Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 71(2) 2017 95–102

Model Based Calibration Approach for Estimating Population Total in Successive Sampling

Nirupam Ghosh, U.C. Sud, Hukum Chandra and V.K. Gupta
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 25 July 2016; Revised 24 October 2016; Accepted 09 December 2016

SUMMARY

The conventional calibration approach is appropriate when study and auxiliary variables are linearly related. However, when study and auxiliary variables are non-linearly related model based calibration technique is appropriate. In this article two model based calibration estimators along with their variances and estimator of variances in two occasion successive sampling are proposed. The performance of the proposed estimators are studied via a simulation study vis-à-vis design based calibration estimator and an estimator which doesn't consider auxiliary information at estimation stage.

Keywords: Calibration estimator, Generalized linear model, Successive sampling, Superpopulation.

1. INTRODUCTION

In sample surveys, auxiliary information on the finite population is often used to increase the precision of estimators of finite population total or mean or distribution function. In the simplest settings, ratio and regression estimators incorporate known finite population means of auxiliary variables. The calibration approach is one of the approaches being used for building estimators based on auxiliary information. The calibration approach focuses on the weights given to the units for the purpose of estimation. Calibration implies that a set of starting weights (usually the sampling design weights) are transformed into a set of new weights, called calibrated weights. The calibrated weight of a unit is the product of its initial weight and a calibration factor. The calibration factors are obtained by minimizing a function measuring the distance between the initial weights and the calibrated weights, subject to the constraint that the calibrated weights yield exact estimates of the known auxiliary population totals. This population total is estimated by a linear estimator whose weights are as close as possible to some benchmark weights and which at the same time satisfy some calibration constraints with respect to some suitable auxiliary variables.

Corresponding author: Nirupam Ghosh
E-mail address: nirugh4846@gmail.com

Deville and Sarndal (1992) formulated the calibration approach to estimate the finite population parameters as (a) a computation of weights that incorporate specified auxiliary information and are restrained by calibration equation(s), (b) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units**,** (c) an objective to obtain nearly design unbiased estimates. It can be said that this procedure adjusts the sampling weights by multipliers known as calibration factors that make the estimates agree with known totals. The resulting weights are called calibration weights or final estimation weights. These calibration weights result in estimates that are design consistent and that have a smaller variance than the Horvitz -Thompson estimator. We have used the notations of Deville and Sarndal (1992) for calibration in sampling, Rueda *et al.* (2009) for calibration in successive sampling and Wu and Sitter (2001) for model based calibration approach in successive sampling. In this paper an attempt is made to estimate the population total using model based calibration approach when study and auxiliary variable are non-linearly related. In the next Section, we introduce the general notations

used under successive sampling design. In Section 3, we discuss design based calibration estimation of population total under successive sampling design. Section 4 presents model based calibration estimation of population total under successive sampling design. In Section 5, simulation results of the developed estimators are provided. Finally, Section 6 presents concluding remarks.

2. GENERAL NOTATIONS UNDER SUCCESSIVE SAMPLING

Consider sampling on two occasions from a finite population $U=\{1,2,...,k,...,N\}$ of size N. We assume that population retains its composition over two time periods but values of units change over occasions. The study variable is observed on two occasions but not necessarily for the same set of elements.

A sample s' of size n' is drawn on the first occasion according to a sampling design d_1, such that $p_{d_1}(s')$ is the probability that s' is chosen. The respective first and second order inclusion probabilities for unit k and pair of units k, l in the sample s' of size n' associated with the design $p_{d_1}(s')$ are $\pi_k(\forall\, k\in U)$ and $\pi_{kl}(\forall\, k\neq l\in U)$. So the sampling weight for the k^{th} unit, $a_{1k}=1/\pi_k$. To the sample s' drawn at the first occasion corresponds a complementary sample, $s'^c=U-s'$ containing all those units of U not surveyed on the first occasion. Let $\pi_k^c(\forall\, k\in U)$ and $\pi_{kl}^c(\forall\, k\neq l\in U)$ be the first and second inclusion probabilities, respectively for unit k and pair of units k, l induced by the sampling design $p_{d_1}(s'^c)$. Here we denote the sampling weight for k^{th} unit into sample s'^c by, $a_{1k}^c=1/\pi_k^c$. For the second occasion, a sample s_m of size m, called matched sample, is drawn from s' with a design d_2, such that $p_{d_2}(s_m\mid s')$ is the conditional probability of choosing s_m. The first and second order inclusion probabilities induced by the sampling design $p_{d_2}(s_m\mid s')$ are denoted by $\pi_{k|s'}(\forall\, k\in s')$ and $\pi_{kl|s'}(\forall\, k\neq l\in s')$, respectively. From the sample s'^c another sample s_u of size u, called unmatched sample, is drawn with a design d_3, such that $p_{d_3}(s_u\mid s'^c)$ is the conditional probability of choosing s_u. The first order and second order inclusion probabilities under the design $p_{d_3}(s_u\mid s'^c)$ are respectively $\pi_{k|s'^c}(\forall\, k\in s'^c)$ and $\pi_{kl|s'^c}(\forall\, k\neq l\in s'^c)$. The overall sampling weights for the selected k^{th} unit in the matched sample s_m will be $a_k=1/\pi_k\pi_{k|s'}$, and that for the unmatched sample $a_{2k}=1/\pi_k^c\pi_{k|s'^c}$.

