TECHNICAL REPORT NO. 937

October 1994

VARIANCE ESTIMATION FOR STRATIFIED MULTISTAGE SAMPLES WITH SIMPLE HOT DECK IMPUTATION

by

Yinzhong Chen & Jun Shao
Department of Statistics
University of Wisconsin-Madison
VARIANCE ESTIMATION FOR STRATIFIED MULTISTAGE SAMPLES
WITH SIMPLE HOT DECK IMPUTATION

YINZHONG CHEN AND JUN SHAO
Department of Statistics
University of Wisconsin-Madison
Madison, WI 53706

Abstract

Hot deck imputation for nonrespondense is often used in surveys. It is a common practice
to treat the imputed values as if they are true values, and compute parameter estimates and
their variance estimates using standard formulas. The variance estimates, however, have
seriously negative biases when the nonrespondense rate is appreciable. Rubin (1978) pro-
posed the multiple imputation method which utilizes the variabilities among independently
imputed data sets. However, the validity of Rubin's method requires that the imputation
procedure be proper. A proper imputation procedure may be difficult or even impossible to
find when the sampling design is complex. Practitioners often use an improper imputation
method for its simplicity and efficiency, but misuse Rubin's method for variance estimation.
In this paper we derive an asymptotically consistent variance estimator in the situation where
a stratified-multistage sampling design is used to collect survey data; the simplest hot deck
multiple imputation (which is simple, efficient but improper) is applied to impute missing
data; and we do not know which units are nonrespondents.

Key words and phrases. Nonrespondense, Multiple imputation, Variance estimation, Multi-
stage sampling.

\footnote{The research was partially supported by the Natural Sciences and Engineering Research Council of Canada.}
1. Introduction

Most survey data are incomplete due to nonresponse. Hot deck imputation is commonly used for item nonresponse because of various practical (not necessarily statistical) reasons (Kalton, 1981; Sedransk, 1985). It is a common practice to treat the imputed values as if they are true values, and then compute point estimates of population parameters and their variance estimates using standard formulas. If imputation is suitably carried out, point estimates of population parameters based on imputed data are asymptotically valid; their variance estimates, however, have seriously negative biases when the proportion of nonrespondents is appreciable, because standard formulas for variance estimation do not account for the inflation in variance due to missing data and/or imputation. Consequently, inference based on these variance estimates can be very misleading.

There exist two types of methods which provide better variance estimators:

(1) Rubin (1978) and Rubin and Schenker (1986) proposed the multiple imputation method which requires to perform $T \geq 2$ independent imputations and computes variance estimators using the variabilities among the $T$ imputed data sets (more details will be provided later).

(2) There are methods based on some adjustments which account for the inflation in the variance due to nonresponse and/or imputation, e.g., the adjusted jackknife method (Rao and Shao, 1992), the adjusted linearization method (Rao, 1993), and the bootstrap method (Shao and Sitter, 1993). These methods provide asymptotically consistent variance estimators and work for both single and multiple imputation, but require identification flags to locate nonrespondents.

In this paper we focus on the situation where we do not know which sampled units are nonrespondents (which may occur when information concerning whether a unit is a nonrespondent is confidential) so that the second type methods are not applicable. Finding a good variance estimator in this situation is much more difficult than that in the case where we know which units are nonrespondents, since many statistics whose computations require knowing which units are nonrespondents cannot be used.

To illustrate Rubin’s multiple imputation method, consider the estimation of the population mean $\bar{Y}$. Let $\bar{y}^{\text{ref}}$ be the point estimate of $\bar{Y}$ and $\text{v}^{\text{ref}}$ be its variance estimate, both calculated based on standard formulas and the $t$-th imputed data set, $t = 1, \ldots, T$, $T \geq 2$. 
Rubin proposed

\[ \bar{y}^* = \frac{1}{T} \sum_{i=1}^{T} y^*_{it} \]

as the final estimate of \( \bar{Y} \) with variance estimate

\[ v_R = \hat{W} + \frac{T+1}{T} \hat{B}, \] \hspace{1cm} (1.1)

where

\[ \hat{W} = \frac{1}{T} \sum_{i=1}^{T} v^*_{it} \] \hspace{1cm} (1.2)

measures the within-imputation variability and

\[ \hat{B} = \frac{1}{T-1} \sum_{i=1}^{T} (y^*_{it} - \bar{y}^*)^2 \] \hspace{1cm} (1.3)

measures the between-imputation variability. The calculation of \( \bar{y}^* \), \( \hat{W} \) and \( \hat{B} \) requires to maintain \( T \) imputed data sets, but does not require to know which units are nonrespondents.

Although \( v_R \) is derived based on a Bayesian perspective, Rubin (1986) and Rubin and Schenker (1986) showed that if the multiple imputation procedure is proper (i.e., it satisfies conditions 1-3 in Rubin (1986, pp.118-119)), then for any fixed \( T \geq 2 \), \( (\bar{y}^* - \bar{Y})/\sqrt{v_R} \) is approximately distributed as a t-variable with

\[ (T-1) \left[ 1 + \left( \frac{T}{T+1} \right) \left( \frac{\hat{W}}{\hat{B}} \right) \right]^2 \]

degrees of freedom. To see what is required for a proper multiple imputation, consider the simplest case where \( n \) units are sampled from the population by simple random sampling and the units have the same response probability (uniform response mechanism). Rubin (1986) showed that the following imputation scheme, called the approximate Bayesian bootstrap (ABB) hot deck imputation, is proper: for each \( t = 1, ..., T \), create \( n \) bootstrap values by drawing \( n \) values at random with replacement from the respondents, and then impute the nonrespondents by drawing values at random with replacement from the bootstrap values. Note that for each \( t \), one has to draw two sets of random variates to impute missing values.

Nowadays, simple random sampling is rarely used because of both practical and theoretical considerations. Stratification and multistage sampling are commonly used (Kish and Frankel, 1974). Suppose that the population has \( L \) strata and \( n_h \) units are sampled from the \( h \)-th stratum, \( h = 1, ..., L \), and that the units still have the same response probability. If all \( n_h \) are large and \( L \) is small, then a proper multiple imputation procedure can be obtain
by simply performing the ABB hot deck within each stratum. However, in the common situation where \( L \) is large and all \( n_h \) are small (Krewski and Rao, 1981), it is hard (may be impossible) to find a proper multiple imputation procedure unless a strong model assumption is imposed to the population. When some \( n_h \) are small (\( n_h = 2 \), for example), applying the ABB hot deck within each stratum does not lead to a proper multiple imputation and is impractical since there may not be enough respondents in some small-size strata or clusters.

