Rao and Bellhouse’s Theorem in the Context of Simple Random Sampling with Response Error

Ruitao Zhang

Introduction

Rao and Bellhouse (1978) state in Theorem 1.1 conditions under which an estimator is optimal in a space. The space in this paper is the class of linear unbiased predictors of a first-stage sample units (PSUs) based on the expanded random variables which identify both unit and position. The idea that we develop here is how this theorem can be applied to a simple random sample with response error to show that a predictor of the first-stage sample units (PSUs), can not be optimally developed from the usual random variables.

Background and Notation

A finite population is defined by a listing of $N$ subjects, indexed by $s = 1, ..., N$, where the $kth$ response for subject $s$ is given by

$$Y_{sk} = y_s + W_{sk}.$$ 

This is a response error model where $y_s$ is a fixed constant representing the expected response for the unit $s$, and $W_{sk}$ represents response error (with zero expected value). We denote the expected value with respect to sampling via a subscript $\xi$, and expected value with respect to response error via a subscript $R$. The response error variance for unit $s$ is given by

$$\text{var}_R (W_{sk}) = \sigma^2_{se} \quad \text{for all } s = 1, ..., N, \quad k = 1, ..., r_s.$$ 

And we assume that
\[
\text{var}_R(W_{sk}, W_{sk^{*}}) = 0 \text{ for all } s = 1, \ldots, N; \ s^{*} = 1, \ldots, N; \ k = 1, \ldots, r_s; \ k^{*}(\neq k) = 1, \ldots, r_s.
\]

In our context, we set \( r_s = 1 \) for all \( s = 1, \ldots, N \) so that \( k = 1 \). (1 measure of response).

We define the one-stage random permutation model as an ordered list of \( N \) random variables, where units are independently permuted. For each permutation, we assign a new label, \( i = 1, \ldots, N \), to the units in a permutation according to their position in the permuted list. We use sampling indicator random variables, \( U_{is} \), which take a value of 1 when the unit in position \( i \) is unit \( s \) to indicate which unit occurs at position \( i \). When there is no measurement error, we denote response for position \( i \) by \( Y_i = \sum_{s=1}^{N} U_{is} y_s \). When measurement error is present, we denote response for position \( i \) by \( Y^{*}_{ik} = \sum_{s=1}^{N} U_{is} Y_{sk} = Y_i + W^{*}_{ik} \), where \( W^{*}_{ik} = \sum_{s=1}^{N} U_{is} W_{sk} \). This model accounts for the one-stage random permutations with response error.

Letting \( U' = (\begin{pmatrix} U_1^T & U_2^T & \cdots & U_N^T \end{pmatrix}) \) denote the \( N \times N \) matrix with column

\[
U_i = \begin{pmatrix} U_{i1} & U_{i2} & \cdots & U_{IN} \end{pmatrix}^T
\]

and letting \( Y = (\begin{pmatrix} Y_1^T \end{pmatrix}) \), \( y = (\begin{pmatrix} y_s \end{pmatrix}) \) then \( Y = Uy \). Similarly, defining

\[
W = \begin{pmatrix} W_{1k} & W_{2k} & \cdots & W_{Nk} \end{pmatrix}^T
\]

and \( W^{*} = \begin{pmatrix} W^{*}_{1k} \end{pmatrix} \), then \( W^{*} = UW \).

Then the model can be written as:

\[
Y^{*} = Y + W^{*}
\]

Let’s define the Vec of a Matrix as a column vector formed by stacking the columns in the matrix, locating one column under another. For example, suppose \( A_{m \times n} = (a_1, a_2, \ldots, a_n) \) is

\[
m \times n \text{ matrix where each vector, } a_j \text{ is of dimension } m \times 1. \text{ Then } vec(A_{m \times n}) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \text{ By using }
\]

above definition, the model can also be expressed as:
Predictors of linear combinations of PSUs of a finite population in a two-stage sampling setting for above model were developed by Stanek (2004).