We consider the auxiliary variable at population level denoted by z. The value of z for the unit k is denoted by z_k. The variables x and y denote the first occasion and second occasion observations, respectively and for k^{th} unit the values are x_k and y_k, respectively.

3. DESIGN BASED CALIBRATION ESTIMATION IN SUCCESSIVE SAMPLING

The purpose of this section is to modify the design weights for different samples using calibration approach and produces calibrated weights $\{w_k: k\in s_m\}$ and $\{w_{2k}: k\in s_u\}$ respectively for design weights a_k and a_{2k} corresponding to matched and unmatched samples respectively. Consider the auxiliary variable at population level denoted by z. The value of z for the unit k is denoted by z_k. The variables x and y denote the first occasion and second occasion observations, respectively and for k^{th} unit the values are x_k and y_k, respectively.

In this case, for the matched sample the weight w_k is obtained by following a two-step procedure. In the first step determine the weight w_{1k} as solution to the weighted least square minimization problem Min $\sum_{s'}(w_{1k}-a_{1k})^2/q_{1k}a_{1k}$ subject to the calibration equation $\sum_{s'}w_{1k}z_k=\sum_U z_k$ and in the second step determine the weight w_k as solution to the weighted least square minimization problem Min $\sum_{s_m}(w_k-a_k)^2/q_ka_k$ subject to the calibration equation $\sum_{s_m}w_kx_k=\sum_{s'}w_{1k}x_k$. For the unmatched sample we obtain the weight w_{2k} by following a single-step procedure. Determine the weight w_{2k} as solution to the weighted least square minimization problem Min $\sum_{s_u}(w_{2k}-a_{2k})^2/q_{2k}a_{2k}$ subject to the calibration equation $\sum_{s_u}w_{2k}z_k=\sum_U z_k$. Here we consider $q_{1k}=q_k=q_{2k}=1$. The estimators constructed with the derived weights are given by.

$$\hat{Y}_{m,cal}=\sum_{s_m}a_ky_k+\left(\sum_U z_k-\sum_{s'}a_{1k}z_k\right)\hat{B}_{(xB:z)}+\left(\sum_{s'}a_{1k}x_k-\sum_{s_m}a_kx_k\right)\hat{B}_{(y:x)} \tag{3.1}$$

$$\hat{Y}_{u,cal}=\sum_{s_u}a_{2k}y_k+\left(\sum_U z_k-\sum_{s_u}a_{2k}z_k\right)\hat{B}_{(y:z)} \tag{3.2}$$

respectively for matched and unmatched sample.

Where,

$$\hat{B}_{(xB:z)} = \frac{\left(\sum_{s'} a_{1k} z_k x_k\right)}{\left(\sum_{s'} a_{1k} z_k^2\right)} \frac{\left(\sum_{s_m} a_k x_k y_k\right)}{\left(\sum_{s_m} a_k x_k^2\right)},$$

$$\hat{B}_{(y:x)} = \frac{\left(\sum_{s_m} a_k x_k y_k\right)}{\left(\sum_{s_m} a_k x_k^2\right)} \text{ and } \hat{B}_{(y:z)} = \frac{\left(\sum_{s_u} a_{2k} z_k y_k\right)}{\left(\sum_{s_u} a_{2k} z_k^2\right)}.$$

Finally, we consider the following combined estimator,

$$\hat{Y}_{cal} = \alpha_1 \hat{Y}_{m,cal} + \alpha_2 \hat{Y}_{u,cal} \tag{3.3}$$

where α_1 and α_2 are non-negative constant weights to be determined such that $\alpha_1 + \alpha_2 = 1$.

4. MODEL BASED CALIBRATION APPROACH IN SUCCESSIVE SAMPLING

This section is devoted to obtain calibration estimators when there is a non-linear relationship between study and auxiliary variable. Assume that the relationship between study variable and auxiliary variable can be described by two superpopulation models for two occasions through the first and second moments,

$$E_\xi(y_{hk}/z_k) = \mu_h(z_k, \theta_h) = \mu_{hk}; \quad V_\xi(y_{hk}/z_k) = v_{hk}\sigma_h^2; \quad h = 1,2;\ k = 1,2,\ldots, N \tag{4.1}$$

where θ_h and σ_h^2 are unknown super population parameters, $\mu_h(z, \theta_h)$ is a known function of z and θ_h, the v_h is a known function of z. E_ξ and V_ξ denote the expectation and variance with respect to the superpopulation model, where h denotes the number of occasion. Note that the first occasion sample and second occasion matched sample are linearly related as they are taken on the same units. The first occasion population is denoted by x and for the second occasion as y.

