



Calibration approach estimators in stratified sampling

Jong-Min Kim^{a,*}, Engin A. Sungur^a, Tae-Young Heo^b

^a*Division of Science and Mathematics, University of Minnesota, Morris, MN 56267, USA*

^b*IT Service Research Division, Electronic Telecommunication Research Institute, Daejeon 305-700, South Korea*

Received 25 August 2004; received in revised form 16 May 2006; accepted 23 May 2006

Abstract

Calibration is commonly used in survey sampling to include auxiliary information to increase the precision of the estimates of population parameter. In this paper, we newly propose various calibration approach ratio estimators and derive the estimator of the variance of the calibration approach ratio estimators in stratified sampling.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Calibration approach; Stratified sampling; Estimation of variance; Ratio and regression-type estimator; Auxiliary information

1. Introduction

Calibration is commonly used in survey sampling to include auxiliary information to increase the precision of the estimators of population parameter. Deville and Särndal (1992) first presented calibration estimators in survey sample and calibration estimation has been studied by many survey statisticians. A few key references are Dupont (1995), Hidioglou and Särndal (1998), Singh et al. (1998, 1999), Singh (2001), Sitter and Wu (2002), and Tracy et al. (2003). Wu and Sitter (2001) and Sitter and Wu (2002) generalize the calibration procedure by means of model calibrations and extend a pseudo-empirical likelihood method to obtain efficient estimator of quadratic and other second-order finite population functions.

There are three major advantages of calibration approach in survey sampling. First, the calibration approach leads to consistent estimates. Second, it provides an important class of technique for the efficient combination of data sources. Third, calibration approach has computational advantage to calculate estimates.

In this paper, we propose new calibration approach estimation to several ratio estimators for improving variance estimator with the aid of auxiliary information in stratified random sampling.

The paper is organized as follows: Section 2.1 reviews the calibration method in survey sampling and suggests calibration approach combined ratio estimator. In Section 2.2, we suggest a new calibration approach ratio estimator of various ratio estimators in stratified sampling and derive the estimator of variance of calibration approach combined ratio estimator. Finally, we present the conclusion remarks in Section 3.

*Corresponding author. Tel.: +1 320 5896341; fax: +1 320 589 6371.

E-mail address: jongmink@mrs.umn.edu (J.-M. Kim).

2. Calibration estimation

2.1. Reviews of calibration estimation

Consider a finite population $\Omega = \{1, \dots, N\}$ consisting of N identifiable units. For each unit in the population the value of a vector x of N auxiliary variables is available. A sample S of size n is drawn without replacement from Ω according to a probabilistic sampling plan with inclusion probabilities $\pi_i = \Pr(i \in S)$ assumed to be strictly positive.

The study variable y is observed for each unit in the sample, hence y_i is known for all $i \in S$ and the values x_1, x_2, \dots, x_N are known for the entire population. To estimate the population total $Y = \sum_{i=1}^N y_i$, [Deville and Särndal \(1992\)](#) first introduce the notion of calibration estimator of Y , which is constructed as $\hat{Y}_c = \sum_{i \in S} p_i y_i$, where the calibration weights p_i 's are chosen to minimize their average distance from the basic design weights $d_i = 1/\pi_i$ that are used in the Horvitz–Thompson estimator

$$\hat{Y}_{HT} = \sum_{i \in S} d_i y_i$$

subject to the constraint $\sum_{i \in S} p_i x_i = X$, where X are the known population totals for the auxiliary variables. The distance measure is most commonly chosen as

$$\Phi = \sum_{i \in S} \frac{(d_i - p_i)^2}{d_i q_i},$$

where the q_i 's are known positive weights unrelated to d_i . The resulting calibration estimator is

$$\hat{Y}_c = \sum_{i \in S} p_i y_i = \hat{Y}_{HT} + (X - \hat{X}_{HT})' \hat{B},$$

where $\hat{B} = [\sum_{i \in S} d_i q_i x_i x_i']^{-1} \sum_{i \in S} d_i q_i x_i y_i$, $\hat{X}_{HT} = \sum_{i \in S} d_i x_i$ and $\hat{Y}_{HT} = \sum_{i \in S} d_i y_i$ are the Horvitz–Thompson estimators. The definition of \hat{Y}_c is equivalent to a generalized regression estimator with the choice of Φ .

