DEPARTMENT OF STATISTICS
University of Wisconsin-Madison
1300 University Ave.
Medical Science Center
Madison, WI 53706

TECHNICAL REPORT NO. 1137r
February 2007

Confidence Intervals Based on Survey Data With Nearest Neighbor Imputation

By

Jun Shao and Hansheng Wang
Confidence Intervals Based On Survey Data With Nearest Neighbor Imputation

Jun Shao\(^1\) and Hansheng Wang\(^2\)

\(^1\)Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.
\(^2\)Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing, 100871, P. R. China

Abstract

Nearest neighbor imputation (NNI) is a popular imputation method used to compensate for item nonresponse in sample surveys. Although previous results showed that the NNI sample mean and quantiles are consistent estimators of the population mean and quantiles, large sample inference procedures, such as asymptotic confidence intervals for the population mean and quantiles, are not available. For the population mean, we establish the asymptotic normality of the NNI sample mean and derive a consistent estimator of its limiting variance, which leads to an asymptotically valid confidence interval. For the quantiles, we obtain consistent variance estimators and asymptotically valid confidence intervals using a Bahadur type representation for NNI sample quantiles. Some limited simulation results are presented to examine the finite-sample performance of the proposed variance estimators and confidence intervals.

Key Words: Bahadur representation; Hot deck; Mean; Quantiles; Variance estimation.

Acknowledgments: This work was partially supported by NSF Grant DMS-0404535. The authors would like to thank referees for their helpful comments.
1 Introduction

Consider a bivariate sample \((x_1, y_1), \ldots, (x_n, y_n)\) with observed \(y_1, \ldots, y_r\) (respondents), missing \(y_{r+1}, \ldots, y_n\) (nonrespondents), and observed \(x_1, \ldots, x_n\). In sample surveys, imputation is commonly applied to compensate for this type of item nonresponse (Sedransk 1985, Kalton and Kasprzyk 1986, Rubin 1987). The nearest neighbor imputation (NNI) method imputes a missing \(y_j\) by \(y_i\), where \(1 \leq i \leq r\) and \(i\) is the nearest neighbor of \(j\) measured by the \(x\)-variable, i.e., \(i\) satisfies \(|x_i - x_j| = \min_{1 \leq l \leq r} |x_l - x_j|\). We focus on the case where \(x\) has a continuous distribution so that there are no tied \(x\)-values. In practice, if there are tied \(x\)-values due to reasons such as rounding, NNI can be applied by randomly selecting a \(j\) from the nearest neighbors from tied \(x\)-values. Also, NNI is often carried out by first dividing the sample into several “imputation classes” and then finding nearest neighbors within each imputation class.

The NNI method has some nice features. First, it imputes a nonrespondent by a respondent from the same variable; the imputed values are actually occurring values, not constructed values, and they may not be perfect substitutes, but are unlikely to be nonsensical values. Second, the NNI method may be more efficient than other popular methods such as the mean imputation and the random hot deck imputation when the \(x\)-variable provides useful auxiliary information. Third, the NNI method does not assume a parametric regression model between \(y\) and \(x\) and, hence, it is more robust against model violations than methods such as ratio imputation and regression imputation that are based on a linear regression model. Finally, under some conditions NNI estimators (i.e., estimators calculated using standard formulas and treating nearest neighbor imputed values as observed data) are asymptotically valid not only for moments of the \(y\)-variable, but also for the distribution and quantiles of the \(y\)-variable, which is a superiority over other non-random imputation methods (such as the mean, ratio, and regression imputation) that lead to valid moment estimators only.

There are some other nonparametric imputation methods (see, e.g., Cheng 1994; Wang and Rao 2002) that are more efficient than NNI, although they impute nonrespondents by constructed values. However, the NNI method is a popular method in many survey agencies.
and it has a long history of applications in surveys such as the Census 2000 and the Current Population Survey conducted by the U.S. Census Bureau (Farber and Griffin 1998; Fay 1999), the Job Openings and Labor Turnover Survey and the Employee Benefits Survey conducted by the U.S. Bureau of Labor Statistics (Montaquila and Ponikowski 1993), and the Unified Enterprise Survey, the Survey of Household Spending, and the Financial Farm Survey conducted by Statistics Canada (Rancourt 1999). In these agencies, it is unlikely that NNI will be replaced by another nonparametric imputation method in a short period of time.

