Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics and Data Analysis 52 (2008) 4253–4271

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

On the estimation of the general parameter

Matthew Stearns[a], Sarjinder Singh[b,*]

[a] Department of Statistics, St. Cloud State University, St. Cloud, MN 56301–4498, USA
[b] Department of Mathematics, University of Texas at Brownsville and Texas Southmost College, Brownsville, TX 78520, USA

Received 4 February 2008; received in revised form 5 February 2008; accepted 7 February 2008
Available online 13 February 2008

Abstract

In this paper, we first discuss the origin, developments and various thoughts by several researchers on the generalized linear regression estimator (GREG) due to Deville and Särndal [Deville, J.C., Särndal, C.E., 1992. Calibration estimators in survey sampling. J. Amer. Statist. Assoc. 87, 376–382]. Then, the problem of estimation of the general parameter of interest considered by Rao [Rao, J.N.K., 1994. Estimating totals and distribution functions using auxiliary information at the estimation stage. J. Official Statist. 10 (2), 153–165], and Singh [Singh, S., 2001. Generalized calibration approach for estimating the variance in survey sampling. Ann. Inst. Statist. Math. 53 (2), 404–417; Singh, S., 2004. Golden and Silver Jubilee Year-2003 of the linear regression estimators. In: Proceedings of the Joint Statistical Meeting, Toronto (Available on the CD), 4382-4380; Singh, S., 2006. Survey statisticians celebrate Golden Jubilee Year-2003 of the linear regression estimator. Metrika 1–18] is further investigated. In addition to that it is shown that the Farrell and Singh [Farrell, P.J., Singh, S., 2005. Model-assisted higher order calibration of estimators of variance. Australian & New Zealand J. Statist. 47 (3), 375–383] estimators are also a special case of the proposed methodology. Interestingly, it has been noted that the single model assisted calibration constraint studied by Farrell and Singh [Farrell, P.J., Singh, S., 2002. Re-calibration of higher order calibration weights. Presented at Statistical Society of Canada conference, Hamilton (Available on CD); Farrell, P.J., Singh, S., 2005. Model-assisted higher order calibration of estimators of variance. Australian & New Zealand J. Statist. 47 (3), 375–383] and Wu [Wu, C., 2003. Optimal calibration estimators in survey sampling. Biometrika 90, 937–951] is not helpful for calibrating the Sen [Sen, A.R., 1953. On the estimate of the variance in sampling with varying probabilities. J. Indian Soc. Agril. Statist. 5, 119–127] and Yates and Grundy [Yates, F., Grundy, P.M., 1953. Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc. Ser. 15, 253–261] estimator of the variance of the linear regression estimator under the optimal designs of Godambe and Joshi [Godambe, V.P., Joshi, V.M., 1965. Admissibility and Bayes estimation in sampling finite populations—I. Ann. Math. Statist. 36, 1707–1722]. Three new estimators of the variance of the proposed linear regression type estimator of the general parameters of interest are introduced and compared with each other. The newly proposed two-dimensional linear regression models are found to be 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 always remains more efficient than the GREG estimator.

* Corresponding author. Tel.: +1 956 8828892.
E-mail address: sarjinder@yahoo.com (S. Singh).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2008.02.004

1. Introduction

The concept of linear weighting of sample survey data can be found in Bethlehem and Keller (1987). Consider a population $\Omega = \{1, 2, \ldots, N\}$, from which a probability sample s ($\subset \Omega$) is drawn with a given sampling design $p(s)$. The inclusion probabilities $\pi_i = P(i \in s)$ and $\pi_{ij} = P(i, j \in s)$ are assumed to be strictly positive and known. Deville and Särndal (1992) used calibration on the known population total X to modify the basic sampling design weights, $d_i = \pi_i^{-1}$, that appear in the Horvitz and Thompson (1952) estimator

$$\widehat{Y}_{\text{HT}} = \sum_{i \in s} d_i y_i. \tag{1.1}$$

An estimator

$$\widehat{Y}_{\text{G}} = \sum_{i \in s} w_i^* y_i \tag{1.2}$$

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

$$\sum_{i \in s} w_i^* x_i = X. \tag{1.3}$$

Minimization of the chi square distance function

$$\text{D} = \sum_{i \in s} \frac{\left(w_i^* - d_i\right)^2}{d_i q_i} \tag{1.4}$$

between the new weights w_i^* and selection weights or design weights d_i leads to the GREG of population total Y, given by

$$\widehat{Y}_{\text{G}} = \sum_{i \in s} d_i y_i + \widehat{\beta}_{\text{ds}} \left(X - \sum_{i \in s} d_i x_i \right) \tag{1.5}$$

where

$$\widehat{\beta}_{\text{ds}} = \frac{\sum_{i \in s} d_i q_i x_i y_i}{\sum_{i \in s} d_i q_i x_i^2} \tag{1.6}$$

and q_i are suitably chosen weights such that the estimator depends upon its choice. As pointed out by Singh (2003, 2006), $\widehat{\beta}_{\text{ds}}$ in (1.6) is not an ordinary least square estimator, and hence GREG can never be as efficient as the linear regression estimator $\overline{y}_{\text{lr}}$ if the regression line does not pass through the origin. If $q_i = 1/x_i$, then the GREG reduces to the ratio estimator due to Cochran (1940). Singh et al. (1998) reported that there is no choice of q_i such that the GREG reduces to the product estimator due to Murthy (1964). Following Särndal et al. (1989), Deville and Särndal (1992), Särndal (1996), Rao (1997), and several others, the GREG can be written as

$$\widehat{Y}_{\text{G}} = \sum_{i \in s} d_i \widehat{e}_i + \widehat{\beta}_{\text{ds}} X \tag{1.7}$$

and the Sen (1953) and Yates and Grundy (1953) form of the first estimator of variance of the GREG is

$$\widehat{V}_{\text{ds}(1)} = \frac{1}{2} \sum_{i \neq j} \sum_{\in s} D_{ij} \left(d_i \widehat{e}_i - d_j \widehat{e}_j \right)^2 \tag{1.8}$$

where

$$D_{ij} = \frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}} \quad \text{and} \quad \widehat{e}_i = y_i - \widehat{\beta}_{\text{ds}} x_i.$$