When the sampling design is complex, practitioners often use an improper but simple imputation method. Note that the use of an improper imputation method does not necessarily result in asymptotically inconsistent point estimators of population parameters, just like the use of an improper prior does not necessarily result in a bad Bayes rule in Bayesian analysis. In fact, \( \bar{y}^* \) based on an improper multiple imputation is often consistent and may be more efficient than that based on a proper multiple imputation. For example, in the case of simple random sampling, a simple and improper imputation procedure, called the **simple hot deck imputation**, is to impute the missing values by drawing a simple random sample from the respondents; \( \bar{y}^* \) based on this simple hot deck imputation has a smaller mean squared error than that based on the ABB hot deck (which is proper), simply because in the ABB hot deck procedure one creates some extra variability by using the bootstrap values as donors (the set of data from which we draw imputed values). The simple hot deck can be easily extended to the case of stratified sampling with large \( L \) and small \( n_h \): one simply imputes the missing values by taking a sample from all the respondents. Since imputation is cutting across strata, there are enough donors even if some small-size strata have no respondents.

However, practitioners often misuse formulas (1.1)-(1.3) to estimate the variances of point estimates based on improper multiple imputation. Under improper multiple imputation, the variance estimator \( v_R \) is usually not asymptotically valid and \( (\bar{y}^* - \bar{Y})/\sqrt{v_R} \) is not approximately distributed as a t-variable.

In short, the use of the variance estimator \( v_R \) defined in (1.1) does not require to know which units are nonrespondents, but the validity of \( v_R \) depends on whether the imputation procedure is proper; a proper multiple imputation method may not exist or is hard to find when the sampling design is complex; and practitioners often use an improper imputation method (because of its simplicity and efficiency or because of the nonexistence of a proper imputation procedure), but misuse the variance estimator \( v_R \) which is valid only for proper multiple imputation. It is then desired to derive asymptotically valid variance estimators for
point estimators (such as \( \hat{y}^* \)) based on simple but improper multiple imputation methods.

The purpose of this paper is to derive asymptotically valid variance estimators for point estimators based on the simple hot deck imputation under stratified multistage sampling with large \( L \) and small \( n_h \), assuming that we do not know which units are nonrespondents.

2. Variance Estimation under Simple Hot Deck Multiple Imputation

Consider a population with \( L \) strata and \( N_h \) first-stage units in the \( h \)-th stratum, \( h = 1, \ldots, L \). Suppose that \( n_h \geq 2 \) first-stage units are sampled from stratum \( h \) with replacement, independently across the strata. Within the \( h \)-th stratum, the \((h, i)\)-th first-stage unit is selected with probability \( p_{hi} > 0 \), \( i = 1, \ldots, N_h \). The special case of \( p_{hi} = \frac{1}{N_h} \) corresponds to simple random sampling within stratum \( h \). If the first-stage units are clusters, then a second-stage sample, a third-stage sample, \ldots, may be selected within each cluster, and the samples are selected independently across the clusters. We do not specify the number of stages and the sampling methods used after the first-stage sampling. For simplicity, we shall index the ultimate units in a first-stage cluster by using a single index, i.e., unit \((h, i, j)\) is the \( j \)-th ultimate unit in the \( i \)-th first-stage cluster of stratum \( h \), \( i = 1, \ldots, n_h, h = 1, \ldots, L \).

This sampling design is called the stratified multistage sampling plan.

We focus on the common case where \( L \) is large and all \( n_h \) are small (i.e., \( n_h \) are bounded by a fixed integer). Although we assume with replacement first-stage sampling within each stratum, our discussions and results are still valid when the first-stage sampling is without replacement, because all \( n_h \) are small and all \( N_h \) are usually large so that the first-stage sampling fractions \( \frac{n_h}{N_h} \) are negligible.

2.1. Univariate case with uniform response

We start with the simplest case. Suppose that \( r \) sampled ultimate units respond on an item \( y \), \( m \) sampled ultimate units do not respond, and \( r + m = n \) is the total number of sampled ultimate units. Let

\[
A_r = \{(h, i, j) : \text{the } (h, i, j)\text{-th unit responds}\},
\]

\[
A_m = \{(h, i, j) : \text{the } (h, i, j)\text{-th unit does not respond}\},
\]

and \( A_n = A_r \cup A_m \). We assume that the units respond with the same probability \( p \) (uniform response mechanism). Extensions to unequal response probability cases will be discussed in Section 2.3.
Let $w_{hij}$ be the survey weight associated with the $(h, i, j)$-th sampled ultimate unit. The survey weights are constructed so that when there is no nonrespondent ($A_m = \emptyset$),

$$\hat{Y} = \sum_{(h,i,j) \in A_n} w_{hij}y_{hij}$$

is an unbiased estimator of the population total $Y$ on item $y$, where $y_{hij}$ is the observed value of the $(h, i, j)$-th sampled ultimate units on item $y$. Since in multistage sampling the total number of ultimate units $M$ is often unknown, the population mean $\bar{Y} = Y/M$ is estimated by a ratio estimate

$$\bar{y} = \hat{Y} / \hat{M},$$

where

$$\hat{M} = \sum_{(h,i,j) \in A_n} w_{hij}.$$  

(2.1)

When there are nonrespondents, $\bar{y}$ cannot be calculated. For various practical reasons (Kalton, 1981; Rubin, 1986), one imputes the missing value by $y^*_{hij}$ for unit $(h, i, j) \in A_m$, and then estimate $\hat{Y}$ by

$$\hat{y}^* = \sum_{(h,i,j) \in A_n} w_{hij}y^*_{hij} / \hat{M},$$

(2.2)

where

$$y^*_{hij} = y_{hij} \quad \text{if} \quad (h, i, j) \in A_r.$$

A simple imputation method, called the simple hot deck imputation, obtains $\{y^*_{hij} : (h, i, j) \in A_m\}$ by taking an i.i.d. sample from the respondents, where each $y_{hij}$, $(h, i, j) \in A_r$, is selected with probability $\pi_{hij} = w_{hij} / \sum_{(h,i,j) \in A_r} w_{hij}$. Under this simple hot deck imputation, $\hat{y}^*$ in (2.2) is asymptotically unbiased, consistent, and normal (Rao and Shao, 1992). Since imputation is carried out by cutting across strata and clusters, it can be done even when some strata or clusters have no respondent.

A variance estimator for $\hat{y}^*$ calculated based on the standard formula (e.g., Cochran, 1977; Krewski and Rao, 1981) is

$$v^* = \sum_{h=1}^{L} \frac{\nu_h}{n_h - 1} \sum_{i=1}^{n_h} (\zeta^*_{hi} - \bar{\zeta}^*)^2,$$

(2.3)

where

$$\zeta^*_{hi} = \frac{1}{M} \sum_{j=1}^{n_{hi}} w_{hij}(y^*_{hij} - \bar{y}^*),$$

and $\bar{\zeta}^* = \sum_{h=1}^{L} \frac{\nu_h}{n_h} \bar{\zeta}^*_{hi}$.
\( n_{hi} \) is the number of sampled ultimate units in the \( i \)-th cluster of stratum \( h \), and

\[
\bar{\zeta}_h^* = \frac{1}{n_h} \sum_{i=1}^{n_h} \zeta_{hi}^*.
\]

Note that the calculation of \( \bar{y}^* \) and \( v^* \) according to (2.2)-(2.3) does not require to know which sampled units are nonrespondents. However, \( v^* \) has a seriously negative bias if the response rate \( p \) is low (Rubin, 1986; Rao and Shao, 1992).

