ESTIMATING PARAMETERS WHEN CONSIDERING THE UNOBSERVED UNITS AS MISSING VALUES IN SIMPLE RANDOM SAMPLING

A Dissertation Presented

By

JINGSONG LU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCES

December 2003

Biostatistics and Epidemiology

ESTIMATING PARAMETERS WHEN CONSIDERING THE UNOBSERVED UNITS AS MISSING VALUES IN SIMPLE RANDOM SAMPLING

A Dissertation Presented

By

JINGSONG LU

Approved as to style and content by

_____

Edward J. Stanek III, Chair

_____

Elaine Puleo, Member

_____

Lisa Chasan-Taber, Graduate Program Director
Biostatistics and Epidemiology
School of Public Health and Health Sciences

DEDICATION

To my family

CONTENTS

ACKNOWLEDGMANTS

I would like to thank my advisor, Edward J. Stanek III, for his thoughtful, patient guidance and support. Thanks are also due to the member of my committee, Elaine Puleo. Together their friendship and selfless contribution to my professional development have been invaluable and will forever be appreciated.

A special thank you to all those whose support and friendship helped me to stay focused on the project and provided me with the encouragement to continue when the going got tough.

ABSTRACT

ESTIMATING PARAMETERS WHEN CONSIDERING THE UNOBSERVED UNITS AS MISSING VALUES IN SIMPLE RANDOM SAMPLING

FEBRUARY 2004

JINGSONG LU, B.S. FUDAN UNIVERSITY

M.S. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Edward J. Stanek III

We developed the estimator of the population mean when some of the values of sample units are missing completely at random (MCAR), under the assumption of simple random sampling (SRS) from a finite super-population using a best linear unbiased predictor (BLUP) approach. The developed best linear unbiased estimator (BLUE) was compared theoretically and with small-scale simulations to the other estimators, which are typically used for handling missing data. A simulation shows the good performance of the new estimator and shows the new estimator is unbiased. The BLUE works more stable especially under small sample size and large proportion of missing values than the estimator ignoring missing values. But the variance of the BLUE is greater than the variance of the estimator ignoring the missing values under the same situation.

KEYWORDS: Simple random sampling, Missing data, MCAR, Super-population, Best linear unbiased estimator (BLUE).

LIST OF TABLES

LIST OF FIGURES

A. INTRODUCTION

We focus on a probability model induced by simple random survey sampling from a finite population represented by a simple random permutation population model. Model-based prediction tools are used to optimally estimate linear combinations of random variables in the model. We outline these methods in the context of simple random sampling in section C. In this thesis, the best linear unbiased estimator (BLUE) is developed and compared theoretically and via small-scale simulations to an ad hoc estimator that ignores the missing values under the same situation.

B. REVIEW OF LITERATURE

The general modeling framework for survey sampling is given by Cassel, Särndal and Wretman (1977); Hedayat and Sinha (1991); Mukhopadhyay (2001); Smith, T.M.F. (1976); Särndal, Swensson, and Wretman (1992); Lehtonen and Pahkinen(1994); Thompson (1997); Valliant, Dorfman, and Royall (2000). Recent reviews of inference in survey sampling are given by Rao (1997, 1999a). Brewer, Hanif, and Tam (1988), Brewer (1999) and Brewer (2002) have discussed reconciling model-based and design based inference.

We focus on a probability model induced by simple random survey sampling from a finite population represented by a simple random permutation population model. We discuss the population approach versus the super-population approach in sample surveys in section B.1 and review three different classes of missingness defined by Little and Rubin (1987) in section B.2. We introduce two methods used for estimating the population mean when there are missing data in section B.3. The two methods include a method discussed by Cochran (1977), and an imputation method due to Rubin (1986). Definitions and notations are defined for a random permutation model in section B.4.

B.1 Population Approach Versus Super-population Approach

In survey sampling a fixed finite population is under consideration, where the population elements are labeled so that each element can be identified. Statistical inference is directed towards a finite population (or descriptive) parameter - a quantity whose value would be known exactly if a census of the survey population was carried out. A different problem is considered in a model-based approach. For a model-based approach, a

super-population is assumed rather than a finite population. The actual finite population is considered to be a realization from the hypothetical super-population (Lehtonen and Pahkinen, 1994). A model is specified for the super-population. Parameters governing this stochastic process are called super-population parameters, and may be of interest to the survey analyst. Alternatively, combinations of finite population values may be the target.

B.2. Missing Data

Statistical analysis in the presence of incomplete or missing data is a pervasive problem in sample surveys. Nonresponse causes missing data, which results in a sample data set whose real size is smaller than the intended sample size. At a minimum, the precision of estimators is reduced. The impact of the missing data on the results of statistical analysis depends on the missing data mechanism, and the way in which the data analyst deals with it. Little and Rubin (1987) define three different classes of missingness:

- ✧ MCAR (Missing Completely at Random)

  The term refers to data where the missing data mechanism does not depend on any variable. This means that the probability of response does not depend on either the auxiliary variables X or the variable of interest Y. When the assumption of MCAR holds, the missing data is missing at random, and observed data is observed at random.

- ✧ MAR (Missing at Random)

  The term MAR is confusing because the missing data mechanism may depend on the realization of some other variables. What MAR means is that the probability of missingness of the 'Y-variable' of interest is not a function of the outcome of Y, but is a

function of some other 'X-variable'. Both MAR and MCAR require that the variable with missing data be unrelated to whether or not the value of response on a person is missing. The difference between MAR and MCAR is whether or not other variables in the data set are associated with the probability of missing response for Y.

✧ NMAR (Not Missing at Random)

This occurs when the missing data mechanism depends on the actual value of the variable Y where response is missing. This is the most difficult condition to model for.

When making sampling distribution based inference about a population parameter, the role of the study design is not completely clear. Rubin (1976) indicated that when making direct-likelihood or Bayesian inference about the parameter, the design can be ignored if the unobserved values are missing at random and the observed values are observed at random. MCAR and MAR are ignorable, for likelihood-based imputation methods. NMAR is not ignorable (Little and Rubin 1987). Multiple imputations, EM imputation and regression imputation all are valid provided the missing data mechanism is not NMAR, and the percentage of missing data is not too large.

This study concentrates on developing an estimator for the population mean considering the unobserved values in a SRS from a finite population as missing values with the assumption of MCAR (Missing Completely at Random). We compare the developed best linear unbiased estimator (BLUE) theoretically and with small-scale simulations to the other estimators, which are typically used for handling missing data.

B.3. Methods For Handling Data With Ignorable Nonresponse

Several methods are used for analysis when there are missing data, like reweighting (Lehtonen and Pahkinen 1994), imputation, expectation maximization (EM) algorithm (Dempster, Laird and Rubin 1977), and the method ignoring the missing values. We discuss two main methods to estimate the population mean when the missing data mechanism is ignorable.

1. The first method was discussed by Cochran (1977).

Let n_1 and n_0 be the numbers of the units observed and missing, respectively, in the sample and let $p_0 = n_0 / n$ be the proportion of missing values in the sample, where $n = n_0 + n_1$. Also, define $\bar{Y}_1^*$ as the average response of observed sample units. The first method ignores the missing values and uses the sample mean $\bar{Y}_1^*$ based on the observed values in the sample. The difference between $\bar{Y}_1^*$ and the mean of all sample units $\bar{Y}$ depends on the p_0 and the difference between the mean of observed values $\bar{Y}_1^*$ and the mean of missing values $\bar{Y}_0$, which is unknown. The difference is

$$\bar{Y}_1^* - \bar{Y} = \bar{Y}_1^* - ((1 - p_0)\bar{Y}_1^* + p_0\bar{Y}_0) = p_0\left(\bar{Y}_1^* - \bar{Y}_0\right).$$

Cochran uses s^{*2}, the sample variance estimate based on the observed sample units to estimate the variance. If the population size N is large enough, the finite population correction factor (fpc) can be ignored when evaluating the variance.

2. The second method that we discuss was introduced by Rubin (1986).

In this method, values of missing observations are imputed $m \geq 2$ imputations for

each value instead of a single imputation. For each set of imputations, the average is estimated. The final estimator is the average of the estimators from each imputation. This method retains the main advantages of the single imputation but avoids its drawbacks by replacing each missing datum with two or more values representing a distribution of likely values. There are several imputation methods, such as Simple Random(SR) Imputation, Bayesian Bootstrap (BB) imputation, and Approximate Bayesian Bootstrap (ABB) imputation described by Rubin (1986).

When the response variable is normally distributed, suppose that Q is a scalar quantity (i.e. Q is a linear function of the data) to be estimated and that if there were no missing values, the inference for Q would be based on the statement that

$$\hat{Q} - Q \sim N(0, V),$$

where $\hat{Q}$ is a statistic giving the estimate of Q and the variance of $\hat{Q}$ is $\hat{V}$, respectively. With m imputations of missing values, there will be m completed data sets and m values of $\hat{Q}$ and V, say $\hat{Q}_{*l}$, $\hat{V}_{*l}$, for $l = 1, 2, \cdots, m$. Inference for Q is based on the statement that

$$\hat{Q}_{*\bullet} - Q \sim N(0, T_*),$$

where

$$\hat{Q}_{*\bullet} = \sum_{l=1}^{m} \hat{Q}_{*l} / m$$

is the estimate of Q based on m imputations and the variance of $\hat{Q}_{*\bullet}$ is

$$T_* = \hat{W} + ((m+1)/m))\hat{B}$$

where $\hat{B} = \sum_{l=1}^{m} \left(\hat{Q}_{*l} - \hat{Q}_{*\bullet} \right)^2 / (m-1)$ is the between-imputation estimated variance of $Q - \hat{Q}_{*\bullet}$ and $\hat{W} = \sum_{l=1}^{m} \hat{V}_{*l} / m$, where $\hat{V}_{*l}$ is the within-imputation variance of $\hat{Q}_{*\bullet}$

B.4. Definitions And Notation For A Random Permutation Model

We describe the definitions and notations that will be used for a random permutation model. This model provides the framework for the research. We follow the definition and notation used by Stanek (2002). We consider the problem of estimating a linear combination of values for a finite population of units under simple random sampling without replacement.

We define a finite population as a collection of a known number, N, of identifiable units labeled $s = 1, 2, \cdots, N$. Associated with unit s is a response y_s. We summarize the set of responses in the vector $\mathbf{y} = \left(y_1, \cdots, y_N \right)'$ and assume that when unit s is observed, the response y_s is known without error. Typically, there is interest in a $p \times 1$ vector of parameters of the form $\boldsymbol{\beta} = \mathbf{G}\mathbf{y}$ where $\mathbf{G}$ is a matrix of known constants. We limit consideration to settings where $p = 1$ and

$$\mathbf{G} = \frac{\mathbf{1}_N'}{N} \qquad \text{(B.1).}$$

We define a probability model that links the population parameters to a vector of random variables, which is induced by a simple random sampling design, and develop estimators of linear functions of these random variables. Assuming simple random sampling without replacement, the random permutation probability model assigns equal

probability to all permutations of the finite population units. We index each unit's position in the permutation by $i = 1, 2, \cdots, N$. We represent the value in position i of a randomly selected permutation by the random variable $Y_i = \sum_{j=1}^{N} U_{is} y_s$, where $U_{is} = 1$ if unit s is in position i and $U_{is} = 0$ otherwise. When all permutations are equally likely, the random vector $\mathbf{Y} = (Y_1, \cdots, Y_N)'$ is a random permutation of the population (Cassel et al., 1977). We can relate $\mathbf{Y}$ to $\mathbf{y}$ such that $\mathbf{Y} = \mathbf{U}\mathbf{y}$, where

$$\mathbf{U} = \begin{pmatrix} U_{11} & \cdots & U_{1N} \\ \vdots & \ddots & \vdots \\ U_{N1} & \cdots & U_{NN} \end{pmatrix}.$$

The prediction-based approach used in sample survey model based inference predicts linear functions of the random variables in the population based on a linear combination of sample random variables. We use the prediction approach that is common in model-based inference to develop the estimators. The prediction approach is based on an underlying probability model for a vector of random variables $\mathbf{Y} = (Y_1, \cdots, Y_N)'$. The vector of random variables is partitioned into a subset, which we call the sample, $\mathbf{Y}_I = (Y_1, \cdots, Y_n)'$ which is indexed by $i = 1, 2, \cdots, n$, and the remainder, $\mathbf{Y}_{II} = (Y_{n+1}, \cdots, Y_N)'$ such that $\mathbf{Y} = \left(\mathbf{Y}_I' \mid \mathbf{Y}_{II}'\right)'$. Inference is solely based on linear models of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \tag{B.2}$$

where $\mathbf{X}$ is a known non-stochastic full rank matrix, $\boldsymbol{\beta}$ is a p-dimensional vector of population parameters and $\mathbf{E}$ is an N-dimensional vector of random errors governed by the

probability model for $\mathbf{E}$ where $E_{\xi}(\mathbf{E}) = \mathbf{0}$. We use the subscript ξ denotes expectation with respect to the random permutation. The parameter of interest is the linear combination $\boldsymbol{\beta} = \mathbf{Gy}$, where $\mathbf{G}$ is defined by expression (B.1). Since $\mathbf{GU} = \frac{\mathbf{1}_N'}{N}$, $\mathbf{GY} = \mathbf{GUy} = \mathbf{Gy} = \beta = \mu$. Expressed as model (B.2), we set $\mathbf{X} = \mathbf{1}_N$ and $\boldsymbol{\beta} = \mu$.

C. METHODS

We develop estimators of the mean and variance when some of the values of sample units are missing completely at random (MCAR), under the assumption of simple random sampling from a finite population. Two missing data mechanisms will be used for ignorable missing data in a random permutation model. The first missing data mechanism is applied to positions. The second missing data mechanism is applied to units. The two missing data mechanisms are presented by separate models. We develop the estimators using a best linear unbiased predictor (BLUP) approach. For each model, we derive the expected value and variance of random variables. Using these expressions, we develop a BLUP based on the sampled data. We illustrate the results in a small artificial data set representing non-response in a survey setting, and via a small scale simulation.

C.1. Two Methods of Specifying Ignorable Missing Data

Under the assumption of MCAR, We specify two models that account for missing data. In each model, we assume each unit in the population has an equal probability, π, of being missing. We define a $2N \times N$ dimensional matrix $\mathbf{M} = \begin{pmatrix} \mathbf{I}_N - \mathbf{M}^* \\ \hline \mathbf{M}^* \end{pmatrix}$ to indicate which units are missing, and $\mathbf{M}^* = \bigoplus_{i=1}^{N} M_i$, where we define

$$\bigoplus_{i=1}^{N} a_i = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_N \end{pmatrix},$$

$M_i = 0$ if response for the unit in position i is not missing, and $M_i = 1$ otherwise where $i = 1, 2, \cdots, N$.

The first model for the missing data mechanism is to assume that the missing data mechanism is associated with measurement. We represent the vector of random variables corresponding to the potentially observable population with missing data as the $2N \times 1$ vector:

$$\mathbf{Z}_1 = \mathbf{MUy} = \mathbf{MY} = \begin{pmatrix} \mathbf{Y}_I^{(o)} \\ \mathbf{Y}_{II}^{(o)} \\ \mathbf{Y}_I^{(m)} \\ \mathbf{Y}_{II}^{(m)} \end{pmatrix}$$

(C.1)

where $\mathbf{Y}_I^{(o)}$ represents the n units which are observed and selected in the sample, $\mathbf{Y}_{II}^{(o)}$ represents $N-n$ are the $N-n$ units which would be observed if the units were selected but are not in the sample, $Y_I^{(m)}$ represents the n units which are missing and selected in the sample, and $\mathbf{Y}_{II}^{(m)}$ represents the $N-n$ units which are missing and not selected in the sample. As an example, the missing data may arise from a simple random sample of n subjects, where each subject is to be interviewed, and different interviewers will randomly have non-response.

