}\left(\hat{Y}_c^{\oplus}\right) = \hat{v}_1 + \hat{v}_2 \text{ where } \hat{v}_1 = \frac{1}{2}\sum_{i\neq j\in s_2}\sum \pi_{1ij}W_{2ij}\left(\tilde{d}_i^{\oplus}\hat{e}_{2i}^* - \tilde{d}_j^{\oplus}\hat{e}_{2j}^*\right)^2,\ \ \hat{v}_2 = \frac{1}{2}\sum_{i\neq j\in s_2}\sum W_{1ij}\left(\tilde{d}_{1i}^{\oplus}\hat{e}_{1i}^* - \tilde{d}_{1j}^{\oplus}\hat{e}_{1j}^*\right)^2 \tag{4.1.1}$$

with $W_{2ij} = \left(\pi_{2i}\pi_{2j} - \pi_{2ij}\right)/\left(\pi_{1ij}\pi_{2ij}\right)$, and $W_{1ij} = \left(\pi_{1i}\pi_{1j} - \pi_{1ij}\right)/\left(\pi_{1ij}\pi_{2ij}\right)$. Note that $\hat{e}_{1i}^* = y_i - \hat{\alpha}_{2(\mathrm{ols})} - \hat{\beta}_{2(\mathrm{ols})}\left\{x_{2i} - \hat{\alpha}_{1(\mathrm{ols})} - \hat{\beta}_{1(\mathrm{ols})}x_{1i}\right\}$, and $\hat{e}_{2i}^* = y_i - \hat{\alpha}_{2(\mathrm{ols})} - \hat{\beta}_{2(\mathrm{ols})}x_{2i}$ are different estimators of the error terms than reported by Särndal, Swensson and Wretman (1992), because of use of ordinary least square estimators of the regression coefficients. Further note that $\hat{\alpha}_{2(\mathrm{ols})}$ and $\hat{\beta}_{2(\mathrm{ols})}$ are the ordinary least square estimates of α_2 and β_2 obtained by minimizing $\sum_{i\in s_2} q_{2i}^{\oplus}d_{1i}d_{2i}\hat{e}_{2i}^{*2}$; whereas $\hat{\alpha}_{1(\mathrm{ols})}$ and $\hat{\beta}_{1(\mathrm{ols})}$ are the ordinary least square estimators of α_1 and β_1 obtained by minimizing $\sum_{i\in s_2} d_{1i}\left(x_{2i} - \hat{\alpha}_{1(\mathrm{ols})} - \hat{\beta}_{1(\mathrm{ols})}x_{1i}\right)^2$.

4.2. Repaired Higher Level Calibrated Estimators of Variance

Again we repair the SHY methodology of Singh (2000) to form a new a higher order calibrated estimator of the variance in two-phase sampling as

$$\hat{V}_{\mathrm{HO}}\left(\hat{Y}_c^{\oplus}\right) = \frac{1}{2}\sum_{i\neq}\sum_{j\in s_2}\pi_{1ij}\Omega_{2ij}^{\oplus}\left(\tilde{d}_{2i}^{\oplus}\hat{e}_{2i}^* - \tilde{d}_{2j}^{\oplus}\hat{e}_{2j}^*\right)^2 + \frac{1}{2}\sum_{i\neq}\sum_{j\in s_2}\Omega_{1ij}^{\oplus}\left(\tilde{d}_{1i}^{\oplus}\hat{e}_{1i}^* - \tilde{d}_{1j}^{\oplus}\hat{e}_{1j}^*\right)^2 \tag{4.2.1}$$

where $\Omega_{1ij}^{\oplus}$ and $\Omega_{2ij}^{\oplus}$ are the weights such that the distance between $\Omega_{1ij}^{\oplus}$ and W_{1ij} and that between $\Omega_{2ij}^{\oplus}$ and W_{2ij} is minimum. Let us define two chi-square type distance functions as:

$$D_1^{\oplus} = \frac{1}{2}\sum_{i\neq j\in s_1}\sum\frac{\left(\Omega_{1ij}^{\oplus} - W_{1ij}\right)^2}{Q_{1ij}^{\oplus}W_{1ij}} \text{ and } D_2^{\oplus} = \frac{1}{2}\sum_{i\neq j\in s_2}\sum\frac{\left(\Omega_{2ij}^{\oplus} - W_{2ij}\right)^2}{Q_{2ij}^{\oplus}W_{2ij}} \tag{4.2.2}$$