If we know which units are nonrespondents, i.e., we know the set \( A_r \) and \( A_m \), then the variance estimation becomes simple. A straightforward calculation leads to

\[
\text{Var}(\bar{y}_r) = \text{Var}(\bar{y}_r) + E \left( \frac{\sum_{(h,i,j) \in A_m} w_{hij}}{\sum_{(h,i,j) \in A_n} w_{hij}} \right)^2 \left( \frac{\sum_{(h,i,j) \in A_r} w_{hij}(y_{hij} - \bar{y}_r)^2}{\sum_{(h,i,j) \in A_r} w_{hij}} \right),
\]

where

\[
\bar{y}_r = \frac{\sum_{(h,i,j) \in A_r} w_{hij} y_{hij}}{\sum_{(h,i,j) \in A_r} w_{hij}}.
\]

Then \( \text{Var}(\bar{y}^*) \) can be consistently estimated by

\[
\sum_{i=1}^{L} \frac{n_h}{n_h - 1} \frac{1}{n_h} \sum_{i=1}^{n_h} (\zeta_{hi}^* - \bar{\zeta}_h^*)^2 + \left( \frac{\sum_{(h,i,j) \in A_m} w_{hij}}{\sum_{(h,i,j) \in A_n} w_{hij}} \right)^2 \left( \frac{\sum_{(h,i,j) \in A_r} w_{hij}(y_{hij} - \bar{y}_r)^2}{\sum_{(h,i,j) \in A_r} w_{hij}} \right),
\]

where

\[
\zeta_{hi}^* = \sum_{j: (h,i,j) \in A_r} w_{hij}(y_{hij} - \bar{y}_r) / \sum_{(h,i,j) \in A_r} w_{hij}
\]

and

\[
\bar{\zeta}_h^* = \frac{1}{n_h} \sum_{i=1}^{n_h} \zeta_{hi}^*.
\]

This estimator cannot be calculated without knowing which units are nonrespondents.

Without knowing which units are nonrespondents, it is generally unknown whether there exists a consistent estimator of \( \text{Var}(\bar{y}^*) \) based on a single imputed data set (in some special cases we can find consistent variance estimators based on a single imputed data set, e.g., when there is no strata and the sampling is single stage).

We now consider multiple imputation, which can be carried out by repeating independently the simple hot deck imputation procedure \( T \geq 2 \) times. Let \( y_{hij}^t, \bar{y}^t, \zeta_{hi}^t, \bar{\zeta}_h^t \) and \( v^t \).
be defined the same as \( y_{hij}, \tilde{y}, \zeta_{hi}, \tilde{\zeta}_h \) and \( v^* \), respectively, but based on the \( t \)-th imputed data set, \( t = 1, \ldots, T \). Then the estimator of \( \bar{Y} \) based on multiple imputation is

\[
\bar{y}^* = \frac{1}{T} \sum_{t=1}^{T} \bar{y}^{*t}.
\]  

(2.4)

This estimator is asymptotically valid for any fixed \( T \). Rubin's variance estimator \( v_R \) can be computed by directly using formulas (1.1)-(1.3), with \( \bar{y}^{*t} \) and \( v^{*t} \) calculated based on (2.2)-(2.3) and the \( t \)-th imputed data set. Note that computing \( \bar{y}^* \) and \( v_R \) does not require to know which units are nonrespondents, but it requires to maintain multiple imputed data sets. Because of the extra work involved in multiple imputation, \( T \) is usually not large in practice.

As we pointed out in Section 1, however, \( v_R \) is not correct since the simple hot deck imputation is not a proper multiple imputation procedure according to Rubin's (1986) definition. In fact, under the asymptotic framework given in Section 3, it can be shown that

\[
\frac{v_R}{\text{Var}(\bar{y}^*)} \xrightarrow{p} \frac{p\lambda + (1 + \frac{T+1}{T}X^2)\gamma}{\frac{\lambda}{p} + \frac{\gamma}{T}}
\]

(2.5)

as \( L \to \infty \), where \( \xrightarrow{p} \) denotes convergence in probability, \((T - 1)X^2\) is distributed as chi-square with \( T - 1 \) degrees of freedom, and \( \lambda \) and \( \gamma \) are unknown and unequal. For example, in the case of stratified one-stage sampling with simple random sampling within each stratum,

\[
\lambda = \lim \frac{L}{N^2} \left[ \sum_{h=1}^{L} \frac{N_h^2 \sigma^2_h}{n_h} + (1 - p) \sum_{h=1}^{L} \frac{N_h^2}{n_h} (\bar{Y}_h - \bar{Y})^2 \right]
\]

(2.6)

and

\[
\gamma = \lim \frac{(1 - p)L}{N} \sum_{h=1}^{L} \frac{N_h^2}{n_h} \left[ \sum_{h=1}^{L} N_h \sigma^2_h + \sum_{h=1}^{L} N_h (\bar{Y}_h - \bar{Y})^2 \right],
\]

(2.7)

where \( N = \sum_h N_h \), \( \bar{Y}_h = E(y_{hij}) \) and \( \sigma^2_h = \text{Var}(y_{hij}) \). As a consequence of result (2.5), \((\bar{y}^* - \bar{Y})/\sqrt{v_R}\) is not approximately distributed as a \( t \)-variable (or a scaled \( t \)-variable).

The variance estimator \( v_R \) essentially depends on two statistics, \( \hat{W} \) and \( \hat{B} \), whose computations do not require to know which units are nonrespondents. Note that \( \hat{B} \) in (1.3) is a measure of the variation among \( \bar{y}^{*t} \)'s, but it ignores the existence of \( L \) strata. The right-hand side of (2.5) does not depend on any unknown quantity (and consequently \((\bar{y}^* - \bar{Y})/\sqrt{v_R} \) is approximately distributed as a scaled \( t \)-variable) if all the strata means and variances are equal (see, for example, (2.6) and (2.7)). Thus, we have to take into account the effect due to the existence of strata. This leads us to consider the following statistic which estimates
the between-imputation variation within stratum \( h \):

\[
\hat{B}_h = \frac{1}{T - 1} \sum_{i=1}^{T} (\bar{y}_h^{*i} - \bar{y}_h^{*})^2,
\]

(2.8)

for \( h = 1, ..., L \), where

\[
\bar{y}_h^{*i} = \frac{1}{M} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} w_{hij} y_{hij}
\]

and

\[
\bar{y}_h^{*} = \frac{1}{T} \sum_{i=1}^{T} \bar{y}_h^{*i}.
\]

Note that \( \hat{B}_h \) is essentially Rubin’s \( \hat{B} \) (estimate of between-imputation variability) based on the imputed data within the \( h \)-th stratum (although the imputation is preformed cutting across strata).

In view of (2.5), we also need a consistent estimator \( \hat{p} \) of the response probability \( p \). For example, we may use \( \hat{p} = r/n \), where \( n \) is the total number of sampled ultimate units and \( r \) is the total number of respondents in the sample. This assumes that we know \( r \), which is not unusual in practice.