In this paper, we use an expanded model to find predictors of linear combinations of PSUs. When there is no measurement error, we define the expanded random variables as

\[
\tilde{Y} = \text{vec}\begin{pmatrix}
U_{11}Y_{1k} & U_{12}Y_{2k} & \cdots & U_{1N}Y_{Nk} \\
U_{21}Y_{1k} & U_{22}Y_{2k} & \cdots & U_{2N}Y_{Nk} \\
\vdots & \vdots & \ddots & \vdots \\
U_{N1}Y_{1k} & U_{N2}Y_{2k} & \cdots & U_{NN}Y_{Nk}
\end{pmatrix}
\]

and

\[
\tilde{y}_i = (U_{i1}Y_{1k}, U_{i2}Y_{2k}, \cdots, U_{iN}Y_{Nk})'.
\]

Similarly, when measurement error is present, we define the expanded random variables as

\[
\tilde{Y}^* = \text{vec}\begin{pmatrix}
U_{11}Y_{1k}^* & U_{12}Y_{2k}^* & \cdots & U_{1N}Y_{Nk}^* \\
U_{21}Y_{1k}^* & U_{22}Y_{2k}^* & \cdots & U_{2N}Y_{Nk}^* \\
\vdots & \vdots & \ddots & \vdots \\
U_{N1}Y_{1k}^* & U_{N2}Y_{2k}^* & \cdots & U_{NN}Y_{Nk}^*
\end{pmatrix}
\]

Defining \( \tilde{y}_i^* = (U_{i1}Y_{1k}^*, U_{i2}Y_{2k}^*, \cdots, U_{iN}Y_{Nk}^*)' \) and \( \tilde{w}_i = \left( \bigoplus_{s=1}^N W_{sk} \right) U_i \), \( \tilde{Y}^* \) can be expressed as:

\[
\tilde{Y}^* = \text{vec}\left( \tilde{y}_1^*, \tilde{y}_2^*, \cdots, \tilde{y}_N^* \right) \text{ or } \tilde{Y}^* = \left[ I_N \otimes \left( \bigoplus_{s=1}^N (y_s + W_{sk}) \right) \right] \text{vec}(U').
\]

We also can express this as:

\[
\tilde{Y}^* = \tilde{Y} + \tilde{W} \text{ or }
\]

\[
\tilde{y}_i^* = \tilde{y}_i + \tilde{w}_i = \left( \bigoplus_{s=1}^N (y_s + W_{sk}) \right) U_i = (U_{i1}(y_{1k} + W_{1k}) \quad U_{i2}(y_{2k} + W_{2k}) \cdots \quad U_{iN}(y_{Nk} + W_{Nk})).
\]
Denote the expected value with respect to sampling via a subscript $\xi$, and the expected value with respect to response error via a subscript $R$. Then

$$E_{\xi}(\bar{Y}) = \left[ \frac{1}{N} \left( I_N \otimes I_N \right) \right] y$$

and

$$\text{var}_{\xi}(\bar{Y}) = P_N \otimes \Delta,$$

where $\Delta = \frac{1}{N-1} \left( \oplus_{s=1}^{N} y_s \right) P_N \left( \oplus_{s=1}^{N} y_s \right)$, and $P_N = I_N - N^{-1}J_N$.

Also, we find that

$$E_{\xi R}(\bar{Y}^*) = \left[ \frac{1}{N} \left( I_N \otimes I_N \right) \right] y$$

and

$$\text{var}_{\xi R}(\bar{Y}^*) = E_{\xi} \left[ \text{var}_{R\xi}(\bar{Y}^*) \right] + \text{var}_{\xi} \left[ E_{R\xi}(\bar{Y}^*) \right]$$

$$= E_{\xi} \left[ \text{var}_{R\xi}(\bar{Y}^*) \right] + \text{var}_{\xi}(\bar{Y}),$$

which can be simplified as:

$$\text{var}_{\xi R}(\bar{Y}^*) = \frac{1}{N} I_N \otimes \left( \oplus_{s=1}^{N} \sigma_{w_s}^2 \right) + P_N \otimes \Delta$$

We are interested in developing an optimal linear unbiased predictor $P$ of a linear combination of the first-stage sample units (PSUs) based on a simple random sample. We can represent the sample data using the random variables in $Y^*$ or using the random variables in $\bar{Y}^*$.

