# Bias Interrupters Developed Estimators' Network (BIDEN) with a mixture of TRUMP Cuts and Jackknifing

Sarjinder Singh and Stephen A. Sedory
Department of Mathematics
Texas A&M University-Kingsville,
Kingsville, TX 78363
**E-mail**:sarjinder@yahoo.com

## Abstract

Quenouille (1956) introduced the idea that Jackknifing can be used to reduce bias resulting from using the ratio estimator due to Cochran (1940). Singh and Sedory (2017a) proposed the Tuned Ratio Unbiased Mean Predictor (TRUMP) where they introduced the idea of TRUMP Cuts. In this paper, we introduce the Bias Interrupters Developed Estimators' Network (BIDEN) which utilises the help of TRUMP Cuts and Jackknifing. We show that proper use of what we call BIDEN Care Coefficients could reduce the bias when using the ratio estimator even more than that obtained when using Quenouille's method. It could also be made more efficient than the sample mean estimator with an appropriated choice of TRUMP Care Coefficient. These new findings are supported with exact numerical computations using a well known set of data available in Horvitz and Thompson (1952).

**Keywords**: Ratio estimator, Bias, Quenouille's method, Relative efficiency, TRUMP Cuts, Jackknifing.

## 1. Introduction

Suppose a population $\Omega$ consisting of $N$ units has a study variable $y$ and auxiliary variable $x$. Let $(y_i, x_i), i = 1,2,...,N$ be the ordered pairs of values of the study variable and the auxiliary variable. Consider a sample $s$ of $n$ units taken from the population $\Omega$ of $N$ units by using SRSWOR design. Let $(y_i, x_i)$, $i = 1,2,...,n$ be the values of the study variable and the auxiliary variable associated with the ith unit in the sample. Select a sub-sample $s_1$ of $n_1$ units from the given sample $s$ of $n$ units by SRSWOR design. In other words, randomly divide the sample $s$ of $n$ into two sub-samples $s_1$ and $s_2$ of sizes $n_1$ and $n_2$ where $n_1 + n_2 = n.$

Let $\overline{Y} = N^{-1} \sum_{i\in\Omega} y_i$ be the population mean of the study variable $y,$ which we intend to estimate, and let $\overline{X} = N^{-1} \sum_{i\in\Omega} x_i$ be the population mean of the auxiliary variable $x,$ which we assume to be known.

Further let

$S_y^2 = (N-1)^{-1} \sum_{i\in\Omega} (y_i - \overline{Y})^2$ be the population mean square for the study variable,

$S_x^2 = (N-1)^{-1} \sum_{i\in\Omega} (x_i - \overline{X})^2$ be the population mean square for the auxiliary variable, and

$S_{xy} = (N-1)^{-1} \sum_{i\in\Omega} (y_i - \overline{Y})(x_i - \overline{X})$ be the population covariance between the study variable and the auxiliary variable.

Similarly let

$\bar{y}_n = \frac{1}{n} \sum_{i \in s} y_i$ be the sample mean estimator, based on sample $s$, of the population mean $\bar{Y}$, $\bar{x}_n = \frac{1}{n} \sum_{i \in s} x_i$ be the sample mean estimator, based on sample $s$ , of the population mean $\bar{X}$, $s_y^2 = (n-1)^{-1} \sum_{i \in s} (y_i - \bar{y}_n)^2$ be the sample variance estimator, based on sample $s$, of the population mean square $S_y^2$, $s_x^2 = (n-1)^{-1} \sum_{i \in s} (x_i - \bar{x}_n)^2$ be the sample variance estimator, based on sample $s$, of the population mean square $S_x^2$, and $s_{xy} = (n-1)^{-1} \sum_{i \in s} (y_i - \bar{y}_n)(x_i - \bar{x}_n)$ be the sample covariance estimator, based on sample $s$, of the population covariance $S_{xy}$.

