# Introduction

This document reports the optimality of the proposed non-monotone estimator. It turns out that the proposed estimator is not optimal. We show a better estimator and present a simulation study.

# Setup

- Let $U = \{1, \dots N\}$ be a finite population.
- Consider the random variables $(X_i, Y_{1i}, Y_{2i}, \pi_i) \overset{\text{ind}}{\sim} F$ for some unknown $F$ for all $i \in U$.
- We assume that $x_i$ is observed throughout the finite population.
- Let $I_i \overset{\text{ind}}{\sim} \pi_i$ for $i \in U$ be the first phase sample inclusion indicator.
- If $i \in \{1, \dots, N\}$ is in the first phase sample ($I_i = 1$) then $\pi_i$ is observed.
- The second phase sampling indicators are $\delta_{00}$, $\delta_{01}$, $\delta_{10}$, and $\delta_{11}$. These mutual exclusive variables encode the following: if both $Y_1$ and $Y_2$ are observed then $\delta_{11} = 1$; if $Y_1$ is observed and $Y_2$ is missing then $\delta_{10} = 1$; if $Y_1$ is missing and $Y_2$ is observed then $\delta_{01} = 1$; and if both $Y_1$ and $Y_2$ are missing then $\delta_{00} = 1$. (For notational purposes, it is sometimes convienent to write $\delta_1 = \delta_{11} + \delta_{10}$ and $\delta_2 = \delta_{11} + \delta_{01}$. A similar notation will be used for $\pi$. We have $\pi_{1+} = \pi_{11} + \pi_{10}$ and $\pi_{2+} = \pi_{11} + \pi_{01}$.)
- All of the sampling indicators were drawn via *Poisson* sampling. So the sample size of each category $A_{ij}$ is random. However, this does ensure independence between observations.
- The goal is to estimate $\theta = E[g(X, Y_1, Y_2)]$ in the population.
- The proposed estimator is

$$\begin{aligned}
\hat{\theta}_{\text{eff}} &= n^{-1} \sum_{i=1}^{n} E[g_i \mid X_i] \\
&+ n^{-1} \sum_{i=1}^{n} \frac{\delta_{1i}}{\pi_{1+}(X_i)} (E[g_i \mid X_i, Y_{1i}] - E[g_i \mid X_i]) \\
&+ n^{-1} \sum_{i=1}^{n} \frac{\delta_{2i}}{\pi_{2+}(X_i)} (E[g_i \mid X_i, Y_{2i}] - E[g_i \mid X_i]) \\
&+ n^{-1} \sum_{i=1}^{n} \frac{\delta_{1i}\delta_{2i}}{\pi_{11}(X_i)} (g_i - E[g_i \mid X_i, Y_{1i}] - E[g_i \mid X_i, Y_{2i}] + E[g_i \mid X_i]) \qquad (1)
\end{aligned}$$

- The goal is to show that this estimator is optimal within the class of estimators:

$$\begin{aligned}\hat{\theta}_{\text{eff}} &= n^{-1}\sum_{i=1}^{n} E[g_i \mid X_i] \\ &+ n^{-1}\sum_{i=1}^{n} \frac{\delta_{1i}}{\pi_{1+}(X_i)}(b(X_i, Y_{1i}) - E[g_i \mid X_i]) \\ &+ n^{-1}\sum_{i=1}^{n} \frac{\delta_{2i}}{\pi_{2+}(X_i)}(a(X_i, Y_{2i}) - E[g_i \mid X_i]) \\ &+ n^{-1}\sum_{i=1}^{n} \frac{\delta_{1i}\delta_{2i}}{\pi_{11}(X_i)}(g_i - b_i - a_i + E[g_i \mid X_i]). \end{aligned} \tag{2}$$

# Results

The first part recaps what Dr. Fuller and I discussed previously.