Let $Y^* = \begin{pmatrix} Y_i^* & Y_{II}^* \end{pmatrix}'$ be partitioned so that $Y_i^*$ is a $n \times 1$ vector representing response for the first $n$ positions in a permutation, i.e. the sample. Similarly, let $\bar{Y}^* = \begin{pmatrix} \bar{Y}_i^* & \bar{Y}_{II}^* \end{pmatrix}'$ be partitioned so that $\bar{Y}_i^*$ is a $nN \times 1$ vector representing response for the first $n$ positions in a permutation, i.e. the sample. The question that we wish to answer is whether or not we can develop an optimal
predictor based on \( Y_i' \) instead of \( \bar{Y}_i' \). To answer this question, we use Theorem 1.1 of Rao and Bellhouse (1978).

**Rao and Bellhouse’s Theorem 1.1**

We now consider Rao and Bellhouse’s theorem. Here is Theorem 1.1 from Rao and Bellhouse (1978), which we restate here. Rao and Bellhouse (1978) define expectations with respect to a probability design, a model, and response error using the subscripts \( p, m, r \).

“Let \( C \) denote a class of pm-unbiased estimators of \( \mu = E_m(\bar{Y}) \), i.e., \( E(e_b) = \mu \).

Let \( C_0 \) be the corresponding class of pm-unbiased estimators, \( e_d \), of zero, i.e., \( E(e_d) = 0 \).

Theorem 1.1. An estimator \( e_b \) in \( C \) is “optimal” for \( \mu \) if and only if for every estimator \( e_d \) of zero belonging to \( C_0 \) it is true that \( E(e_b - \mu) e_d = 0 \).

We now express the target random variables we want to estimate, and extend Rao and Bellhouse’s theorem to our setting. We define \( I_i' = (e_{1d} \quad e_{1l}') \), where \( e_i \) is an \( N \times 1 \) vector with element \( i \) equal to one, and all other elements equal to zero. We also define

\[
g' = (g'_1 \quad g'_N) = (e'_{1d} \otimes I_N' \quad e'_{1l} \otimes I_N') \)

Then we define a ‘target’ as \( P_i = g' \bar{Y} = g' \left( \bar{Y}_I \quad \bar{Y}_I \right) \).

Recalling that \( g' = (g'_1 \quad g'_N) \), then \( P_i = g'_I \bar{Y}_I + g'_{II} \bar{Y}_{II} \). Rao and Bellhouse specify when we can develop an optimal predictor of \( P_i \) based solely on \( Y_i' \) instead of \( \bar{Y}_I' \).

There is a simple relationship between \( Y^* = \left( I_N \otimes (y + W) \right) vec(U') \) and

\[
\bar{Y}^* = \left[ I_N \otimes \left( \oplus_{s=1}^{N} (y_s + W_{sk}) \right) \right] vec(U') \]

given by \( Y^* = (I_N \otimes I_N') \bar{Y}^* \). Let \( X' = (I_N \otimes I_N') \). We want
to express \( \tilde{Y}^* \) as a linear combination of two terms, one of which is a linear function of \( Y^* \), and the other that has expected value equal to zero. To do so, notice that \( XX' = (I_N \otimes J_N) \), and using the identity that \( I_N \otimes I_N = \frac{1}{N} XX' + (I_N \otimes P_N) \), we can express

\[
\tilde{Y}^* = \frac{1}{N} XX' \tilde{Y}^* + (I_N \otimes P_N) \tilde{Y}^* 
= \frac{1}{N} XY^* + (I_N \otimes P_N) \tilde{Y}^* .
\]