Therefore, a theoretical study of the properties of NNI is important. Although NNI is the same as regression imputation using \( k \)-nearest neighbor regression with \( k = 1 \), the existing theoretical (asymptotic) results for \( k \)-nearest neighbor regression (see, e.g., Härdle 1990) are all for the case where \( k \to \infty \) as the sample size increases. Theoretical studies of the NNI methodology started with Lee, Rancourt and Särndal (1994) and Rancourt (1999) who showed some properties of NNI estimators when \( y_i \) and \( x_i \) are assumed to follow a simple linear regression model. The first piece of theoretical work for NNI under a general nonparametric setting is given in Chen and Shao (2000) who established the consistency of NNI estimators such as the sample mean and sample quantile. A year later, Chen and Shao (2001) investigated jackknife variance estimators for the sample mean. In practice, statistical inference such as setting an approximate confidence interval for a population parameter is often needed. Asymptotic result on confidence intervals based on survey data with NNI, however, is not available and the purpose of this paper is to fill this gap.

The most important population parameter in surveys is the population mean (or a function of several population means). Although empirical results (e.g., Chen and Shao 2001) showed that a confidence interval of the form

\[
\text{NNI sample mean} \pm z_\alpha \sqrt{\text{variance estimator for NNI sample mean}} \quad (1)
\]

works well, where \( \alpha \) is a fixed nominal confidence level and \( z_\alpha \) is the \( 100(1 - \alpha/2) \)th normal percentile, the use of (1) lacks statistical justification, since it has not been shown that the interval in (1) is asymptotically valid in the sense of

\[
P(\text{confidence interval covers the true population parameter}) \to 1 - \alpha \quad (2)
\]
(under some limiting process as the sample size increases). After an introduction to some notation and assumptions in Section 2, we establish result (2) in Section 3 by first showing the asymptotic normality of NNI sample means and then finding a variance estimator in (1) that is consistent for the variance in the limiting distribution of the NNI sample mean.

Estimation or inference on population quantiles has become more and more important in modern survey statistics (Rao, Kovar and Mantel, 1990; Francisco and Fuller, 1991). For income variables, for example, the median income or other quantiles are as important as the mean income. In children with cystic fibrosis, the 10th percentiles of height and weight are important clinical boundaries between healthy and possibly nutritionally compromised patients (Kosorok 1999). Although Chen and Shao (2000) showed that NNI sample quantiles are consistent for population quantiles, variance estimation for NNI sample quantiles were not discussed. Note that the jackknife cannot be directly applied to sample quantiles (Efron 1982). In Section 4, we establish some Bahadur type representations for NNI sample quantiles, which not only shows the asymptotic normality of NNI sample quantiles, but also provides consistent variance estimators for NNI sample quantiles and asymptotically valid confidence intervals (in the sense of (2)) for population quantiles.

To complement the theoretical results, some simulation results are presented in Section 5 to examine the performance of the proposed estimators and confidence intervals. Some discussions are given in the last section.

2 Notation and Assumption

This section introduces some notation and general assumptions used throughout the paper. Let $P$ be a finite population containing $M$ units indexed by $i$. A sample $S$ of size $n$ is taken from $P$ according to some sampling design. Let $w_i$ be the survey weight for unit $i$, which is equal to the inverse of the probability that unit $i$ is selected. For any set of values $\{z_i : i \in P\}$,

$$E_s\left(\sum_{i \in S} w_i z_i\right) = \sum_{i \in P} z_i$$

(3)
(i.e., \( \sum_{i \in S} w_i z_i \) is a Horvitz-Thompson type estimator), where \( E_s \) is the expectation with respect to sampling. Although \( w_i \) is defined for any \( i \in \mathcal{P} \), we only need \( w_i \) for \( i \in S \) in applications.

We consider one stage sampling without clusters. Some discussion about cluster sampling and multistage sampling is given in the last section.