Further let

$\bar{y}_{n_1} = \frac{1}{n_1} \sum_{i \in s_1} y_i$ be the sample mean estimator, based on the sub-sample $s_1$, of the population mean $\bar{Y}$, and $\bar{x}_{n_1} = \frac{1}{n_1} \sum_{i \in s_1} x_i$ be the sample mean estimator, based on the sub-sample $s_1$, of the population mean $\bar{X}$, and let $\bar{y}_{n_2} = \frac{1}{n_2} \sum_{i \in s_2} y_i$ be the sample mean estimator, based on the sub-sample $s_2$, of the population mean $\bar{Y}$, and $\bar{x}_{n_2} = \frac{1}{n_2} \sum_{i \in s_2} x_i$ be the sample mean estimator, based on the sub-sample $s_2$, of the population mean $\bar{X}$.

In 1940, Cochran first introduced the ratio estimator into the field of survey sampling.

**Fig. 1.1.** Professor W.G. Cochran (1909-1980)

Let

$$\bar{y}_{\text{RAT}_1} = \bar{y}_{n_1}\left(\frac{\bar{X}}{\bar{x}_{n_1}}\right)$$

be the ratio estimator of the population mean $\bar{Y}$ based on the first sub-sample $s_1$ of $n_1$ units;

$$\bar{y}_{\text{RAT}_2} = \bar{y}_{n_2}\left(\frac{\bar{X}}{\bar{x}_{n_2}}\right)$$

be the ratio estimator of the population mean $\bar{Y}$ based on the second sub-sample $s_2$ of $n_2 = (n - n_1)$ units, and

$$\bar{y}_{\text{RAT}} = \bar{y}_{n}\left(\frac{\bar{X}}{\bar{x}_{n}}\right)$$

be the ratio estimator of the population mean $\bar{Y}$ based on the full sample $s$ of $n$ units.

In 1965, Quenouille considered the estimator of the population mean $\bar{Y}$ given by

$$\bar{y}_Q = a\left(\bar{y}_{\text{RAT}_1} + \bar{y}_{\text{RAT}_2}\right) + (1 - 2a)\bar{y}_{\text{RAT}}$$

**Fig. 1.2.** A scientist smiling on a success

We imagine a smile crossing his face as he showed the ratio estimator is unbiased, to the second order of approximation, if

$$a = \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{2}{n} - \frac{1}{n_1} - \frac{1}{n_2}\right)^{-1}$$

In particular, for the case of $n_1 = n_2 = n$, the value of $a$ becomes

$$a = -\frac{(N - 2n)}{2N} = Q \text{ (say)}$$

which is called the pioneer Quenouille (1956) bias adjusting constant.

In the next section, we introduce the new idea of Bias Interrupters Developed Estimators' Network (BIDEN) which, with the right choice of TRUMP Care Coefficient, helps to reduce bias in the ratio estimator and minimize the variance.

## 2. The BIDEN

We define the Bias Interrupters Developed Estimators' Network (BIDEN) as:

$$\bar{y}^{*}_{\text{BIDEN}} = b_0 \bar{y}_{RAT} + b_1 \bar{y}^{J}_{RAT(TC)} + b_2 \bar{y}_{POWER}$$

where $b_0$, $b_1$ and $b_2$ are what we refer to as the three BIDEN Care Coefficients, which respect the natural constraint of unity:

$$b_0 + b_1 + b_2 = 1 .$$

The network of associated ratio estimators consists of:

$$\bar{y}_{\text{RAT}} = \bar{y}_n \left( \frac{\bar{X}}{\bar{x}_n} \right),$$

$$\bar{y}^{J}_{\text{RAT(TC)}} = \bar{y}^{TC}_{n_1} \left( \frac{\bar{X}}{\bar{x}^{J}_{n_1}} \right),$$

and

$$\bar{y}_{\text{POWER}} = \bar{y}_{n_2} \left( \frac{\bar{X}}{\bar{x}_{n_2}} \right)^{g} .$$

Following Singh and Sedory (2017a), we define a very special kind of TRUMP Cuts as:

$$\bar{y}^{TC}_{n_1} = \frac{n^{g} \bar{y}_{n-n_1} - (n-n_1)^{g} \bar{y}_n}{n^{g} - (n-n_1)^{g}}$$

where $g \neq 0$ is called the TRUMP Care Coefficient.