After examining asymptotic behaviors of \( \hat{W} \) and \( \hat{B}_h, \ h = 1, ..., L \), and substituting unknown quantities in \( \text{Var}(\bar{y}^{*}) \), we obtain the following variance estimator:

\[
v_s = \frac{\hat{W}}{\hat{p}^2} + \left( \frac{1}{T} - \frac{1}{\hat{p}^2} \right) \hat{B},
\]

(2.9)

where

\[
\hat{B} = \sum_{h=1}^{L} \hat{B}_h.
\]

(2.10)

This estimator can be computed without knowing which units are nonrespondents. It will be shown in Section 3 that this estimator is asymptotically consistent, i.e.,

\[
\frac{v_s}{\text{Var}(\bar{y}^{*})} \rightarrow_p 1
\]

for any fixed \( T \geq 2 \).

2.2. Multivariate case with uniform response

Survey data are usually multivariate, i.e., each ultimate unit has a vector of responses. We focus on the two-dimensional case: \((y_{hij}, z_{hij}) \) is the response of the \((h, i, j)-th \) ultimate unit if it responds to both item \( y \) and item \( z \). Extensions of our results to three or more dimensional cases are straightforward.
If there is no nonrespondent, then the population mean vector \((\bar{Y}, \bar{Z})\) is estimated by \((\bar{y}, \bar{z})\), where \(\bar{z}\) is calculated according to (2.1) with \(y_{hij}\) replaced by \(z_{hij}\). Note that the same survey weight \(w_{hij}\) is applied to both \(y\) and \(z\) variables.

In practice a sampled ultimate unit cooperates in the survey but often fails to provide answers to some (not all) of the questions. This is referred to as item nonresponse. Define

\[
A_{mm} = \{(h, i, j) \in A_n : \text{both } y_{hij} \text{ and } z_{hij} \text{ are missing}\},
\]

\[
A_{rr} = \{(h, i, j) \in A_n : \text{both } y_{hij} \text{ and } z_{hij} \text{ are observed}\},
\]

\[
A_{rm} = \{(h, i, j) \in A_n : y_{hij} \text{ is observed but } z_{hij} \text{ is missing}\},
\]

and

\[
A_{mr} = \{(h, i, j) \in A_n : y_{hij} \text{ is missing but } z_{hij} \text{ is observed}\}.
\]

Then all these four subsets of \(A_n\) may be nonempty and have appreciable sizes.

If the imputation is carried out jointly, i.e., for any unit in \(A_{rm} \cup A_{mr} \cup A_{mm}\), its \(y\) and \(z\) values are imputed by using \((y_{hij}, z_{hij})\), \((h, i, j) \in A_{rr}\), irrespective of whether both \(y\) and \(z\) values are missing or only one of these values is missing, then the extension of the results in Section 2.1 to the multivariate case is trivial: We only need to view \(y_{hij}\) as a vector and change the squares to vector products in appropriate places. However, using joint imputation we throw away the data in \(A_{rm} \cup A_{mr}\), which is not desirable. Furthermore, \(A_{rr}\) may be of a small size. Because of these considerations, in practice imputation is often carried out marginally, i.e., missing \(y\) values are imputed using the respondents \(y_{hij}\) with \((h, i, j) \in A_{rr} \cup A_{rm}\), missing \(z\) values are imputed using \(z_{hij}\) with \((h, i, j) \in A_{rr} \cup A_{mr}\), and the \(y\) and \(z\) values are imputed independently.

Marginal imputation is simple and does not require any model assumption (between \(y\) and \(z\) variables). A limitation of the marginal imputation is that it does not preserve the relation between the \(y\) and \(z\) variables so that we cannot estimate any parameter which measures how \(y\) and \(z\) are related (e.g., the correlation coefficient between the two variables). However, in this paper we focus on the situation where the parameter of interest is \(\theta = g(\bar{Y}, \bar{Z})\), a function of the population mean vector (e.g., \(\theta = \bar{Y} / \bar{Z}\)). In such cases marginal imputation provides asymptotically valid estimator of \(\theta\).

We still assume uniform response, but allow the probabilities of responses to item \(y\) and item \(z\) to be different. Let \(p_y\) be the response probability to \(y\) variable and \(p_z\) be the response probability to \(z\) variable.
For variable \( z \), let \( z_{hij}, \bar{z}, \bar{z}^{*}, \bar{z}^{*}, v_{z}^{*} \), and \( \hat{B}_{zh} \) be analogs to \( y_{hij}, \bar{y}, \bar{y}^{*}, \bar{y}^{*}, v_{y}, \) and \( \hat{B}_{yh} \), respectively, which are defined previously for variable \( y \) (see formulas (2.2), (2.3), (2.4) and (2.8)). A naive estimator of the variance-covariance matrix of \((\bar{y}^{*}, \bar{z}^{*})\), calculated based on the standard formula, is

\[
v^{*} = \begin{pmatrix}
  v_{y}^{*} & v_{yz}^{*} \\
  v_{yz}^{*} & v_{z}^{*}
\end{pmatrix},
\]

where

\[
v_{yz}^{*} = \frac{L}{n_{h}} \frac{n_{h}}{n_{h} - 1} \sum_{i=1}^{n_{h}} (\zeta_{hi}^{y*} - \bar{z}_{h}^{y*}) (\zeta_{hi}^{z*} - \bar{z}_{h}^{z*}),
\]

\[
\zeta_{hi}^{y*} = \frac{1}{M} \sum_{j=1}^{n_{hi}} w_{hij} (y_{hij}^{*} - \bar{y}^{*}), \quad \bar{z}_{h}^{y*} = \frac{1}{n_{h}} \sum_{i=1}^{n_{h}} \zeta_{hi}^{y*},
\]

and

\[
\zeta_{hi}^{z*} = \frac{1}{M} \sum_{j=1}^{n_{hi}} w_{hij} (z_{hij}^{*} - \bar{z}^{*}), \quad \bar{z}_{h}^{z*} = \frac{1}{n_{h}} \sum_{i=1}^{n_{h}} \zeta_{hi}^{z*}.
\]

Similar to the estimator in (2.3), this estimator is inconsistent when \( p_{y} < 1 \) or \( p_{z} < 1 \).