- Let $g = E[Y_2]$, and suppose that we know the distribution of $X$ and the covariance structure of $[X, Y_1, Y_2, \pi]$. Then instead of modeling the relationship between $X$ and $Y_k$, we can use difference estimators: $Z_1 \equiv Y_1 - b_1\tilde{X}$ and $Z_2 =\equiv Y_2 - b_2\tilde{X}$, where $\tilde{X} = X - E[X]$. To minimize the variance of $Z_k$ we can choose the optimal value of $b_k$, which is

$$b_k = \frac{\text{Cov}(Y_k, X)}{\text{Var}(X)}.$$

  Since these covariance values are known, $b_k$ is known. This means that we now have the following table:

| | $Z_1$ | $Z_2$ |
|---|---|---|
| $A_{11}$ | ✓ | ✓ |
| $A_{10}$ | ✓ | |
| $A_{01}$ | | ✓ |
| $A_{00}$ | | |

  Because we assumed that the distribution of $X$ is known, the section $A_{00}$ contains no additional information about $Y_1$ or $Y_2$. Let $\mu_k = E[Y_k]$.

- We now consider a normal model:

$$\begin{pmatrix} x_i \\ e_{1i} \\ e_{2i} \end{pmatrix} \overset{ind}{\sim} N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma_{11} & \sigma_{12} \\ 0 & \sigma_{12} & \sigma_{22} \end{bmatrix}, \right)$$

and define $y_{1i} = \theta_1 + x_i + e_{1i}$ and $y_{2i} = \theta_2 + x_i + e_{2i}$. Then $b_1 = b_2 = 1$. We define $\bar{z}_k^{(ij)}$ as the mean of $y_k$ in segment $A_{ij}$. This means that we have means $\bar{z}_1^{(11)}$, $\bar{z}_2^{(11)}$, $\bar{z}_1^{(10)}$, and $\bar{z}_2^{(01)}$. Let $W = [\bar{z}_1^{(11)}, \bar{z}_2^{(11)}, \bar{z}_1^{(10)}, \bar{z}_2^{(01)}]'$, then for $n_{ij} = |A_{ij}|$, we have

$$Z - M\mu \sim N(\vec{0}, V)$$

where

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } V = \begin{bmatrix} \frac{\sigma_{11}}{n_{11}} & \frac{\sigma_{12}}{n_{11}} & 0 & 0 \\ \frac{\sigma_{12}}{n_{11}} & \frac{\sigma_{22}}{n_{11}} & 0 & 0 \\ 0 & 0 & \frac{\sigma_{11}}{n_{10}} & 0 \\ 0 & 0 & 0 & \frac{\sigma_{22}}{n_{01}} \end{bmatrix}.$$

Thus, the BLUE for $\mu = [\mu_1, \mu_2]'$ is

$$\hat{\mu} = (M'V^{-1}M)^{-1}M'V^{-1}W. \tag{3}$$

- Since this is the BLUE, we would expect it to be at least as good as the proposed estimator. And if the proposed estimator is optimal it should be equivalent to the BLUE in the case that $X$, $Y_1$ and $Y_2$ are normal. So I ran a simulation study to test this. For $i = \{1, \ldots, n = 1000\}$,

$$\begin{pmatrix} x_i \\ e_{1i} \\ e_{2i} \end{pmatrix} \overset{ind}{\sim} N\left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma_{12} \\ 0 & \sigma_{12} & 1 \end{bmatrix}, \right) \text{ and } y_{1i} = \theta_1 + x_i + e_{1i}, y_{2i} = \theta_2 + x_i + e_{2i}.$$

Each observation $i$ was then assigned a segment $A_{11}$, $A_{10}$, $A_{01}$ or $A_{00}$ independently with each draw having probability $p_{11} = 0.4$, $p_{10} = 0.2$, $p_{01} = 0.2$ and $p_{00} = 0.2$ respectively. This means that $\pi_{11} = 0.4$ and $\pi_{1+} = 0.6 = \pi_{2+}$. We let $g = E[Y_2]$ and we test the following estimators:

  - Oracle: This computes the average value of $Y_2$ in all observations (even when $Y_2$ is not supposed to be observed).

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} y_{2i}.$$

  - OracleX: This computes the average value of $Y_2 - X$ in all observations (even when $Y_2$ is not supposed to be observed.)