Let \( C \) denote the class of linear unbiased predictors of \( P_i = g'\tilde{Y} \) based on \( Y_i^* \), such that the optimal predictor in this class is given by \( \hat{P}_i = (e_{i}\tilde{a} + \tilde{a})Y_i^* \). Since the predictor of \( P_i \) given by \( \hat{P}_i = (e_{i}\tilde{a} + \tilde{a})Y_i^* \) is unbiased and \( P_i = (e_{i}\tilde{a} + \tilde{a}) \left( \begin{array}{c} Y_i^* \\ Y_{ii} \end{array} \right) \), \( \tilde{a} \) must satisfy the equation

\[
E\left((e_{i}\tilde{a} + \tilde{a})Y_i^* - (e_{i}\tilde{a} + \tilde{a})Y \right) = 0 , \text{ or }
E\left((e_{i}\tilde{a} + \tilde{a})Y_i^* - (e_{i}\tilde{a} + \tilde{a})Y \right) = E\left( \begin{array}{c|c} (e_{i}\tilde{a} + \tilde{a}) & -e_{i}\tilde{a} \\ \hline -e_{i}\tilde{a} & Y_i^* \\ -e_{i}\tilde{a} & Y_{ii} \end{array} \right) , \text{ implying that }
\tilde{a}'1_n - e_{i}'1_{N\times n} = 0 .
\]

Let \( C_0 \) denote the class of linear unbiased predictors of \( B = b'[ (I_N \otimes P_N) \tilde{Y} ] = 0 \) be given by \( \hat{B} = b'[ (I_N \otimes P_N) \tilde{Y} ] \). For this predictor to be unbiased, we require

\[
E\left[ b'[ (I_n \otimes P_N) \tilde{Y}_i^* - B \right] = 0 , \text{ or } b'[ (I_n \otimes P_N) E\left( \tilde{Y}_i^* \right) ] = 0 . \text{ Using } E\left( \tilde{Y}_i^* \right) = \frac{1}{N} 1_n \otimes y , \text{ this implies that } b'[ (I_n \otimes P_N) y = 0 . \text{ For this relationship to hold for all } y , \text{ we require the unbiased constraint given by } \left( 1_n' \otimes P_N \right) \hat{b} = 0 .
Rao and Bellhouse state that an estimator $\hat{P}_i = \hat{a}^T Y_i$ is optimal for $P_i = L^T \tilde{Y}$ if and only if

for every estimator $\hat{B} = \hat{b}' (I_n \otimes P_N) \tilde{Y}_I^* = \tilde{Y}_I^* (I_n \otimes P_N) \hat{b}$ , $E_{\xi R} \left[ \left( \hat{P}_i - P_i \right) \hat{B} \right] = 0$.

Now

$$E_{\xi R} \left[ \left( \hat{P}_i - P_i \right) \hat{B} \right] = E_{\xi R} \left[ \left( (e_i' + \hat{a}') \mid -e_i' \mid -e_i' \right) \left( \frac{Y_i^*}{Y_I^* \ Y_H^*} \right) \left( \tilde{Y}_I^* (I_n \otimes P_N) (\hat{b}) \right) \right]$$

$$= \left( (e_i' + \hat{a}') \mid -e_i' \mid -e_i' \right) E_{\xi R} \left[ \left( \frac{Y_i^*}{Y_I^* \ Y_H^*} \right) \left( I_n \otimes P_N \right) (\hat{b}) \right]$$

Since

$$\left( \begin{array}{c} Y_i^* \\ Y_I^* \\ Y_H^* \end{array} \right) = \left( \begin{array}{cc} I_n \otimes I_N' & 0_{n \times nN} \\ 0_{N \times nN} & I_N \otimes I_N' \end{array} \right) \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) ,$$

then

$$E_{\xi R} \left[ \left( \begin{array}{c} Y_i^* \\ Y_I^* \\ Y_H^* \end{array} \right) \tilde{Y}_I'' \right] = E_{\xi R} \left[ \left( \begin{array}{cc} I_n \otimes I_N' & 0_{n \times nN} \\ 0_{N \times nN} & I_N \otimes I_N' \end{array} \right) \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right]$$

$$= \left( \begin{array}{cc} I_n \otimes I_N' & 0_{n \times nN} \\ 0_{N \times nN} & I_N \otimes I_N' \end{array} \right) E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right]$$

We determine an expression for $E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right]$.