If $g = -1$, then

$$\bar{y}^{TC}_{n_1} = \frac{n^{-1} \bar{y}_{n-n_1} - (n-n_1)^{-1} \bar{y}_n}{n^{-1} - (n-n_1)^{-1}} = \bar{y}_{n_1}$$

that is, jackknifing is a special case of TRUMP Cuts.

We also define a jackknifed mean as:

$$\bar{x}^{J}_{n_1} = \frac{n \bar{x}_n - (n-n_1) \bar{x}_{n-n_1}}{n_1}$$

Clearly $\bar{y}^{J}_{\text{RAT(TC)}}$ is a mixture of TRUMP Cuts and Jackknifing, and $\bar{y}_{\text{POWER}}$ is a type of power transformation estimator introduced by Srivastava (1967).

**Fig. 2.1**. TRUMP Cuts, Jackknifing and assembling workshops.

Following Singh, Sedory, Rueda, Arcos and Arnab (2016, p. 283-4), we also imagine that TRUMP Cuts and Jackknifing of a sample can make precise cogs for an estimator. The cogs assembled together with a proper choice of TRUMP Care Coefficient helps the BIDEN to reduce bias from the ratio estimator, and also provides a Jury of Estimators (JOE) where BIDEN is more efficient than the sample mean estimator. In business food companies similarly cut long vegetables (radish, carrots, etc.) on the bias to expose more surface area to increase interaction between food and tongue by choosing a good-flavour $(g)$.

**Fig. 2.2**. Office work

**Theorem 1**. The Bias Interrupters' Developed Estimators' Network (BIDEN) given by:

$$\bar{y}^*_{BIDEN} = b_0\bar{y}_{RAT} + b_1\bar{y}^J_{RAT(TC)} + b_2\bar{y}_{\text{POWER}}$$

is unbiased, to the second order of approximation, for the three BIDEN Care Coefficients given by:

$$b_0 = \frac{\left(\frac{1}{n_1}-\frac{1}{N}\right)\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g-1)}{2}+\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g+1)}{2}\left\{\frac{1}{n}-\frac{1}{n_1}-\frac{n^{g-1}}{n^g-(n-n_1)^g}\right\}}{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g-1)}{2}\left(\frac{1}{n_1}-\frac{1}{n}\right)+\left\{\frac{1}{n}-\frac{1}{n_1}-\frac{n^{g-1}}{\{n^g-(n-n_1)^g\}}\right\}\left\{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g+1)}{2}-\left(\frac{1}{n}-\frac{1}{N}\right)\right\}}$$

$$b_1 = \frac{-\left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g-1)}{2}}{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g-1)}{2}\left(\frac{1}{n_1}-\frac{1}{n}\right)+\left\{\frac{1}{n}-\frac{1}{n_1}-\frac{n^{g-1}}{\{n^g-(n-n_1)^g\}}\right\}\left\{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g+1)}{2}-\left(\frac{1}{n}-\frac{1}{N}\right)\right\}}$$

and

$$b_2 = \frac{-\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\frac{1}{n}-\frac{1}{n_1}-\frac{n^{g-1}}{n^g-(n-n_1)^g}\right\}}{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g-1)}{2}\left(\frac{1}{n_1}-\frac{1}{n}\right)+\left\{\frac{1}{n}-\frac{1}{n_1}-\frac{n^{g-1}}{\{n^g-(n-n_1)^g\}}\right\}\left\{\left(\frac{1}{n_2}-\frac{1}{N}\right)\frac{g(g+1)}{2}-\left(\frac{1}{n}-\frac{1}{N}\right)\right\}}$$

**Proof**. See Singh and Sedory (2017b).

In the next section, we consider finding those values of the TRUMP Care Coefficients which help the proposed BIDEN reducing the bias.