A second model for the missing data mechanism is to assume that the missing data mechanism is related to the subjects. In this case, $\mathbf{H} = \bigoplus_{s=1}^{N} H_s$, $\mathbf{M} = \begin{pmatrix} \mathbf{I}_N - \mathbf{H} \\ \hline \mathbf{H} \end{pmatrix}$, and $H_s = 1$ if response for subject s is missing and $H_s = 0$ otherwise. The potentially observable random variables representing the population with this missing data mechanism is

$$\mathbf{Z}_2 = (\mathbf{I_2} \otimes \mathbf{U})\mathbf{M}\mathbf{y} = \begin{pmatrix} \mathbf{Y}_I^{(o)} \\ \mathbf{Y}_{II}^{(o)} \\ \mathbf{Y}_I^{(m)} \\ \mathbf{Y}_{II}^{(m)} \end{pmatrix} \quad . \tag{C.2}$$

C.2 Expected Value and Variance of Random Variable

Under the random permutation model (B.1) augmented for missing data using (C.1), we use the subscript ξ_1 to denote expectation with respect to the missing data mechanism, and the subscript ξ_2 to denote expectation with respect to permutation of units. Since $\mathbf{Z} = \mathbf{MUy}$, we get the expected value of $\mathbf{Z}$ and variance of $\mathbf{Z}$. Expressions can be derived for the variance using expressions similar to those given by Stanek (2001). With the same procedure, we obtain the expected value and variance for (C.2).

C.3 Deriving the Best Linear Unbiased Estimator of Population Mean

We derive the best linear unbiased estimator based on the expected value and variance for (C.1) and (C.2). We use three criteria for developing the estimator making use of the prediction framework. First, we represent the estimator as a linear function of the sample data, $p = \mathbf{L}'\mathbf{Y}_1$ where $\underset{n\times 1}{\mathbf{L}}$ is a vector of unknown constants and $\mathbf{Y}_1 = \mathbf{Y}_I^{(o)}$ is a vector

of selected sample with size n. Next, we require the estimator to be unbiased. This implies that $E_{\xi_1\xi_2}\left(\mathbf{L}'\mathbf{Y}_1 - \mu\right) = 0$. Third, we want to find values of $\mathbf{L}'$ that minimize the variance.

C.4. Comparisons of methods

We propose to compare the developed best linear unbiased estimator to the other estimators, which are typically used for handling missing data, such as the two methods mentioned in section B.3. We conduct the comparison both theoretically and with a small-scale simulation with different assumptions made for π, σ^2, N, n.

D. RESULTS

D.1 Development of expressions for the Expected Value and Variance

<u>D.1.1. Development of the expressions for the expected value and variance of</u> $\mathbf{Z}_1 = \mathbf{MUy}$

We assume each unit in the population has an equal probability, π, of being missing. Under the random permutation model (B.2) augmented for missing data using (C.1) $\mathbf{Z}_1 = \mathbf{MUy}$. We use the subscript ξ_1 to denote the expectation with respect to the missing data mechanism, and the subscript ξ_2 to denote expectation with respect to permutation of units. We define $\mathbf{I}_N$ matrix as an $N \times N$ matrix with 1 at diagonal and 0 anywhere else. We define $\mathbf{J}_N$ matrix as a $N \times N$ matrix with all elements equal to 1. We represent population mean as $\mu = \frac{1}{N}\sum_{s=1}^{N} y_s = \frac{1}{N}\mathbf{1}'\mathbf{y}$ and population variance as $\sigma^2 = \frac{\sum_{s=1}^{N}(y_s - \mu)^2}{N-1}$.

Since $E_{\xi_2}(U_{is}) = \frac{1}{N}$ (only one element could be 1 in each row of $\mathbf{U}$),

then $E_{\xi_2}(\mathbf{U}) = \frac{1}{N}\mathbf{J}_N$. Since $E_{\xi_1}(M_j) = \pi$, then $E_{\xi_1}(\mathbf{M}) = \begin{pmatrix} 1-\pi \\ \pi \end{pmatrix} \otimes \mathbf{I}_N$, where $\otimes$ is the Kronecker product. The vector $\mathbf{y}$ is non stochastic. Thus, the expected value of $\mathbf{Z}_1$ with respect to ξ_1 and ξ_2 is

$$
\begin{aligned}
E_{\xi_1\xi_2}(\mathbf{Z}_1) &= E_{\xi_1}\left[E_{\xi_2|\xi_1}(\mathbf{MUy})\right] = E_{\xi_1}\left(\mathbf{M}\left[E_{\xi_2|\xi_1}(\mathbf{U})\right]\mathbf{y}\right) \\
&= E_{\xi_1}\left(\mathbf{M}\left[\frac{1}{N}\mathbf{J}_N\right]\mathbf{y}\right) = E_{\xi_1}(\mathbf{M})\mathbf{1}_N\mu = \mu\left[\begin{pmatrix} 1-\pi \\ \pi \end{pmatrix} \otimes \mathbf{I}_N\right]\mathbf{1}_N \\
&= \mu\left[\begin{pmatrix} 1-\pi \\ \pi \end{pmatrix} \otimes \mathbf{1}_N\right]
\end{aligned}
$$

.

Next, we evaluate $Var_{\xi_1\xi_2}(\mathbf{Z}_1)$. From Stanek (2002),

$$Var_{\xi_2}(\mathbf{Y}) = Var_{\xi_2}(\mathbf{Uy}) = \sigma^2\left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right),$$

then variance of $\mathbf{Z}$ is

$$Var_{\xi_1\xi_2}(\mathbf{Z}) = Var_{\xi_1}\left[E_{\xi_2|\xi_1}(\mathbf{MUy})\right] + E_{\xi_1}\left[\left(Var_{\xi_2|\xi_1}(\mathbf{MUy})\right)\right] = Var_{\xi_1}\left(\mathbf{M}\frac{\mathbf{J}_N}{N}\mathbf{y}\right) + E_{\xi_1}\left(\mathbf{M}\left[Var_{\xi_2|\xi_1}(\mathbf{Uy})\right]\mathbf{M}'\right)$$

$$= Var_{\xi_1}\left(\mathbf{M}\mathbf{1}_N\mu\right) + E_{\xi_1}\left[\mathbf{M}\left(\sigma^2\left[\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right]\right)\mathbf{M}'\right] = \mu^2 Var_{\xi_1}\begin{pmatrix} 1-M_1 \\ 1-M_2 \\ \vdots \\ 1-M_N \\ M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix} + \sigma^2 E_{\xi_1}\left(\mathbf{M}\left[\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right]\mathbf{M}'\right).$$

We assume that the missing data random variables are independent and identically distributed.

As a result,

$Var(M_i) = \pi(1-\pi)$, $Cov(M_i, 1-M_i) = -\pi(1-\pi)$ and $Cov(M_i M_{i^*}) = 0$, for $i \neq i^*$.

Let $\mathbf{M}^{**} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{pmatrix}$, then $Var_{\xi_1}\begin{pmatrix} \mathbf{1}_N - \mathbf{M}^{**} \\ \mathbf{M}^{**} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes Var_{\xi_1}\left(\mathbf{M}^{**}\right)$.

Since, $Var_{\xi_1}\left(\mathbf{M}^{**}\right) = \pi(1-\pi)\mathbf{I}_N$, then $Var_{\xi_1}\begin{pmatrix} \mathbf{1}_N - \mathbf{M}^{**} \\ \mathbf{M}^{**} \end{pmatrix} = \pi(1-\pi)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{I}_N$.

We evaluate $E_{\xi_1}\left(\mathbf{M}\left[\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right]\mathbf{M}'\right) = E_{\xi_1}\left(\mathbf{MM}'\right) - \frac{1}{N}E_{\xi_1}\left(\mathbf{M}\mathbf{1}_N\mathbf{1}_N'\mathbf{M}'\right)$ next. First,

$$E_{\xi_1}\left(\mathbf{MM}'\right) = E_{\xi_1}\left[\left(\begin{array}{c} \mathbf{I}_N - \mathbf{M}^* \\ \hline \mathbf{M}^* \end{array}\right)\left(\mathbf{I}_N - \mathbf{M}^* \;\vdots\; \mathbf{M}^*\right)\right] = E_{\xi_1}\left(\begin{array}{c|c} \left(\mathbf{I}_N - \mathbf{M}^*\right)\left(\mathbf{I}_N - \mathbf{M}^*\right) & \left(\mathbf{I}_N - \mathbf{M}^*\right)\mathbf{M}^* \\ \hline \left(\mathbf{I}_N - \mathbf{M}^*\right)\mathbf{M}^* & \mathbf{M}^*\mathbf{M}^* \end{array}\right)$$

$$= E_{\xi_1}\left(\begin{array}{c|c} \mathbf{I}_N - \mathbf{M}^* & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{M}^* \end{array}\right) = \begin{pmatrix} 1-\pi & 0 \\ 0 & \pi \end{pmatrix} \otimes \mathbf{I}_N.$$

Also,

$$E_{\xi_1}\left(\mathbf{M}\mathbf{1}_N\mathbf{1}_N'\mathbf{M}'\right)=E_{\xi_1}\left[\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}'\right].$$

$$=Var_{\xi_1}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}+E_{\xi_1}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}E_{\xi_1}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}'$$

Now, $E_{\xi_1}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}=\begin{pmatrix}1-\pi\\ \pi\end{pmatrix}\otimes\mathbf{1}_N$, hence

$$E_{\xi_1}\left(\mathbf{M}\mathbf{1}_N\mathbf{1}_N'\mathbf{M}'\right)=\pi\left(1-\pi\right)\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}_N+\begin{pmatrix}1-\pi\\ \pi\end{pmatrix}\begin{pmatrix}1-\pi\\ \pi\end{pmatrix}'\otimes\mathbf{J}_N.$$

Combining the results,

$$Var_{\xi_1\xi_2}\left(\mathbf{Z}\right)=\mu^2 Var_{\xi_1}\begin{pmatrix}\mathbf{1}_N-\mathbf{M}^{**}\\ \hline \mathbf{M}^{**}\end{pmatrix}+\sigma^2 E_{\xi_1}\left(\mathbf{M}\left[\mathbf{I}_N-\frac{\mathbf{J}_N}{N}\right]\mathbf{M}'\right)$$

$$=\mu^2\left(1-\pi\right)\pi\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}_N$$

$$+\sigma^2\left[\left(\begin{bmatrix}1-\pi & 0\\ 0 & \pi\end{bmatrix}\otimes\mathbf{I}_N\right)-\frac{1}{N}\left(\pi\left(1-\pi\right)\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}_N+\begin{bmatrix}\left(1-\pi\right)^2 & \pi\left(1-\pi\right)\\ \pi\left(1-\pi\right) & \pi^2\end{bmatrix}\otimes\mathbf{J}_N\right)\right]$$

$$=\left[\mu^2\left(1-\pi\right)\pi\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}+\frac{N-1}{N}\sigma^2\left(1-\pi\right)\pi\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right]\otimes\mathbf{I}_N$$

$$+\sigma^2\left(\left(\begin{bmatrix}1-\pi & 0\\ 0 & \pi\end{bmatrix}-\left[\pi\left(1-\pi\right)\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right]\right)\otimes\mathbf{I}_N-\begin{bmatrix}\left(1-\pi\right)^2 & \pi\left(1-\pi\right)\\ \pi\left(1-\pi\right) & \pi^2\end{bmatrix}\otimes\frac{\mathbf{J}_N}{N}\right)$$

$$=\left(\sigma^2\begin{bmatrix}\left(1-\pi\right)^2 & \pi\left(1-\pi\right)\\ \pi\left(1-\pi\right) & \pi^2\end{bmatrix}\otimes\left(\mathbf{I}_N-\frac{\mathbf{J}_N}{N}\right)\right)+\left[\pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2+\mu^2\right)\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}_N\right].$$

<u>D.1.2. Development of the expressions for the expected value and variance of</u>

$\mathbf{Z}_2=(\mathbf{I_2}\otimes\mathbf{U})\mathbf{M}\mathbf{y}$

Under the random permutation model (B.2) augmented for missing data using (C.2) $\mathbf{Z}_2 = (\mathbf{I_2} \otimes \mathbf{U})\mathbf{M}\mathbf{y}$, then the expected value of $\mathbf{Z}_2$ with respect to ξ_1 and ξ_2 is

$$
\begin{aligned}
E_{\xi_1\xi_2}\left(\mathbf{Z}_2\right) &= E_{\xi_1}\left[E_{\xi_2|\xi_1}\left(\left[\mathbf{I_2} \otimes \mathbf{U}\right]\mathbf{M}\mathbf{y}\right)\right] = E_{\xi_1}\left[E_{\xi_2}(\mathbf{I_2} \otimes \mathbf{U})\mathbf{M}\mathbf{y}\right] \\
&= E_{\xi_1}\left(\left[\mathbf{I_2} \otimes \frac{\mathbf{J}_N}{N}\right]\mathbf{M}\mathbf{y}\right) = \left(\mathbf{I_2} \otimes \frac{\mathbf{J}_N}{N}\right)E_{\xi_1}\left(\mathbf{M}\right)\mathbf{y} \\
&= \left[\mathbf{I_2} \otimes \frac{\mathbf{J}_N}{N}\right]\left[\begin{pmatrix}1-\pi \\ \pi\end{pmatrix} \otimes \mathbf{I}_N\right]\mathbf{y} = \mu\begin{pmatrix}1-\pi \\ \pi\end{pmatrix} \otimes \mathbf{1}_N
\end{aligned}
,
$$

and variance of $\mathbf{Z}_2$ is

$$
\begin{aligned}
Var_{\xi_1\xi_2}\left(\mathbf{Z}_2\right) &= Var_{\xi_2}\left(E_{\xi_1|\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\mathbf{M}\mathbf{y}\right]\right) + E_{\xi_2}\left(Var_{\xi_1|\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\mathbf{M}\mathbf{y}\right]\right) \\
&= Var_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left(E_{\xi_1|\xi_2}\left[\mathbf{M}\right]\right)\mathbf{y}\right] + E_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left[Var_{\xi_1|\xi_2}(\mathbf{M}\mathbf{y})\right](\mathbf{I_2} \otimes \mathbf{U})'\right] \\
&= Var_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{I}_N\right)\mathbf{y}\right] + E_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left[Var_{\xi_1|\xi_2}(\mathbf{I}_{2N}\mathbf{M}\mathbf{y})\right](\mathbf{I_2} \otimes \mathbf{U})'\right]
\end{aligned}
$$

For the first term,

$$
Var_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{I}_N\right)\mathbf{y}\right] = Var_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\begin{pmatrix}[1-\pi]y_1 \\ [1-\pi]y_2 \\ \vdots \\ [1-\pi]y_N \\ \pi y_1 \\ \pi y_2 \\ \vdots \\ \pi y_N\end{pmatrix}\right]
$$

$$
= Var_{\xi_2}\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{Y}\right) = E\left[\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{Y}\right)\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{Y}\right)'\right] - E\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{Y}\right)E\left(\begin{bmatrix}1-\pi \\ \pi\end{bmatrix} \otimes \mathbf{Y}\right)'
$$

$$
\begin{aligned}
&= \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes E(\mathbf{YY}') - \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes E(\mathbf{Y})E(\mathbf{Y}') \\
&= \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes \left[E(\mathbf{YY}') - E(\mathbf{Y})E(\mathbf{Y}') \right] \\
&= \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes Var(\mathbf{Y}) \\
&= \sigma^2 \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes \left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N} \right)
\end{aligned}
$$