[Singh et al. \(1998\)](#) introduce the calibration estimation in stratified sampling design based on the calibration approach by [Särndal \(1996\)](#). Suppose the population consists of K strata with N_j units in the j th stratum from which a simple random sample of size n_j is taken without replacement. Let total population size be $N = \sum_{j=1}^K N_j$ and sample size be $n = \sum_{j=1}^K n_j$, respectively. Associated with the i th unit of the j th stratum there are two values y_{ji} and x_{ji} with $x_{ji} > 0$ being the covariate. For the j th stratum, let $w_j = N_j/N$ be the stratum weights, $f_j = n_j/N_j$ the sample fraction, $\bar{y}_j, \bar{x}_j, \bar{Y}_j, \bar{X}_j$ the y -sample and x -sample and population means, respectively. Assume $\bar{X} = \sum_{j=1}^K w_j \bar{X}_j$ is known. The purpose of [Singh et al.'s \(1998\)](#) work is to estimate $\bar{Y} = \sum_{j=1}^K w_j \bar{Y}_j$, possibly by incorporating the covariate information x .

2.2. Calibration approach combined ratio estimator in stratified sampling

In this paper, we suggest the calibration approach combined ratio estimator using auxiliary information in stratified sampling. Calibration ratio estimator under the stratified sampling is given by

$$\bar{y}_{c,st}^* = \sum_{j=1}^K w_j^* \bar{y}_j \quad (1)$$

with new weights w_j^* . The new weights w_j^* are chosen such that chi-square-type distance given by

$$\Phi_{\text{new}} = \sum_{j=1}^K \frac{(w_j^* - w_j)^2}{w_j^* q_j} \quad (2)$$

is minimum subject to calibration constraint

$$\sum_{j=1}^K w_j^* \bar{x}_j = \bar{X}. \quad (3)$$

Minimizing the chi-square-type distance measure (2) subject to the calibration constraint (3) leads to the calibration weight for stratified sampling followed by

$$w_j^* = w_j + \frac{w_j q_j \bar{x}_j}{\sum_{j=1}^K w_j q_j \bar{x}_j^2} \left[\bar{X} - \sum_{j=1}^K w_j \bar{x}_j \right],$$

where q_j are known positive numbers and \bar{X} is the Horvitz–Thompson-type estimator. Therefore, the calibration approach combined regression estimator in stratified sampling is

$$\bar{y}_{c, \text{st}}^* = \sum_{j=1}^K w_j \bar{y}_j + \frac{\sum_{j=1}^K w_j q_j \bar{x}_j \bar{y}_j}{\sum_{j=1}^K w_j q_j \bar{x}_j^2} \left[\bar{X} - \sum_{j=1}^K w_j \bar{x}_j \right]. \quad (4)$$

Property 1. If $q_j = \bar{x}_j^{-1}$, then (4) reduces to the well-known combined ratio estimator in stratified sampling

$$\bar{y}_{\text{st}} = \frac{\sum_{j=1}^K w_j \bar{y}_j}{\sum_{j=1}^K w_j \bar{x}_j} \bar{X}. \quad (5)$$

Next we consider the estimator of variance of calibration approach combined ratio estimator in stratified sampling. The estimator of variance of combined regression estimator is given by

$$\text{Var}(\bar{y}_{\text{st}}) = \sum_{j=1}^K \frac{w_j^2 (1 - f_j)}{n_j} s_{ej}^2, \quad (6)$$

where $s_{ej}^2 = (n_j - 1)^{-1} \sum_{i=1}^{n_j} \hat{e}_{ji}^2$ is the j th stratum sample variance, $\hat{e}_{ji} = y_{ji} - \bar{y}_j - b(x_{ji} - \bar{x}_j)$ and $b = (\sum_{j=1}^K w_j q_j \bar{x}_j \bar{y}_j / \sum_{j=1}^K w_j q_j \bar{x}_j^2)$ have their usual meaning.

The general estimator of variance of the calibration approach combined regression estimator (Singh et al., 1998) is as follows:

$$\text{Var}_c(\bar{y}_{\text{st}}^*) = \sum_{j=1}^K \frac{D_j w_j^{*2}}{w_j} s_{ej}^2, \quad (7)$$

where $D_j = w_j^2 (1 - f_j) / n_j$ and w_j^* is given new weight.

