PRAJÑĀ-Journal of Pure and Applied Sciences, Vol. 19: 97 - 99 (2011)
ISSN 0975 - 2595

ON VARIANCE ESTIMATION FOR THE GREG ESTIMATOR

P.A. Patel[1] and R.D. Chaudhari[2]

[1]*Department of Statistics, Sardar Patel University,Vallabh Vidhyanagar-388120, Gujarat, India*
[2]*Department of Statistics, M. G. Science College, Ahmedabad, Gujarat, India*

ABSTRACT

This article extends the π-weighted ratio-type variance estimator of the Horvitz–Thompson estimator, suggested by Patel and Chaudhari (2006), for the generalized regression (GREG) estimation. The suggested variance estimator based on the empirical mean squared error yields some gain over the available variance estimators in simulations when the underlying assumptions are satisfied.

Key words: *Finite population, Generalized regression estimator, Variance estimation.*

INTRODUCTION

Recently more attention has been given to the generalized regression estimator of the finite population total. Some of the reasons are given in Särndal (1996). Important references on GREG estimation and on its variance estimation are Särndal (1980, 1996), Robinson and Särndal (1983), Wright (1983), Särndal et al. (1989, 1992), Deville and Särndal (1992), Kott (1990), Chaudhuri and Maiti (1994), Singh et al. (1999) and Duchesne (2000) and references cited there.

Following the notations of Patel and Chaudhari (2006), we postulate the model

$$\left.\begin{array}{l} y_i = \beta x_i + \varepsilon_i \quad i \in U = \{1, \ldots, N\} \\ E_\xi(\varepsilon_i / x_i) = 0 \\ V_\xi(\varepsilon_i / x_i) = \sigma_i^2 = \sigma^2 x_i^\gamma, 0 \le \gamma \le 2 \\ C_\xi(\varepsilon_i, \varepsilon_j / x_i, x_j) = 0 \;\; (i \ne j) \end{array}\right\} \qquad (1)$$

Where β, $\sigma^2 > 0$ are the parameters. Here $E_\xi(.)$, $V_\xi(.)$ and $C_\xi(.)$ denote ξ- expectation, ξ- variance and ξ- covariance. Under Model (1) the GREG- estimator of $Y = \Sigma_{i \in U} y_i$ is given by

$$\hat{Y}_{GREG} = \sum_s \frac{g_i y_i}{\pi_i} \qquad (2)$$

where $g_i = 1 + (X - \hat{X}_{HT}) \frac{x_i q_i \pi_i}{\sum_s x_i^2 q_i}$ is the g-adjustment factors, $\hat{x}_{HT} = \sum_s \frac{x_i}{\pi_i}$ is the Horvitz. Thompson estimater of $x = \Sigma_U x_i$ and q_i is chosen by the user.

The estimator $\hat{Y}_{GREG}$ is optimal in the following way: Starting from the basic estimator $\hat{Y}_{HT} = \sum_s a_i y_i$ with $a_i = 1/\pi_i$, create a new estimator $\hat{Y} = \sum_s w_i y_i$ with weight w_i lying as close as possible to the basic weights a_i, subject to the calibration constraint $\sum_s w_i x_i = X$. when the distance to minimize is given as $\sum c_i (w_i - a_i)^2 / a_i$, where c_i's are constants, the optimal weights w_i are precisely $w_i = a_i g_i$.

The Taylor expansion variance for the GREG – estimator (See, Särndal et al. 1992) is

$$V_T = \sum\sum_{i<j \in U} \Delta_{ij} \left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j} \right)^2$$

where

$$\Delta_{ij} = \pi_i \pi_j - \pi_{ij}, \; E_i = y_i - B_Q x_i \;\text{ and }\; B = \frac{\sum_U y_i x_i Q_i}{\sum_U x_i^2 Q_i}.$$

*Corresponding author: patelpraful_a@indiatimes.com

Two versions of Yates-Grundy type variance estimators of V_T are

$$v_s = \sum\sum_{i<j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2 \quad \text{and} \quad v_g = \sum\sum_{i<j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \left(\frac{e_i g_i}{\pi_i} - \frac{e_j g_j}{\pi_j} \right)^2 \qquad (3)$$

where $e_i = y_i - \hat{\beta}\, x_i$ and $\hat{\beta} = \frac{\Sigma_s x_i y_i q_i}{\Sigma_s x_i^2 q_i}$

Kott (1990) proposed an estimator of V_T (see Appendix A) that is design-consistent and model-unbiased. Chaudhari and Maiti (1994), following Kott's estimator, suggested various model-assisted estimator of V_T.