To consider asymptotics, we assume that the finite population \( \mathcal{P} \) is a member of a sequence of finite populations indexed by \( \nu \). All limiting processes in this paper are understood to be as \( \nu \to \infty \). As \( \nu \to \infty \), the population size \( M \) and the sample size \( n \) increase to infinity. In sample surveys, the following regularity conditions on \( w_i \)'s are typically imposed:

\[
\max_{i \in \mathcal{P}} \frac{n w_i}{M} \leq b_0 \tag{4}
\]

and

\[
\frac{n}{M^2} \text{Var}_s \left( \sum_{i \in S} w_i \right) \leq b_1, \tag{5}
\]

where \( b_0 \) and \( b_1 \) are some positive constants and \( \text{Var}_s \) is the variance with respect to sampling. Condition (4) ensures that none of the weights \( w_i \) is disproportionately large (see Krewski and Rao 1981). Condition (5) means that \( \text{Var}_s(\sum_{i \in S} w_i/M) \) is at most of the order \( n^{-1} \). Conditions (4)-(5) are satisfied for stratified simple random sampling designs.

Note that (3) and (5) imply that \( \sum_{i \in S} w_i / M \to_p 1 \), where \( \to_p \) is convergence in probability as \( \nu \to \infty \). Furthermore, \( E_s(\sum_{i \in S} w_i^3) = \sum_{i \in \mathcal{P}} w_i^3 \leq (Mb_0/n)^3 \) under condition (4). Hence, it follows from Liapunov's central limit theorem that

\[
\left( \frac{\sum_{i \in S} w_i}{M} - 1 \right) / \sqrt{\text{Var}_s \left( \sum_{i \in S} w_i / M \right)} \to_d N(0,1), \tag{6}
\]

where \( \to_d \) is convergence in distribution as \( \nu \to \infty \).

Let \( (x, y) \) be a bivariate characteristic from a given unit in the population, where \( y \) is the main variable of interest and \( x \) is a covariate. Let \( a \) be the response indicator, i.e., \( a = 1 \) if \( y \) is observed and \( a = 0 \) if \( y \) is a nonrespondent. Note that we define \( a \) for every unit in \( \mathcal{P} \) (see, e.g., Shao and Steel, 1999). For unit \( i \in \mathcal{P} \), we denote \( (x, y, a) \) by \( (x_i, y_i, a_i) \). To study asymptotic validity of NNI, we need some assumptions.
Assumption A. Each finite population $\mathcal{P}$ is divided into $K$ (a fixed integer) imputation classes $\mathcal{P}_k$, $k = 1, \ldots, K$, such that the population and sample sizes of each imputation class increase to infinity and, within imputation class $k$, $(x_i, y_i, a_i)$'s are independent and identically distributed (i.i.d.) from a superpopulation with $P(a_i = 1|x_i, y_i, k) = P(a_i = 1|x_i, k)$, and $P(a_i = 1|k) = p_k > 0$. Sampling is independent of the superpopulation. $(x_i, y_i, a_i)$'s from different imputation classes are independent. NNI is carried out within each imputation class.

Throughout this paper, the probability, expectation, and variance with respect to sampling and the superpopulation in Assumption A are denoted by $P$, $E$, and $\text{Var}$, respectively. When $\mathcal{S}$ is not present, however, $P$, $E$, and $\text{Var}$ reduce to the probability, expectation, and variance with respect to the superpopulation only. For example, $E(y|x)$ is the conditional expectation of $y$ given $x$ with respect to the superpopulation; $E(\sum_{i \in \mathcal{S}} y_i)$ is the expectation with respect to both sampling and the superpopulation. Furthermore, i.i.d. is always with respect to the superpopulation.

The assumption on the response probability means that the response indicator $a$ is independent of $y$, given the covariate $x$ and the imputation class $k$. That is, within an imputation class, the response mechanism is covariate-dependent (Little 1995) or unconfounded (Lee, Rancourt and Särndal, 1994), an assumption made for the validity of many other popular imputation methods. Although $(x, y, a)$'s within an imputation class are assumed i.i.d., the response mechanism is still not completely at random, since $P(a = 1|x)$ depends on the covariate $x$. In particular, the conditional distribution of $(x, y)$ given $a = 1$ may be different from the conditional distribution of $(x, y)$ given $a = 0$.

Imputation classes are usually constructed using a categorical variable whose values are observed for all sampled units; for example, under stratified sampling, strata or unions of strata are often used as imputation classes. Each imputation class should contain a large number of sampled units. When there are many strata of small sizes, imputation classes are often obtained through poststratification (Valliant 1993) and/or combining small strata.