Under the models (4.1), auxiliary information should be used through the fitted values $\mu_h(z_k, \hat{\theta}_h)$; $h = 1, 2, k = 1, 2\ldots, N$. To do this we define the calibration estimator for both matched and unmatched samples and then combine those estimators to obtain the desired estimator. For matched sample in the first step we obtain the weights w_{1k} as minimizing the distance function in $\sum_{s'}(w_{1k} - a_{1k})^2/q_{1k}a_{1k}$ subject to the constraints, $\sum_{s'} w_{1k} = N$ and $\sum_{s'} w_{1k}\mu_1(z_k, \hat{\theta}_1) = \sum_U \mu_1(z_k, \hat{\theta}_1)$. One should note that in the original formulation of calibration estimator, the constraint $\sum_{s'} w_{1k} = N$ is not present. If this constraint is added, the resulting estimator in no auxiliary information is, $\hat{Y} = \sum_s a_k y_k / \sum_s a_k$ and not $\hat{Y}_{HT} = \sum_s a_k y_k$. It was illustrated in Rao (1966) and later in the more well-known Basu (1971) elephant example that even though the first estimator estimates the population size N and the second uses its known quantity, the first has the better properties. This is true for calibration generally. In the second step, we obtain the weights w_k as minimizing the distance function given in $\sum_{s_m}(w_k - a_k)^2/q_k a_k$ subject to the constraint $\sum_{s_m} w_k x_k = \sum_{s'} w_{1k} x_k$. For unmatched sample we obtain the weights w_{2k} as minimizing the distance function given in $\sum_{s_u}(w_{2k} - a_{2k})^2/q_{2k}a_{2k}$ subject to the constraints $\sum_{s_u} w_{2k} = N$ and $\sum_{s_u} w_{2k}\mu_2(z_k, \hat{\theta}_2) = \sum_U \mu_2(z_k, \hat{\theta}_2)$. Here we consider $q_{1k} = q_k = q_{2k} = 1$. The estimators constructed with the derived weights are given by

$$\hat{Y}_{m,cal} = \sum_{s_m} a_k y_k + \left(\sum_U \hat{\mu}_{1k} - \sum_{s'} a_{1k}\hat{\mu}_{1k}\right)\hat{B}_{(xB:\hat{\mu}_1)} + \left(\sum_{s'} a_{1k} x_k - \sum_{s_m} a_k x_k\right)\hat{B}_{(y:x)} \tag{4.2}$$

$$\hat{Y}_{u,cal} = \sum_{s_u} a_{2k} y_k + \left(\sum_U \hat{\mu}_{2k} - \sum_{s_u} a_{2k}\hat{\mu}_{2k}\right)\hat{B}_{(y:\hat{\mu}_2)} \tag{4.3}$$

where,

$$\hat{B}_{(xB:\hat{\mu}_1)} = \frac{\sum_{s'} a_{1k}(\hat{\mu}_{1k} - \bar{\mu}_1)(x_k - \bar{x})}{\sum_{s'} a_{1k}(\hat{\mu}_{1k} - \bar{\mu}_1)^2} \frac{\sum_{s_m} a_k x_k y_k}{\sum_{s_m} a_k x_k^2},$$

$$\hat{B}_{(y:\hat{\mu}_2)} = \frac{\sum_{s_u} a_{2k}(\hat{\mu}_{2k} - \bar{\mu}_2)(y_k - \bar{y})}{\sum_{s_u} a_{2k}(\hat{\mu}_{2k} - \bar{\mu}_2)^2}, \text{ and } \hat{B}_{(y:x)} \text{ are}$$

defined earlier.

If the constraint $\sum_{s'} w_{1k} = N$ is dropped, then we have the calibration estimators as follows,

$$\hat{Y}^*_{m,cal} = \sum_{s_m} a_k y_k + \left(\sum_U \hat{\mu}_{1k} - \sum_{s'} a_{1k}\hat{\mu}_{1k}\right)\hat{B}^*_{(xB:\hat{\mu}_1)} + \left(\sum_{s'} a_{1k} x_k - \sum_{s_m} a_k x_k\right)\hat{B}_{(y:x)} \tag{4.4}$$

$$\hat{Y}^*_{u,cal} = \sum_{s_u} a_{2k} y_k + \left(\sum_U \hat{\mu}_{2k} - \sum_{s_u} a_{2k}\hat{\mu}_{2k}\right)\hat{B}^*_{(y:\bar{\mu}_2)} \tag{4.5}$$

where,

$$\hat{B}^{*}_{(xB:\hat{\mu}_1)} = \frac{\sum_{s'} a_{1k}\hat{\mu}_{1k}x_k}{\sum_{s'} a_{1k}\hat{\mu}^2_{1k}} \frac{\sum_{s_m} a_k x_k y_k}{\sum_{s_m} a_k x_k^2},$$

$$\hat{B}^{*}_{(y:\hat{\mu}_2)} = \frac{\sum_{s_u} a_{2k}\hat{\mu}_{2k}y_k}{\sum_{s_u} a_{2k}\hat{\mu}^2_{2k}}, \text{ and } \hat{B}_{(y:x)} \text{ are defined earlier}$$

Finally we consider the following combined estimators,

$$\hat{Y}_{MC,A} = \beta_1\hat{Y}_{m,cal} + \beta_2\hat{Y}_{u,cal} \tag{4.6}$$

where β_1 and β_2 are non-negative constant weights to be determined such that $\beta_1 + \beta_2 = 1$, and

$$\hat{Y}_{MC,B} = \gamma_1\hat{Y}^{*}_{m,cal} + \gamma_2\hat{Y}^{*}_{u,cal} \tag{4.7}$$

where γ_1 and γ_2 are non-negative constant weights to be determined such that $\gamma_1 + \gamma_2 = 1$.