For the second term,

$Var_{\xi_1|\xi_2}(\mathbf{I}_{2N}\mathbf{My}) = Var_{\xi_1|\xi_2}\left[(\mathbf{y}' \otimes \mathbf{I}_{2N}) Vec(\mathbf{M}) \right] = (\mathbf{y}' \otimes \mathbf{I}_{2N}) Var_{\xi_1|\xi_2}\left[Vec(\mathbf{M}) \right] (\mathbf{y} \otimes \mathbf{I}_{2N})$, where

$$
Var_{\xi_1|\xi_2}\left[Vec(\mathbf{M}) \right] = Var_{\xi_1|\xi_2} \begin{pmatrix} 1-H_1 \\ \mathbf{0}_{N-1} \\ \hline H_1 \\ \mathbf{0}_{N-1} \\ \hline 0 \\ 1-H_2 \\ \mathbf{0}_{N-2} \\ \hline 0 \\ H_2 \\ \mathbf{0}_{N-2} \\ \hline \vdots \\ \hline \mathbf{0}_{N-1} \\ 1-H_N \\ \hline \mathbf{0}_{N-1} \\ H_N \end{pmatrix} =
$$

$$
\left(\begin{array}{c|c|c|c|c}
\begin{array}{c|c|c|c}
\pi(1-\pi) & \mathbf{0}'_{N-1} & -\pi(1-\pi) & \mathbf{0}'_{N-1} \\
\hline
\mathbf{0}_{N-1} & \mathbf{0} & \mathbf{0}_{N-1} & \mathbf{0} \\
\hline
-\pi(1-\pi) & \mathbf{0}'_{N-1} & \pi(1-\pi) & \mathbf{0}'_{N-1} \\
\hline
\mathbf{0}_{N-1} & \mathbf{0} & \mathbf{0}_{N-1} & \mathbf{0}
\end{array}
& \mathbf{0} & \cdots & \mathbf{0} \\
\hline
\mathbf{0} & \begin{array}{c|c|c|c|c|c}
0 & 0 & \mathbf{0}'_{N-2} & 0 & 0 & \mathbf{0}'_{N-2} \\
\hline
0 & \pi(1-\pi) & 0 & 0 & -\pi(1-\pi) & 0 \\
\hline
\mathbf{0}_{N-2} & 0 & \mathbf{0} & \mathbf{0}_{N-2} & 0 & \mathbf{0} \\
\hline
0 & 0 & \mathbf{0}'_{N-2} & 0 & 0 & \mathbf{0}'_{N-2} \\
\hline
0 & -\pi(1-\pi) & 0 & 0 & \pi(1-\pi) & 0 \\
\hline
\mathbf{0}_{N-2} & 0 & \mathbf{0} & \mathbf{0}_{N-2} & 0 & \mathbf{0}
\end{array}
& \ddots & \vdots \\
\hline
\vdots & \ddots & & \ddots & \mathbf{0} \\
\hline
\mathbf{0} & \cdots & & \mathbf{0} & \begin{array}{c|c|c|c}
\mathbf{0} & \mathbf{0}'_{N-1} & \mathbf{0} & \mathbf{0}'_{N-1} \\
\hline
\mathbf{0}_{N-1} & \pi(1-\pi) & \mathbf{0}_{N-1} & -\pi(1-\pi) \\
\hline
\mathbf{0} & \mathbf{0}'_{N-1} & \pi(1-\pi) & \mathbf{0}'_{N-1} \\
\hline
\mathbf{0}_{N-1} & -\pi(1-\pi) & \mathbf{0}_{N-1} & \pi(1-\pi)
\end{array}
\end{array}\right)
$$

$= \bigoplus_{i=1}^{N} \mathbf{A}_i$, where

$$
\mathbf{A}_i = \pi(1-\pi)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \left(\begin{array}{c|c|c}
0_{(i-1)\times(i-1)} & 0 & 0_{(i-1)\times(N-1)} \\
\hline
0 & 1 & 0 \\
\hline
0_{(N-1)\times(i-1)} & 0 & 0_{(N-i)\times(N-i)}
\end{array}\right) = \pi(1-\pi)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{e}_i \mathbf{e}_i' \ ,
$$

where we define $\mathbf{e}_i = \begin{pmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}'$ as an $N \times 1$ vector with a value of one in the ith row and 0 otherwise.

Using this result,

$$
\begin{aligned}
& E_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U}) \left[Var_{\xi_1|\xi_2}(\mathbf{I}_{2N}\mathbf{M}\mathbf{y}) \right] (\mathbf{I_2} \otimes \mathbf{U})' \right] \\
& = E_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U}) \left(\mathbf{y}' \otimes \mathbf{I}_{2N} \right) Var_{\xi_1|\xi_2}\left[Vec(\mathbf{M}) \right] \left(\mathbf{y} \otimes \mathbf{I}_{2N} \right) (\mathbf{I_2} \otimes \mathbf{U})' \right] \\
& = E_{\xi_2}\left[\left(\begin{array}{c|c} \mathbf{U} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{U} \end{array} \right) \left(\mathbf{I}_{2N} y_1 \mid \mathbf{I}_{2N} y_2 \mid \cdots \mid \mathbf{I}_{2N} y_N \right) Var_{\xi_1|\xi_2}\left[Vec(\mathbf{M}) \right] \left(\mathbf{y} \otimes \mathbf{I}_{2N} \right) (\mathbf{I_2} \otimes \mathbf{U})' \right] \\
& = E_{\xi_2}\left(\left[\mathbf{y}' \otimes \left(\mathbf{I_2} \otimes \mathbf{U} \right) \right] Var_{\xi_1|\xi_2}\left[Vec(\mathbf{M}) \right] \left[\mathbf{y} \otimes (\mathbf{I_2} \otimes \mathbf{U}') \right] \right) \\
& = E_{\xi_2}\left(\left[\mathbf{y}' \otimes \left(\mathbf{I_2} \otimes \mathbf{U} \right) \right] \left[\bigoplus_{i=1}^{N} \mathbf{A}_i \right] \left[\mathbf{y} \otimes (\mathbf{I_2} \otimes \mathbf{U}) \right]' \right)
\end{aligned}
$$

$$= E_{\xi_2}\left[\left(y_1 \otimes [\mathbf{I_2} \otimes \mathbf{U}][\mathbf{A}_1] \mid y_2 \otimes [\mathbf{I_2} \otimes \mathbf{U}][\mathbf{A}_2] \mid \cdots \mid y_N \otimes [\mathbf{I_2} \otimes \mathbf{U}][\mathbf{A}_N]\right)\begin{pmatrix} y_1 \otimes [\mathbf{I_2} \otimes \mathbf{U}'] \\ \hline y_2 \otimes [\mathbf{I_2} \otimes \mathbf{U}'] \\ \hline \vdots \\ \hline y_N \otimes [\mathbf{I_2} \otimes \mathbf{U}'] \end{pmatrix}\right]$$

$$= E_{\xi_2}\left[\sum_{i=1}^{N}\left(y_i^2 [\mathbf{I_2} \otimes \mathbf{U}][\mathbf{A}_i][\mathbf{I_2} \otimes \mathbf{U}']\right)\right]$$

$$= \pi(1-\pi) E_{\xi_2} \sum_{i=1}^{N}\left[y_i^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{U}\mathbf{e}_i\mathbf{e}_i'\mathbf{U}'\right]$$

Since

$$\mathbf{U}\mathbf{e}_i\mathbf{e}_i'\mathbf{U}' = \left[\begin{pmatrix} U_{11} & \cdots & U_{1i} & \cdots & U_{1N} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ U_{i1} & \cdots & U_{ii} & \cdots & U_{iN} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ U_{N1} & \cdots & U_{Ni} & \cdots & U_{NN} \end{pmatrix}\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}\right]\left[\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}'\begin{pmatrix} U_{11} & \cdots & U_{1i} & \cdots & U_{1N} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ U_{i1} & \cdots & U_{ii} & \cdots & U_{iN} \\ \vdots & \cdots & & \cdots & \vdots \\ U_{N1} & \cdots & U_{Ni} & \cdots & U_{NN} \end{pmatrix}'\right],$$

$$= \begin{pmatrix} U_{1i} \\ \vdots \\ U_{ii} \\ \vdots \\ U_{Ni} \end{pmatrix}\begin{pmatrix} U_{1i} \\ \vdots \\ U_{ii} \\ \vdots \\ U_{Ni} \end{pmatrix}' = \begin{pmatrix} U_{1i} & 0 & \cdots & 0 \\ 0 & U_{2i} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & U_{Ni} \end{pmatrix} = \bigoplus_{j=1}^{N} U_{ji}$$

we obtain

$$E_{\xi_2}\left[(\mathbf{I_2} \otimes \mathbf{U})\left[Var_{\xi_1|\xi_2}(\mathbf{I}_{2N}\mathbf{M}\mathbf{y})\right](\mathbf{I_2} \otimes \mathbf{U})'\right]$$

$$= \pi(1-\pi) E_{\xi_2}\left(\sum_{i=1}^{N}\left[y_i^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \left(\bigoplus_{j=1}^{N} U_{ji}\right)\right]\right) .$$

$$= \pi(1-\pi)\sum_{i=1}^{N}\left[y_i^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \left(\frac{\mathbf{I}_N}{N}\right)\right] = \pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{I}_N$$

Hence,

$$Var_{\xi_1\xi_2}(\mathbf{Z}_2) = \left(\sigma^2 \begin{bmatrix} (1-\pi)^2 & \pi(1-\pi) \\ \pi(1-\pi) & \pi^2 \end{bmatrix} \otimes \left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right)\right) + \left[\pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{I}_N\right]$$

Summarily, the two missing data mechanisms lead to the same expected value and

the variance for the expected value, that are

$$E_{\xi_1\xi_2}(\mathbf{Z}) = \mu\left[\begin{pmatrix}1-\pi\\ \pi\end{pmatrix}\otimes\mathbf{1}_N\right], \tag{D.1}$$

and

$$Var_{\xi_1\xi_2}(\mathbf{Z}_2) = \left(\sigma^2\begin{bmatrix}(1-\pi)^2 & \pi(1-\pi)\\ \pi(1-\pi) & \pi^2\end{bmatrix}\otimes\left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N}\right)\right) + \left[\pi(1-\pi)\left(\frac{N-1}{N}\sigma^2+\mu^2\right)\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}_N\right]. \tag{D.2}$$

<u>D.1.2. Verify the expressions for the expected value and variance via simulation</u>.

1. Simulation results for $\mathbf{Z}_1 = \mathbf{MUy}$

We use the SAS procedure to verify the expressions for the expected value and variance of expected value. First, we create a population of size 4 with fixed mean of 50 and variance of 10. Table D-1 list the values of the population units that we used.

Table D-1. List of Population

Population
46
49
53
52

Second, we randomly permute the units in the population and store them in a vector $\mathbf{Y}$. Next, we create a matrix that indicates the missing value pattern, $\mathbf{M}$ of dimension 8×4, where the probability of a value being missing is π. The product of the two matrices, $\mathbf{MY} = \mathbf{Z}_1$ produces a random permuted population with missing values. This process of permuting units and then assigning missing values corresponds to a trial. We repeat this

process for 100,000 trials. We get the average mean and variance for all simulated random permuted populations. We use the theoretic expressions (D.1) and (D.2) to calculated estimation of the population mean and variance and compare to the simulation results. Table D-2 list the permutations of the population from the first ten trials of simulation.

Table D-2. List of Ten Permutations of The Population $N = 4$

Trials	$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	49	46	52	53
2	49	52	53	46
3	49	46	53	52
4	52	46	53	49
5	53	49	46	52
6	52	46	49	53
7	49	46	53	52
8	52	46	53	49
9	46	49	52	53
10	49	53	52	46

Source: th03lj06.sas

Table D-3 list the first ten simulated $\mathbf{Z}_1$ vectors, which are created from the ten permutations of the population listed in table D-2.

Table D-3. The First Ten Simulated $\mathbf{Z}_1$ vectors with $\pi = 0.1$

	Potential Observed Data				**Missing Data**			
Trials	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	0	46	52	53	49	0	0	0
2	49	52	53	46	0	0	0	0
3	49	46	53	52	0	0	0	0
4	52	46	53	49	0	0	0	0
5	53	49	46	52	0	0	0	0
6	52	46	0	53	0	0	49	0
7	49	46	0	52	0	0	53	0
8	52	46	53	49	0	0	0	0
9	46	0	52	53	0	49	0	0

10	0	53	52	0	49	0	0	46

Source: th03lj06.sas

Table D-4 presents theoretical $\mathbf{Z}_1$ calculated using expression (D.1) and simulated mean for each position of $\mathbf{Z}_1$ and the standard error for the mean. To evaluate if the theoretical one is consistent with the simulation results, we plot the standardized difference between theoretical mean and simulated mean. The plot is presented as figure D-1.

Table D-4. Difference Between Theoretical Mean and Simulated Mean for $\mathbf{Z}_1$

Data	Position (i)	Theoretical mean	Simulated mean	Standard Error	Difference	Standardized Difference
Potential Observed Data	1	45	45.002	0.048	-0.002	-0.038
	2	45	45.032	0.048	-0.032	-0.666
	3	45	45.001	0.048	-0.001	-0.029
	4	45	45.043	0.048	-0.043	-0.888
Missing Data	1	5	5.003	0.048	-0.003	-0.059
	2	5	4.957	0.047	0.043	0.914
	3	5	4.999	0.047	0.001	0.030
	4	5	4.964	0.047	0.036	0.760

Source: th03lj06.sas

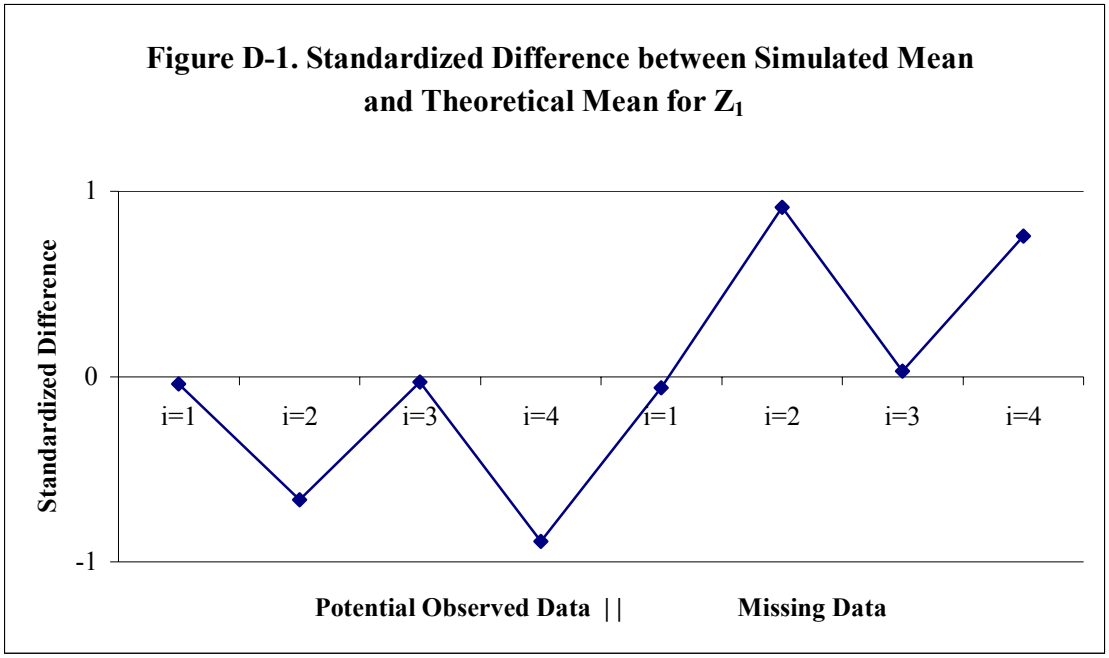


Figure D-1 shows plots of the standardized difference between simulated mean and theoretical expected value of vector $\mathbf{Z}_1$. Note that all standardized difference fall between –1 and 1. The simulation results are consistent with the theoretical expected values.