The superpopulation assumption on $(x, y, a)$ within each imputation class is natural, since NNI requires some exchangeability of units within each imputation class. The NNI
method is a model-based approach, rather than a design-based approach. However, the model assumption is nonparametric (i.e., we only assume that \((x, y, a)\)'s are i.i.d. within each imputation class) and is much weaker than a parametric linear model assumption on the conditional mean of \(y\) given the covariate \(x\), which is typically assumed for a regression type imputation.

Let \(\psi(x)\) be a continuous, bounded, and strictly increasing function of \(x\). Then \(E(y|x) = E(y|\psi(x))\). Since the NNI method recovers information about nonrespondents using \(x\)-covariates through \(E(y|x)\), without loss of generality we can always apply the transformation \(\psi(x)\). Hence, we assume in the rest of this paper that \(x\) is bounded, i.e., \(-\infty < x_- \leq x \leq x_+ < \infty\). However, we do not need to know the values of \(x_-\) and \(x_+\).

**Assumption B.** For any fixed imputation class \(k\), \(P(a = 1|x, k) > 0\) for all \(x \in [x_-\), \(x_+\)\) and is a continuous function of \(x\); conditional on \(a\), \(x\) has a bounded and continuous Lebesgue density \(f_{k,a}\); and \(E(y|x, k)\) is Lipschitz continuous of \(x\).

We now introduce more notation. In imputation class \(k\), let \(R_k\) be the set of indices of \(y\)-respondents, \(\tilde{R}_k\) be the set of indices of nonrespondents, and \(S_k = R_k \cup \tilde{R}_k\). Define \(\mu_k(x) = E(y|x, k), \mu_{k,a} = E(y|k, a), \mu_k = E(y|k) = p_k \mu_{k,1} + (1 - p_k) \mu_{k,0}\), and \(\mu = E(y) = \sum_{k=1}^{K} \frac{M_k}{M} \mu_k\), where \(M_k\) is the size of \(P_k\) and \(M = \sum_k M_k\). Conditional on \(R_k\) and \(X_k = \{x_i : i \in R_k\}\) (the covariates from the respondents), let

\[
q_{k,i} = P\left(\left|x - x_i\right| = \min_{j \in \tilde{R}_k} \left|x - x_j\right|\bigg| a = 0, k, R_k, X_k, S_k\right)
\]

be the probability that \(i \in \tilde{R}_k\) will be selected as the nearest neighbor for a nonrespondent within \(\tilde{R}_k\), where \(P\) is with respect to \(x\) conditional on \(a = 0, k, R_k, X_k, S_k\).

When there is only one imputation class (\(K = 1\)), the subscript \(k\) will be dropped for simplicity, i.e., \(S, R, X, p, \mu, \mu_a, q, f, a\), etc. are the same as \(S_k, R_k, X_k, p_k, \mu_k, \mu_{k,a}, q_{k,i}, f_{k,a}\), etc. when \(K = 1\).

### 3 Confidence Intervals for Means

Without nonresponse, the superpopulation mean \(\mu\) and the finite population mean \(\bar{Y} = M^{-1} \sum_{i \in P} y_i\) are estimated by a Horvitz-Thompson estimator \(\sum_{i \in S} w_i y_i / M\). After NNI, our
estimator of \( \mu \) or \( \bar{Y} \) is

\[
\hat{\mu} = \frac{1}{M} \sum_{k} \left( \sum_{i \in R_k} w_i y_i + \sum_{i \in \tilde{R}_k} w_i \tilde{y}_i \right) = \sum_{k} \frac{M_k}{M} \left( \sum_{i \in R_k} \tilde{w}_{k,i} y_i + \sum_{i \in \tilde{R}_k} \tilde{w}_{k,i} \tilde{y}_i \right),
\]

(7)

where \( \tilde{y}_i \) is the imputed value for the nonrespondent \( y_i, i \in \tilde{R}_k \), and \( \tilde{w}_{k,i} = w_i / M_k \) when \( i \in S_k \). The estimator \( \hat{\mu} \) will be referred to as NNI sample mean, although it is a weighted average of respondents and imputed values. In some cases, \( M \) is unknown. From result (6), \( \hat{M} = \sum_{i \in S} w_i \) is a consistent estimator of \( M \). We can estimate \( \mu \) or \( \bar{Y} \) by a ratio estimator

\[
\frac{1}{M} \sum_{k} \left( \sum_{i \in R_k} w_i y_i + \sum_{i \in \tilde{R}_k} w_i \tilde{y}_i \right) = \frac{\hat{\mu}}{M / \hat{M}}.
\]

An example of using this ratio estimator is the estimator given by (25) in Section 4. The asymptotic property of this estimator can be obtained using result (6), the result for \( \hat{\mu} \), and the delta-method. Similarly, if the parameter of interest is a differentiable function of several population means, the point estimator is the same function of sample means and its asymptotic property can be derived using the delta-method. Thus, in what follows we focus on the asymptotic property of \( \hat{\mu} \).