4.1 Asymptotic Properties of the Model Based Calibration Estimators

In this section, we establish the asymptotic behavior of the estimators proposed in the previous sections. The design based calibration estimators in (3.1) and (3.2) are asymptotically unbiased, and their asymptotic variances and covariance between them are given by,

$$V\left(\hat{Y}_{m,cal}\right) = \sum\sum_{U}\Delta_{kl}\frac{e_{1k}}{\pi_k}\frac{e_{1l}}{\pi_l} + E_1\left(\sum\sum_{s'}\Delta_{kl|s'}\frac{e_k}{\pi_k\pi_{k|s'}}\frac{e_l}{\pi_l\pi_{l|s'}}\right) \tag{4.1.1}$$

$$V\left(\hat{Y}_{u,cal}\right) = \sum\sum_{U}\Delta^{c}_{kl}\frac{e_{2k}}{\pi^{c}_k}\frac{e_{2l}}{\pi^{c}_l} + E_1\left(\sum\sum_{s^{rc}}\Delta_{kl|s^{rc}}\frac{e_{2k}}{\pi^{c}_k\pi_{k|s^{rc}}}\frac{e_{2l}}{\pi^{c}_l\pi_{l|s^{rc}}}\right) \tag{4.1.2}$$

$$C\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right) = -\sum\sum_{U}\Delta_{kl}\frac{e_{1k}}{\pi_k}\frac{e_{2l}}{\pi^{c}_l} \tag{4.1.3}$$

respectively. With $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$, $\Delta_{kl|s'} = \pi_{kl|s'} - \pi_{k|s'}\pi_{l|s'}$, $\Delta^{c}_{kl} = \Delta_{kl}$, $\Delta_{kl|s^{rc}} = \pi_{kl|s^{rc}} - \pi_{k|s^{rc}}\pi_{l|s^{rc}}$, $e_k = y_k - B_{(y:x)}x_k$, $B_{(y:x)} = \frac{\left(\sum_U x_k y_k\right)}{\left(\sum_U x_k^2\right)}$, $e_{1k} = y_k - B_{(xB:z)}z_k$, $e_{2k} = y_k - B_{(y:z)}z_k$, $B_{(xB:z)} = \frac{\sum_U z_k x_k}{\sum_U z_k^2}\frac{\sum_U x_k y_k}{\sum_U x_k^2}$, $B_{(y:z)} = \frac{\sum_U z_k y_k}{\sum_U y_k^2}$.

Here, E_1 denotes the expectation value with respect to the first occasion design d_1. The estimators of the variances and covariance are given by,

$$\hat{V}\left(\hat{Y}_{m,cal}\right) = \sum\sum_{s_m}\frac{\Delta_{kl}}{\pi_{kl}\pi_{kl|s'}}\frac{\hat{e}_{1k}}{\pi_k}\frac{\hat{e}_{1l}}{\pi_l} + \sum\sum_{s_m}\frac{\Delta_{kl|s'}}{\pi_{kl|s'}}\frac{\hat{e}_k}{\pi_k\pi_{k|s'}}\frac{\hat{e}_l}{\pi_l\pi_{l|s'}} \tag{4.1.4}$$

$$\hat{V}\left(\hat{Y}_{u,cal}\right) = \sum\sum_{s_u}\frac{\Delta^{c}_{kl}}{\pi^{c}_{kl}\pi_{kl|s^{rc}}}\frac{\hat{e}_{2k}}{\pi^{c}_k}\frac{\hat{e}_{2l}}{\pi^{c}_l} + \sum\sum_{s_u}\frac{\Delta_{kl|s^{rc}}}{\pi_{kl|s^{rc}}}\frac{\hat{e}_{2k}}{\pi^{c}_k\pi_{k|s^{rc}}}\frac{\hat{e}_{2l}}{\pi^{c}_l\pi_{l|s^{rc}}} \tag{4.1.5}$$

$$\hat{C}\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right) = -\sum\sum_{s_m}\frac{\Delta_{kl}}{\pi_{kl}\pi_{kl|s'}}\frac{\hat{e}_{1k}}{\pi_k}\frac{\hat{e}_{2l}}{\pi^{c}_l} \tag{4.1.6}$$

with $\hat{e}_k = y_k - \hat{B}_{(y:x)}x_k$, $\hat{e}_{1k} = y_k - \hat{B}_{(xB:z)}z_k$ and $\hat{e}_{2k} = y_k - \hat{B}_{(y:z)}z_k$ where $\hat{B}_{(y:x)}$, $\hat{B}_{(xB:z)}$ and $\hat{B}_{(y:z)}$ are defined earlier.