Table D-5 present the theoretical variance matrix and the simulated matrix. From the expression (D.2), the theoretical variance matrix for $\mathbf{Z}_1$ has a pattern like

$$\left(\begin{array}{cccc|cccc} p_1 & p_2 & \cdots & p_2 & p_3 & p_4 & \cdots & p_4 \\ p_2 & p_1 & \ddots & \vdots & p_4 & p_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_2 & \vdots & \ddots & \ddots & p_4 \\ p_2 & \cdots & p_2 & p_1 & p_4 & \cdots & p_4 & p_3 \\ \hline p_3 & p_4 & \cdots & p_4 & p_5 & p_6 & \cdots & p_6 \\ p_4 & p_3 & \ddots & \vdots & p_6 & p_5 & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_4 & \vdots & \ddots & \ddots & p_6 \\ p_4 & \cdots & p_4 & p_3 & p_6 & \cdots & p_6 & p_5 \end{array}\right).$$

Table D-5. Variance Matrix of $\mathbf{Z}_1 = \mathbf{MUy}$

Theoretical Variance Matrix Using Expression (D.2)							
231.75	-2.025	-2.025	-2.025	-225	-0.225	-0.225	-0.225

-2.025	231.75	-2.025	-2.025	-0.225	-225	-0.225	-0.225
-2.025	-2.025	231.75	-2.025	-0.225	-0.225	-225	-0.225
-2.025	-2.025	-2.025	231.75	-0.225	-0.225	-0.225	-225
-225	-0.225	-0.225	-0.225	225.75	-0.025	-0.025	-0.025
-0.225	-225	-0.225	-0.225	-0.025	225.75	-0.025	-0.025
-0.225	-0.225	-225	-0.225	-0.025	-0.025	225.75	-0.025
-0.225	-0.225	-0.225	-225	-0.025	-0.025	-0.025	225.75
Simulated Variance Matrix							
231.850	-3.240	-2.311	-1.340	-225.14	0.911	0.177	-0.914
-3.240	230.03	-1.922	-2.292	0.979	-223.22	-0.373	0.033
-2.311	-1.922	231.85	-2.091	-0.186	-0.327	-224.95	-0.064
-1.340	-2.292	-2.091	230.29	-1.097	0.086	0.042	-223.59
-225.14	0.979	-0.186	-1.097	225.90	-1.147	-0.194	0.880
0.911	-223.22	-0.327	0.086	-1.147	223.93	0.113	-0.353
0.177	-0.373	-224.95	0.042	-0.194	0.113	225.55	-0.370
-0.914	0.033	-0.064	-223.59	0.880	-0.353	-0.370	224.38

Source: th03lj06.sas

In order to compare the theoretical variance matrix and the simulated matrix, we compare the theoretical values and simulated values via p_1 to p_6, where simulated values are average values for all correspond values for a particular position (i.e., to calculate simulated values for p_1, I use all the diagonal elements in upper-left part of matrix, which are labeled as p_1). We run the simulation ten times by changing the starting value in the random number generate function and get ten mean values for each p_1 though p_6. Then we calculated the overall mean and standard deviation for p_1 to p_6 based on the ten means. Table D-6 present the results.

Table D-6: Comparison between the Simulated Variance Matrix and Theoretical Variance Matrix for $\mathbf{Z}_1 = \mathbf{MUy}$

Part	Theoretical Value	Simulated Average	Standard Deviation	Difference	Standardized Difference

		Value			
p_1	231.750	231.559	0.177	-0.191	-1.081
p_2	-2.025	-1.935	0.090	0.090	0.998
p_3	-225.000	-224.845	0.189	0.155	0.817
p_4	0.225	0.302	0.086	0.077	0.903
p_5	225.750	225.632	0.203	-0.118	-0.583
p_6	-0.025	0.040	0.082	0.065	0.795

Source: th03lj06.sas

We can note that all the standardized differences between theoretical and simulated values except for p_1 fall within one standard error from the simulated average value. This also shows the variance from our expressions are very close to the results from simulation.

2. Simulation results for $\mathbf{Z}_2 = (\mathbf{I_2} \otimes \mathbf{U})\mathbf{My}$.

We repeat the same process for $\mathbf{Z}_2 = (\mathbf{I_2} \otimes \mathbf{U})\mathbf{My}$. The simulation results are consistent with theoretical values from expressions. Simulation results are listed below.

Table D-7. Difference Between Theoretical Mean and Simulated Mean for $\mathbf{Z}_2$

Data	Position (i)	Theoretical mean	Simulated mean	Standard Error	Difference	Standardized Difference
Potential Observed Data	1	45	44.985	0.048	0.015	0.302
	2	45	44.967	0.048	0.033	0.677
	3	45	45.006	0.048	-0.006	-0.122
	4	45	45.048	0.048	-0.048	-0.992
Missing Data	1	5	5.037	0.048	-0.037	-0.775
	2	5	5.021	0.048	-0.021	-0.439
	3	5	4.983	0.047	0.017	0.358
	4	5	4.953	0.047	0.047	0.996

Source: th03lj07.sas

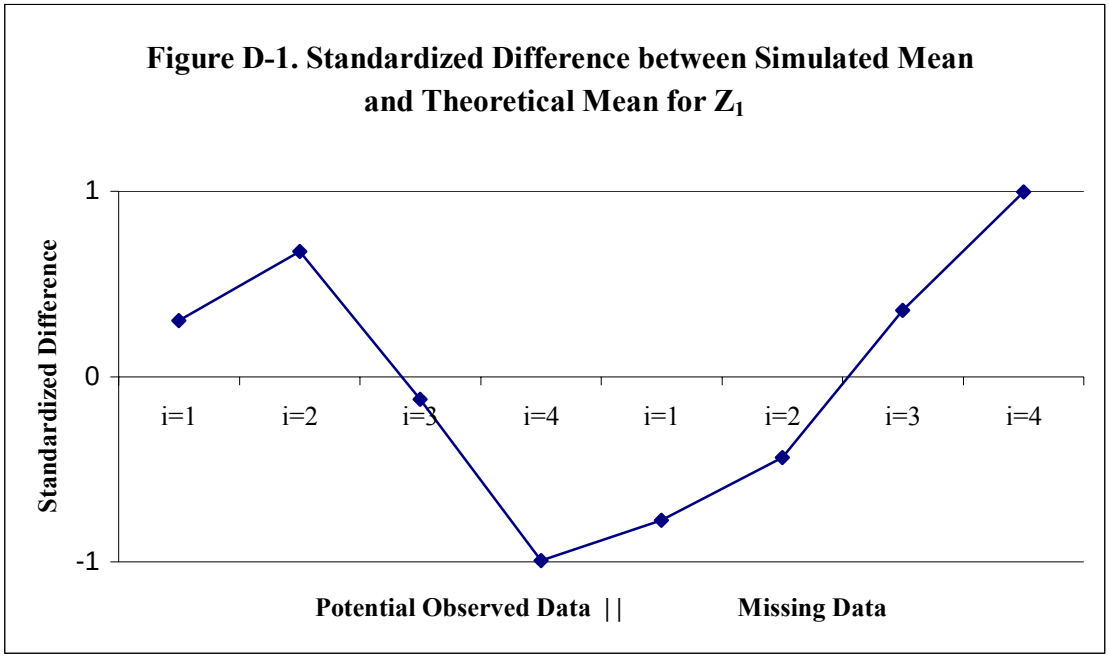


Table D-8. Variance Matrix of $Z_2 = (I_2 \otimes U)My$

Theoretical Variance Matrix Using Expression (D.2)							
231.75	-2.025	-2.025	-2.025	-225	-0.225	-0.225	-0.225
-2.025	231.75	-2.025	-2.025	-0.225	-225	-0.225	-0.225
-2.025	-2.025	231.75	-2.025	-0.225	-0.225	-225	-0.225
-2.025	-2.025	-2.025	231.75	-0.225	-0.225	-0.225	-225
-225	-0.225	-0.225	-0.225	225.75	-0.025	-0.025	-0.025
-0.225	-225	-0.225	-0.225	-0.025	225.75	-0.025	-0.025
-0.225	-0.225	-225	-0.225	-0.025	-0.025	225.75	-0.025
-0.225	-0.225	-0.225	-225	-0.025	-0.025	-0.025	225.75
Simulated Variance Matrix							
232.28	-0.761	-3.049	-2.474	-225.50	-1.280	0.813	-0.033
-0.761	232.43	-0.881	-1.415	-1.503	-225.61	-1.663	-0.598
-3.049	-0.881	236.44	-2.317	0.710	-1.598	-229.40	0.095
-2.474	-1.415	-2.317	231.18	0.258	-0.634	-0.144	-224.46
-225.50	-1.503	0.710	0.258	226.25	1.042	-1.000	-0.257
-1.280	-225.61	-1.598	-0.634	1.042	226.30	1.640	0.146
0.813	-1.663	-229.40	-0.144	-1.000	1.640	229.85	-0.098
-0.033	-0.598	0.095	-224.46	-0.257	0.146	-0.098	225.20

Source: th03lj07.sas

Table D-9: Comparison between the Simulated Variance Matrix and Theoretical Variance Matrix for $\mathbf{Z}_2 = (\mathbf{I}_2 \otimes \mathbf{U})\mathbf{My}$

Part	Theoretical Value	Simulated Average Value	Standard Deviation	Difference	Standardized Difference
p_1	231.750	232.111	0.358	0.361	0.358
p_2	-2.025	-2.180	0.063	-0.155	0.063
p_3	-225.000	-225.368	0.358	-0.368	0.358
p_4	0.225	-0.067	0.063	-0.292	0.063
p_5	225.750	226.126	0.359	0.376	0.359
p_6	-0.025	-0.185	0.063	-0.160	0.063

Source: th03lj07.sas

D.2.Deriving the Best Linear Unbiased Estimator (BLUE) of Population Mean

Let $\mathbf{Z} = \begin{pmatrix} \mathbf{Y}_1 \\ \hline \mathbf{Y}_2 \end{pmatrix}$, where $\mathbf{Y}_1 = \mathbf{Y}_I^{(o)}$ is a vector of selected sample with size n, and $\mathbf{Y}_2$ is a vector of the remaining random variables in vector $\mathbf{Z}$ with size $2N - n$.

We want to estimate the population mean, defined by $\mu = \frac{\mathbf{1}'_N}{N}\mathbf{y}$. Under model (B.2), we also could express it as $\mu = \frac{\mathbf{1}'_{2N}}{N}\mathbf{Z}$, or $\mu = \frac{1}{N}\left(\mathbf{1}'_n \mathbf{Y}_1 + \mathbf{1}'_{2N-n}\mathbf{Y}_2\right)$. Thus, the parameter of

interest, μ, is a linear combination of the random variables, $\frac{\mathbf{1}'_{2N}}{N}\mathbf{Z}$.

We use the three criteria for developing the estimator making use of the prediction framework. First, we represent the estimator as a linear function of the sample data, $P = \mathbf{L}'\mathbf{Y}_1$, where $\underset{n\times 1}{\mathbf{L}}$ is a vector of unknown constants. Next, we require the estimator to be unbiased. This implies that $E\left(\mathbf{L}'\mathbf{Y}_1 - \mu\right) = 0$.

We simplify this expression by substituting the expression for $\mu = \frac{1}{N}\left(\mathbf{1}'_n\mathbf{Y}_1 + \mathbf{1}'_{2N-n}\mathbf{Y}_2\right)$. Then, using the expected value of $\mathbf{Z}$ given by $E_{\xi_1\xi_2}\left(\mathbf{Z}\right) = \mu\left[\begin{pmatrix}1-\pi\\ \pi\end{pmatrix}\otimes\mathbf{1}_N\right]$, we get

$$\begin{aligned}
&\left(1-\pi\right)\mathbf{L}'\mathbf{1}_n\mu - \frac{1}{N}\left(\left(1-\pi\right)\mathbf{1}'_n\mathbf{1}_n\mu + \left(1-\pi\right)\mathbf{1}'_{N-n}\mathbf{1}_{N-n}\mu + \pi\mathbf{1}'_N\mathbf{1}_N\mu\right) = 0\\
&\left(1-\pi\right)\mathbf{L}'\mathbf{1}_n\mu - \frac{1}{N}\left(\left(1-\pi\right)n\mu + \left(1-\pi\right)(N-n)\mu + \pi N\mu\right) = 0\\
&\left(1-\pi\right)\mathbf{L}'\mathbf{1}_n\mu - \mu = 0
\end{aligned}.$$

Combining terms, the unbiasedness requirement implies that $\left[\left(1-\pi\right)\mathbf{L}'\mathbf{1}_n - 1\right]\mu = 0$. In order for this constraint to hold for all possible values of μ, we require that $\left(1-\pi\right)\mathbf{L}'\mathbf{1}_n = 1$.

The variance of the estimator is given by $\operatorname{var}\left(P-\mu\right) = \operatorname{var}\left(P\right)$ since the estimator is unbiased. Since $P = \mathbf{L}'\mathbf{Y}_1$ and defining $\mathbf{V} = Var\left(\mathbf{Y}_1\right)$ where $Var\left(\mathbf{Y}_1\right) = \left(\sigma^2\left(1-\pi\right)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right) + \left[\pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right]$ as a result of partitioning the variance matrix for $\mathbf{Z}$, we can express $\operatorname{var}\left(P\right) = \mathbf{L}'\left(\left[\sigma^2\left(1-\pi\right)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right] + \left[\pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right]\right)\mathbf{L} = \mathbf{L}'\mathbf{V}\mathbf{L}$.

We find values of $\mathbf{L}'$ that minimize the variance, and are unbiased. To do so, we form a Lagrangian function, where λ is the Lagrangian multiplier,

$$F = \mathbf{L}'\mathbf{V}\mathbf{L} + \left[(1-\pi)\mathbf{L}'\mathbf{1}_n - 1\right]2\lambda .$$

By differentiating this function with respect to $\mathbf{L}$ and λ, and setting the resulting derivatives to zero, and we obtain the coefficients that will result in the minimum variance.

$$\frac{\partial F}{\partial \mathbf{L}} = 2\mathbf{V}\mathbf{L} + \left[(1-\pi)\mathbf{1}_n\right]2\lambda$$

$$\text{and } \frac{\partial F}{\partial \lambda} = \left[(1-\pi)\mathbf{L}'\mathbf{1}_n - 1\right]2 .$$

Setting these two equations to zero,

$$\mathbf{V}\hat{\mathbf{L}} + \left[(1-\pi)\mathbf{1}_n\right]\hat{\lambda} = 0$$

$$(1-\pi)\hat{\mathbf{L}}'\mathbf{1}_n - 1 = 0 .$$

We transpose the terms in the second equation:

$$(1-\pi)\hat{\mathbf{L}}\mathbf{1}_n' = 1 ,$$

and represent the two equations simultaneously as

$$\left(\begin{array}{c|c} \mathbf{V} & (1-\pi)\mathbf{1}_n \\ \hline (1-\pi)\mathbf{1}_n' & 0 \end{array}\right)\left(\begin{array}{c} \hat{\mathbf{L}} \\ \hline \hat{\lambda} \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline 1 \end{array}\right).$$

The matrix $\left(\begin{array}{c|c} \mathbf{V} & (1-\pi)\mathbf{1}_n \\ \hline (1-\pi)\mathbf{1}_n' & 0 \end{array}\right)$ is of full rank, resulting in

$$\left(\begin{array}{c} \hat{\mathbf{L}} \\ \hline \hat{\lambda} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{V} & (1-\pi)\mathbf{1}_n \\ \hline (1-\pi)\mathbf{1}_n' & 0 \end{array}\right)^{-1}\left(\begin{array}{c} 0 \\ \hline 1 \end{array}\right).$$

Since $\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \left(\begin{array}{c|c} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}^{-1} \\ \hline -\mathbf{C}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}^{-1} \end{array}\right)$ where $\mathbf{C} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$,

$$\mathbf{C} = 0 - \left[\left(1-\pi\right)\mathbf{1}_n'\right]\mathbf{V}^{-1}\left[\left(1-\pi\right)\mathbf{1}_n\right] = -\left(1-\pi\right)^2\mathbf{1}_n'\mathbf{V}^{-1}\mathbf{1}_n$$

and

$$\begin{aligned}\hat{\mathbf{L}} &= \left[-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}^{-1}\right]\times 1 = -\mathbf{V}^{-1}\left[\left(1-\pi\right)\mathbf{1}_n\right]\left[-\left(1-\pi\right)^2\mathbf{1}_n'\mathbf{V}^{-1}\mathbf{1}_n\right]^{-1} \\ &= -\frac{1}{1-\pi}\left[\mathbf{V}^{-1}\mathbf{1}_n\right]\left[-\mathbf{1}_n'\mathbf{V}^{-1}\mathbf{1}_n\right]^{-1}\end{aligned}.$$