We suggest in the next section an estimator of V_T. To study the repeated sampling properties relative to standard one a limited simulation study is conducted in Section 3. The conclusions and recommendations are given in section 4.

THE RATIO-TYPE VARIANCE ESTIMATOR

Patel and Chaudhari (2006) suggested a π- weighted ratio- type estimator ($\upsilon_{\pi wr}$) for the variance of Horvitz- Thompson estimator of population total. They showed that this estimator is asymptotically design-unbiased (ADU) and asymptotically design consistent ADC). Empirically this estimator has performed very well when the relationship between *y* and *x* is linear passing through the origin and the $V_\xi(y_i) \propto x_i^\gamma$, $\gamma \in [1,2]$, under fixed size or non-fixed size sampling design. Motivated by this we suggest the following estimator for V_T:

$$v_\pi = \frac{\sum_s \phi_{ii} g_i^2 e_i^2 / \pi_i}{\sum_s \phi_{ii} g_i^2 x_i^2 / \pi_i} \sum_U \phi_{ii} x_i^2 + \frac{\sum\sum_s \phi_{ij} g_i e_i g_j e_j / \pi_{ij}}{\sum\sum_s \phi_{ij} g_i x_i g_j x_j / \pi_{ij}} \sum\sum_U \phi_{ij} x_i x_j$$

where $\phi_{ij} = \frac{1}{\pi_i} - 1$, if $i = j$ and $= \frac{\pi_{ij}}{\pi_i \pi_j} - 1$ if $i \neq j$.

Remark : The construction of v_π would suggest that v_π performs well if the ratios y_i/x_i is more or less constant and the variance of y_i is proportional to x_i.

SIMULATION

A finite population of size N =400 was created.The characteristics x and y for the i^{th} unit were generated using the model

$$y_i = \beta x_i + x_i^{\gamma/2} \varepsilon_i \quad , i = 1, \ldots, N$$

for specified values of β, γ, g, h and σ_ε^2, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ were independent of $x_i \sim Gamma(g, h)$. Thus the mean, variance and coefficient of variation of x_i are given by $\mu_x = gh$, $\sigma_\varepsilon^2 = gh^2$ and $C_x = \sigma_x / \mu_x = g^{1/2}$.
Further the mean of y_i is $\mu_y = \beta \mu_x$, variance of x_i is $\sigma_y^2 = \beta^2 \sigma_x^2 + \sigma_\varepsilon^2 E(x_i^\gamma)$

and the correlation coefficient $Corr(x_i, y_i) = \rho = \beta\sigma_x/\sigma_y$ vary depending on the choice of γ. Here $\gamma = 0$, 1 and 2 were considered so that $E(x_i^{\gamma/2}) = 1$, μ_x and $\mu_x^2 + \sigma_x^2$; for each of these cases σ_ε^2 and σ_x^2 were then chosen to match various values of ρ and C_x. The three (β, μ_x, γ) combinations were: (a) $\beta = 1, \mu_x = 100, \gamma = 0$; (b) $\beta = 1, \mu_x = 100, \gamma = 1$; and (c) $\beta = 1, \mu_x = 100, \gamma = 2$.