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} (y_{2i} - x_i).$$

- CC: This computes the average value of $Y_2$ in all observations in which $Y_2$ is observed.
$$\hat{\theta} = \frac{\sum_{i=1}^{n} \delta_{2i} y_{2i}}{\sum_{i=1}^{n} \delta_{2i}}.$$
- Proposed: This is the proposed estimator from Equation 1.
- WLS: This is the weighted linear estimator from Equation 3. Note, since $g = E[Y_2]$ this only contains the second element from Equation 3.

The results are shown in the table below. This simulation was run with the number of observations $n = 1000$ and the Monte Carlo sample size of $B = 3000$.

Table 1: This table shows the estimators of $\theta = E[Y_2]$. The true value of $\theta$ is 5 and the true value of $\sigma_{12} = 0.5$. The bias column shows the average bias of the estimator and the actual value of $\theta = 5$ across the $B = 3000$ simulations. The SD column shows the average standard deviation across the $B = 3000$ simulations for each algorithm. The Tstat column displays the t-statistic of a t-test comparing the estimator to the actual value. The value of this column is computed via $\frac{\bar{\hat{\theta}} - \theta}{\sqrt{\text{Var}\hat{\theta}/B}}$. The Pval column displays the p-value of the t-statistic.

| Algorithm | Bias | SD | Tstat | Pval |
|---|---|---|---|---|
| Oracle | 0.001 | 0.044 | 1.546 | 0.061 |
| OracleX | 0.000 | 0.032 | 0.479 | 0.316 |
| CC | 0.002 | 0.058 | 1.549 | 0.061 |
| WLS | 0.001 | 0.040 | 1.207 | 0.114 |
| Prop | 0.002 | 0.051 | 1.918 | 0.028 |

# Computing the Variance by Hand

Table  demonstrates that the WLS estimator outperforms the Oracle estiamator. I was initially confused about why this is the case, but realized that I could approximate the standard error of both estimators reasonably well. It turns out that I was able to get reasonable estimates for the variance of all of the proposed estimators includeing $\hat{\theta}_{prop}$. In this part of the report, I detail my computations and show how the simulated results are accurate.

$$\begin{aligned} \text{Var}(\hat{\theta}_{Oracle}) &= \text{Var}\left( n^{-1} \sum_{i=1}^{n} y_{2i} \right) \\ &= n^{-1} \text{Var}(y_2) \\ &= n^{-1}(1 + \sigma_{22}) \qquad \text{because Var}(x) = 1. \end{aligned} \tag{4}$$

$$\begin{aligned}\text{Var}(\hat{\theta}_{OracleX}) &= \text{Var}\left(n^{-1}\sum_{i=1}^{n} y_{2i} - x_i\right) \\ &= n^{-1}\text{Var}(e_2) \\ &= n^{-1}\sigma_{22}. \end{aligned} \tag{5}$$

Using a Taylor expansion,

$$\text{Var}(\hat{\theta}_{cc}) = \text{Var}\left(\theta_2 + \pi_{2+}^{-1}\left(n^{-1}\sum_{i=1}^{n}\delta_{2i}y_{2i} - \pi_{2+}\theta_2\right) - \frac{\theta_2\pi_{2+}}{\pi_{2+}^2}\left(n^{-1}\sum_{i=1}^{n}\delta_{2i} - \pi_{2+}\right) + o_p(1)\right)$$

(I can add more steps later if necessary.) (6)

$$= \frac{1+\sigma_{22}}{n\pi_{2+}} + o_p(1). \tag{7}$$

To compute the variance of $\hat{\theta}_{WLS}$, we have,

$$\begin{aligned}\text{Var}(\hat{\theta}_{WLS}) &= \text{Var}((M'V^{-1}V)^{-1}M'V^{-1}Z) \\ &= E[\text{Var}((M'V^{-1}V)^{-1}M'V^{-1}Z \mid n_1)] \\ &= E[(M'V^{-1}M)] \end{aligned} \tag{8}$$

$$\tag{9}$$

By Slutky's Theorem, $\text{Var}(\hat{\theta}_{WLS}) \xrightarrow{P} (M'E[V]^{-1}M)^{-1}$.

