Pak. J. Statist.
2016 Vol. 32(5), 405-407

A REPLY TO, "A NOTE ON REGRESSION ESTIMATION WITH UNKNOWN POPULATION SIZE by Hidiroglou, Kim and Nambeu (2016)"

Sarjinder Singh
Department of Mathematics, Texas A&M University-Kingsville
Kingsville, TX 78363
Email: sarjinder@yahoo.com

ABSTRACT

Hidiroglou, Kim and Nambeu (2016), referred to hereafter as HKN, claimed to find through a simulation study that the Singh and Raghunath (2011) estimator performs "worst" for Poisson sampling, with insufficient understanding of their own work, and without knowing the limitations of the Poisson sampling. They claimed to show Singh and Raghunath's estimator does not work, but their own estimator is not even computable. Curiously, they assert there is a parameter, which cannot be computed, but can be estimated by HKN methodology. This reply offers some needed clarification.

KEYWORDS

Regression estimator, GREG, multicollinearity, calibration, estimation of total, ratio and regression estimators.

REPLY AND CLARIFICATIONS

Please note the following important points and clarifications.

a) Setting

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} d_i \qquad \text{(R.1)}$$

does not imply that Singh and Arnab (2011) are estimating N. This constraint was designed by Singh (2003) to ***control extra variation in the resultant estimator***, but Singh and Arnab (2011) are not finally interested in estimating population size N.

Perhaps HKN confused this with the Wu and Sitter (2001) constraint. Kim (2010: Survey Methodology) claimed to use the constraint (R.1) if the population size is unknown, otherwise the Wu and Sitter (2001) constraint was used. So perhaps the Kim (2010: Survey Methodology) claim is incorrect, not Singh and Raghunath (2011) as mentioned in HKN note. Incidentally, Singh and Arnab (2011) was incorrectly cited in the HKN note as Singh and Raghunath (2011), indicating lack of attention.

In Singh and Arnab (2011) on page 190, a very simple special case, which every layman can understand, of the estimator of total is given in (2.2.7), using known population size N:

$$\hat{Y}_{LR} = N\left[\bar{y} + \frac{s_{xy}}{s_x^2}\left(\bar{X} - \bar{x}\right)\right]. \tag{R.2}$$

The population size N is clearly the coefficient (in front) of the estimator in (R.2).

(R.1) does not depend on the population size, and it implicitly estimates the population size N. There is no need to estimate N when it is known in (R.2). If the Deville and Särndal (1992) estimator can be used when N is unknown, then HKN could compare their estimator with the Deville and Särndal (1992) estimator when N is unknown. The Singh and Arnab (2011) estimator does not apply when N is unknown. The HKN note qualifying the Singh and Raghunath (2011) estimator as “worst” is based on the erroneous assumption by HKN that the population size N is unknown in Singh and Arnab (2011).

Note that the well-known Horvitz-Thompson estimator is only convertible to the estimator $N\bar{y}$ if the population size N is known; otherwise it is a biased estimator like Hazek’s estimator. Singh and Arnab (2011) never mentioned that the Hazek ratio type estimator is a special case of their estimator, a misunderstanding created by the HKN paper. Singh and Arnab (2011) have not claimed that their estimator can be useful when the population total of the auxiliary variable X is known from administrative sources without knowing the population size. Simulation study using Poisson sampling is full of limitations. Singh and Arnab (2011) never assumed unknown population size.

b) When one sets

$$\sum_{i=1}^{n} w_i x_i = X \tag{R.3}$$

where $X = \sum_{i=1}^{N} X_i$ is the population total of the auxiliary variable, it indicates the population size N is already used in computing the population total of an auxiliary variable. Its reuse, as in Wu and Sitter (2001), will not help. Note that SRS is a special case of (R.1) where N disappears while deriving the calibrated estimator, see Singh (2003).

c) In HKN, the model, $y_i = \pi_i\beta_1 + x_i^{*T}\beta_2 + e_i$, is same as the model, $y_i = z_i^{*T}\beta + e_i$, where $z_i^* = \left(\pi_i, x_i^{*T}\right)$, simply treating π_i as an auxiliary variable in the model. It will not add anything new to the model except one more auxiliary variable. It will again make the intercept equal to zero and leads to ratio type estimators.

d) Another point noted by HKN themselves is that B_{KOPT} ***cannot be computed*** because their variance-co-variance matrix is not invertible. This is obvious due to the problem of multicollinearity. Hence the new estimator proposed in this note by HKN is non-functional. The same problem of multicollinearity will make it impossible to estimate B_{KOPT} through several iterations in a simulation study. Due to the multicollinearity problem, the relative efficiency of HKN estimator in (3.4) with $\hat{B}_{\text{KOPT}}$ ***should be undefined with respect to its competitors***. The HKN simulation study is doubtful, but would be illuminated if they published the code used in their simulation study. However, no published code is provided, (Singh and Arnab (2011) have published their code), and further and sufficient evidence of doubtfulness of their simulation study is that a non-computable parameter B_{KOPT} is being estimated through the simulation study.

ACKNOWLEDGEMENTS

The author is heartily thankful to The Chief Editor, Prof. Munir Ahmad, and Dr. Zahoor Ahmed for the compliment, “It is a very important observation that you had made” and for accepting this reply for publication. The author would also like to thank Dr. Polly D. Allred, Department of Mathematics, Texas A&M University-Kingsville, for editing the original version of this reply. At the end, the author would like to request that use of insulting language for someone’s work like the word ‘worst’ should be avoided in professional journals.

REFERENCES

1. Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
2. Hidiroglou, M.A., Kim, J.K. and Nambeu, C.O. (2016). A note on regression estimation with unknown population size. *Survey Methodology*, 42(1), 121-135, Statistics Canada, Catalogue No. 12-001-X.
3. Kim, J.K. (2010). Calibration estimation using exponential tilting in sample surveys. *Survey Methodology*, 36, 145-155.
4. Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael selected Amy.* Vol. 1 and 2, Kluwer Academic Publisher, The Netherlands.
5. Singh, S. and Arnab, R. (2011). On the calibration of design weights. *Metron*, LXIX(2), 185-205.
6. Wu, C. and Sitter, R.R. (2001). A model calibration approach to using complete auxiliary information from survey data. *J. Amer. Statist. Assoc.*, 96, 185-193.