A finite population was created for each of (a) - (c) and each combination of (ρ, C_x) and a sample of size n =30 was drawn using Sunter's (1986) sampling design. The variance estimators were computed from each sample. This process was repeated M = 10,000 times. For each of these samples, we computed the estimators v_g, v_k and v_π corresponding to different values of q_i, i = 1,2,3 given by

Choice of q_i	**Form of Estimator**
$q_1 = 1/x_i$	$\hat{Y}_{GREG} = \hat{Y}_{HT} + \frac{\bar{y}}{\bar{x}}(X - \hat{X}_{HT})$
$q_2 = 1/\pi_i x_i$	$\hat{Y}_{GR} = \frac{\hat{Y}_{HT}}{\hat{X}_{HT}} X$
$q_3 = (1-\pi_i)/\pi_i x_i$	$\hat{Y}_B = \hat{Y}_{HT} + \frac{\sum_s \left(\frac{1}{\pi_i} - 1\right) y_i}{\sum_s \left(\frac{1}{\pi_i} - 1\right) x_i}(X - \hat{X}_{HT})$

The performance of the different variance estimators w s measures and compared in terms of relative bias in percenta e (RB), relative efficiency (RE) and empirical coverage ra e (ECR). The simulated values of RB and RE for a particul r variance estimator were computed as

$$RB(v) = 100 \times \frac{\bar{v} \quad V}{V} \quad \text{where} \quad \bar{v} = \frac{1}{M}\sum_{j=1}^{M} v_{(j)}$$

The relative efficiency of v is given by

$$RE(v) = \frac{MSE(v_{YG})}{MSE(v)} = \frac{RSE(v_{YG})}{RSE(v)}^{2}$$

where $MSE(v) = \frac{1}{M-1}\sum_{j=1}^{M}(v_{(j)} - V)^2$ and $RSE(v) = 100 \times \sqrt{\frac{MSE(v)}{V}}$

The REs of the estimators is presented in Table 1 whereas the RBs in the table 2 given in Appendix A.

Table 1 RE under Sunter's Sampling Scheme

q_i	γ		0			1			2		
	Est.	$\rho \backslash C_x$	1.5	0.75	0.33	1.5	0.75	0.33	1.5	0.75	0.33
q_1	v_g	0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	v_k	0.9	1.057	1.047	1.047	1.049	1.033	1.093	1.024	1.039	1.034
		0.8	1.041	1.030	1.022	1.050	1.028	1.053	1.020	1.069	1.026
		0.7	1.041	1.033	1.029	1.056	1.052	1.064	1.023	1.082	1.034
	v_π	0.9	1.112	1.968	1.065	**1.131**	**4.519**	**1.221**	1.077	1.033	1.075
		0.8	1.124	1.741	1.150	**1.106**	**3.163**	**1.149**	1.073	2.954	1.028
		0.7	1.137	1.836	1.102	**1.198**	**4.652**	**1.189**	1.001	3.827	1.035
q_2	v_g	0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	v_k	0.9	1.048	0.933	1.048	1.039	0.740	1.073	1.019	0.927	1.028
		0.8	1.026	0.923	0.978	1.044	0.763	1.037	1.016	0.813	1.025
		0.7	1.024	0.911	1.010	1.041	0.740	1.039	1.019	0.800	1.031
	v_π	0.9	1.101	1.287	1.056	**1.120**	**3.957**	**1.207**	1.066	0.528	1.062
		0.8	1.112	1.251	1.143	**1.098**	**2.357**	**1.145**	1.063	1.200	1.025
		0.7	1.120	1.402	1.091	**1.177**	**2.650**	**1.169**	0.996	1.591	1.030
q_3	v_g	0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	v_k	0.9	1.048	0.932	1.049	1.038	0.704	1.071	1.018	0.926	1.028
		0.8	1.025	0.921	0.971	1.043	0.730	1.034	1.016	0.796	1.025
		0.7	1.022	0.904	1.008	1.040	0.704	1.036	1.019	0.777	1.031
	v_π	0.9	1.099	1.218	1.054	**1.119**	**3.862**	**1.205**	1.065	0.487	1.060
		0.8	1.111	1.197	1.142	**1.097**	**2.255**	**1.145**	1.062	1.045	1.024
		0.7	1.117	1.345	1.090	**1.174**	**2.316**	**1.167**	0.996	1.361	1.030