Finally, we compute the variance for the proposed estimator. First, we estimate the variance for a similar estimator $\tilde{\theta}_{prop}$. This estimator is the same as the proposed estimator except that we assume that the estimates from $E[Y_{2i} \mid X, Y_1]$ and $E[Y_{2i} \mid X]$ are population coefficients and independent of the rest of the sample. This could occur if there is a separate sample in which we can estimate these expectations. Since the population coefficients of the regression $Y = \beta_0 + \beta_1 X$ are

$$\beta_0 = E[Y] - E[X]\text{Var}(X)^{-1}Cov(X'Y) \text{ and } \beta_1 = \text{Var}(X)^{-1}\text{Cov}(X, Y),$$

we have that,

$$E[Y_2 \mid X] = E[Y_2] - \beta_1 E[X] + \beta_1 x = \theta_2 - x$$

and

$$\begin{aligned}
E[Y_2 \mid X, Y_1] &= \beta_0 + [X, Y_1][\beta_1 \beta_2]' \\
&= E[Y_2] - E[[X, Y_1]]\text{Var}([X, Y_1])^{-1}\text{Cov}([X, Y_1], Y_2) + [X, Y_1]\text{Var}([X, Y_1])^{-1}\text{Cov}([X, Y_1], Y_2) \\
&= \theta_2 + [X, Y_1]\sigma_{11}^{-1} \begin{bmatrix} 1+\sigma_{11} & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1+\sigma_{12} \end{bmatrix} \\
&= \theta_2 + x\left(\frac{\sigma_{11} - \sigma_{12}}{\sigma_{11}}\right) + y_1\left(\frac{\sigma_{12}}{\sigma_{11}}\right) \\
&= \theta_2 + x(1 - \sigma_{12}) + y_1(\sigma_{12}).
\end{aligned}$$

This means that

$$E[Y_2 \mid X, Y_1] - E[Y_2 \mid X] = \sigma_{12}(y_1 - x) = \sigma_{12}e_1.$$

Hence, whereas for $\theta = E[Y_2]$,

$$\hat{\theta}_{prop} = n^{-1}\sum_{i=1}^{n}\left\{E[Y_2 \mid X] + \left(\frac{\delta_{1i}}{\pi_{1+}} - \frac{\delta_{11i}}{\pi_{11}}\right)(E[Y_2 \mid X, Y_1] - E[Y_2 \mid X]) + \frac{\delta_{2i}}{\pi_{2+}}(y_{2i} - E[Y_2 \mid X])\right\}$$

for the similar estimator we have

$$\tilde{\theta}_{prop} = n^{-1}\sum_{i=1}^{n}\left(\theta_2 + x_i + \left(\frac{\delta_{1i}}{\pi_{1+}} - \frac{\delta_{11i}}{\pi_{11}}\right)\sigma_{12}e_{1i} + \frac{\delta_{2i}}{\pi_{2+}}(e_{2i})\right)$$

This means that

$$\begin{aligned}
\text{Var}(\tilde{\theta}_{prop}) &= n^{-1}\text{Var}\left(x + \left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\sigma_{12}e_1 + \frac{\delta_2}{\pi_{2+}}(e_2)\right) \\
&= n^{-1}\left(\text{Var}(x) + \text{Var}\left(\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\sigma_{12}e_1\right) + \text{Var}\left(\frac{\delta_2}{\pi_{2+}}(e_2)\right)\right. \\
&\qquad \left. + 2\text{Cov}\left(\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\sigma_{12}e_1, \frac{\delta_2}{\pi_{2+}}(e_2)\right)\right) \\
&\equiv A + B + C + D.
\end{aligned}$$

$$A = \text{Var}(x) = 1.$$

$$\begin{aligned}
B &= \sigma_{12}^2\sigma_{11}\text{Var}\left(\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\right) \\
&= \sigma_{12}^2\sigma_{11}\left(\text{Var}(\delta_1/\pi_{1+}) + \text{Var}(\delta_{11}/\pi_{11}) - 2\text{Cov}(\delta_1/\pi_{1+}, \delta_{11}/\pi_{11})\right) \\
&= \sigma_{12}^2\sigma_{11}(\pi_{1+}^{-1} + \pi_{11}^{-1} - 2 - 2(\pi_{11} - \pi_{1+}\pi_{11})/(\pi_{1+}\pi_{11})) \\
&= \sigma_{12}^2\sigma_{11}(\pi_{11}^{-1} - \pi_{1+}^{-1}).
\end{aligned}$$