The distance $D_1^{\oplus}$ is minimized subject to the two-calibration constraints defined as:

$$\sum_{i\neq j\in s_1}\sum\Omega_{1ij}^{\oplus} = \sum_{i\neq j\in s_1}\sum W_{1ij} \text{ and } \frac{1}{2}\sum_{i\neq j\in s_1}\sum\Omega_{1ij}^{\oplus}\left(\frac{x_{1i}}{\pi_{1i}} - \frac{x_{1j}}{\pi_{1j}}\right)^2 = V\left(\hat{X}_1\right) \quad \text{where } V\left(\hat{X}_1\right) = \frac{1}{2}\sum_{i\neq j\in\Omega}\sum\left(\pi_{1i}\pi_{1j} - \pi_{1ij}\right)\left(\frac{x_{1i}}{\pi_{1i}} - \frac{x_{1j}}{\pi_{1j}}\right)^2 \tag{4.2.3}$$

The distance $D_2^{\oplus}$ is minimized subject to the two-calibration constraints defined as:

$$\sum_{i\neq j\in s_2}\sum\Omega_{2ij}^{\oplus} = \sum_{i\neq j\in s_2}\sum W_{2ij} \text{ and } \frac{1}{2}\sum_{i\neq j\in s_2}\sum\Omega_{2ij}^{\oplus}\left(\frac{x_{2i}}{\pi_{2i}} - \frac{x_{2j}}{\pi_{2j}}\right)^2 = \hat{V}\left(\hat{X}_2\right) \quad \text{where } \hat{V}\left(\hat{X}_2\right) = \frac{1}{2}\sum_{i\neq j\in s_1}\sum\frac{\left(\pi_{1i}\pi_{1j} - \pi_{1ij}\right)}{\pi_{1ij}}\left(\frac{x_{2i}}{\pi_{1i}} - \frac{x_{2j}}{\pi_{2j}}\right)^2 \tag{4.2.4}$$

Note that the conditions (4.2.3) and (4.2.4) are the same conditions of chi-square test as suggested by Sir R.A. Fisher, and the higher level calibration estimator of variance of the repaired calibrated estimator in two-phase sampling becomes:

$$\hat{V}_{HO}\left(\hat{Y}_c^{\oplus}\right)=\hat{v}_1+\hat{B}_1^{\oplus}\left\{V\left(\hat{X}_1\right)-\hat{V}\left(\hat{X}_1\right)\right\}+\hat{v}_2+\hat{B}_2^{\oplus}\left\{\hat{V}\left(\hat{X}_2\right)-\hat{\hat{V}}\left(\hat{X}_2\right)\right\}$$

where

$$\hat{B}_1^{\oplus}=\frac{\sum\sum_{i\neq j\in s_2}\pi_{1ij}Q_{1ij}^{\oplus}W_{1ij}\left(d_{1i}x_{1i}-d_{1j}x_{1j}\right)^2\left(\tilde{d}_{2i}^{\oplus}\hat{e}_{2i}^{*}-\tilde{d}_{2j}^{\oplus}\hat{e}_{2j}^{*}\right)^2\sum\sum_{i\neq j\in s_1}Q_{1ij}^{\oplus}W_{1ij}-\sum\sum_{i\neq j\in s_1}Q_{1ij}^{\oplus}W_{1ij}\left(d_{1i}x_{1i}-d_{1j}x_{1j}\right)^2\sum\sum_{i\neq j\in s_2}Q_{1ij}^{\oplus}W_{1ij}\left(\tilde{d}_{2i}^{\oplus}\hat{e}_{2i}^{*}-\tilde{d}_{2j}^{\oplus}\hat{e}_{2j}^{*}\right)^2}{\left(\sum\sum_{i\neq j\in s_1}Q_{1ij}^{\oplus}W_{1ij}\right)\left(\sum\sum_{i\neq j\in s_1}Q_{1ij}^{\oplus}W_{1ij}\left(d_{1i}x_{1i}-d_{1j}x_{1j}\right)^4\right)-\left\{\sum\sum_{i\neq j\in s_1}Q_{1ij}^{\oplus}W_{1ij}\left(d_{1i}x_{1i}-d_{1j}x_{ij}\right)^2\right\}^2}$$