Denoting the values $V\left(\hat{Y}_{m,cal}\right)$, $V\left(\hat{Y}_{u,cal}\right)$ and $C\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right)$ by V_1, V_2 and C respectively, the variance of (3.3) can be expressed as,

$$\begin{aligned} V\left(\hat{Y}_{cal}\right) &= V\left(\alpha_1\hat{Y}_{m,cal} + \alpha_2\hat{Y}_{u,cal}\right) \\ &= \left[V_1 + V_2 - 2C\right]\left[\alpha_1 - \frac{V_2 - C}{V_1 + V_2 - 2C}\right]^2 + \frac{V_1V_2 - C^2}{V_1 + V_2 - 2C} \\ &\geq \frac{V_1V_2 - C^2}{V_1 + V_2 - 2C} = V_{\min}\left(\hat{Y}_{cal}\right) \end{aligned} \tag{4.1.7}$$

Because $V_1 + V_2 - 2C = V\left(\hat{Y}_{cal} - Y\right) \geq 0$, equality holds if and only if

$$\alpha_1 = 1 - \alpha_2 = \frac{V_2 - C}{V_1 + V_2 - 2C} \tag{4.1.8}$$

This value can be estimated by using the estimators given by (4.1.4), (4.1.5) and (4.1.6).

The model based calibration estimators (4.2) and (4.3) are asymptotically unbiased, and their asymptotic variances and covariance between them are given by,

$$V\left(\hat{Y}_{m,cal}\right) = \sum\sum_{U}\Delta_{kl}\frac{e_{1k}}{\pi_k}\frac{e_{1l}}{\pi_l} + E_1\left(\sum\sum_{s'}\Delta_{kl|s'}\frac{e_k}{\pi_k\pi_{k|s'}}\frac{e_l}{\pi_l\pi_{l|s'}}\right) \tag{4.1.9}$$

$$V\left(\hat{Y}_{u,cal}\right) = \sum\sum_{U}\Delta^{c}_{kl}\frac{e_{2k}}{\pi^{c}_k}\frac{e_{2l}}{\pi^{c}_l} + E_1\left(\sum\sum_{s^{rc}}\Delta_{kl|s^{rc}}\frac{e_{2k}}{\pi^{c}_k\pi_{k|s^{rc}}}\frac{e_{2l}}{\pi^{c}_l\pi_{l|s^{rc}}}\right) \tag{4.1.10}$$

$$C\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right)=-\sum\sum_{U}\Delta_{kl}\frac{e_{1k}}{\pi_k}\frac{e_{2l}}{\pi_l^c} \qquad (4.1.11)$$

with $\Delta_{kl}=\pi_{kl}-\pi_k\pi_l$, $\Delta_{kl|s'}=\pi_{kl|s'}-\pi_{k|s'}\pi_{l|s'}$, $\Delta_{kl}^c=\Delta_{kl}$, $\Delta_{kl|s'^c}=\pi_{kl|s'^c}-\pi_{k|s'^c}\pi_{l|s'^c}$, $e_k=y_k-B_{(y:x)}x_k$, $B_{(y:x)}=\dfrac{\left(\sum_U x_k y_k\right)}{\left(\sum_U x_k^2\right)}$, $e_{1k}=y_k-B_{(xB:\mu_1)}\mu_{1k}$, $B_{(xB:\mu_1)}=\dfrac{\sum_U(\mu_{1k}-\bar{\mu}_1)(x_k-\bar{x})}{\sum_U(\mu_{1k}-\bar{\mu}_1)^2}\dfrac{\sum_U x_k y_k}{\sum_U x_k^2}$, $e_{2k}=y_k-B_{(y:\mu_2)}\mu_{2k}$, $B_{(y:\mu_2)}=\dfrac{\sum_U(\mu_{2k}-\bar{\mu}_2)(y_k-\bar{y})}{\sum_U(\mu_{2k}-\bar{\mu}_2)^2}$.

Here, E_1 denotes the expectation value with respect to the first occasion design d_1. The estimators of variances and covariance are given by,

$$\hat{V}\left(\hat{Y}_{m,cal}\right)=\sum\sum_{s_m}\frac{\Delta_{kl}}{\pi_{kl}\pi_{kl|s'}}\frac{\hat{e}_{1k}}{\pi_k}\frac{\hat{e}_{1l}}{\pi_l}+\sum\sum_{s_m}\frac{\Delta_{kl|s'}}{\pi_{kl|s'}}\frac{\hat{e}_k}{\pi_k\pi_{k|s'}}\frac{\hat{e}_l}{\pi_l\pi_{l|s'}} \qquad (4.1.12)$$

$$\hat{V}\left(\hat{Y}_{u,cal}\right)=\sum\sum_{s_u}\frac{\Delta_{kl}^c}{\pi_{kl}^c\pi_{kl|s'^c}}\frac{\hat{e}_{2k}}{\pi_k^c}\frac{\hat{e}_{2l}}{\pi_l^c}+\sum\sum_{s_u}\frac{\Delta_{kl|s'^c}}{\pi_{kl|s'^c}}\frac{\hat{e}_{2k}}{\pi_k^c\pi_{k|s'^c}}\frac{\hat{e}_{2l}}{\pi_l^c\pi_{l|s'^c}} \qquad (4.1.13)$$