Table 1 reveals the following comments

- The absolute values of RBs of v_g for C_x = 0.33, 1.5 and $\gamma = 2$ are all within reasonable range of 1%, whereas for C_x = 0.33, $\gamma = 2$ it has large absolute values of RBs ranging from 1-19%. However, for $\gamma = 2$ v_g has performed well.
- For $\gamma = 1$, C_x = 0.75 the absolute RBs of v_π are small compare to v_g and has performed extremely well. This improvement of v_π over v_g is from 125-350%.
- For $\gamma = 1$, C_x = 0.33 1.5 the absolute RBs of v_π and v_g are in the reasonable range with the largest occurring as 3% and 4.67% respectively. But v_π has 10-20% more efficiency compared to v_g.
- For $\gamma = 1$, C_x = 1.5 the estimator v_π should be avoided. However, in case for the other values of C_x, v_π is fractionally more efficient and has reasonable absolute RBs than v_g.
- Overall the variance estimator v_k of $\hat{Y}_{GREG} = \hat{Y}_{HT} + \frac{\bar{y}}{\bar{x}}(X - \hat{X}_{HT})$ is slightly more efficient than v_g but less efficient than v_π when γ = 0,1,2 and C_x = 0.33, 0.75,1.5. In rest of all cases it should be avoidable.

Remark 3. In our simulation study it was borned out that the estimators suggested by Chaudhari and Maiti (1994) performed poorly. Therefore, the results corresponding to these estimators were not presented in respective tables.

Appendix A

The following estimator is included for the comparison.

Kott's Estimator: A variance estimator somewhat similar in spirit to that in equation (3) was proposed by Kott (1990). His point of departure is to create a variance estimator that is unbiased with respect to the model but is still design consistent. The objective is achieved by attaching a ratio adjustment to the estimator (3). Kott's variance estimator is given as

$$v_k = \frac{v_g v_1}{v_2}$$

where $v_1 = v(\hat{Y}_{GREG} - Y)$

$$= \sigma^2 \left[\frac{\sum_s f_i x_i^2 Q_i^2}{\left(\sum_s x_i^2 Q_i\right)^2} \left(X - \hat{X}_{HT}\right)^2 + \sum_s \frac{f_i}{\pi_i^2} + \sum_U f_i + \frac{2}{\sum_s x_i^2 Q_i}\left(X - \hat{X}_{HT}\right) \sum_s \frac{1-\pi_i}{\pi_i} f_i x_i Q_i - 2\sum_s \frac{f_i}{\pi_i} \right]$$

and

$$v_2 = \sigma^2 \left[\sum\sum_{i<j\in s} \frac{\Delta_{ij}}{\pi_{ij}} \left\{ \frac{f_i g_i^2}{\pi_i^2} + \frac{f_j g_j^2}{\pi_j^2} + \frac{\sum_s f_i x_i Q_i^2}{\left(\sum_s x_i^2 Q_i\right)^2} \left(\frac{x_i g_i}{\pi_i} - \frac{x_j g_j}{\pi_j} \right)^2 - \frac{2}{\left(\sum_s x_i^2 Q_i\right)} \left(\frac{x_i g_i}{\pi_i} - \frac{x_j g_j}{\pi_j} \right) \left(\frac{x_i Q_i f_i g_i}{\pi_i} - \frac{x_j Q_j f_j g_j}{\pi_j} \right) \right\} \right]$$