Notice that

$$\text{var}_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right] = E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right] - E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right] E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right]^{-1} E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right] E_{\xi R} \left[ \left( \begin{array}{c} \tilde{Y}_I^* \\ Y_I^* \\ Y_H^* \end{array} \right) \right]^{-1} ,$$
so that \( E_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* & \tilde{Y}_{l}^* \\ \tilde{Y}_{H} & \tilde{Y}_{H} \end{bmatrix} \right] = \text{var}_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] + E_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] \). As a result,

\[
E_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] \tilde{Y}_{l}^{**} \]
corresponds to the first \( nN \) columns of \( \text{var}_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] + E_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right]. \)

By the results in c06rz15.doc, we know that \( \text{var}_{\xi R} \left( \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right) = \left( \begin{bmatrix} V_{I} & V_{I, II} \\ V_{H, I} & V_{II} \end{bmatrix} \right), \) where

\[
\left( \begin{bmatrix} V_{I} & V_{I, II} \\ V_{H, I} & V_{II} \end{bmatrix} \right) = \left( \begin{bmatrix} I_n - \frac{1}{N} J_n \odot \Delta & -\frac{1}{N} J_n \odot \Delta \\ -\frac{1}{N} J_n \odot \Delta & (I_{N-n} - \frac{1}{N} J_{N-n}) \odot \Delta \end{bmatrix} \right),
\]

Since \( E_{\xi R} \left( \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right) = \frac{1}{N} \left( {1}_n \otimes y \right) \) and

\[
E_{\xi R} \left( \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right) = \frac{1}{N^2} \left[ \begin{bmatrix} (I_{n-1} \otimes yy') (I_n \otimes yy') \\ (I_{n-1} \otimes yy') (I_n \otimes yy') \end{bmatrix} \right] \),

then

\[
\text{var}_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] + E_{\xi R} \left[ \begin{bmatrix} \tilde{Y}_{l}^* \\ \tilde{Y}_{l}^* \\ \tilde{Y}_{H} \\ \tilde{Y}_{H} \end{bmatrix} \right] = \left( \begin{bmatrix} V_{I} & V_{I, II} \\ V_{I, II} & V_{II} \end{bmatrix} \right) + \frac{1}{N^2} \left[ \begin{bmatrix} (I_{n-1} \otimes yy') (I_n \otimes yy') \\ (I_{n-1} \otimes yy') (I_n \otimes yy') \end{bmatrix} \right] \). \]

As a result,
Then,

\[
E_{\xi R} \left[ \begin{array}{c} \bar{Y}_I' \\ \bar{Y}_I \\ \bar{Y}_H \\
\end{array} \right] = \left( \begin{array}{ccc} I_n & 0_{N \times nN} & 0_{nN} \\
0_{N \times nN} & I_N & 0 \\
0_{nN} & I_N & I_N' \\
\end{array} \right) \left( \begin{array}{c} \bar{Y}_I' \\ \bar{Y}_I \\ \bar{Y}_H \\
\end{array} \right)
\]

Next, we determine \( E[\hat{T} - T]\), subject to the unbiased constraints that \( \hat{a}'1_n - e_{eff}1_{N-n} = 0 \) and \( (1'_n \otimes P_N)\hat{b} = 0 \). We express
\[ E_{\xi R} \left[ \left( \hat{P}_i - P_i \right) \hat{B} \right] \]

\[ = E_{\xi R} \left[ \left( (e_{II} + \hat{a}') \mid -e_{II} \mid -e_{III} \right) \left( Y_i^* \begin{bmatrix} Y_i \\ Y_{II} \end{bmatrix} \hat{Y}_\iota^* (I_n \otimes P_N) \left( \hat{b} \right) \right) \right] \]

\[ = \left( (e_{II} + \hat{a}') \mid -e_{II} \mid -e_{III} \right) E_{\xi R} \left[ \left( Y_i^* \begin{bmatrix} Y_i \\ Y_{II} \end{bmatrix} \hat{Y}_\iota^* (I_n \otimes P_N) \left( \hat{b} \right) \right) \right] \]