Property 2. If $q_j = \bar{x}_j^{-1}$ in (7), then (7) reduces to

$$\text{Var}_c(\bar{y}_{\text{st}}^*) = \left(\frac{\bar{X}}{\bar{x}_{\text{st}}} \right)^2 \sum_{j=1}^K \frac{w_j^2 (1 - f_j)}{n_j} s_{ej}^2. \quad (8)$$

It follows that (8) is a special case ($g = 2$) of estimator for estimating the variance of combined ratio estimator given by Wu (1985) as

$$\text{Var}_w(\bar{y}_{\text{st}}^*) = \left(\frac{\bar{X}}{\bar{x}_{\text{st}}} \right)^g \sum_{j=1}^K \frac{w_j^2 (1 - f_j)}{n_j} s_{ej}^2.$$

Therefore, we derive the estimator of variance of calibration approach combined ratio estimator as follows:

$$\text{Var}_c(\bar{y}_{\text{st}}^*) = \left(\frac{\bar{X}}{\bar{x}_{\text{st}}} \right)^2 \text{Var}(\bar{y}_{\text{st}}). \quad (9)$$

Using Equation (A.1) in Wu (1985), it can be shown that

$$\left(\frac{\bar{X}^2}{\bar{x}_{\text{st}}} \right) = 1 - 2\bar{X}^{-1}(\bar{x}_{\text{st}} - \bar{X}) + 3[\bar{X}^{-1}(\bar{x}_{\text{st}} - \bar{X})]^2 + O_p(n^{-3/2}).$$

Also we can rewrite (9) as follows:

$$\text{Var}_c(\bar{y}_{\text{st}}^*) = \text{Var}(\bar{y}_{\text{st}}) - \text{Var}(\bar{y}_{\text{st}}) \{ 2\bar{X}^{-1}(\bar{x}_{\text{st}} - \bar{X}) - 3[\bar{X}^{-1}(\bar{x}_{\text{st}} - \bar{X})]^2 - O_p(n^{-3/2}) \}.$$

Table 1
Ratio estimator and calibration approach ratio estimator in stratified sampling

Method	Ratio estimator	Calibration approach ratio estimator
SD	$\bar{y}_{SD} = \bar{y} \left(\frac{\bar{X} + C_x}{\bar{x} + C_x} \right)$	$\bar{y}_{SD}^* = \left(\frac{\sum_{j=1}^K w_j \bar{y}_j}{\sum_{j=1}^K w_j (\bar{x}_j + C_{xj})} \right) \sum_{j=1}^K w_j (\bar{X}_j + C_{xj})$
SK	$\bar{y}_{SK} = \bar{y} \left(\frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)} \right)$	$\bar{y}_{SK}^* = \left(\frac{\sum_{j=1}^K w_j \bar{y}_j}{\sum_{j=1}^K w_j (\bar{x}_j + \beta_{2j}(x))} \right) \sum_{j=1}^K w_j (\bar{X}_j + \beta_{2j}(x))$
US ₁	$\bar{y}_{US_1} = \bar{y} \left(\frac{\bar{X} \beta_2(x) + C_x}{\bar{x} \beta_2(x) + C_x} \right)$	$\bar{y}_{US_1}^* = \left(\frac{\sum_{j=1}^K w_j \bar{y}_j}{\sum_{j=1}^K w_j (\bar{x}_j \beta_{2j}(x) + C_{xj})} \right) \sum_{j=1}^K w_j (\bar{X}_j \beta_{2j}(x) + C_{xj})$
US ₂	$\bar{y}_{US_2} = \bar{y} \left(\frac{\bar{X} C_x + \beta_2(x)}{\bar{x} C_x + \beta_2(x)} \right)$	$\bar{y}_{US_2}^* = \left(\frac{\sum_{j=1}^K w_j \bar{y}_j}{\sum_{j=1}^K w_j (\bar{x}_j C_{xj} + \beta_{2j}(x))} \right) \sum_{j=1}^K w_j (\bar{X}_j C_{xj} + \beta_{2j}(x))$

Table 2
Estimator of variance of calibration approach combined ratio estimator

Method	q_j	Estimators of variance
SD	$(\bar{x}_j + C_{xj})^{-1}$	$\text{Var}_c(\bar{y}_{SD}^*) = \left(\frac{\bar{X} + C_x}{\bar{x}_{st}} \right)^2 \sum_{j=1}^K \frac{w_j^2 (1 - f_i)}{n_j} s_{ej}^2$
SK	$(\bar{x}_j + \beta_{2j}(x))^{-1}$	$\text{Var}_c(\bar{y}_{SK}^*) = \left(\frac{\bar{X} + \beta_2(x)}{\bar{x}_{st}} \right)^2 \sum_{j=1}^K \frac{w_j^2 (1 - f_i)}{n_j} s_{ej}^2$
US ₁	$(\bar{x}_j \beta_{2j}(x) + C_{xj})^{-1}$	$\text{Var}_c(\bar{y}_{US_1}^*) = \left(\frac{\bar{X} \beta_2(x) + C_x}{\bar{x}_{st}} \right)^2 \sum_{j=1}^K \frac{w_j^2 (1 - f_i)}{n_j} s_{ej}^2$
US ₂	$(\bar{x}_j C_{xj} + \beta_{2j}(x))^{-1}$	$\text{Var}_c(\bar{y}_{US_2}^*) = \left(\frac{\bar{X} C_x + \beta_2(x)}{\bar{x}_{st}} \right)^2 \sum_{j=1}^K \frac{w_j^2 (1 - f_i)}{n_j} s_{ej}^2$

Under condition $q_j = \bar{x}_j^{-1}$, we claim that the Singh et al. (1998) calibration estimator of variance of the combined ratio estimator can be found by the combinations of the estimator of variance of combined ratio estimator.