$$\begin{aligned}
C &= \text{Var}(\delta_2 e_2/\pi_{2+}) \\
&= \pi_{2+}^{-2}(\text{Var}(\delta_2)\text{Var}(e_2) + \text{Var}(\delta_2)E[e_2]^2 + \text{Var}(e_2)E[\delta_2]^2) \\
&= \pi_{2+}^{-2}(\pi_{2+}(1-\pi_{2+})\sigma_{22} + \sigma_{22}\pi_{2+}^2) \\
&= \sigma_{22}/\pi_{2+}.
\end{aligned}$$

$$\begin{aligned}
D &= 2\text{Cov}\left(\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\sigma_{12}e_1, \frac{\delta_2}{\pi_{2+}}(e_2)\right) \\
&= 2E\left[\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\sigma_{12}e_1\frac{\delta_2}{\pi_{2+}}(e_2)\right] \\
&= 2E\left[\left(\frac{\delta_1}{\pi_{1+}} - \frac{\delta_{11}}{\pi_{11}}\right)\frac{\delta_2}{\pi_{2+}}\right]\sigma_{12}^2 \\
&= 2E\left[\frac{\delta_{11}}{\pi_{1+}\pi_{2+}} - \frac{\delta_{11}}{\pi_{2+}}\right]\sigma_{12}^2 \\
&= 2\frac{\pi_{11}}{\pi_{2+}}(\pi_{1+}^{-1} - 1)\sigma_{12}^2.
\end{aligned}$$

Hence,

$$\text{Var}(\tilde{\theta}_{prop}) = n^{-1}(1 + \sigma_{12}^2\sigma_{11}(\pi_{11}^{-1} - \pi_{1+}^{-1}) + \sigma_{22}\pi_{2+}^{-1}2\sigma_{12}^2(\pi_{1+}^{-1} - 1)\pi_{11}/\pi_{2+}). \tag{10}$$

Since $\hat{\beta} - \beta = O(n)$ from Theorem 2.2.1 of [1], $\text{Var}(\hat{\theta}_{prop}) \to \text{Var}(\tilde{\theta}_{prop})$. (I think that I also need to show that the dependency between the estimated coefficients and the observed data is bounded and that this bound goes to zero as the sample size increases.)

These estimated variances are almost exactly in line with the estimated standard deviations from the simulation study. With values of $n = 1000$, $\sigma_{11} = 1$, $\sigma_{22} = 1$, $\sigma_{12} = 0.5$, $\theta_2 = 5$, $\pi_{11} = 0.4$, $\pi_{10} = 0.2$, $\pi_{01} = 0.2$, and $\pi_{00} = 0.2$ the standard deviations from the previous calculations and from the Monte Carlo simulation are listed in Table 2:

Table 2: This table compares the calculated and estimated standard errors for each of the algorithms listed. The calculated values are computed as the square root of the variance derived in the previous section. We use Equations 4, 5, 6 8, and 10 to calculate the variance of Oracle, OracleX, CC, WLS, and Prop respectively and then take the square root of the answer. For the estimated standard error, we take the standard deviation of the $B$ estimators for each algorithm.

| Algorithm | Calculated Standard Error | Estimated Standard Error |
|---|---|---|
| Oracle | 0.045 | 0.044 |
| OracleX | 0.032 | 0.032 |
| CC | 0.058 | 0.058 |
| WLS | 0.040 | 0.040 |
| Prop | 0.056 | 0.051 |

# Non-linear Estimation

- While we have previously shown that in the case of the linear estimator, that the weighted least squares (WLS) estimator is superior to the proposed estimator. However, we also want to see if this holds for a non-linear function. Consider the estimating equation $g(X, Y_1, Y_2) = Y_1^2 Y_2$. Using the same setup as the previous simulation, we have the same definitions of $X$, $Y$, $n$, $\pi$, and $B$.