Now

$$\begin{aligned}\mathbf{V} &= \left(\sigma^2\left(1-\pi\right)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right) + \left[\pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right] \\ &= \left[\sigma^2\left(1-\pi\right)^2 + \pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\right]\mathbf{I}_n - \frac{\sigma^2\left(1-\pi\right)^2}{N}\mathbf{J}_n \\ &= a\mathbf{I}_n + b\mathbf{J}_n = a\left(\mathbf{I}_n + \frac{b}{a}\mathbf{J}_n\right),\end{aligned}$$

Where

$$a = \left[\sigma^2\left(1-\pi\right)^2 + \pi\left(1-\pi\right)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\right] \text{ and } b = \frac{\sigma^2\left(1-\pi\right)^2}{N}$$

From patterered inverses,

$$\mathbf{V}^{-1} = \frac{1}{a}\left(\mathbf{I}_n - \frac{\frac{b}{a}}{1+\left(\frac{b}{a}\right)n}\mathbf{J}_n\right) = \frac{1}{a}\left(\mathbf{I}_n - \frac{nb}{a+nb}\frac{\mathbf{J}_n}{n}\right).$$

Also, $\mathbf{V}^{-1}\mathbf{1}_n = \frac{1}{a}\left(\mathbf{1}_n - \frac{nb}{a+nb}\mathbf{1}_n\right) = \left(\frac{1}{a+nb}\right)\mathbf{1}_n$,

$$\mathbf{1}_n'\mathbf{V}^{-1}\mathbf{1}_n = \frac{n}{a+nb}.$$

Thus

$$\hat{\mathbf{L}} = -\frac{1}{1-\pi}\left[\mathbf{V}^{-1}\mathbf{1}_n\right]\left[-\mathbf{1}_n'\mathbf{V}^{-1}\mathbf{1}_n\right]^{-1} = \frac{1}{1-\pi}\left(\frac{1}{a+nb}\right)\mathbf{1}_n\frac{a+nb}{n}$$

$$= \frac{1}{n(1-\pi)} \mathbf{1}' .$$

As a result, the BLUE of μ is given by $\hat{\mathbf{L}}'\mathbf{Y}_1$ which is

$$\hat{P} = \frac{1}{n(1-\pi)} \mathbf{1}_n' \mathbf{Y}_1 . \tag{D.3}$$

The variance of the estimator is given by

$$\begin{aligned} \operatorname{var}\left(\hat{P}\right) &= \hat{\mathbf{L}}'\left(\left[\sigma^2(1-\pi)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right] + \left[\pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right]\right)\hat{\mathbf{L}} \\ &= \frac{1}{n^2(1-\pi)^2}\mathbf{1}_n'\left(\left[\sigma^2(1-\pi)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right] + \left[\pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right]\right)\mathbf{1}_n \end{aligned}$$

Since $\mathbf{1}_n'\left(\left[\sigma^2(1-\pi)^2\left(\mathbf{I}_n - \frac{\mathbf{J}_n}{N}\right)\right] + \left[\pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\mathbf{I}_n\right]\right)\mathbf{1}_n$ is the sum of the elements in $\mathbf{V}$, hence,

$$\operatorname{var}\left(\hat{P}\right) = \frac{1}{n^2(1-\pi)^2}\left(n\left[\frac{N-1}{N}\sigma^2(1-\pi)^2 + \pi(1-\pi)\left(\frac{N-1}{N}\sigma^2 + \mu^2\right)\right] - \frac{\left(n^2-n\right)}{N}\sigma^2(1-\pi)^2\right)$$ Th

at is

$$\operatorname{var}\left(\hat{P}\right) = \frac{1}{n(1-\pi)}\left[\pi\mu^2 + \frac{N - n(1-\pi) - \pi}{N}\sigma^2\right] \tag{D.4}$$

D.3. Comparisons of Methods

<u>D.3.1. We compare the methods assuming we know</u> $\pi, \sigma^2, \mathrm{N}, n$

1. Theoretical results:

(1) We use the method ignoring the missing values suggested by Cochran (1977), where the estimate of the population mean is given by $\bar{Y}_1^* = \frac{1}{n_1}\mathbf{1}_n'\mathbf{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n} Y_i$, where $Y_i = \sum_{s=1}^{N} M_i U_{is} y_s$

or $Y_i = \sum_{s=1}^{N} U_{is} H_i y_s$, and $n_1 = \sum_{i=1}^{n}\sum_{s=1}^{N} M_i U_{is}$ or $n_1 = \sum_{i=1}^{n}\sum_{s=1}^{N} U_{is} H_s$. The variance of the estimate is given by $\sigma_c^2 = \left(1 - \frac{n_1}{N}\right)\frac{\sigma^2}{n_1}$. We estimate σ^2 with $s^{*2} = \sum_{i=1}^{n_1} \frac{\left(Y_i - \overline{Y}_1^*\right)^2}{n_1 - 1}$, thus

$$\hat{\sigma}_c^2 = \left(1 - \frac{n_1}{N}\right)\frac{s^{*2}}{n_1} \qquad \text{(D.5).}$$

(2) BLUE:

Estimate of mean: $\hat{P} = \frac{1}{n(1-\pi)} \mathbf{1}_n' \mathbf{Y}_1$

Variance of estimate:

$$Var\left(\hat{P}\right) = \frac{1}{n(1-\pi)}\left[\pi\hat{\mu}^2 + \frac{N - n(1-\pi) - \pi}{N}\sigma^2\right].$$

Under this assumption, the two methods have different estimate of population mean and the variance of the estimator.

2. SAS simulation results (Program see appendix II):

We simulate a population of 50 subjects and each subject is associated with a response real number. The population has a fixed mean of 50 and variance of 4. We randomly permute the population units and select the first n units as the sample. Each sample unit is missing at random with probability π. We run the simulation of 10,000 trials and get the expected value of mean and expected variance of the estimator for different n and π. We also calculate the theoretical variance of BLUE using the expression (D.4) and theoretical sample variance using expression (D.5). The results of simulation are presented in table D-10.

Table D-10. Simulation Results Assuming Known π, σ^2, N, n

Method	Sample Size n	π	*Simulated* $\hat{\mu}$	*Simulated* $\sigma^2\{\hat{\mu}\}$	*Theoretical* $\sigma^2\{\hat{\mu}\}$
BLUE	15	0.1	49.9752	18.856	18.716
	15	0.2	50.0408	41.877	41.987
	15	0.4	49.9353	113.950	111.185
	15	0.6	50.1102	250.008	251.682
	30	0.1	49.9945	9.285	9.325
	30	0.2	49.9971	21.256	20.917
	30	0.4	49.9887	55.550	55.671
	30	0.6	50.0046	127.158	125.272
Method	Sample Size n	π	*Simulated* $\hat{\mu}$	*Simulated* $\sigma^2\{\hat{\mu}\}$	*Theoretical Variance*
Ignore the Missing Values	15	0.1	50.0024	0.218	0.21833
	15	0.2	50.0020	0.261	0.26054
	15	0.4	50.0024	0.394	0.39325
	15	0.6	49.9667	1.938	0.66162
	30	0.1	49.9971	0.069	0.06851
	30	0.2	49.9996	0.089	0.08804
	30	0.4	49.9988	0.145	0.14729
	30	0.6	49.9899	0.274	0.27139

Source: th03lj05.sas

From table D-10, we can note that the variance of the BLUE is greater than the variance of the estimator ignoring the missing values under the same situation. So the performance of BLUE is poor. When sample size becomes larger, then both variances become smaller. When the probability of being missing increases, the variances of both estimators become larger. We also can find that the estimated the variances of BLUE are very close to the theoretical ones and the variances of estimators ignoring the missing values are very similar to the estimated sample variance. This simulation also verified the expression (D.4) using estimates versus actual variance from simulation.

To evaluate if the developed BLUE is unbiased, we plot the absolute bias, which equals the estimated mean minus the true mean, versus the probability of being missing (π) in order to see the performance of BLUE at different π. The figure D-3 and figure D-4 correspond to different sample sizes. We can see all the absolute bias fall within the range of one standard error of estimated mean from zero. The plots also distribute closer to zero and smoothly when the sample size is bigger (figure D-4) compared to that of smaller sample size (Figure D-3).

Figure D-3. Absolute Bias of BLUE methods when N=50 and n=15

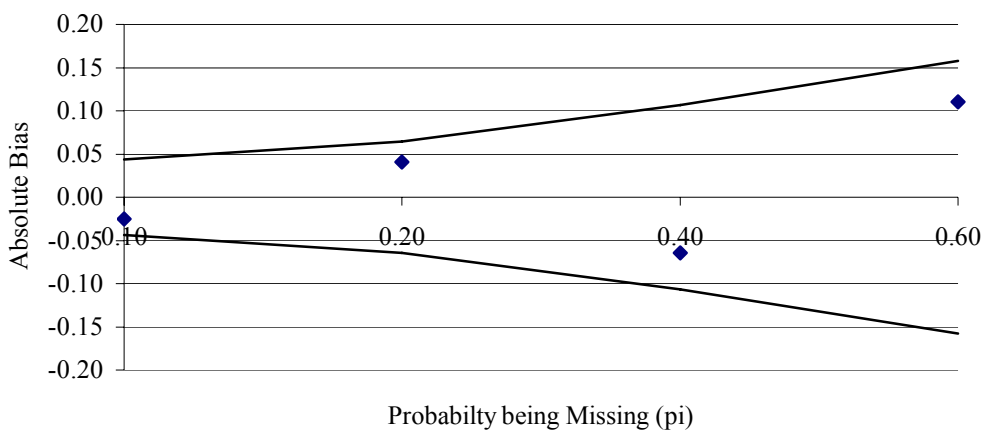
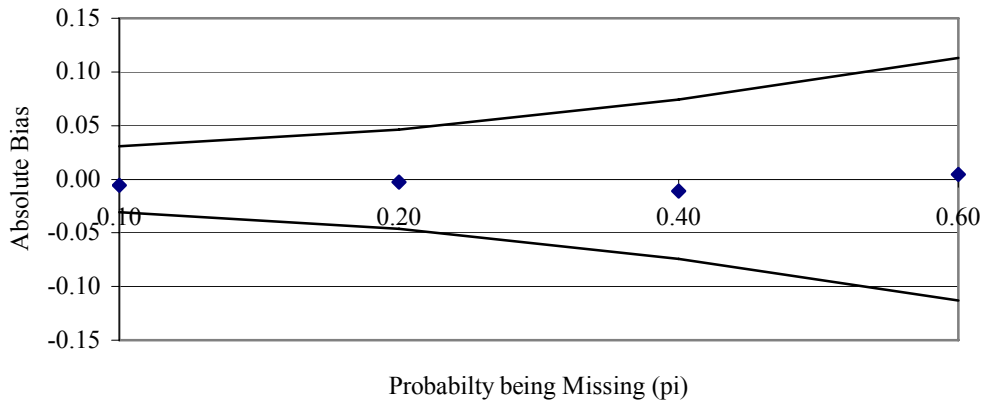


Figure D-4. Absolute Bias of BLUE methods when N=50 and n=30



We compared the relative bias (the absolute bias divided by standard error) for the two estimations. From figure D-5 and D-6, we can see all the relative bias of BLUE (Black dots) fall between the –1 and 1, but some of the relative bias of estimate ignoring missing values (Red stars) fall out of the range of –1 and 1. This shows the BLUE is unbiased and the ad hoc estimator ignoring missing values is somewhat biased.

Figure D-5. Relative Bias of two methods when N=50 and n=15

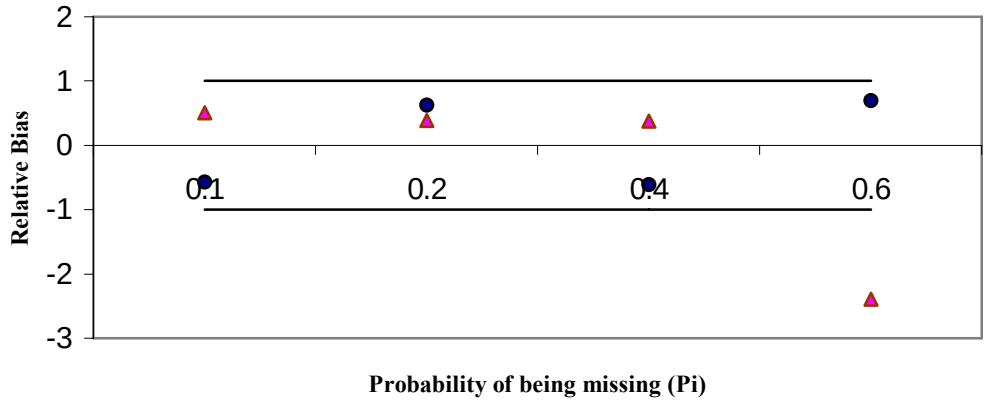
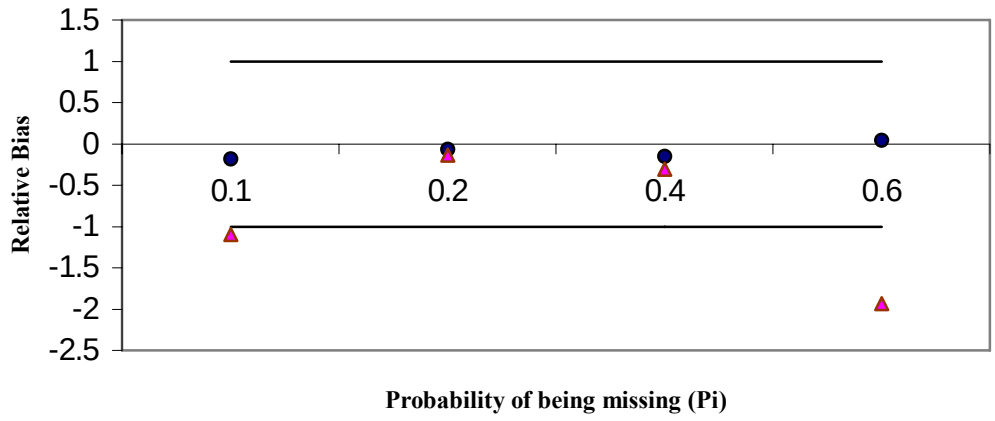


Figure D-6. Relative Bias of two methods when N=50 and n=30



From Table D-10 we have seen that the variances of estimators increase when the proportion of missing values increase. To compare the if the increases are the same, we plot the variances versus proportion of missing values for both estimations, which are shown as figure D-7 and D-8. The curves for BLUE in figure D-7 are more smoothly and increase stably. For estimation ignoring missing values, the variance has a more rapid increase when the sample size is small.