$$\hat{C}\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right)=-\sum\sum_{s_m}\frac{\Delta_{kl}}{\pi_{kl}\pi_{kl|s'}}\frac{\hat{e}_{1k}}{\pi_k}\frac{\hat{e}_{2l}}{\pi_l^c} \qquad (4.1.14)$$

with $\hat{e}_{1k}=y_k-\hat{B}_{(xB:\hat{\mu}_1)}\hat{\mu}_{1k}$, $\hat{e}_k=y_k-\hat{B}_{(y:x)}x_k$, $\hat{e}_{2k}=y_k-\hat{B}_{(y:\hat{\mu}_2)}\hat{\mu}_{2k}$ and $\hat{B}_{(xB:\hat{\mu}_1)}$, $\hat{B}_{(y:x)}$, $\hat{B}_{(y:\hat{\mu}_2)}$ are defined earlier. Similar results hold for $\hat{Y}_{MC,B}$ if we replace $B_{(xB:\mu_1)}$ by $B^*_{(xB:\mu_1)}$, $\hat{B}_{(xB:\hat{\mu}_1)}$ by $\hat{B}^*_{(xB:\hat{\mu}_1)}$, $B_{(y:\mu_2)}$ by $B^*_{(y:\mu_2)}$ and $\hat{B}_{(y:\hat{\mu}_2)}$ by $\hat{B}^*_{(y:\hat{\mu}_2)}$.

The variance of the two model based calibration estimators (4.6) and (4.7) can be easily obtained in the similar way as given by (4.1.7) and (4.1.8).

Remark 4.1: Let us consider the design used in a simulation study. The design consists of drawing samples using SRSWOR at each occasion. Suppose that the first occasion sample s' is drawn from the population U with simple random sampling without replacement (SRSWOR) with sample size n'. Thus

$$\pi_k'=\frac{n'}{N};\ \pi_{kl}'=\frac{n'(n'-1)}{N(N-1)}.$$

Also the complementary sample $s'^c=U-s'$ is a simple random sample without replacement of size $N-n'$. Then

$$\pi_k'^c=\frac{N-n'}{N};\ \pi_{kl}'^c=\frac{(N-n')(N-n'-1)}{N(N-1)}.$$

Suppose the matched sample s_m is drawn from s' with SRSWOR of size m and so

$$\pi_{k|s'}=\frac{m}{n'};\ \pi_{kl|s'}=\frac{m(m-1)}{n'(n'-1)}.$$

Finally, the unmatched sample s_u is drawn from s'^c with SRSWOR of size u. Thus

$$\pi_{k|s'^c}=\frac{u}{N-n'};\ \pi_{kl|s'^c}=\frac{u(u-1)}{(N-n')(N-n'-1)}.$$

It is easy to show that for this design the approximate variances and covariance for $\hat{Y}_{cal}$ are given as follows,

$$V_1=N^2\left[\left(\frac{1}{m}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n'}-\frac{1}{N}\right)\left\{B_{(xB:z)}^2S_z^2-2B_{(xB:z)}S_{yz}\right\}+\left(\frac{1}{m}-\frac{1}{n'}\right)\left\{B_{(y:x)}^2S_x^2-2B_{(y:x)}S_{yx}\right\}\right]$$

$$V_2=N^2\left(\frac{1}{u}-\frac{1}{N}\right)\left[S_y^2+B_{(y:z)}^2S_z^2-2B_{(y:z)}S_{yz}\right]$$

$$C=-N\left[S_y^2+B_{(xB:z)}B_{(y:z)}S_z^2-\left\{B_{(xB:z)}+B_{(y:z)}\right\}S_{yz}\right]$$

with $S_y^2=\sum_U\frac{(y_k-\bar{y})^2}{N-1}$, $S_x^2=\sum_U\frac{(x_k-\bar{x})^2}{N-1}$, $S_z^2=\sum_U\frac{(z_k-\bar{z})^2}{N-1}$, $S_{yx}=\sum_U\frac{(y_k-\bar{y})(x_k-\bar{x})}{N-1}$, $S_{yz}=\sum_U\frac{(y_k-\bar{y})(z_k-\bar{z})}{N-1}$.

For this design the approximate variances and covariance for $\hat{Y}_{MC,A}$ are given as follows,

$$V\left(\hat{Y}_{m,cal}\right)=N^2\left[\left(\frac{1}{m}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{n'}-\frac{1}{N}\right)\left\{B_{(xB:\mu_1)}^2S_{\mu_1}^2-2B_{(xB:\mu_1)}S_{y\mu_1}\right\}+\left(\frac{1}{m}-\frac{1}{n'}\right)\left\{B_{(y:x)}^2S_x^2-2B_{(y:x)}S_{yx}\right\}\right]$$

$$V\left(\hat{Y}_{u,cal}\right)=N^2\left(\frac{1}{u}-\frac{1}{N}\right)\left[S_y^2+B_{(y:\mu_2)}^2S_{\mu_2}^2-2B_{(y:\mu_2)}S_{y\mu_2}\right]$$

$$C\left(\hat{Y}_{m,cal},\hat{Y}_{u,cal}\right)=-N\Big[S_y^2+B_{(xB:\mu_1)}B_{(y:\mu_2)}S_{\mu_1\mu_2} -B_{(xB:\mu_1)}S_{y\mu_1}-B_{(y:\mu_2)}S_{y\mu_2}\Big]$$

with $S_{\mu_1}^2=\sum_U\frac{(\mu_{1k}-\bar{\mu}_1)^2}{N-1}$, $S_{\mu_2}^2=\sum_U\frac{(\mu_{2k}-\bar{\mu}_2)^2}{N-1}$,