## 3. Reproducible Exact Numerical Evidences for the BIDEN

For our numerical illustration, we have selected all possible SRSWOR samples $s$ of size $n=6$ from the well known Horvitz and Thompson (1952) population of size $N=20$.

**Fig. 3.1**. Hardworking for crunching numbers.

We split the sample of size $n=6$ into two sub-samples $s_1$ and $s_2$ each of size $n_1=n_2=3$ units. Thus there is a total of $NSPLITS=\binom{n}{n_1}=\binom{6}{3}=20$ possibilities within each sample. Total number of main samples will be $NITR=\binom{N}{n}=\binom{20}{6}=38{,}760$.

Then we computed the exact biases in the sample mean and the ratio estimators, respectively, as:

$$B(0)=B(\bar{y}_s)=\frac{1}{NITR}\sum_{k=1}^{NITR}\bar{y}_{s|k}-\bar{Y}$$

and

$$B(1)=B(\bar{y}_{RAT})=\frac{1}{NITR}\sum_{k=1}^{NITR}\bar{y}_{RAT|k}-\bar{Y}$$

We also computed the exact values of the biases in the Quenouille's and the BIDEN estimators, respectively, as:

$$B(2)=B(\bar{y}_Q)=\frac{1}{NSPLITS}\sum_{kk=1}^{NSPLITS}\frac{1}{NITR}\sum_{k=1}^{NITR}(\bar{y}_{Q|k})_{kk}-\bar{Y}$$

and

$$B(3)=B\left(\bar{y}^*_{\text{BIDEN}}\right)=\frac{1}{NSPLITS}\sum_{kk=1}^{NSPLITS}\frac{1}{NITR}\sum_{k=1}^{NITR}\left(\bar{y}^*_{\text{BIDEN}|k}\right)_{kk}-\bar{Y}$$

for a given value of $g$.

Further, we also computed the exact percent relative efficiencies of the ratio, the Quenouille's and the BIDEN estimators with respect to the sample mean estimator, respectively, as:

$$RE(1)=\frac{\sum_{k=1}^{NITR}\left(\overline{y}_{s|k}-\overline{Y}\right)^2}{\sum_{k=1}^{NITR}\left(\overline{y}_{RAT|k}-\overline{Y}\right)^2}\times 100\%$$

$$RE(2)=\frac{\sum_{k=1}^{NITR}\left(\overline{y}_{s|k}-\overline{Y}\right)^2}{\frac{1}{NSPLITS}\sum_{kk=1}^{NSPLITS}\sum_{k=1}^{NITR}\left[\left(\overline{y}_{Q|k}\right)_{kk}-\overline{Y}\right]^2}\times 100\%$$

and

$$RE(3)=\frac{\sum_{k=1}^{NITR}\left(\overline{y}_{s|k}-\overline{Y}\right)^2}{\frac{1}{NSPLITS}\sum_{kk=1}^{NSPLITS}\sum_{k=1}^{NITR}\left[\left(\overline{y}^{*}_{\text{BIDEN}|k}\right)_{kk}-\overline{Y}\right]^2}\times 100\%$$

for a given value of $g$.

We investigated two different situations by using the same Horvitz and Thompson (1952) population but reversing the roles of the variables. This we did to investigate more situations while using the same data set available in the public domains.

**Situation-I.** We treated number of households on the ith block estimated by eye as he study variable $y$ and the actual number of households on the ith block (known from past records) as the auxiliary variable $x$. The value of $B(0)=0$, (Sample mean) $B(1)=0.08926778$, (Ratio estimator), $B(2)=0.01959317$, (Quenouille's estimator) are free from the value of $g$. The value of $RE(1)=345.31\%$ (Ratio estimator) and $RE(2)=303.09\%$ (Quenouille's estimator) are also free from the value of $g$. The values of the three BIDEN Care Coefficients $b_0, b_1$ and $b_2$, $RE(3)$ and $B(3)$ are functions of the TRUMP Care Coefficient $g$. The results obtained for the BIDEN are given in Table 1.