Figure D—7. Plot Variance of Estimator for BLUE method

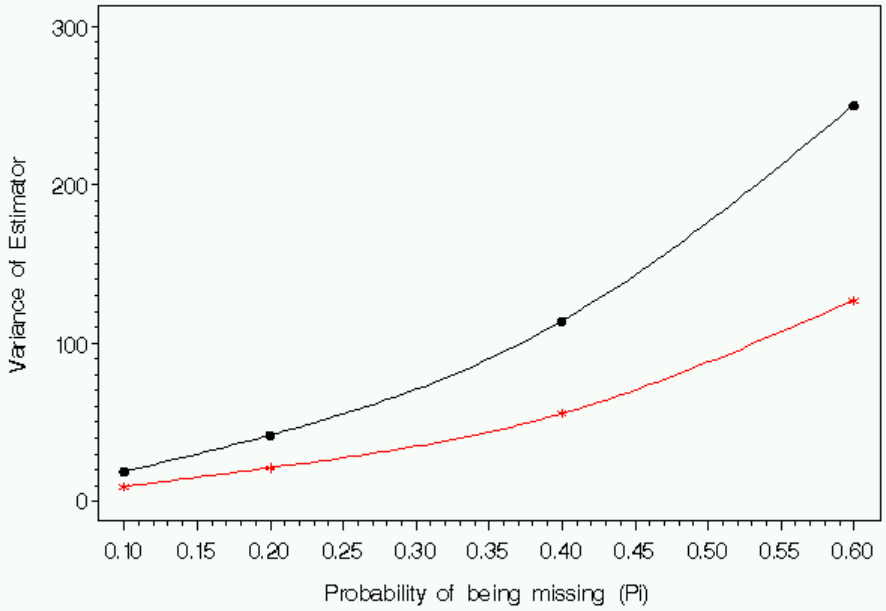
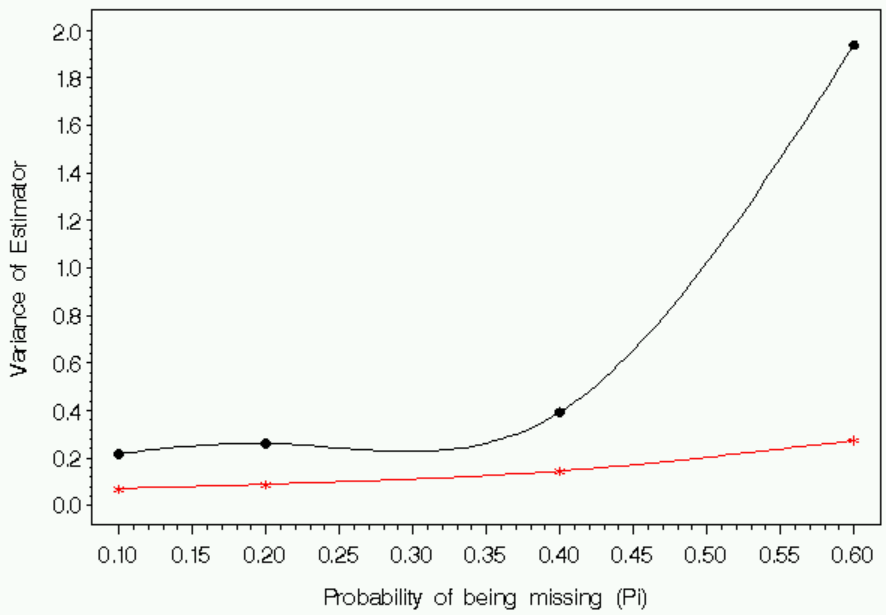


Figure D—8. Plot Variance of Estimator for method Ignoring Missing Values



D.3.2 We compare the methods assuming we know N, n, σ^2, but we don't know π:

1. Theoretical results.

(1). We use the method ignoring the missing values suggested by Cochran (1977), and get the

same expression as in section D.3.1, where the estimate of the population mean is given by

$\bar{Y}_1^* = \frac{1}{n_1}\mathbf{1}_n'\mathbf{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n} Y_i$, the variance $\hat{\sigma}_c^2 = \left(1 - \frac{n_1}{N}\right)\frac{s^{*2}}{n_1}$

(2) BLUE:

Use $\hat{\pi} = \frac{n_0}{n}$

Estimate of mean: $\hat{\mu} = \frac{1}{n\left(1 - \frac{n_0}{n}\right)}\mathbf{1}_n'\mathbf{Y}_1 = \frac{1}{n_1}\mathbf{1}_n'\mathbf{Y}_1$

We can find that under this assumption, the BLUE is equivalent to the estimator ignoring missing values. This also was verified by SAS simulation, which results are presented in table D-11.

2. Simulation results.

Table D-11 Simulated Results Assuming N, n, σ^2 Are Known, But π Is Unknown

Method	Sample Size n	π	*Simulated* $\hat{\mu}$	*Simulated* $\sigma^2\{\hat{\mu}\}$	*Theoretical Variance*
BLUE	20	0.1	50.0003	0.14159	0.14335
	20	0.2	50.0000	0.17607	0.17350
Ignore the Missing Values	20	0.1	50.0003	0.14159	0.14335
	20	0.2	50.0000	0.17607	0.17350

Source: th03lj04.sas

The table D-11 shows, the estimated mean and the variance of the estimator are the same for both methods. And the variance of the estimator is close to the estimated sample variance.

<u>D.3.3 We compare the methods assuming we know</u> N, n <u>, but we don't know</u> π <u>and</u> σ^2:

We can get the same results that the BLUE is equivalent to the methods ignoring missing values as assuming we know N, n, σ^2, but π is unknown.

E. DISCUSSION

In the survey estimation, source of variation could be sampling error and non-sampling error. We measure the sampling error with the standard error of an estimator. Other sources of variation is called non-sampling error. The important types of non-sampling errors includes nonresponse(cause missing data), measurement error, and outlying observations (Lehtonen and Pahkinen1994). Nonresponse results in a sample data set whose real size n_1 is smaller than the intended sample size n, thus increasing the standard errors of the estimates and resulting biased estimate. Therefore, the amount and structure of missing data should be carefully examined.

We tend to develop a new method to estimate the population mean under the assumption that sample units are missing completely at random (MCAR) in simple random sampling (SRS) from a finite super-population. Recent years, best linear unbiased predictor (BLUP) based on mixed model theory is typically used for realized random effects (Goldberger1962; Robinson 1991; Stanek, Well, and Ockene 1999; Stanek 2002). Scott and Smith (1969) developed Predictors of linear combinations of elements of a finite population in a two-stage sampling setting using a super-population model. This research extends the finite population random permutation model framework to missing data and we develop the estimator using a best linear unbiased predictor approach. We got the expressions of the best linear unbiased estimator of the population mean based on a simple random sample with missing values and its variance. A simulation shows the poor performance of the new estimator (table D-1) even though the plots (figure D-3, D-4, D-5, and D-6) of absolute

bias and relative bias from the simulation show the new estimator is unbiased assuming we know $\pi, \sigma^2, \mathrm{N}, n$.

Some typical methods are also used to handle miss data, such as reweighting (Lehtonen and Pahkinen 1994), imputation, expectation maximization (EM) algorithm (Dempster, Laird and Rubin 1977), and the method ignoring the missing values. Rubin, D.B. (1986) introduced the method of multiple imputation. J.L.Schafer (2002) presented a EM algorithm for parameter estimation for multivariate linear mixed-effects models with missing values. Haziza and Rao (2003) discuss the problem of unweighted imputation for missing survey data. They show that unweighted imputation, unlike weighted imputation, generally leads to biased estimators under the design-based approach (i.e., uniform response). They propose a bias-adjusted estimator, which is simple to obtain and has the desirable property that it is approximately unbiased under both the design-based and the model-based approaches.

We compared the best linear unbiased estimator to the estimator ignoring missing values. Assuming we know $\pi, \sigma^2, \mathrm{N}, n$, a simulation shows the BLUE works more stable especially under small sample size and large proportion of missing values than the estimator ignoring missing values. The plots (figure D-3, D-4) show the relative bias of BLUE fall between the –1 and 1, but some of the relative bias of estimate ignoring missing values fall out of the range of –1 and 1, and we may think the estimator ignoring missing values is biased and the bias increase when the proportion of missing data increase. But the variance of the BLUE is greater than the variance of the estimator ignoring the missing values under the

same situation. In the cases when we don't know π and we need to estimate π using $\hat{\pi} = \frac{n_0}{n}$, the two methods are equivalent.

The proposed research may provide the groundwork for subsequent development of methods that can address missing data in more complex setting. Since the time is limited, we can't compare the BLUE to other estimator from the methods of handling missing data, such as multiple imputation, EM imputation and regression imputation. We only consider two types of missing data mechanisms and one-stage simple random sample from a finite population and one measurement on per subject without measurement error.

APPENDICES

Appendix I. Verify the expressions of the expected value and variance of Z matrix

1. For $Z = MUy$

SAS Program:

```
OPTIONS LINESIZE=100 PAGESIZE=80 NOCENTER NODATE NONUMBER NOFMTERR;
*********************************************************************
*;
* PROJECT NAME: Simulation of samples for Thesis        ;
*           PROGRAM NAME       DATE                 PROGRAMMER    ;
TITLE1 "Source:th03lj06.sas  12/08/03 LJS "   ;
* Description: Simulation macro to evaluate expected value and variance ;
*             For first missing mechanism z=MUy                   ;
*     INPUT: none                                              ;
*********************************************************************
*;
LIBNAME thesis 'C:\My Documents\jingsonglu\thesis\data';
DATA POP;
 input population;
 cards;
 46
 49
 53
 52
 ;
RUN;
PROC PRINT noobs;
title2 "Table D-1 List of Population with Mean of 50";
title3 "         and Variance of 9";
RUN;
PROC MEANS DATA=POP MEAN VAR NWAY;
RUN;
%Macro SRS(N,nsim,pi,data,data1,data2,type,c,m1);
PROC IML;
 USE POP;
 READ ALL INTO pop;
 CLOSE POP;
  N=&n;                    /*N=population size*/
```

```
MEAN=SUM(POP)/N;                    /*POPULATION MEAN*/
VAR=(SSQ(POP)-(N*(MEAN**2)))/(N-1);  /*population VARIANCE*/

N2=2*&N;NP1=&N+1;/*N2=2*N, NP1=N+1*/
ZZ=J(N2,&nsim,0);/*column of zz contains each simulation of MUy */
s=J(N2,N2,0);/*VARIANCE MATRIX */
ssqx=s;
PI=&pi;
permu=J(N,&nsim,0);

Do sim=1 to &nsim;/*start of simulation*/
/*PERMUTATION*/
 check = j(&N,1, 0);/*vector to check 'un-replacement'*/
 comb=check;/*vector to save permutated units COMB=Uy*/
 do i=1 to &N;
  label:u=ranuni(&c);
  k=int(u*&N+1);
  kk=0;
  label2:kk=kk+1;
   if k=check[kk] then goto label;
   if kk=i then go to skip;
  goto label2;
  skip:check[i]=k;
  comb[i]=pop[k];/*randomly pick for each position without replacement*/
  permu[i,sim]=pop[k];
 end;/*end of permutation*/


 Z=j(N2,1,0);
 M=J(N2,&N,0);
 M1=I(&N);          /*M is the matrix indicate of missing*/
 do i=1 to &N;
  M1[i,i]=RANBIN(0,1,pi);
 end;
 /*set the matrix M*/
 I=I(&N);
 M2=I-M1;
 do i= 1 to &N;
  M[i,i]=m2[i,i];
 end;
 do i=NP1 to N2;
```

```
  ii=i-&N;
  M[i,ii]=m1[ii,ii];
 end;
 z=M*comb;

 do i= 1 to N2;
  zz[i,sim]=z[i];
 end;
end;/*end of simulation*/
/*mean of Z*/
nz=ncol(zz);
meanz=j(N2,1,0);/*simulation mean for each element of Z*/
stdz=meanz;
do j=1 to N2;
 meanz[j]=zz[j,+]/nz;
 stdz[j]=(SSQ(zz[j,])-(&nsim*(MEANz[j]**2)))/(&Nsim-1);
end;
expZ=j(N2,1,0);/*expected value of z=MUy*/
do i=1 to &N;
 expz[i]=(1-&pi)*mean;
end;
do i= NP1 to N2;
 expz[i]=&pi*mean;
end;

/*variance*/
 /*From simulation*/
do i=1 to NZ;
  ssqx=ssqx+zz[,i]*t(zz[,i]);
end;
s=(ssqx-nz*meanz*t(meanz))/(nz-1);
 /*From expression*/
vp1=J(&n,&n,0);/*part1 of the matrix*/
vp2=vp1;
vp3=vp1;
IN=I(&N);
JN=J(&N,&N,1);

vp1=(var*((1-pi)**2))@(IN-JN/&N)+(pi*(1-pi)*((mean**2)+var*((&n-1)/&n))
)@IN;
```

```
vp2=(var*pi*(1-pi))@(IN-JN/&N)-(pi*(1-pi)*((mean**2)+var*((&n-1)/&n)))@
IN;

vp3=(var*(pi**2))@(IN-JN/&N)+(pi*(1-pi)*((mean**2)+var*((&n-1)/&n)))@IN
;
 print /"Simulation variance Matrix" s
     /"part of variance matrix from expression" /vp1 vp2 vp3;
  if &type=1 then do;
 permut=t(permu);
 names={'Posit1','Posit2','Posit3','Posit4'};
 create &data from permut[colname=names];
 append from permut;
 zzz=t(zz);
 names1={'Posit1','Posit2','Posit3','Posit4',
        'Posit5','Posit6','Posit7','Posit8'};
 create &data1 from zzz[colname=names1];
 append from zzz;
 end;
 If &type=0 then do;
  meanzz=t(meanz);
  stdzz=t(stdz);
  expzz=t(expz);
  ez=J(3,N2,0);
  do J=1 to n2;
    ez[1,j]=meanzz[j];
    ez[2,j]=stdzz[j];
    ez[3,j]=expzz[j];
  end;
  names2={'Posit1','Posit2','Posit3','Posit4',
        'Posit5','Posit6','Posit7','Posit8'};
  create &data from ez[colname=names2];
  append from ez;
  tez=t(ez);
  names3={'Sim_mean','Var_est','Theo_mean'};
  create &data1 from tez[colname=names3];
  append from tez;
 end;
 /*compare variance*/
 od1=J(6,1,0);od3=od1;od2=J(12,1,0);
 d1=J(4,1,0);d2=d1;d3=d1;
 do i=1 to 4;
```

```
 i2=i+4;
 d1[i]=s[i,i];
 d2[i]=s[i,i2];
 d3[i]=s[i2,i2];
end;
do i=1 to 3;
   i2=i+1;
   i3=i+5;i4=i+6;
   od1[i]=s[1,i2];
   od2[i]=s[1,i3];od2[i4]=s[5,i2];
   od3[i]=s[5,i3];

end;
do i=4 to 5;
   i2=i-1;
   i3=i+3;i4=i+6;
   od1[i]=s[2,i2];
   od2[i]=s[2,i3];od2[i4]=s[6,i2];
   od3[i]=s[6,i3];
end;
od1[6]=s[3,4];
od2[6]=s[3,8];od2[12]=s[7,4];
od3[6]=s[7,8];
Out=J(6,2,0);
out[1,1]=sum(d1)/4;
out[1,2]=1;
out[2,1]=sum(od1)/6;
out[2,2]=2;
out[3,1]=sum(d2)/4;
out[3,2]=3;
out[4,1]=sum(od2)/12;
out[4,2]=4;
out[5,1]=sum(d3)/4;
out[5,2]=5;
out[6,1]=sum(od3)/6;
out[6,2]=6;
print out;
 names3={"&m1","Position"};
 create &data2 from out[colname=names3];
 append from out;
RUN;
```

```
QUIT;
%mend;
%SRS(4,10,0.1,d1,d2,s1,1,22434,vm1);
%SRS(4,100000,0.1,d3,d4,s2,0,4444443,vm2);
%SRS(4,100000,0.1,d3,d4,s3,0,646464,vm3);
%SRS(4,100000,0.1,d3,d4,s4,0,646464,vm4);
%SRS(4,100000,0.1,d3,d4,s5,0,555355,vm5);
%SRS(4,100000,0.1,d3,d4,s6,0,57544755,vm6);
%SRS(4,100000,0.1,d3,d4,s7,0,47364363,vm7);
%SRS(4,100000,0.1,d3,d4,s8,0,373477,vm8);
%SRS(4,100000,0.1,d3,d4,s9,0,65636,vm9);
%SRS(4,100000,0.1,d3,d4,s10,0,62634,vm10);
%SRS(4,100000,0.1,d3,d4,s11,0,97987,vm11);
data permut;
 set d1;
 Trials=_N_;
run;
proc print data=permut noobs;
title2 "Table D-2 List of Permutations of The Population";
Title3 "        From the First Ten Trials";
Footnote "Source:th03lj06.sas";
var trials posit1-posit4;
run;

data z;
 set d2;
 Trials=_N_;
run;
proc print data=z noobs;
title2 "Table D-3 The First Ten Simulated Z1 vectors";
Title3 "        and Expected Value of Z1";
Footnote "Source:th03lj06.sas";
var trials posit1-posit8;
run;
proc print data=d3 noobs;
title2 "Table D-3 The First Ten Simulated Z1 vectors";
Title3 "        and Expected Value of Z1";
Footnote "Source:th03lj06.sas";
run;
proc print data=d4 noobs;
run;
```

```
proc format;
value posif 0="Potential Observed Data"
          1="Missing Data";
data plot;
  set d4;
  bias=Theo_mean-Sim_mean;
  sigma=sqrt(Var_est/100000);
  relbias=bias/sigma;
  up=1;low=-1;
  Posit=_N_;
  dat=0;
  if posit >=5 then dat=1;
  if dat=1 then posit=posit-4;
  format dat posif.;
run;
proc print data=plot noobs; var dat posit Theo_mean Sim_mean sigma bias
relbias;
Title2 "Table D-4. Difference Between Theoretical Mean and Simulated Mean";
run;
goptions ftext=SWISS ctext=BLACK htext=1 cells;
 axis1 width=1 offset=(3 pct) label=(a=90 'Relative Bias' r=0);
 axis2 width=1 offset=(3 pct) label=('Position');
symbol1 i=join c=blue;
symbol2 v=dot i=none  c=black;
symbol3 v=star i=none  c=red;
proc gplot data=plot;
 plot up*posit=1 low*posit=1 relbias*posit=2 /overlay
                                          vaxis = axis1
                                          haxis = axis2;
 Title " Figure D-1. Standardized diffrence between Simulated mean and
Theoretical mean";
 Footnote "Source:th03lj06.sas";
 run;
proc sort data=s2;
by position;
proc sort data=s3;
by position;
proc sort data=s4;
by position;
proc sort data=s5;
by position;
```

```
proc sort data=s6;
by position;
proc sort data=s7;
by position;
proc sort data=s8;
by position;
proc sort data=s9;
by position;
proc sort data=s10;
by position;
proc sort data=s11;
by position;
RUN;
data s;
 merge s2 s3 s4 s5 s6 s7 s8 s9 s10 s11;
 by position;
 drop position;
run;
proc iml;
use s;
read all into s;
close s;
n=ncol(s);
d=j(6,2,0);

do i=1 to 6;
 d[i,1]=s[i,+]/n;
 d[i,2]=(SSQ(s[i,])-(n*(d[i,1]**2)))/(n-1);
 end;
 print d;
  names ={"Sim_mean","Var_est"};
  create d5 from d[colname=names];
  append from d;
 run;quit;
 data d6;
   set d5;
   Part=_n_;
   if part=1 then Theo_mean=231.75;
   if part=2 then Theo_mean=-2.025;
   if part=3 then Theo_mean=-225;
   if part=4 then Theo_mean=--0.225;
```

```
  if part=5 then Theo_mean=225.75;
  if part=6 then Theo_mean=-0.025;
  std_est=sqrt(Var_est);
  diff=Sim_mean-Theo_mean;
 run;
 proc print data=d6 noobs;
 title2 "Table D-6: Compare the simulated Variance Matrix to Theoretical
Variance Matrix";
 var Part Theo_mean Sim_mean Var_est std_est diff;
 run;

proc print data=s;
run;
```