$S_{y\mu_1}=\sum_U\frac{(y_k-\bar{y})(\mu_{1k}-\bar{\mu}_1)}{N-1}$, $S_{y\mu_2}=\sum_U\frac{(y_k-\bar{y})(\mu_{2k}-\bar{\mu}_2)}{N-1}$,

$S_{\mu_1\mu_2}=\sum_U\frac{(\mu_{1k}-\bar{\mu}_1)(\mu_{2k}-\bar{\mu}_2)}{N-1}$.

Similar results hold for $\hat{Y}_{MC,B}$ if we replace $B_{(xB:\mu_1)}$ by $B^*_{(xB:\mu_1)}$ and $B_{(y:\mu_2)}$ by $B^*_{(y:\mu_2)}$.

5. SIMULATION RESULTS

In this section, a simulation study has been performed to reveal the behavior of the proposed estimators. Our intention is to point out the most efficient estimator. Let us consider the behavior of the proposed estimators with baseline estimator as Horvitz-Thompson estimator in successive sampling.

For this, an artificial population is created as follows: For each unit k, we generate Z_k~Gamma(1,1) and $\varepsilon_{1k}\sim\text{Normal}\left(0,\sigma_{e_1}^2\right)$; $k=1,2,\ldots,1000$. Then for the first occasion a finite population consisting of N=1000 was generated as an iid sample from $\log(x)=\theta_0+\theta_1 z+\varepsilon_{1k}$. We choose $\theta_0=\theta_1=1$. Then for the second occasion we computed $\varepsilon_{2k}=\varepsilon_{1k}+\delta_k$, k=1,2,…,1000 where $\delta_k:\text{Normal}\left(0,\sigma_\delta^2\right)$ and again a finite population consisting of N=1000 was generated as an iid sample from $\log(y)=\theta_0+\theta_1 z+\varepsilon_{2k}$, where $\theta_0=\theta_1=1$. Five different finite populations were generated for different values of $\sigma_{e_1}^2$, σ_δ^2 and $\sigma_{e_3}^2$ we get different correlation 0.9, 0.8, 0.7, 0.6 and 0.5 respectively between log(y), log(x) and z.

For each finite population, a simple random sample of size 300 was taken from population size 1000 at the first occasion and a log-linear model $\log(\mu_{1k})=\alpha_1+\beta_1 z_k$, $V(\mu_1)=\nu_1^2$ was fitted using maximum likelihood estimation. First occasion calibrated weights were computed using the sample values and all the fitted values. For the second occasion matched sample a sample of sizes 20, 50, 80 respectively were selected from first occasion sample and design based calibration approach was used to derive the calibrated weights. For the second occasion unmatched sample, simple random samples of sizes 20, 50, 80 respectively were taken from remaining population of size 700 and for each sample a log-linear model $\log(\mu_{2k})=\alpha_2+\beta_2 z_k$, $V(\mu_2)=\nu_2^2$ was fitted using maximum likelihood estimation. Therefore we have the three combinations of second occasion samples as (20, 20), (50, 50) and (80, 80) respectively. This total process was repeated for 10000 times. The following figure shows the relationship between study and auxiliary variable in the population.

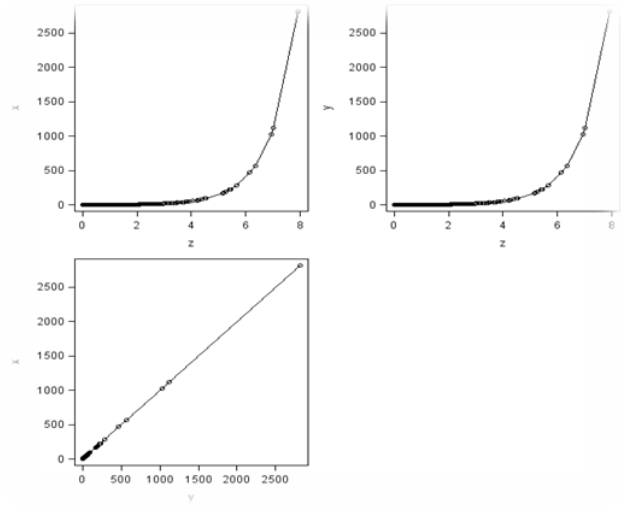

Fig. 5.1. Relationship between study and auxiliary variables

Fig. 5.1 shows the relationships between study variable and auxiliary variable. It can be seen that study variable on both the occasion x and y respectively are non-linearly related with the auxiliary variable z. But relationship between the study variables on two occasion are linear. To compare the estimators we compute Percentage Relative Efficiency (PRE). The results are shown in Tables (5.1) and (5.2). The expression for Percentage Relative Efficiency is as follows:

$$PRE=\frac{MSE(\cdot)}{MSE(\hat{\theta})}\times 100$$

where i indexes the i^{th} simulation run. $MSE(\hat{\theta})=\frac{1}{M}\sum_{i=1}^{M}(\hat{\theta}_i-\theta)^2$, defines the Mean Square Error of the proposed calibration estimators and $MSE(\cdot)$ similarly defines the Mean Square Error for the estimator which cannot use auxiliary information (Horvitz-Thompson estimator).