Appendix B

Table 2. RB (%) under Sunter's Sampling Scheme

q_i	γ		0			1			2		
	Est	$\rho \backslash C_x$	1.5	0.75	0.33	1.5	0.75	0.33	1.5	0.75	0.33
q_1	v_g	0.9	0.44	7.34	-9.67	0.58	95.28	1.59	0.49	2.07	0.95
		0.8	3.12	2.45	7.91	1.18	80.35	4.67	0.10	12.99	0.15
		0.7	1.72	9.67	0.13	1.63	45.41	2.32	0.46	18.28	0.72
	v_k	0.9	0.33	7.78	-9.77	0.40	94.79	1.46	-0.60	0.78	-0.61
		0.8	3.09	3.03	8.02	1.04	80.39	4.46	-0.84	11.20	-1.19
		0.7	1.68	10.05	0.19	1.52	45.32	2.20	-0.65	16.27	-0.96
	v_π	0.9	-1.40	-27.27	-10.93	-0.96	27.44	-0.36	-0.71	-15.75	-0.27
		0.8	0.95	-23.75	4.85	-0.29	21.90	3.02	-1.48	-8.33	-1.17
		0.7	-0.91	-18.37	-2.13	0.11	3.69	0.22	-1.34	-5.54	0.22
q_2	v_g	0.9	-0.82	-18.34	-10.15	-0.18	44.20	0.86	0.17	-2.39	0.67
		0.8	1.46	-17.24	4.86	0.61	39.94	3.82	-0.20	2.26	0.09
		0.7	-0.33	-11.66	-1.67	0.64	16.39	1.13	0.26	5.78	0.57
	v_k	0.9	-0.59	-10.73	-10.02	-0.15	59.65	1.08	-0.80	-0.26	-0.81
		0.8	1.91	-10.59	6.05	0.65	55.13	3.91	-1.05	5.16	-1.21
		0.7	0.20	-4.80	-0.95	0.82	27.25	1.54	-0.73	9.02	-1.06
	v_π	0.9	-2.57	-39.61	-11.35	-1.62	2.13	-1.14	-1.12	-19.02	-0.43
		0.8	-0.59	-33.28	2.03	-0.89	2.41	2.01	-1.86	-13.27	-1.47
		0.7	-2.76	-30.48	-3.79	-0.84	-9.28	-0.89	-1.96	-11.21	0.07
q_3	v_g	0.9	-1.02	-21.77	-10.22	-0.30	37.33	0.72	0.12	-2.78	0.60
		0.8	1.19	-20.16	4.34	0.53	33.98	3.68	-0.25	1.00	0.08
		0.7	-0.66	-14.68	1.97	0.50	11.76	0.95	0.24	4.10	0.55
	v_k	0.9	-0.74	-13.57	-10.07	-0.23	54.19	1.01	-0.83	-0.21	-0.86
		0.8	1.72	-12.93	5.70	0.59	50.73	3.81	-1.08	4.36	-1.21
		0.7	-0.06	-7.17	-1.15	0.72	23.91	1.44	-0.73	7.87	-1.07
	v_π	0.9	-2.77	-41.36	-11.43	-1.72	-1.55	-1.30	-1.19	-19.53	-0.47
		0.8	-0.85	-34.79	1.53	-0.98	-0.86	1.82	-1.92	-13.91	-1.54
		0.7	-3.08	-32.27	-4.08	-0.97	-11.60	-1.07	-2.07	-12.07	0.04

References

[1] Patel, P. A. and Chaudhari, R. D. (2006). Design-Based Horvitz-Thompson Variance Estimation: π-Weighted Ratio Type estimator. Statistics in Transition, 7(6), 1277-1293.

[2] Särndal, C.E. (1996). Efficient estimators with simple variance in unequal probability sampling. Journal of the American Statistical Association, 91,1289-1300.

[3] Särndal, C.E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. Biometrika, 67, 639-650.

[4] Robinson, P. M. and Särndal, C.E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. Sankhya, Series B, 45, 240-248.

[5] Wright, R. L. (1983). Finite population sampling with multivariate auxiliary information, Journal of the American Statistical Association, 78, 879-884.

[6] 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, 527-537.

[7] Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). Model Assisted Survey Sampling. Springer Verlag, New York.

[8] Deville, J.C., Särndal, C.E. (1992). Calibration using auxiliary information. Journal of the American Statistical Association, 78, 117-123.

[9] Chaudhari, A. and Maiti, T. (1994). Variance estimation in model assisted survey sampling. Communication is Statistical Theory and Methods, 23(4), 1203-1214.

[10] Singh, S., Horn, S., Chowdhury, S. and Yu. S. (1999). Calibration of the estimators of variance. Australian and New Zealand Journal of Statistics., 41, 199-212.

[11] Duchesne, P. (2000). A note on jackknife variance estimation for the general regression estimator. Journal of Official Statistics, 16(2), 133-138.

[12] Kott, P. S. (1990). Estimating the conditional variance of a design consistent regression estimator. Journal of Statistical Planning and Inference, 24, 287-296.

[13] Sunter, A. B. (1986). Solution to the problem unequal probability sampling without replacement. International Statistical Review, 54, 33-50.