Suppose now that \( T \geq 2 \) independent imputed data sets are available. Let \( v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t} \) and \( v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t}, v_{y}^{*t} \), respectively, but based on the \( t \)-th imputed data set, and let \( \hat{p}_{y} \) and \( \hat{p}_{z} \) be estimators of \( p_{y} \) and \( p_{z} \), respectively, constructed using the same method for constructing \( \hat{p} \) in (2.9). Define

\[
\hat{B}_{h} = \begin{pmatrix}
  \hat{B}_{yh} & 0 \\
  0 & \hat{B}_{zh}
\end{pmatrix},
\]

\[
v^{*t} = \begin{pmatrix}
  v_{y}^{*t} & v_{yz}^{*t} \\
  v_{yz}^{*t} & v_{z}^{*t}
\end{pmatrix},
\]

and

\[
\hat{P} = \begin{pmatrix}
  \hat{p}_{y} & 0 \\
  0 & \hat{p}_{z}
\end{pmatrix}.
\]

A multivariate analog of \( v_{S} \) in (2.9) is

\[
v_{S} = \hat{P}^{-1} \hat{W} \hat{P}^{-1} + \frac{\hat{B}}{T} - \hat{p}^{-1} \hat{B} \hat{p}^{-1},
\]

(2.11)

where \( \hat{W} \) and \( \hat{B} \) are still defined by (1.2) and (2.10), i.e.,

\[
\hat{W} = \frac{1}{T} \sum_{t=1}^{T} v_{y}^{*t} \quad \text{and} \quad \hat{B} = \sum_{h=1}^{L} \hat{B}_{h}.
\]

The consistency of \( v_{S} \) in (2.11) will be established in Section 3.
If we estimate \( \theta = g(\bar{Y}, \bar{Z}) \) by \( \hat{\theta}^* = g(\bar{y}^*, \bar{z}^*) \), then a linearization variance estimator for \( \hat{\theta}^* \) is

\[
[\nabla g(\bar{y}^*, \bar{z}^*)]^t v_S \nabla g(\bar{y}^*, \bar{z}^*) ,
\]

where \( \nabla g \) is the vector of partial derivatives of \( g \).

2.3. Multiple imputation classes

The assumption of uniform response probability is often impractical. Our results can be extended to the following more realistic case (see, e.g., Schenker and Welsh, 1988; Rao and Shao, 1992). The population can be divided into \( K \) classes, according to the value of a variable \( x \) observed for all the sampled units. Within each class, the sampled units response with equal probability. These classes are called imputation classes and imputation is then performed independently within each imputation class.

Let \( v^k_S \) be the variance estimator calculated according to (2.9) or (2.11) but based on the data in the \( k \)-th imputation class, \( k = 1, ..., K \). Then an asymptotically consistent variance estimator is

\[
v_S = \sum_{k=1}^{K} v^k_S .
\]

Extensions of our method to more general non-uniform response cases rely on whether we can obtain an explicit asymptotic formula for \( \text{Var}(\bar{y}^* ) \) and whether we can use statistics such as \( \hat{W} \) and \( \hat{B}_h \) to provide consistent estimators of the unknown quantities in \( \text{Var}(\bar{y}^* ) \). These extensions have to be handled case by case and will not be further discussed here.

3. Properties of the Proposed Variance Estimators

3.1. Asymptotic results

A framework for the development of asymptotic theory is provided by the concept of a sequence of populations \( \{ P_\nu, \nu = 1, 2, ... \} \), where each \( P_\nu \) contains \( L_\nu \) strata. The population under consideration is a member of this sequence of populations. Note that \( L, N_h, w_{hij}, \) and \( y_{hij} \) etc. described in the previous sections depend on \( \nu \), but \( \nu \) is omitted for simplicity. We assume that the \( n_h \) are bounded and that \( L \to \infty \) as \( \nu \to \infty \). All limiting process will be understood to be as \( \nu \to \infty \).

Since it is convenient to work with \( \bar{w}_{hij} = w_{hij}/M \), we substitute \( \bar{w}_{hij} \) for \( w_{hij} \) in all places. All the estimators under consideration are not affected by this change.
We need some more notations. Let \( n_T \) denote the number of first-stage sampled clusters, i.e., \( n_T = \sum_{h=1}^{L} n_h \). Let \( a^y_{hi,j} \) be the response indicator for item \( y \), i.e.,

\[
    a^y_{hi,j} = \begin{cases} 
    1 & \text{if } y_{hi,j} \text{ is observed} \\
    0 & \text{if } y_{hi,j} \text{ is missing}
    \end{cases}
\]

To establish the consistency of \( v_S \) in (2.9), (2.11) or (2.12), we need the following regularity conditions.

1. \( n_T^{1+\delta} \sum_h \sum_i E|r_{hi} - E(r_{hi})|^{2+\delta} = O(1) \) for some fixed \( \delta > 0 \), where \( r_{hi} = \sum_j \bar{w}_{hi,j} a^y_{hi,j} y_{hi,j} \), \( \sum_j \bar{w}_{hi,j} a^y_{hi,j} \), or \( \sum_j \bar{w}_{hi,j} \).

2. \( n_T \text{ covariance matrix of } \sum_{(h,i,j) \in A_h} \bar{w}_{hi,j} y_{hi,j}, \sum_{(h,i,j) \in A_n} \bar{w}_{hi,j} \) and \( \sum_{(h,i,j) \in A_n} \bar{w}_{hi,j} \) have eigenvalues bounded away from 0 and \( \infty \).

3. \( \sum_{(h,i,j) \in A_n} \bar{w}_{hi,j} y_{hi,j} - \bar{Y}^{1+\delta} = O_p(1) \) for some \( \delta > 0 \).

4. \( n_T \text{ max}_h \sum_j \bar{w}_{hi,j} = O_p(1) \).

The following is our main asymptotic result. Its proof is given in the Appendix.

**Theorem 1** Let \( T \geq 2 \) be a fixed integer.

(i) Under conditions (1)-(4),

\[
    \frac{v_S}{\text{Var}(\bar{Y}^*)} \xrightarrow{p} 1.
\]

(ii) If conditions (1)-(3) also hold with \( y \) replaced by \( z \), then

\[
    v_S \left( \begin{array}{cc}
    \text{Var}(\bar{y}^*) & \text{Cov}(\bar{y}^*, \bar{z}^*) \\
    \text{Cov}(\bar{y}^*, \bar{z}^*) & \text{Var}(\bar{z}^*)
    \end{array} \right)^{-1} \xrightarrow{p} I_2,
\]

where \( I_2 \) is the \( 2 \times 2 \) identity matrix.

Since \( (\bar{y}^*, \bar{z}^*) \) is asymptotically normal under conditions (1)-(4) (e.g., Schenker and Welsh, 1988; Rao and Shao, 1992), the result in Theorem 1 implies that the distribution of

\[
    v_S^{-1/2} \left( \begin{array}{c}
    \bar{y}^* - \bar{Y} \\
    \bar{z}^* - \bar{Z}
    \end{array} \right)
\]

converges to the two-dimensional standard normal distribution, and the distribution of

\[
    (\bar{\theta}^* - \theta) / \sqrt{[\nabla g(\bar{y}^*, \bar{z}^*)] v_S \nabla g(\bar{y}^*, \bar{z}^*)}
\]

converges to the chi-squared distribution with \( 2k \) degrees of freedom, where \( k \) is the number of parameters for the model.
converges to the standard normal distribution, where \( \theta = g(\bar{Y}, \bar{Z}) \), \( \dot{\theta}^* = g(\bar{y}^*, \bar{z}^*) \), and \( \nabla g \)

is assumed to be continuous at \((\bar{Y}, \bar{Z})\).

All these results can be easily extended to three or more dimensional cases.