  For the definition of each estimator we have the following algorithms:

  - Oracle: This computes the average value of $Y_1^2 Y_2$ in all observations (even when $Y_1$ or $Y_2$ is not supposed to be observed).

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} y_{1i}^2 y_{2i}.$$

  - OracleX: This computes the average value of $Y_1^2(Y_2 - X)$ in all observations (even when $Y_1$ or $Y_2$ is not supposed to be observed.)

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} y_{1i}^2 (y_{2i} - x_i).$$

  - CC: This computes the average value of $Y_1^2 Y_2$ in all observations in which $Y_1$ and $Y_2$ are observed.

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \delta_{1i}\delta_{2i} y_{1i}^2 y_{2i}}{\sum_{i=1}^{n} \delta_{1i}\delta_{2i}}.$$

  - Proposed: This is the proposed estimator from Equation 1. Note that since,

$$\begin{aligned} Y_1^2 Y_2 &= (x + e_1)^2(\theta_2 + x + e_2) \\ &= x^3 + x^2(\theta_2 + 2e_1 + e_2) + x(2\theta_2 e_1 + 2e_1 e_2 + e_1^2) + e_1^2\theta_2 + e_1^2 e_2 \end{aligned}$$

    for the conditional expectations, we have

$$\begin{aligned} E[Y_1^2 Y_2 \mid X] &= x^3 + x^2\theta_2 + x(2\sigma_{12} + 1) + \theta_2 \\ E[Y_1^2 Y_2 \mid X, Y_1] &= Y_1^2(x + \theta_2) \text{ and} \\ E[Y_1^2 Y_2 \mid X, Y_2] &= (x^2 + 1)Y_2. \end{aligned}$$

  - WLS: To derive the WLS estimator we can view the sample as consisting of four independent parts: $A_{11}, A_{10}, A_{01}$, and $A_{00}$, where $A_{ij}$ indicates the missingness levels of $Y_1$ (if i = 0) or $Y_2$ (if $j = 0$). We know that the correctly specified regression model for $g$ in each of these segments is the following:

$$\begin{aligned} &A_{11} : Y_1^2 Y_2 \\ &A_{10} : Y_1^2(x + \theta_2) \\ &A_{01} : (x^2 + 1)Y_2 \\ &A_{00} : x^3 + x^2\theta_2 + x(2\sigma_{12} + 1) + \theta_2 \end{aligned}$$

The estimator is then to estimate $\theta_2$ using Equation 3. Then, we can plug in the estimated value of $\theta_2$ into the previous equations to get estimates of $g$. The overall estimate is the estimated variance weighted average of the mean estimate from each segment.

- WLSTT: Since the previous estimator is biased, we use the same estimation procedure with the true value of $\theta_2$ to the estimator $\hat{\theta}_{WLSTT}$.

Table 3: True g is 10. $\text{Cov}(e_1, e_2) = 0.5$

| Algorithm | Bias | SD | Tstat | Pval |
|---|---|---|---|---|
| Oracle | 0.007 | 0.529 | 0.741 | 0.229 |
| Oraclex | 0.004 | 0.475 | 0.510 | 0.305 |
| CC | 0.023 | 0.824 | 1.518 | 0.065 |
| WLS | -0.131 | 0.387 | -18.504 | 0.000 |
| WLSTT | -0.132 | 0.380 | -19.007 | 0.000 |
| Prop | 0.011 | 0.625 | 0.952 | 0.171 |

In this scenario, the WLS estimator is biased. Other than the oracle estimators, the proposed estimator is the best unbiased estimator.

# Next Steps

- I am slightly unclear about how to proceed. The proposed estimator seemed to work quite well for the non-linear experiment. However, it was easily out-performed by the WLS estimator in the linear case. The WLS estimator is a relatively intuitive estimator that is optimal in that situation. However, it seems to have worse performance in the non-linear case.

# References

[1] Wayne A Fuller. *Sampling statistics*. John Wiley & Sons, 2009.