and

$$\hat{B}_2^{\oplus}=\frac{\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\left(d_{2i}x_{2i}-d_{2j}x_{2j}\right)^2\left(\tilde{d}_{1i}^{\oplus}\hat{e}_{1i}^{*}-\tilde{d}_{1j}^{\oplus}\hat{e}_{1j}^{*}\right)^2\left\{\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\right\}-\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\left(\tilde{d}_{1i}^{\oplus}\hat{e}_{1i}^{*}-\tilde{d}_{1j}^{\oplus}\hat{e}_{1j}^{*}\right)^2\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\left(d_{2i}x_{2i}-d_{2j}x_{2j}\right)^2}{\left\{\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\left(d_{2i}x_{2i}-d_{2j}x_{2j}\right)^2\right\}\left\{\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\right\}-\left\{\sum\sum_{i\neq j\in s_2}Q_{2ij}^{\oplus}W_{2ij}\left(d_{2i}x_{2i}-d_{2j}x_{2j}\right)^2\right\}^2}$$

and $\hat{\hat{V}}\left(\hat{X}_2\right)-\frac{1}{2}\sum\sum_{i\neq j\in s_2}W_{2ij}\left(\frac{x_{2i}}{\pi_{2i}}-\frac{x_{2j}}{\pi_{2j}}\right)^2$ If $q_{1i}^{\oplus}=q_{2i}^{\oplus}=Q_{1i}^{\oplus}=Q_{2i}^{\oplus}=1\ \forall\ i\&j$ then we consider chain regression type estimators of mean and variance. Note that all these result are analogous of Gupta, Singh and Mangat (1992-93), Singh (1991), and Upadhyaya, Singh and Singh (2004). **Caution!** Be careful while handling calibrated weights $\Omega_{1ij}^{\oplus}$.

5. FURTHER STUDY

Thus looking at new results the concern is that we have to check all the estimates we are producing at the US Bureau of Census, Statistics Canada, Australian Bureau of Statistics, and Statistics Sweden etc. from the last 12 years based on Deville and Särndal (1992) technique implemented especially in the statistical package GES developed by Statistics Canada and **sold** to different organizations across the world. Who knows that some of the results obtained using GES may be inaccurate? Those estimates based on the defective methodology may harm a lot of innocents in the world, and change the Govt. policies or thinking towards certain sensitive estimates such as drug used estimates, abortion estimates, and alcohol consumption estimates, and all other such estimates which have major social implications. In the same way, the other distance functions need also to be investigated with the new set of constraints considered in the present investigation. Similar repair of Estevao and Sarndal (2000, 2002), Lundstrom and Sarndal (1999), and Wu and Luan (2003) are obvious. It is not clear, but still there is possibility of improving the repaired two-phase calibration methodology of this paper by following the recently discovered model-assisted chi-square distance function by Singh (2004b).

ACKNOWLEGEMENTS

Thanks are due to all nears and dears who are reading my papers. All the results/opinions are author's own.

REFERENCES

Ahmed, M.S. (1997). The general class of chain estimators for the ratio of two means using double sampling. *CSTM.,* 26, 2247-2254.

Chand, L. (1975). *Some ratio-type estimators based on two or more auxiliary variables.* Ph.D. thesis, Iowa State University.