### 3.2. Some simulation results

In this section we present the results from a simulation study comparing the true variance and our variance estimator in the stratified one-stage simple random sampling case.

The population we used is similar to those in Kovar, Rao and Wu (1988) and Sitter (1992). There are \( L = 32 \) strata in the population. In the \( h \)-th stratum, the \( y \)-values of the population are generated according to

\[
y_{hi} \overset{i.i.d.}{\sim} N(\bar{Y}_h, \sigma_h^2), \quad i = 1, \ldots, N_h,
\]

where the population parameters \( N_h, \bar{Y}_h, \) and \( \sigma_h \) are given in Table 1.

After the population was generated, a simple random sample of size \( n_h \) was drawn from stratum \( h \), independently across the 32 strata. The sample sizes \( n_h \) are also listed in Table 1. The total number of sampled units is \( n = 75 \). After the samples were generated, the respondents \( \{y_{hi}, (h, i) \in A_r\} \) were obtained by assuming that the sampled units responded with equal probability \( p \); and the missing values \( \{y_{hi}, (h, i) \in A_m\} \) were imputed independently \( T \) times by taking i.i.d. samples from \( \{y_{hi}, (h, i) \in A_r\} \), with selection probability \( \pi_{hi} = w_{hi}/\sum_{A_r} w_{ij} \) for \( y_{hi}, (h, i) \in A_r \), where the survey weight \( w_{hi} = w_h = N_h/n_h \) in this special case. This process was repeated 10,000 times in the simulation.

In each simulation iteration, we calculated the following statistics based on the \( t \)-th imputed data set:

\[
\bar{y}_h^{*t} = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}^{*t}, \quad \bar{y}^{*t} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{y}_h^{*t}
\]

(note that \( M = N \) in the stratified one-stage case), and

\[
v^{*t} = \sum_{h=1}^{L} \frac{N_h^2}{N^2 n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi}^{*t} - \bar{y}_h^{*t})^2.
\]

Then we computed

\[
\bar{y}^* = \frac{1}{T} \sum_{t=1}^{T} \bar{y}_h^{*t}, \quad \bar{y}_h^* = \frac{1}{T} \sum_{t=1}^{T} \bar{y}_h^{*t},
\]

\[
\hat{B}_h = \frac{1}{T - 1} \sum_{t=1}^{T} (\bar{y}_h^{*t} - \bar{y}_h^* )^2.
\]
\[ \hat{W} = \frac{1}{T} \sum_{t=1}^{T} v^{*t}, \quad \hat{B} = \sum_{h=1}^{L} \hat{B}_h, \]

and \( v_S \) according to (2.9).

Table 2 lists, for some combinations of values of \( p \) and \( T \), the variance of \( \hat{y}^* \) (approximated by the sample variance of the 10,000 simulated values of \( \hat{y}^* \)), and the relative bias (RB) and mean square error (MSE) of \( v_S \) (based on 10,000 simulated values of \( v_S \)). In addition, Table 2 also lists the empirical coverage probabilities (NCP and TCP) of 95\% confidence intervals, where NCP is the coverage probability of the confidence interval obtained by treating \( (\hat{y}^* - \bar{Y}) / \sqrt{v_S} \) as the standard normal random variable, whereas TCP is the coverage probability of the confidence interval obtained by treating \( (\hat{y}^* - \bar{Y}) / \sqrt{v_S} \) as the t-random variable with \( r \) (the number of respondents) degrees of freedom.

The computation was done in an IBM-PC, using IMSL subroutines DRNNOA, RNUND and RNUN for random number generations.

The results in Table 2 indicates that the variance estimator \( v_S \) performs well. Its relative bias is under 3\% (most of time 2\%) for all cases considered. The coverage probabilities of the confidence intervals are close to the nominal level when the response probability \( p > 50\% \). When \( p \leq 50\% \), the coverage probabilities tend to be lower than the nominal level; but this is probably because the sample size is not large enough. Note that the mean number of respondents is \( pn = 37.5 \) when \( p = 50\% \) and \( pn = 30 \) when \( p = 40\% \). The number of data we actually observed may be much lower than 30.

4. Conclusions

A variance estimator for the survey estimator of a function of population means is derived in the situation where (1) the sampling plan is stratified multistage; (2) there are nonrespondents and we do not know which sampled units are nonrespondents; and (3) the nonrespondents are imputed by using the simple hot deck multiple imputation which is simple and efficient, but not proper so that Rubin’s variance estimator cannot be used.

The proposed variance estimator is shown to be asymptotically consistent and performs well in an empirical study.
Appendix: The Proof of Theorem 1

We first state without proof the following useful lemma.

**Lemma 1** Let \( \{x_n, n = 1, 2, \ldots\} \) be a sequence of independent random variables. If

\[
\frac{1}{n} \sum_{k=1}^{n} E|x_k|^{1+\delta} = O(1)
\]

as \( n \to \infty \) for some \( \delta \geq 0 \). Then

\[
\frac{1}{n} \left( \sum_{k=1}^{n} x_k - \sum_{k=1}^{n} E x_k \right) \to_p 0.
\]

We only prove Theorem 1(i) in the case of \( K = 1 \) (uniform response case). The proofs for the other cases are similar. Let \((E_*, V_*)\) represents the (asymptotic) expectation and variance with respect to the randomness in the imputation process, and let \((E, V)\) represents the overall (asymptotic) expectation and variance. Note that

\[
V(\tilde{y}^*) = V(E_* \tilde{y}^*) + EV_*(\tilde{y}^*) = V(\tilde{y}_r) + \frac{EV_*(\tilde{y}^*)}{T}
\]

and

\[
v_S = \frac{p^2}{\hat{p}^2} \frac{(\hat{W} - \hat{B})}{p^2} + \frac{\hat{B}}{T}.
\]

Since \( p^2/\hat{p}^2 \to_p 1 \), it suffices to show that

\[
n_T[\hat{B} - EV_*(\tilde{y}^*)] \to_p 0 \quad (A.1)
\]

and

\[
n_T[(\hat{W} - \hat{B}) - p^2 V(\tilde{y}_r)] \to_p 0. \quad (A.2)
\]

Since \( n_T[V_*(\tilde{y}^*) - EV_*(\tilde{y}^*)] \to_p 0 \) and

\[
E_* \hat{B} = E_* \left[ \sum_{h=1}^{L} \frac{1}{T-1} \sum_{t=1}^{T} \left( \tilde{y}^*_h - \bar{y}^*_h \right) \right] = V_*(\tilde{y}^*),
\]

result (A.1) follows from

\[
n_T[\hat{B} - V_*(\tilde{y}^*)] \to_p 0. \quad (A.3)
\]

By appealing to Lemma 1, (A.3) is implied by

\[
n_T^{1+\delta} E_* \left[ \sum_{h=1}^{L} \frac{1}{T-1} \sum_{t=1}^{T} \left| \tilde{y}^*_h - E_* \tilde{y}^*_h \right|^{1+\delta} \right] = O_p(1). \quad (A.4)
\]
From Rao and Shao (1992),
\[ E_*|\tilde{y}_h^* - E_*\tilde{y}_h^*|^2 + \delta \leq O_p(1) \left( \sum_{j=1}^{n_{hi}} \tilde{w}_{hij} / \sum_{A_n} \tilde{w}_{hij} \right)^{2+\delta}. \]

Therefore, (A.4) follows from condition (4) and the fact that \( \sum_{A_n} \tilde{w}_{hij} \to_p 1 \). This proves (A.1).