**Table 1**. Results for the new BIDEN under Situation-I.

| $g$ | $b_0$ | $b_1$ | $b_2$ | $\Sigma b_i$ | $RE(3)$ | $B(3)$ | RRB |
|---|---|---|---|---|---|---|---|
| 1.6 | 1.18768 | 0.09136 | -0.27904 | 1 | 170.83 | 0.01785232 | 8.88 |
| 1.7 | 1.14720 | 0.10390 | -0.25110 | 1 | 159.25 | 0.01725104 | 11.95 |
| 1.8 | 1.11170 | 0.11597 | -0.22767 | 1 | 148.99 | 0.01651546 | 15.71 |
| 1.9 | 1.08013 | 0.12764 | -0.20777 | 1 | 139.91 | 0.01563670 | 20.19 |
| 2.0 | 1.05175 | 0.13891 | -0.19066 | 1 | 131.88 | 0.01460708 | 25.45 |

The optimal choice of the TRUMP Care Coefficient is now the BIDEN's choice while compromising between bias and relative efficiency.

**Fig. 3.2**. Optimal TRUMP Care Coefficient.

The value of the percent relative reduction in bias (RRB) is computed as:

$$RRB = \frac{|RB(2)| - |RB(3)|}{|RB(2)|} \times 100\%$$

If the value of the TRUMP Care Coefficient $g$ is 2.0 then the BIDEN $\bar{y}^*_{\text{BIDEN}}$ has the minimum bias value of 0.01460708, which is a relative reduction in bias of 25.45%, and the exact relative efficiency is 131.88%. This is minimum optimal bias in the $\bar{y}^*_{\text{BIDEN}}$ with the help of appropriate TRUMP Care Coefficient. We also claim that when the TRUMP Care Coefficient $g$ with a value of 2.0 is in power $(\bar{y}_{POWER})$ then the BIDEN is quite successful at reducing the bias from the ratio estimator and has almost has 31.88% gain in efficiency over the sample mean estimator. The choice of TRUMP Care Coefficients $g \in [1.6, 2.0]$ , with a skip of 0.1, which we call the First Basic Information (FBI), forms a Jury of Estimators (JOE) of five members where the proposed BIDEN has less bias in magnitude than Quenouille's estimator with relative reduction in the range of 8.88% to 25.45%; and each member of the JOE has relative efficiency more than the sample mean estimator in the range 131.88% and 170.83%. Thus, the TRUMP Care Coefficient helps BIDEN more flexible for the investigator who must compromise between bias and relative efficiency. Further note that the value of RE(3) is less than the RE(2) and RE(1) which illustrated the cost of reducing the bias of the ratio estimator due to Cochran (1940).

**Situation-II**. Here we treat the actual number of households on the ith block as the study variable $y$, and the estimated (by eye) number of households as the auxiliary variable $x$. The value of $B(0) = 0$, $B(1) = 0.01917783,$, $B(2) = 0.00411917$, are free from the value of $g$. The value of $RE(1) = 404.52\%$ and $RE(2) = 369.23\%$ are again free from the value of $g$. The values of the three BIDEN Care Coefficients $b_0$, $b_1$ and $b_2$, $RE(3)$ and $B(3)$ are functions of the TRUMP Care Coefficient $g$. The results obtained for the BIDEN are given in Table 2.