Some noteworthy results of the tables are as follows:

Table 5.1. Percentage Relative Efficiency with Constant Correlation between First Occasion Population and Auxiliary Variable

ρ_{xz}	ρ_{yz}	ρ_{yx}	s_m	s_u	$\hat{Y}_{MC,A}$	$\hat{Y}_{MC,B}$	$\hat{Y}_{cal}$	$\hat{Y}_{HT}$
0.7	0.5	0.5	20	20	145.40	124.58	117.09	100.00
			50	50	151.03	131.48	119.60	100.00
			80	80	175.94	162.23	121.99	100.00
	0.6	0.6	20	20	168.10	139.50	123.90	100.00
			50	50	175.07	148.73	130.01	100.00
			80	80	217.53	172.69	136.48	100.00
	0.7	0.7	20	20	172.15	139.83	127.76	100.00
			50	50	188.41	151.44	138.69	100.00
			80	80	210.50	201.17	151.92	100.00
	0.8	0.8	20	20	201.33	167.58	131.79	100.00
			50	50	275.36	223.41	151.11	100.00
			80	80	290.82	245.03	180.24	100.00
	0.9	0.9	20	20	262.74	168.87	134.92	100.00
			50	50	277.98	234.92	167.74	100.00
			80	80	292.81	274.29	238.14	100.00

Table 5.2. Percentage Relative Efficiency with Constant Correlation between Second Occasion Population and Auxiliary Variable

ρ_{yz}	ρ_{yx}	ρ_{xz}	s_m	s_u	$\hat{Y}_{MC,A}$	$\hat{Y}_{MC,B}$	$\hat{Y}_{cal}$	$\hat{Y}_{HT}$
0.7	0.7	0.5	20	20	112.69	107.63	106.17	100.00
			50	50	115.66	113.49	109.89	100.00
			80	80	119.56	116.39	112.98	100.00
		0.6	20	20	123.78	112.39	109.43	100.00
			50	50	127.22	116.94	111.39	100.00
			80	80	131.13	122.08	113.22	100.00
		0.7	20	20	137.16	122.26	112.16	100.00
			50	50	139.72	122.57	114.44	100.00
			80	80	142.10	125.85	117.71	100.00
		0.8	20	20	140.90	123.23	113.49	100.00
			50	50	142.04	124.68	115.66	100.00
			80	80	148.41	127.26	119.27	100.00
		0.9	20	20	159.26	125.05	117.09	100.00
			50	50	162.47	131.39	119.60	100.00
			80	80	177.34	140.73	121.99	100.00

(i) $\hat{Y}_{MC,A}$ has the highest percentage relative efficiency than any other estimators in all the cases followed by $\hat{Y}_{MC,B}$. $\hat{Y}_{cal}$ is the least efficient among the calibration estimators as it doesn't consider the underlying working model.

(ii) $\hat{Y}_{cal}$ never outperforms any of the model based calibration estimators. However, it performs well than $\hat{Y}_{HT}$ when there is a strong correlation between study and auxiliary variable.

(iii) All estimators has increasing efficiency with increasing correlation between study and auxiliary variable and also with the increasing sample sizes.

6. CONCLUDING REMARKS

In practice, it is most important to make the best use of the auxiliary information. In case of linear relationship between study and auxiliary information, design based calibration approach is appropriate to use. But when there is a non-linear relationship exists between study and auxiliary variable, there is no compelling reason to use design based calibration approach as it could be inefficient. So a possible solution is model based calibration approach which incorporates the non-linear relationship through the fitted values based on a sample. It can be noted that, in case of linear working-model, it is only necessary to know the total of the auxiliary variables for the entire finite population to construct efficient estimators of population total. However, for a non-linear working-model auxiliary information for all the population units are necessary to be known.

We have proposed model-calibration approach to the use of complete auxiliary information in two-occasion successive sampling to estimate population total. The idea involves fitting a general working model and then calibrating on the resulting fitted values as opposed to on the auxiliary variables themselves. With the proposed methodology, we obtained two different estimators which incorporates the non-linear relationship between study and auxiliary variable. The proposed estimators resulting in higher efficiency than the estimator which doesn't consider the non-linear relationship. However, the gain in efficiency depends upon the appropriateness of the underlying working-model.

REFERENCES

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

Rueda, M., Martinez, S., Arcos, A. and Munoz, J.F. (2009). Mean estimation under successive sampling with calibration estimators. *Comm. Statist. – Theory Methods*, **38**, 808-827.

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

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