For result (A.2), it suffices to show that
\[ n_T [E_*(\tilde{W} - \tilde{B}) - p^2 V(\tilde{y}_r)] \to_p 0 \]  
(A.5)
and
\[ n_T [(\tilde{W} - \tilde{B}) - E_*(\tilde{W} - \tilde{B})] \to_p 0. \]  
(A.6)

Define
\[ u_{hi} = \sum_{j=1}^{n_i} \tilde{w}_{hij} a_{hij}^y, \quad \bar{u}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} u_{hi}, \]
\[ v_{hi} = \sum_{j=1}^{n_i} \tilde{w}_{hij} a_{hij}^y, \quad \bar{v}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} v_{hi}. \]

Since \( E_* \tilde{W} = E_* v^* \) and
\[ E_* v^* = V_*(\tilde{y}^*) + \frac{M^2}{2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)^2 \bar{y}_r^2 \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (v_{hi} - \bar{v}_h)^2 \]
\[ -2\bar{y}_r \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)(v_{hi} - \bar{v}_h), \]
we obtain that
\[ E_* (\tilde{W} - \tilde{B}) = \frac{M^2}{2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)^2 \bar{y}_r \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (v_{hi} - \bar{v}_h)^2 \]
\[ -2\bar{y}_r \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)(v_{hi} - \bar{v}_h). \]

From Krewski and Rao (1981),
\[ n_T \left[ \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)^2 - V \left( \sum_{A_n} \tilde{w}_{hij} a_{hij}^y y_{hij} \right) \right] \to_p 0 \]
and
\[ n_T \left[ \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (v_{hi} - \bar{v}_h)^2 - V \left( \sum_{A_n} \tilde{w}_{hij} a_{hij}^y \right) \right] \to_p 0. \]
Note that

\[ V(\tilde{y}_r) = V\left( \frac{\sum_{A_n} \tilde{w}_{hij} a_{hij} y_{hij}}{\sum_{A_n} \tilde{w}_{hij} a_{hij}^2} \right) \]

\[ = \frac{1}{p^2} \left[ V\left( \sum_{A_n} \tilde{w}_{hij} a_{hij} y_{hij} \right) + Y^2 V\left( \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 \right) \right] - 2Y \text{Cov}\left( \sum_{A_n} \tilde{w}_{hij} a_{hij} y_{hij}, \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 \right) \]

Since \( \tilde{M}/\tilde{M} \to_p 1 \) and \( \tilde{y}_r - \tilde{Y} \to_p 0 \), (A.5) follows from

\[ n_T \left[ \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)(v_{hi} - \bar{v}_h) - \text{Cov}\left( \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 y_{hij}, \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 \right) \right] \to_p 0. \]  \hspace{1cm} (A.7)

Note that

\[ E\left[ \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)(v_{hi} - \bar{v}_h) \right] = \text{Cov}\left( \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 y_{hij}, \sum_{A_n} \tilde{w}_{hij} a_{hij}^2 \right). \]

Therefore, (A.7) follows from Lemma 1 and

\[ \frac{1}{n_T} E\left[ \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} n_T(u_{hi} - Eu_{hi}) n_T(v_{hi} - Ev_{hi}) \right]^{1+\delta} = O_p(1) \] \hspace{1cm} (A.8)

for some \( \delta \geq 0 \). But the left-hand side of (A.8) is bounded by

\[ \frac{1}{2} n_T^{1+2\delta} \left[ E\sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} |u_{hi} - Eu_{hi}|^{2+2\delta} + E\sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} |v_{hi} - Ev_{hi}|^{2+2\delta} \right], \]

which is \( O_p(1) \) under condition (1). Hence (A.7) holds and therefore (A.5) is proved.

It remains to show (A.6). Since

\[ \hat{W} = \frac{1}{T} \sum_{i=1}^T v^* \]

and we have proved that

\[ n_T(\hat{B} - E_* \hat{B}) \to_p 0, \]

we only need to show that

\[ n_T(v^* - E^* v^*) \to_p 0. \] \hspace{1cm} (A.9)

From the definition of \( v^* \), we have

\[ v^* = \frac{M^2}{M^2} \sum_{h=1}^L \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{j=1}^{n_{hi}} \tilde{w}_{hij} (y_{hij}^* - \bar{y}^*) - \frac{1}{n_h} \sum_{j=1}^{n_{hi}} \tilde{w}_{hij} (y_{hij}^* - \bar{y}^*) \right]^2 \]
\[
\begin{align*}
&= \frac{M^2}{\bar{M}^2} \left\{ \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij} - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij}^* - \bar{y}_r) \right]^2 \right. \\
&\quad + 2(\bar{y}_r - \bar{y}^*) \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij}^* - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij}^* - \bar{y}_r) \left[ \right. \\
&\quad \times \left. \sum_{j=1}^{r_{hi}} \bar{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} \right] \\
&\quad \left. \left. + (\bar{y}_r - \bar{y}^*)^2 \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} \right)^2 \right\}.
\end{align*}
\]

Since \( E \bar{y}^* = \bar{y}_r \) and \( V_\epsilon(\bar{y}^*) = o_p(\frac{1}{n_T}) \), we have \( \bar{y}^* - \bar{y}_r \to_p 0 \). By condition (4),

\[
\sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{j=1}^{r_{hi}} \bar{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} \right]^2 \leq \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \sum_{j=1}^{r_{hi}} \bar{w}_{hij} \right)^2 = o_p\left( \frac{1}{n_T} \right).
\]

Therefore,

\[
(\bar{y}_r - \bar{y}^*)^2 \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{j=1}^{r_{hi}} \bar{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} \right]^2 = o_p\left( \frac{1}{n_T} \right).
\]  
(A.10)

Define

\[
t^*_h = \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (1 - a_{hij}^\epsilon) (y_{hij}^* - \bar{y}_r) \quad \text{and} \quad \bar{t}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} t^*_h.
\]