2. For $\mathbf{Z}_2 = (\mathbf{I}_2 \otimes \mathbf{U})\mathbf{My}$

```
OPTIONS LINESIZE=100 PAGESIZE=80 NOCENTER NODATE NONUMBER NOFMTERR;
*********************************************************************
*;
* PROJECT NAME: Simulation of samples for Thesis        ;
*           PROGRAM NAME       DATE                 PROGRAMMER    ;
TITLE1 "Source:th03lj07.sas  12/011/03 LJS "   ;
* Description: Simulation macro to evaluate expected value and variance ;
*             For first missing mechanism z2=I2@UMy                    ;
*     INPUT: none                                                     ;
*********************************************************************
*;
LIBNAME thesis 'C:\My Documents\jingsonglu\thesis\data';
DATA POP;
 input population;
 cards;
 46
 49
 53
 52
 ;
RUN;

%Macro SRS(N,nsim,pi,data,data1,data2,type,c,m1);
```

```
PROC IML;
 USE POP;
 READ ALL INTO pop;
 CLOSE POP;
  N=&n;                    /*N=population size*/
 MEAN=SUM(POP)/N;                 /*POPULATION MEAN*/
 VAR=(SSQ(POP)-(N*(MEAN**2)))/(N-1);  /*population VARIANCE*/

 N2=2*&N;NP1=&N+1;/*N2=2*N, NP1=N+1*/
 ZZ=J(N2,&nsim,0);/*column of zz contains each simulation of MUy */
 s=J(N2,N2,0);/*VARIANCE MATRIX */
 ssqx=s;
 PI=&pi;
 permu=J(N,&nsim,0);

  Do sim=1 to &nsim;/*start of simulation*/

  Z=j(N2,1,0);
  M=J(N2,&N,0);
  H=I(&N);          /*M is the matrix indicate of missing*/
  do i=1 to &N;
   H[i,i]=RANBIN(0,1,pi);
  end;
  /*set the matrix M*/
  I=I(&N);
  H2=I-H;
  do i= 1 to &N;
   M[i,i]=H2[i,i];
  end;
  do i=NP1 to N2;
   ii=i-&N;
   M[i,ii]=H[ii,ii];
  end;
  my=M*pop;
  /*PERMUTATION*/
  check = j(&N,1, 0);/*vector to check 'un-replacement'*/
  z=j(N2,1, 0);/*vector to save permutated units z=I2@UMy*/
  do i=1 to &N;/*For first N position*/
   label:u=ranuni(0);
   k=int(u*&N+1);
   kk=0;
```

```
  label2:kk=kk+1;
   if k=check[kk] then goto label;
   if kk=i then go to skip;
  goto label2;
  skip:check[i]=k;
  z[i]=my[k];/*randomly pick for each position without replacement*/
  z[&N+i]=my[k+&n];
 end;/*end of permutation*/



 do i= 1 to N2;
  zz[i,sim]=z[i];
 end;
end;/*end of simulation*/
/*mean of Z*/
nz=ncol(zz);
meanz=j(N2,1,0);/*simulation mean for each element of Z*/
stdz=meanz;
do j=1 to N2;
 meanz[j]=zz[j,+]/nz;
 stdz[j]=(SSQ(zz[j,])-(&nsim*(MEANz[j]**2)))/(&Nsim-1);
end;
expZ=j(N2,1,0);/*expected value of z=MUy*/
do i=1 to &N;
 expz[i]=(1-&pi)*mean;
end;
do i= NP1 to N2;
 expz[i]=&pi*mean;
end;

/*variance*/
 /*From simulation*/
do i=1 to NZ;
  ssqx=ssqx+zz[,i]*t(zz[,i]);
end;
s=(ssqx-nz*meanz*t(meanz))/(nz-1);
 /*From expression*/
vp1=J(&n,&n,0);/*part1 of the matrix*/
vp2=vp1;
vp3=vp1;
```

```
 IN=I(&N);
 JN=J(&N,&N,1);

vp1=(var*((1-pi)**2))@(IN-JN/&N)+(pi*(1-pi)*((mean**2)+var*((&n-1)/&n))
)@IN;

vp2=(var*pi*(1-pi))@(IN-JN/&N)-(pi*(1-pi)*((mean**2)+var*((&n-1)/&n)))@
IN;

vp3=(var*(pi**2))@(IN-JN/&N)+(pi*(1-pi)*((mean**2)+var*((&n-1)/&n)))@IN
;
 print "Expected Z" expz "mean of z from simulation" meanz
     /"Simulation variance Matrix" s
     /"part of variance matrix from expression" /vp1 vp2 vp3;
 RUN;
  if &type=1 then do;
 permut=t(permu);
 names={'Posit1','Posit2','Posit3','Posit4'};
 create &data from permut[colname=names];
 append from permut;
 zzz=t(zz);
 names1={'Posit1','Posit2','Posit3','Posit4',
        'Posit5','Posit6','Posit7','Posit8'};
 create &data1 from zzz[colname=names1];
 append from zzz;
 end;
 If &type=0 then do;
  meanzz=t(meanz);
  stdzz=t(stdz);
  expzz=t(expz);
  ez=J(3,N2,0);
  do J=1 to n2;
    ez[1,j]=meanzz[j];
    ez[2,j]=stdzz[j];
    ez[3,j]=expzz[j];
  end;
  names2={'Posit1','Posit2','Posit3','Posit4',
        'Posit5','Posit6','Posit7','Posit8'};
  create &data from ez[colname=names2];
  append from ez;
  tez=t(ez);
```

```
 names3={'Sim_mean','Var_est','Theo_mean'};
 create &data1 from tez[colname=names3];
 append from tez;
end;
/*compare variance*/
od1=J(6,1,0);od3=od1;od2=J(12,1,0);
d1=J(4,1,0);d2=d1;d3=d1;
do i=1 to 4;
 i2=i+4;
 d1[i]=s[i,i];
 d2[i]=s[i,i2];
 d3[i]=s[i2,i2];
end;
do i=1 to 3;
   i2=i+1;
   i3=i+5;i4=i+6;
   od1[i]=s[1,i2];
   od2[i]=s[1,i3];od2[i4]=s[5,i2];
   od3[i]=s[5,i3];

end;
do i=4 to 5;
   i2=i-1;
   i3=i+3;i4=i+6;
   od1[i]=s[2,i2];
   od2[i]=s[2,i3];od2[i4]=s[6,i2];
   od3[i]=s[6,i3];
end;
od1[6]=s[3,4];
od2[6]=s[3,8];od2[12]=s[7,4];
od3[6]=s[7,8];
Out=J(6,2,0);
out[1,1]=sum(d1)/4;
out[1,2]=1;
out[2,1]=sum(od1)/6;
out[2,2]=2;
out[3,1]=sum(d2)/4;
out[3,2]=3;
out[4,1]=sum(od2)/12;
out[4,2]=4;
out[5,1]=sum(d3)/4;
```

```
 out[5,2]=5;
 out[6,1]=sum(od3)/6;
 out[6,2]=6;
 print out;
  names3={"&m1","Position"};
  create &data2 from out[colname=names3];
  append from out;
 RUN;
QUIT;
%mend;
%SRS(4,100000,0.1,d3,d4,s2,0,444444,vm2);
%SRS(4,100000,0.1,d3,d4,s3,0,64646,vm3);
%SRS(4,100000,0.1,d3,d4,s4,0,64646,vm4);
%SRS(4,100000,0.1,d3,d4,s5,0,55535,vm5);
%SRS(4,100000,0.1,d3,d4,s6,0,575445,vm6);
%SRS(4,100000,0.1,d3,d4,s7,0,473663,vm7);
%SRS(4,100000,0.1,d3,d4,s8,0,37347,vm8);
%SRS(4,100000,0.1,d3,d4,s9,0,6563,vm9);
%SRS(4,100000,0.1,d3,d4,s10,0,6264,vm10);
%SRS(4,100000,0.1,d3,d4,s11,0,9797,vm11);

proc print data=d3 noobs;
title2 "Table D-7 The First Ten Simulated Z1 vectors";
Title3 "        and Expected Value of Z1";
Footnote "Source:th03lj07.sas";
run;
proc print data=d4 noobs;
run;
proc format;
value posif 0="Potential Observed Data"
          1="Missing Data";
data plot;
  set d4;
  bias=Theo_mean-Sim_mean;
  sigma=sqrt(Var_est/100000);
  relbias=bias/sigma;
  up=1;low=-1;
  Posit=_N_;
  dat=0;
  if posit >=5 then dat=1;
  if dat=1 then posit=posit-4;
```

```
  format dat posif.;
run;
proc print data=plot noobs; var dat posit Theo_mean Sim_mean sigma bias
relbias;
Title2 "Table D-8. Difference Between Theoretical Mean and Simulated Mean";
run;
goptions ftext=SWISS ctext=BLACK htext=1 cells;
 axis1 width=1 offset=(3 pct) label=(a=90 'Relative Bias' r=0);
 axis2 width=1 offset=(3 pct) label=('Position');
symbol1 i=join c=blue;
symbol2 v=dot i=none  c=black;
symbol3 v=star i=none  c=red;
proc gplot data=plot;
 plot up*posit=1 low*posit=1 relbias*posit=2 /overlay
                                        vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-2. Standardized diffrence between Simulated mean and
Theoretical mean";
 Footnote "Source:th03lj06.sas";
 run;
proc sort data=s2;
by position;
proc sort data=s3;
by position;
proc sort data=s4;
by position;
proc sort data=s5;
by position;
proc sort data=s6;
by position;
proc sort data=s7;
by position;
proc sort data=s8;
by position;
proc sort data=s9;
by position;
proc sort data=s10;
by position;
proc sort data=s11;
by position;
RUN;
```

```
data s;
 merge s2 s3 s4 s5 s6 s7 s8 s9 s10 s11;
 by position;
 drop position;
run;
proc iml;
use s;
read all into s;
close s;
n=ncol(s);
d=j(6,2,0);

do i=1 to 6;
 d[i,1]=s[i,+]/n;
 d[i,2]=(SSQ(s[i,])-(n*(d[i,1]**2)))/(n-1);
 end;
 print d;
  names ={"Sim_mean","Var_est"};
  create d5 from d[colname=names];
  append from d;
 run;quit;
 data d6;
   set d5;
   Part=_n_;
   if part=1 then Theo_mean=231.75;
   if part=2 then Theo_mean=-2.025;
   if part=3 then Theo_mean=-225;
   if part=4 then Theo_mean=--0.225;
   if part=5 then Theo_mean=225.75;
   if part=6 then Theo_mean=-0.025;
   std_est=sqrt(Var_est);
   diff=Sim_mean-Theo_mean;
 run;
 proc print data=d6 noobs;
 title2 "Table D-9: Compare the simulated Variance Matrix to Theoretical
Variance Matrix";
 var Part Theo_mean Sim_mean Var_est std_est diff;
 run;

proc print data=s;
run;
```