Notice that

\[ (I_n \otimes I_N' \otimes I_N' \otimes I_N') (\hat{b}) = (I_n \otimes I_N' \otimes I_N' \otimes I_N') (\hat{b}) , \]

\[ = 0 \]

\[ (J_n \otimes I_N' \otimes I_N' \otimes I_N') (\hat{b}) = (I_n \otimes I_N' \otimes I_N' \otimes I_N') (\hat{b}) , \] and

\[ \left( J_n \otimes \Delta \right) (I_n \otimes P_N) (\hat{b}) = (I_{N-n} \otimes \Delta) (I_n \otimes P_N) (\hat{b}) . \]

As a result,

\[ E_{\xi R} \left[ \left( \hat{P}_i - P_i \right) \hat{B} \right] = \left( (e_{II} + \hat{a}') \left[ I_n \otimes I_N' \otimes I_N' \left( \otimes \sigma_{se}^2 \right) \right] (I_n \otimes P_N) (\hat{b}) + \left. \right| \left( e_{II} \otimes I_N' \otimes I_N' \left( \otimes \sigma_{se}^2 \right) \right] (I_n \otimes P_N) (\hat{b}) \right] \]

\[ = \left( \hat{a}' \otimes I_N' \otimes \Delta \right) + \left. \right| \left( e_{II} + \hat{a}' \otimes I_N' \left( \otimes \sigma_{se}^2 \right) \right] (I_n \otimes P_N) (\hat{b}) . \]
Will this expression be equal to zero? Recall that if we are predicting the response for a position, the optimal predictor when \( i \leq n \) is \( \hat{a}' = 0 \). As a result, in this setting,

\[
E_{\xi R} \left[ \left( \hat{P} - P \right) \hat{B} \right] = \frac{1}{N} e''_d \otimes I_N' \left( \bigoplus_{s=1}^{S} \sigma^2_{se} \right) (I_n \otimes P_N) \left( \hat{b} \right) .
\]

This may not be zero. Thus, the usual random variables are not sufficient for optimal prediction of a realized sample unit.

Are the collapsed random variables sufficient for optimal prediction of a non-sampled unit? When \( i > n \), the predictor of a position is given by \( \hat{P}_i = \bar{Y} = \sum_{i=1}^{n} \frac{1}{n} Y_i \). Since \( e''_d = 0 \) when \( i > n \), \( \hat{a}' = \frac{1}{n} I_n' \).

As a result,

\[
E_{\xi R} \left[ \left( \hat{P} - P \right) \hat{B} \right] = \left[ (\hat{a}') \otimes I_N' \Delta + \frac{1}{N} (e''_d + \hat{a}') \otimes I_N' \left( \bigoplus_{s=1}^{S} \sigma^2_{se} \right) \right] (I_n \otimes P_N) \left( \hat{b} \right)
\]

\[
= \left[ (\hat{a}') \otimes I_N' \Delta + \frac{1}{N} (\hat{a}') \otimes I_N' \left( \bigoplus_{s=1}^{S} \sigma^2_{se} \right) \right] (I_n \otimes P_N) \left( \hat{b} \right)
\]

\[
= \left[ \left( \frac{1}{n} I_n \right) \otimes I_N' \Delta + \frac{1}{N} \left( \frac{1}{n} I_n \right) \otimes I_N' \left( \bigoplus_{s=1}^{S} \sigma^2_{se} \right) \right] (I_n \otimes P_N) \left( \hat{b} \right)
\]

\[
= 0
\]

As a result, the predictor is not optimal.