Then

\[
\frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij}^* - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{r_{hi}} \bar{w}_{hij} (y_{hij}^* - \bar{y}_r) \right]^2
\]

\[
= \frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ (t^*_h + u_{hi} - \bar{y}_r v_{hi}) - (\bar{t}_h + \bar{u}_h - \bar{y}_r \bar{v}_h) \right]^2
\]

\[
= \frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ (u_{hi} - \bar{u}_h) - \bar{y}_r (v_{hi} - \bar{v}_h) \right]^2
\]

\[
+ 2 \frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ (u_{hi} - \bar{u}_h) - \bar{y}_r (v_{hi} - \bar{v}_h) \right] (t^*_h - \bar{t}_h)
\]

\[
+ \frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (t^*_h - \bar{t}_h)^2
\]

\[
= E v^* + \frac{M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ (t^*_h - \bar{t}_h)^2 - V_\epsilon(\bar{y}^*) \right]
\]

\[
+ \frac{2M^2}{\bar{M}^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left[ (u_{hi} - \bar{u}_h) - \bar{y}_r (v_{hi} - \bar{v}_h) \right] (t^*_h - \bar{t}_h),
\]
where the last equation follows from
\[
E_*v^* = V_*(\bar{y}^*) + \frac{M^2}{M^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [(u_{hi} - \bar{u}_h) - \bar{y}_r(u_{hi} - \bar{v}_h)]^2.
\]

From Rao and Shao (1992),
\[
n_T \left[ \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (t_{hi}^* - \bar{t}_h^*)^2 - V_*(\bar{y}^*) \right] \to_p 0
\]
and
\[
n_T \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} [(u_{hi} - \bar{u}_h) - \bar{y}_r(u_{hi} - \bar{v}_h)](t_{hi}^* - \bar{t}_h^*) \to_p 0.
\]

Thus,
\[
n_T \left[ \frac{M^2}{M^2} \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r))^2 - E_*v^* \right] \to_p 0. \quad (A.11)
\]

Note that
\[
2 \left| \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r) \right| \left[ \sum_{j=1}^{n_{hi}} \hat{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij} \right]
\]
\[
\leq \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r) - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij}(y_{hij}^* - \bar{y}_r))^2
\]
\[
+ \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_{hi}} \hat{w}_{hij} \right) = O_p \left( \frac{1}{n_T} \right)
\]
by (A.5) and (A.11). This result, together with results (A.10) and (A.11), imply (A.9). This completes the proof.

References


Table 1. Population Parameters and Sample Sizes

<table>
<thead>
<tr>
<th>$h$</th>
<th>$N_h$</th>
<th>$Y_h$</th>
<th>$\sigma_h$</th>
<th>$n_h$</th>
<th>$h$</th>
<th>$N_h$</th>
<th>$Y_h$</th>
<th>$\sigma_h$</th>
<th>$n_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>8.6</td>
<td>4.00</td>
<td>3</td>
<td>17</td>
<td>34</td>
<td>8.6</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>8.7</td>
<td>4.00</td>
<td>3</td>
<td>18</td>
<td>34</td>
<td>8.4</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>8.5</td>
<td>4.00</td>
<td>3</td>
<td>19</td>
<td>34</td>
<td>8.5</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>38</td>
<td>8.3</td>
<td>4.00</td>
<td>3</td>
<td>20</td>
<td>34</td>
<td>8.8</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>8.9</td>
<td>4.00</td>
<td>3</td>
<td>21</td>
<td>34</td>
<td>8.4</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>8.8</td>
<td>4.00</td>
<td>3</td>
<td>22</td>
<td>22</td>
<td>8.7</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>38</td>
<td>8.2</td>
<td>4.00</td>
<td>3</td>
<td>23</td>
<td>22</td>
<td>8.6</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>38</td>
<td>8.6</td>
<td>4.00</td>
<td>3</td>
<td>24</td>
<td>22</td>
<td>8.5</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>38</td>
<td>8.6</td>
<td>4.00</td>
<td>3</td>
<td>25</td>
<td>22</td>
<td>8.4</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>8.4</td>
<td>4.00</td>
<td>3</td>
<td>26</td>
<td>22</td>
<td>8.8</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>38</td>
<td>8.4</td>
<td>4.00</td>
<td>3</td>
<td>27</td>
<td>22</td>
<td>8.9</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>34</td>
<td>8.5</td>
<td>0.25</td>
<td>2</td>
<td>28</td>
<td>22</td>
<td>8.3</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>34</td>
<td>8.1</td>
<td>0.25</td>
<td>2</td>
<td>29</td>
<td>22</td>
<td>8.2</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>34</td>
<td>8.4</td>
<td>0.25</td>
<td>2</td>
<td>30</td>
<td>22</td>
<td>8.9</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>34</td>
<td>8.3</td>
<td>0.25</td>
<td>2</td>
<td>31</td>
<td>22</td>
<td>8.4</td>
<td>1.00</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>34</td>
<td>8.6</td>
<td>0.25</td>
<td>2</td>
<td>32</td>
<td>22</td>
<td>8.6</td>
<td>1.00</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 2. Simulation Results

<table>
<thead>
<tr>
<th>$p$</th>
<th>$T$</th>
<th>$Var(\hat{y}^*)$</th>
<th>RB(%)</th>
<th>MSE</th>
<th>NCP(%)</th>
<th>TCP(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>2</td>
<td>0.1911</td>
<td>1.03</td>
<td>0.1498</td>
<td>91.63</td>
<td>92.60</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1865</td>
<td>-2.68</td>
<td>0.1188</td>
<td>91.84</td>
<td>92.62</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.1760</td>
<td>-1.31</td>
<td>0.0869</td>
<td>93.15</td>
<td>93.98</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>0.1544</td>
<td>-0.59</td>
<td>0.0857</td>
<td>92.75</td>
<td>93.28</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1494</td>
<td>-1.55</td>
<td>0.0714</td>
<td>92.78</td>
<td>93.33</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.1406</td>
<td>-1.90</td>
<td>0.0544</td>
<td>93.13</td>
<td>93.85</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>0.1275</td>
<td>0.34</td>
<td>0.0544</td>
<td>93.83</td>
<td>94.36</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1217</td>
<td>0.05</td>
<td>0.0478</td>
<td>93.75</td>
<td>94.33</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.1140</td>
<td>0.07</td>
<td>0.0398</td>
<td>93.82</td>
<td>94.38</td>
</tr>
<tr>
<td>0.7</td>
<td>2</td>
<td>0.1067</td>
<td>0.65</td>
<td>0.0378</td>
<td>94.20</td>
<td>94.47</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.1024</td>
<td>0.56</td>
<td>0.0337</td>
<td>93.89</td>
<td>94.42</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0964</td>
<td>0.30</td>
<td>0.0295</td>
<td>93.89</td>
<td>94.50</td>
</tr>
<tr>
<td>0.8</td>
<td>2</td>
<td>0.0910</td>
<td>0.03</td>
<td>0.0271</td>
<td>94.10</td>
<td>94.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0887</td>
<td>-0.84</td>
<td>0.0255</td>
<td>94.02</td>
<td>94.40</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0836</td>
<td>0.86</td>
<td>0.0234</td>
<td>94.06</td>
<td>94.42</td>
</tr>
<tr>
<td>0.9</td>
<td>2</td>
<td>0.0773</td>
<td>0.47</td>
<td>0.0209</td>
<td>94.08</td>
<td>94.45</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0768</td>
<td>-0.96</td>
<td>0.0200</td>
<td>94.04</td>
<td>94.42</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0752</td>
<td>-1.33</td>
<td>0.0190</td>
<td>93.83</td>
<td>94.29</td>
</tr>
</tbody>
</table>