Appendix II. Simulation Program for results in section D.3.1

Program:

```
OPTIONS LINESIZE=100 PAGESIZE=80 NOCENTER NODATE NONUMBER NOFMTERR;
*********************************************************************
*;
*** PROJECT NAME: Simulation of comparison for Thesis                ;
***          PROGRAM NAME       DATE               PROGRAMMER    ;
TITLE1 "Source:th03lj05.sas  11/19/03 LJS "   ;
*** Description: Simulation macro to compare BLUE to ignore          ;
***            Missing values method   with known pi and variance   ;
***    INPUT: none                                                  ;
*********************************************************************
*;
LIBNAME thesis 'C:\My Documents\jingsonglu\thesis\';
* N=size of Population;
  *Nsim= Number simulations;
  *mu=Population mean;
  *PVAR=  Variance of population;
  *pi=proportion of non-response;
   *ns=sample size;
%Macro SRS(nsim,N,ns,pi,mu,pvar,data);
DATA POP;
 DO N= 1 TO &N;/*N = population size*/
  y=ROUND(&mu+sqrt(&pvar)*rannor(555456));
  OUTPUT;
 END;
 KEEP Y;
RUN;
proc iml;
*********************************;
*** Generate Population        ;
*********************************;
PI=&pi;N=&n;
N2=2*&N;NP1=&N+1;/*N2=2*N, NP1=N+1*/
y1=j(&n,1,0);
ys=j(&n,1,0);
s=J(&ns,1,0);/*column of sample contains
                  sample units of each simulation*/
n1=J(&nsim,1,0);    /*save the number of observed units for
```

```
                  each simulated sample*/
mub1=J(&nsim,1,0);/*save each BLU estimate */
L=J(&ns,1,1);
mui1=J(&nsim,1,0);vari1=mui1;
 USE POP;
    READ ALL INTO y;
 CLOSE POP;
   MEAN1=SUM(y)/N;
   VAR1=(SSQ(y)-(N*(MEAN1**2)))/(N-1);
   DO i= 1 TO N;   /*center variance*/
     y1[i]=y[i]*sqrt(&pvar/var1);
   END;
  DO i= 1 TO &N;   /*adjust population mean to designed value*/
     ys[i]=y1[i]-(sum(y1)/N-&mu);
  END;
  MEAN=SUM(ys)/N;/*POPULATION MEAN*/
  VAR=(SSQ(ys)-(N*(MEAN**2)))/(N-1);  /*population VARIANCE*/

Do sim=1 to &nsim;  /*start of simulation*/
********************;
***PERMUTATION   ***;
********************;
  check = j(&N,1, 0);/*vector to check 'un-replacement'*/
  comb=check;/*vector to save permutated units */
  do i=1 to &N;
   label:u=ranuni(26215);
   k=int(u*&N+1);
   kk=0;
   label2:kk=kk+1;
    if k=check[kk] then goto label;
    if kk=i then go to skip;
   goto label2;
   skip:check[i]=k;
   comb[i]=ys[k];/*randomly pick for each position without replacement*/
  end;/*end of permutation*/

   M=J(&ns,1,0);

             /*M is the matrix indicate of missing*/
  do i=1 to &Ns;
   M[i]=RANBIN(0,1,pi);
```

```
 end;

 do i= 1 to &ns;
   s[i]=comb[i]*(1-m[i]);
 end;
 n1[sim]=&ns-sum(m);
 /*BLUE*/
 MUB1[sim]=1/(&ns*(1-pi))*sum(s);
 /*ignore*/
 MUi1[sim]=sum(s)/n1[sim];

vari1[sim]=(ssq(s)-n1[sim]*mui1[sim]*t(mui1[sim]))/(n1[sim]-1)*(1-n1[si
m]/N)/n1[sim];
 end;/*end of simulation*/

*******************************;
***N, ns, pi, variance known***;
*******************************;

/*BLUE*/
meanb=sum(mub1)/&nsim;
MSEB=(SSQ(mub1)-(&nsim*(MEANB**2)))/(&Nsim-1);
msebt=1/(&ns*(1-pi))*(pi*meanb*meanb+((N-&ns*(1-pi)-pi)/N*Var));

/*ignore*/
meani=sum(mui1)/&nsim;
MSEi=(SSQ(mui1)-(&nsim*(MEANi**2)))/(&Nsim-1);
MSEit=sum(vari1)/&nsim+(meani-mean)**2;
vari=sum(vari1)/&nsim;

/*out put results*/
out=J(2,10,0);
out[1,1]=1;out[1,2]=n;out[1,3]=&ns;out[1,4]=mean;out[1,5]=var;out[1,6]=
pi;
out[1,7]=meanb;out[1,8]=mseb;out[1,9]=msebt;out[1,10]=vari;
out[2,1]=2;out[2,2]=n;out[2,3]=&ns;out[2,4]=mean;out[2,5]=var;out[2,6]=
pi;
out[2,7]=meani;out[2,8]=msei;out[2,9]=mseit;out[2,10]=vari;
/*create external file*/
names={'Method','Pop_N','Samp_N','Pop_mu','Pop_var','Pi',
     'Est_mu','Var_Est','V_E_thero','Theo_V'};
```

```
create &data from out[colname=names];
append from out;
RUN;
quit;
%mend;
%SRS(10000,50,15,0.1,50,4,d1);
%SRS(10000,50,15,0.2,50,4,d2);
%SRS(10000,50,15,0.4,50,4,d3);
%SRS(10000,50,15,0.6,50,4,d4);
%SRS(10000,50,30,0.1,50,4,d5);
%SRS(10000,50,30,0.2,50,4,d6);
%SRS(10000,50,30,0.4,50,4,d7);
%SRS(10000,50,30,0.6,50,4,d8);
proc format;
 value mdf 1='BLUE'
         2='Ignore';
data thesis.results;
 set d1 d2 d3 d4 d5 d6 d7 d8;
 format method mdf.;
 bias=Est_Mu-Pop_mu;
 sigma=sqrt(Var_est/10000);
 relbias=bias/sigma;
 run;
 data results;
  set thesis.results;
 run;
proc print data=results;
 run;
 proc sort data=results;by   samp_n pi;

 proc transpose data=results out=a prefix=relbias;
    by  samp_n pi;
    var relbias ;
    run;
data b;
    set a;
    up=1;low=-1;
goptions ftext=SWISS ctext=BLACK htext=1 cells;
 axis1 width=1 offset=(3 pct) label=(a=90 'Relative Bias' r=0);
 axis2 width=1 offset=(3 pct) label=('Probability of being missing (Pi)');
symbol1 i=join c=blue;
```

```
symbol2 v=dot i=none  c=black;
symbol3 v=star i=none  c=red;

proc gplot data=b;
 plot up*pi=1 low*pi=1 relbias1*pi=2 relbias2*pi=3/overlay
                                        vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-5. Relative Bias of two methods when N=50 and n=15";
 where samp_n=15;
 run;
 proc gplot data=b;
 plot up*pi=1 low*pi=1 relbias1*pi=2 relbias2*pi=3/overlay
                                        vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-6. Relative Bias of two methods when N=50 and n=30";
 where samp_n=30;
 run;
proc transpose data=results out=a prefix=bias;
    by   samp_n pi;
    var bias ;
    run;

 data b;
 set results(keep=method samp_n pi sigma);
 if method=1;
 run;

 data c;
 merge a b;
 by samp_n pi;drop method;
 sigma2=0-sigma;
 run;
 goptions ftext=SWISS ctext=BLACK htext=1 cells;
 axis1 width=1 offset=(3 pct) label=(a=90 'Absolute Bias' r=0);
 axis2 width=1 offset=(3 pct) label=('Probability of being missing (Pi)');
 symbol1 i=join c=blue;
 symbol2 v=dot i=none  c=black;
 symbol3 v=star i=none  c=red;

 proc gplot data=c;
 plot sigma*pi=1 sigma2*pi=1 bias1*pi=2 /overlay vaxis = axis1
```

```
                                        haxis = axis2;
 Title " Figure D-3. Absolute Bias of BLUE methods when N=50 and n=15";
 where samp_n=15;
 run;
 proc gplot data=c;
 plot sigma*pi=1 sigma2*pi=1 bias1*pi=2 /overlay vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-4. Absolute Bias of BLUE methods when N=50 and n=30";
 where samp_n=30;
 run;
 proc sort data=results;by method;

 Proc Print data=results;
 title ' Table d-10 Simulation results with known N, n, Pi and variance';
 run;

 proc sort data=results;by method pi;

 proc transpose data=results out=d prefix=mse;
    by  method pi;
    var var_est ;
    run;

goptions ftext=SWISS ctext=BLACK htext=1 cells;
 axis1 width=1 offset=(3 pct) label=(a=90 'Variance of Estimator' r=0);

 axis2 width=1 offset=(3 pct) label=('Probability of being missing (Pi)');
 symbol2 v=dot i=spline  c=black ;
 symbol3 v=star i=spline  c=red ;

 proc gplot data=d;
 plot mse1*pi=1  mse2*pi=2 /overlay vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-7. Plot Variance of Estimator for BLUE method ";
 where method=1;
 run;
 proc gplot data=d;
 plot mse1*pi=1  mse2*pi=2 /overlay vaxis = axis1
                                        haxis = axis2;
 Title " Figure D-8. Plot Variance of Estimator for method Ignoring Missing
Values ";
```

```
where method=2;

run;
```

Appendix III. Simulation Program for results in section D.3.2

Program:

```
OPTIONS LINESIZE=100 PAGESIZE=80 NOCENTER NODATE NONUMBER NOFMTERR;
*********************************************************************
*;
*** PROJECT NAME: Simulation of comparison for Thesis                ;
***          PROGRAM NAME       DATE               PROGRAMMER    ;
TITLE1 "Source:th03lj03.sas  11/15/03 LJS "   ;
*** Description: Simulation macro to compare BLUE to ignore          ;
***            Missing values method   with known pi and variance   ;
***    INPUT: none                                                  ;
*********************************************************************
*;
LIBNAME thesis 'C:\My Documents\jingsonglu\thesis\data';
* N=size of Population;
  *Nsim= Number simulations;
  *mu=Population mean;
  *PVAR=  Variance of population;
  *pi=proportion of non-response;
   *ns=sample size;
%Macro SRS(nsim,N,ns,pi,mu,pvar,data);
DATA POP;
 DO N= 1 TO &N;/*N = population size*/
  y=ROUND(&mu+sqrt(&pvar)*rannor(262448));
  OUTPUT;
 END;
 KEEP Y;
RUN;
proc iml;
*********************************;
*** Generate Population        ;
*********************************;
PI=&pi;N=&n;
N2=2*&N;NP1=&N+1;/*N2=2*N, NP1=N+1*/
y1=j(&n,1,0);
ys=j(&n,1,0);
s=J(&ns,1,0);/*column of sample contains
                  sample units of each simulation*/
n1=J(&nsim,1,0);    /*save the number of observed units for
```

```
                  each simulated sample*/
mub2=J(&nsim,1,0);/*save each BLU estimate */
L=J(&ns,1,1);
mui1=J(&nsim,1,0);vari1=mui1;pihat=n1;
 USE POP;
    READ ALL INTO y;
 CLOSE POP;
   MEAN1=SUM(y)/N;
   VAR1=(SSQ(y)-(N*(MEAN1**2)))/(N-1);
   DO i= 1 TO N;   /*center variance*/
     y1[i]=y[i]*sqrt(&pvar/var1);
   END;
  DO i= 1 TO &N;   /*adjust population mean to designed value*/
     ys[i]=y1[i]-(sum(y1)/N-&mu);
  END;
  MEAN=SUM(ys)/N;/*POPULATION MEAN*/
  VAR=(SSQ(ys)-(N*(MEAN**2)))/(N-1);  /*population VARIANCE*/

Do sim=1 to &nsim;  /*start of simulation*/
********************;
***PERMUTATION   ***;
********************;
  check = j(&N,1, 0);/*vector to check 'un-replacement'*/
  comb=check;/*vector to save permutated units */
  do i=1 to &N;
   label:u=ranuni(565644);
   k=int(u*&N+1);
   kk=0;
   label2:kk=kk+1;
    if k=check[kk] then goto label;
    if kk=i then go to skip;
   goto label2;
   skip:check[i]=k;
   comb[i]=ys[k];/*randomly pick for each position without replacement*/
  end;/*end of permutation*/

   M=J(&ns,1,0);

             /*M is the matrix indicate of missing*/
  do i=1 to &Ns;
   M[i]=RANBIN(0,1,pi);
```

```
 end;

do i= 1 to &ns;
  s[i]=comb[i]*(1-m[i]);
end;
n1[sim]=&ns-sum(m);pihat[sim]=1-n1[sim]/&ns;
/*BLUE*/
 MUB2[sim]=1/(&ns*(1-pihat[sim]))*sum(s);
/*ignore*/
MUi1[sim]=sum(s)/n1[sim];

vari1[sim]=(ssq(s)-n1[sim]*mui1[sim]*t(mui1[sim]))/(n1[sim]-1)*(1-n1[si
m]/N)/n1[sim];
 end;/*end of simulation*/
*******************************;
***N, ns, pi, variance known***;
*******************************;

/*ignore*/

meani=sum(mui1)/&nsim;
MSEi=(SSQ(mui1)-(&nsim*(MEANi**2)))/(&Nsim-1);
MSEit=sum(vari1)/&nsim+(meani-mean)**2;
vari=sum(vari1)/&nsim;


/*BLUE*/
pi1=sum(pihat)/&nsim;
meanb2=sum(mub2)/&nsim;
MSEB2=(SSQ(mub2)-(&nsim*(MEANB2**2)))/(&Nsim-1);
msebt=1/(&ns*(1-pi1))*(pi1*meanb2*meanb2+((N-&ns*(1-pi1)+pi)/N*Var));
/*out put results*/
/*out put results*/
out=J(2,10,0);
out[1,1]=1;out[1,2]=n;out[1,3]=&ns;out[1,4]=mean;out[1,5]=var;out[1,6]=
pi;
out[1,7]=meanb2;out[1,8]=mseb2;out[1,9]=msebt;out[1,10]=vari;
out[2,1]=2;out[2,2]=n;out[2,3]=&ns;out[2,4]=mean;out[2,5]=var;out[2,6]=
pi;
out[2,7]=meani;out[2,8]=msei;out[2,9]=mseit;out[2,10]=vari;
/*create external file*/
```

```
names={'Method','Pop_N','Samp_N','Pop_mu','Pop_var','Pi',
     'Est_mu','Var_Est','V_E_thero','Theo_V'};
create &data from out[colname=names];
append from out;
RUN;
quit;
%mend;
%SRS(10000,50,20,0.1,50,4,d1);
%SRS(10000,50,20,0.2,50,4,d2);
proc format;
 value mdf 1='BLUE'
         2='Ignore';
data thesis.results2;
 set d1 d2 ;
 format method mdf.;
 run;
 proc sort; by method;
 proc print;
 run;
```

BIBLIOGRAPHY

Brewer, K.R.W., Hanif, M., and Tam, S.M. (1988). How nearly can model-based prediction and design-based estimation be reconciled? *J. Am. Stat. Assoc.* 83: 128-132.

Brewer, K.R.W. (1999). Design-based or prediction-based inference? Stratified random vs stratified balanced sampling. *Int. Stat. Rev.*, 67:35-47.

Brewer, K.R.W. (2002). *Combined Survey Sampling Inference*. Oxford University Press, New York.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling.* Wiley, New York.

Cochan, W.G. *Sampling Techniques*. Wiley, New York.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood From Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society*, Ser. B, 39: 1-22.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *American Statistical Association Journal*, 57: 369-375.

Haziza D. and Rao J.N.K. (2003). Inference for Population Means Under Unweighted Imputation for Missing Survey Data. *Survey Methodology*, 29(1).

Hedayat, A. and Sinha, B.K. (1991). *Design and Inference in Finite Population Sampling*. Wiley, New York.

Lehtonen R.J. and Pahkinen E.J. (1994) Practical Methods for design and analysis of complex surveys. Wiley, Chichester.

Little, R.J.A. (1982). Models for non-response in sample survey. *American Statistical Association Journal* **77**:237-250

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley.

Mukhopadhyay, P. (2001). *Topics in Survey Sampling*, Springer Lecture Notes in Statistics 153, New York.

Rao, J.N.K. and Bellhouse, D.R. (1978). Estimation of finite population mean under generalized random permutation model. *Journal of Statistical Planning and inference*, 2: 125-141.

Rao, J.N.K. (1997). Developments in Sample Survey Theory: An Appraisal. *Can. J. Stat.* 25:1-21.

Rao, J.N.K. (1999a). Some Current Trends in Sample Survey Theory and Methods. *Sankhyã B*, 61:1-25.

Rubin, D.B. (1976). Inference and missing data. *Biometrika,* 63 (3):581-591.

Rubin, D.B. and Schenker, N. (1986), Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse, *Journal of the American Statistical Association,* 81: 366-374.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science***,** 6(1): 15-51.

Särndal, C-E, Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*.Springer-Verlag, New York.

Scott, A. and T. M. F. Smith (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association,* 64(327): 830-840.

Stanek, E. J. III., A. Well and I. Ockene (1999). Why not routinely use best linear unbiased predictors (BLUPs) as estimates of cholesterol, percent fat from Kcal and physical activity? *Statistics in Medicine* 18: 2943-2959.

Stanek, E.J.III. (2002). A unified approach to estimation and prediction under simple random sampling. *unpublished paper*. (http://www-unix.oit.umass.edu/ ~cluster/ ed/outline/yr2002/C02ed04v1.pdf)

Stanek, E.J.III (2001). Optimal estimators for the finite population parameters in a Single stage sampling. *unpublished paper*. (http://www-unix.oit.umass.edu/~cluster/ed/dialogcom/c01viv05.PDF).

Smith, T.M.F. (1976). The Foundations of Survey Sampling: a Review. *Journal of the Royal Statistical Society A*, 139:183-204.

Thompson, M.E. (1997). *Theory of Sample Surveys*. Chapman and Hall, London.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference:A Prediction Approach.* John Wiley, New York.