From Wu (1985, p. 151) paper, we find the fact that the MSE of the estimator of variance of the calibration approach combined ratio estimator is more efficient than the MSE of the estimator of variance of combined ratio estimator in stratified sampling. Therefore, in general, calibration approach ratio estimator is more efficient than ratio estimator in stratified sampling.

3. Various calibration approach ratio estimators

In this section, we suggest a calibration approach for various ratio estimators in stratified sampling. Kadilar and Cingi (2003) proposed various ratio estimators in stratified sampling. Based on this work, we derived calibration approach for ratio estimators for Sisodia–Dwivedi estimator (SD) (Sisodia and Dwivedi, 1981), Singh–Kakran ratio-type estimator (SK) (Singh and Kakran, 1993), Upadhyaya and Singh (1999) estimator 1 (US₁) and Upadhyaya and Singh (1999) estimator 2 (US₂). Following Section 2.2, we proposed and summarized four ratio estimators and calibration approach ratio estimators in stratified sampling in Table 1. We also derived and summarized estimators of variance of calibration approach combined ratio estimator in Table 2.

4. Concluding remarks

This paper applied calibration estimation to ratio-type estimators in stratified sampling. We proposed and studied calibration approach in four estimators to the use of complete auxiliary information to estimate

ratio-type estimator in stratified sampling and derived the estimator of variance of calibration approach ratio estimators. We also showed that the estimator of variance of the combined ratio estimator in stratified sampling using the calibration approach is more efficient than the standard estimator of variance of combined ratio estimator in stratified sampling. Consequently, we found that the new calibration approach estimators in stratified sampling are very attractive for survey researchers to get consistent estimates and provide more precise estimates of population parameters.

Acknowledgements

The authors are grateful to an anonymous referee and the executive editor Dr. Richard A. Johnson for several valuable comments and suggestions.

References

- Deville, J.-C., Särndal, C.-E., 1992. Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* 87, 376–382.
- Dupont, F., 1995. Alternative adjustments where there are several levels of auxiliary information. *Survey Methodology* 21, 125–135.
- Hidioglou, M.A., Särndal, C.-E., 1998. Use of auxiliary information for two-phase sampling. *Survey Methodology* 24, 11–20.
- Kadilar, C., Cingi, H., 2003. Ratio estimators in stratified random sampling. *Biometrical J.* 45 (2), 218–225.
- Särndal, C.-E., 1996. Efficient estimators with simple variance in unequal probability sampling. *J. Amer. Statist. Assoc.* 91, 1289–1300.
- Singh, H.P., Kakran, M.S., 1993. A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. Unpublished paper.
- Singh, S., 2001. Generalized calibration approach for estimating variance in survey sampling. *Ann. Inst. Statist. Math.* 53, 404–417.
- Singh, S., Horn, S., Yu, F., 1998. Estimation of variance of the general regression estimator: higher level calibration approach. *Survey Methodology* 24, 41–50.
- Singh, S., Horn, S., Chowdhury, S., Yu, F., 1999. Calibration of the estimators of variance. *Austral. New Zealand J. Statist.* 41, 199–212.
- Sisodia, B.V.S., Dwivedi, V.K., 1981. A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Indian Soc. Agricultural Statist.* 33, 13–18.
- Sitter, R.R., Wu, C., 2002. Efficient estimation of quadratic finite population functions. *J. Amer. Statist. Assoc.* 97, 535–543.
- Tracy, D.S., Singh, S., Arnab, R., 2003. Note on calibration in stratified and double sampling. *Survey Methodology* 29, 99–104.
- Upadhyaya, L.N., Singh, H.P., 1999. Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical J.* 41, 627–636.
- Wu, C., 1985. Variance estimation for combined ratio and combined regression estimators. *J. Roy. Statist. Soc. Ser. B: Statist. Methodology* 47, 147–154.
- Wu, C., Sitter, R., 2001. A model-calibration to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.* 96, 185–193.