Deville, J.C. and Särndal, C.E.(1992). Calibration estimators in survey sampling. *JASA*, 87(418), 376-382.

Estevao, V.M. and Särndal, C.E. (2000). A functional form approach to calibration. *JOS*, 16(4), 379-399.

Estevao, V.M. and Särndal, C.E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *JOS*, 233-255.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in sample from an indefinitely large populations. *Biometrika*, 10, 507—521.

Fisher, R.A. (1922). On the interpretation of chi-square from contingency tables. *J. Roy. Statist. Soc.*, 85, 87—94.

Gupta, R.K, Singh, S. and Mangat, N.S. (1992-93). Some chain ratio type estimators for estimating finite population variance. *Aligarh J. Statist.*,12&13,65-69.

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

Hidiroglou, M. A. and Särndal, C.E. (1995). Use of auxiliary information for two-phase sampling. *JSM proceedings*, 873-878.

Hidiroglou, M. A. and Särndal, C.E. (1998). Use of auxiliary information for two-phase sampling. *SM*, 24, 11-20.

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

Kiregyera, B.(1980).A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika,* 217-223.

Lundstrom, S. and Sarndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. *JOS*, 305-327.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable, in the case of a correlated system of variables, is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, 50, 157—175.

Pearson, K. (1922). On the chi-square test of goodness of fit. *Biometrika*, 14, 186—418.

Prasad, B., Singh, R.S. and Singh, H.P. (1996). Some chain ratio-type estimators for ratio of two population means using two auxiliary characters in two-phase sampling. *Metron,* 95-113.

Sahoo, L.N. and Swain, A.K.P.C. (1983). Chain ratio estimators. *J. Indian Soc. Agril. Statist*., 35, 70-79.

Sahoo,L.N. and Swain, A.K.P.C. (1986). Chain product estimators. *Aligarh J. Statist.*, 6, 53-58.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model assisted survey sampling.* New York : Springer-Verlag.

Singh, G.N. and Upadhyaya, L.N. (1995). A class of modified chain-type estimators using two auxiliary variables in two-phase sampling. *Metron*, 117-125.

Singh, S. (1991). Estimation of finite population variance using double sampling. *Aligarh J. Statist*., 11, 53-65.

Singh, S.(2000). Estimation of variance of regression estimator in two-phase sampling. *Calcutta Statist. Assoc. Bull*., 50, 49-63.

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

Singh, S. (2004a). Survey statisticians celebrate Golden Jubilee Year –2003 of the linear regression estimator. *Metrika* (Accepted).

Singh, S. (2004b). Golden and Silver jubilee year-2003 of the linear regression estimators. *Presented at JSM-2004, Toronto (To appear on the CD)*.

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

Singh, V.K. and Singh, G.N. (1991). Chain type regression estimator with two auxiliary variables under double sampling scheme. *Metron,* 279-283.

Singh, V.K., Singh, H.P. and Singh, H.P. (1994). A general class of chain estimators for ratio and product of two means of a finite population. *CSTM*., 23(5), 1341-1355.

Srivastava, S.R., Khare, B.B. and Srivastava, S.R. (1990). A generalized chain ratio estimator for mean of finite population. *J. Indian Soc. Agril. Statist*.,42,108-117.

Tikkiwal, B.D. (1960). On the theory of classical regression and double sampling estimation. *J. R. Statist. Soc.*, B, 22, 131-138.

Upadhyaya, L.N., Khushwaha, K.S. and Singh, H.P. (1990). A modified chain ratio-type estimator in two-phase sampling using multi auxiliary information. *Metron*, 381-393.

Upadhyaya, L.N., Singh, H.P. and Singh, S. (2004). A class of estimators for estimating the variance of the ratio estimator**.** *J. Japan Statist. Soc.* ,34, 47--63.

Wu, C. and Luan, Y. (2003). Optimal calibration estimators under two-phase sampling. *J. Off. Statist.* 19(2), 119-131.