**Table 2**. Results for the new BIDEN under Situation II.

| $g$ | $b_0$ | $b_1$ | $b_2$ | $\Sigma b_i$ | $RE(3)$ | $B(3)$ | RRB |
|---|---|---|---|---|---|---|---|
| 0.86 | 2.04459 | -0.03211 | -1.01248 | 1 | 135.92 | 0.00205136 | 50.20 |
| 0.88 | 1.98001 | -0.02682 | -0.95319 | 1 | 152.11 | 0.00242816 | 41.05 |
| 0.90 | 1.92186 | -0.02183 | -0.90003 | 1 | 168.10 | 0.00277229 | 32.70 |
| 0.92 | 1.86920 | -0.01708 | -0.85212 | 1 | 183.55 | 0.00308773 | 25.04 |
| 0.94 | 1.82126 | -0.01255 | -0.80871 | 1 | 198.14 | 0.00337770 | 18.00 |

Again the optimal choice of TRUMP Care Coefficient depends on the relative importance of bias and the relative efficiency. If the value of the TRUMP Care Coefficient $g$ is 0.86 then the BIDEN $\bar{y}^*_{\text{BIDEN}}$ has minimum bias value of 0.00205136, which is a relative reduction in bias of 50.20%, and the value of the exact relative efficiency is 135.92%. This is minimum bias in the $\bar{y}^*_{\text{BIDEN}}$ with the help of appropriate TRUMP Care Coefficient. We also observe that when the TRUMP Care Coefficient $g$ with a value of 0.86 is in power $(\bar{y}_{POWER})$ then the BIDEN is quite successful in reducing the bias from the ratio estimator and almost 35.92% gain in efficiency over the sample mean estimator. The choice of TRUMP Care Coefficients $g \in [0.86, 0.94]$, with a skip of 0.02, which we say the First Basic Information (FBI), forms a Jury of Estimators (JOE) of five members where the proposed BIDEN has less bias in magnitude than Quenouille's estimator with relative reduction in bias in the range of 18.00% to 50.20% and each member of the JOE has relative efficiency more than the sample mean estimator, in the range 135.92% and 198.14%. Again the TRUMP Care Coefficient helps BIDEN be flexible when one must make a compromise between bias and relative efficiency. Further note that the value of RE(3) is again less than the RE(2) and RE(1) which reveals the cost of reducing the bias from the ratio estimator.

## 4. Conclusion

**Fig. 4.1.** Opening a door

In situation-I the ratio of population mean of the study variable to that of the auxiliary variable is less than one, and in situation-II it is greater than one. That may be one reason that FBI has different range of the TRUMP Care Coefficient for both situations. The proposed BIDEN with a mixture of TRUMP Cuts and Jackknifing opens a big-door to future research for those who are interested in reducing the bias in the ratio type estimators. The Internet is full of such ratio type estimators. Refer to Singh and Sedory (2017b) for more interesting relevant work. R-codes used in producing these results and detailed acknowledgements are also given in Singh and Sedory (2017b).

## Acknowledgements

*The authors are thankful to the R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. All the opinions are of authors' and do not belong to any institute or organization. Any resemblance to actual persons, living or dead, or actual events is purely coincidental.*

## References

Cochran, W.G. (1940). The estimation of yields of cereal experiments by sampling for the ratio of grains to total produce. *Journal of Agricultural Science*, 30, 262-275.

Horvitz, D.G. and Thompson, D.J. (1952). A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663-685.

Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.

Singh, S., Sedory, S.A., Rueda, M. Arcos, A. and Arnab, R. (2016). *A New Concept for Tuning Design Weights in Survey Sampling,* Elsevier Ltd., London, ISBN:978-0-08-100594-1

Singh, S. and Sedory, S.A. (2017a). TRUMP: Tuned Ratio Unbiased Mean Predictor. *Proceedings of the Joint Statistical Meeting, Survey Method Section, pp.1746-1759.*

Singh, S. and Sedory, S.A. (2017b). TRUMP: Tuned Ratio Unbiased Mean Predictor. *Working monograph.*

Srivastava, S.K. (1967). An estimator using auxiliary information in sample surveys. *Calcutta Statistical Association Bulletin*, 16, 121-132.