Note that, conditional on \( S \) and all observed \( (y_i, x_i) \), imputed values within imputation class \( k \) are i.i.d. taking the value \( y_i \) with probability \( q_{k,i}, i \in R_k \). Hence,

\[
E(\hat{\mu}|\text{sample and respondents}) = \sum_{k} \frac{M_k}{M} \left( \sum_{i \in R_k} \tilde{w}_{k,i} y_i + \sum_{i \in \tilde{R}_k} \tilde{w}_{k,i} \sum_{i \in \tilde{R}_k} q_{k,i} y_i \right).
\]

Applying part (iii) of the following lemma to each imputation class, we conclude that \( \sum_{i \in R_k} q_{k,i} y_i \) converges to \( \mu_{k,0} \) (the mean of \( y \)-nonrespondents in imputation class \( k \)), which shows how NNI recovers information about \( y \)-nonrespondents using \( y \)-respondents and \( x \)-values under Assumptions A-B.

**Lemma 1.** Suppose that Assumptions A-B hold with \( K = 1 \).

(i) If \( g \) is a function of \( x \) with \( E[g(x_i)]^2 < \infty \) then, for any \( m = 1, 2, \ldots \),

\[
E \left\{ r^{m-1} \sum_{i \in R} q_i^m g(x_i) \right\} - \frac{(m + 1)!}{2^m} E \left[ \frac{g(x_i) f_0^{m-1}(x_i)}{f_1^{m-1}(x_i)} \bigg| a_i = 0 \right] \to 0,
\]

where \( r \) is the size of \( R \) (the number of respondents).

(ii) If \( E[g(x_i)]^4 < \infty \), then, for any \( m = 1, 2, \ldots \),

\[
E \left\{ r^{m-1} \sum_{i \in R} q_i^m g(x_i) - \frac{(m + 1)!}{2^m} E \left[ \frac{g(x_i) f_0^{m-1}(x_i)}{f_1^{m-1}(x_i)} \bigg| a_i = 0 \right] \right\}^2 \to 0.
\]
(iii) If \( E(y_i^{dl}) < \infty \) for a positive integer \( l \), then

\[
\sum_{i \in R} q_i y_i^l \rightarrow_p E(y_i^l | a_i = 0). \tag{10}
\]

The proof of Lemma 1 is given in the Appendix. The following is a heuristic argument on why NNI and any other type of regression imputation can use the value of \( x \) to impute a missing \( y \) and produce an almost unbiased estimator of \( \mu \), under Assumption A. Assume that \( K = 1 \) and \( \mu(x) \) is a known function. Then a missing \( y_i \) is imputed as \( \mu(x_i) \) and the resulting estimator of \( \mu \) is \( \tilde{\mu} = \sum_{i \in S} (\tilde{w}_i a_i y_i + \tilde{w}_i (1 - a_i) \mu(x_i)) \). Under Assumption A,

\[
E(\tilde{\mu}|x_1, \ldots, x_n) = \sum_{i \in S} \tilde{w}_i [E(a_i y_i|x_i) + E((1-a_i)\mu(x_i)|x_i)]
\]

\[
= \sum_{i \in S} \tilde{w}_i [E(a_i|x_i) E(y_i|x_i) + E(1-a_i|x_i) \mu(x_i)]
\]

\[
= \sum_{i \in S} \tilde{w}_i \mu(x_i)
\]

\[
= E \left( \sum_{i \in S} \tilde{w}_i y_i | x_1, \ldots, x_n \right),
\]

where the second equality follows from \( P(a_i = 1|x_i, y_i) = P(a_i = 1|x_i) \) in Assumption A. Hence, \( \tilde{\mu} \) has the same asymptotic mean as \( \sum_{i \in S} \tilde{w}_i y_i \), the estimator without nonresponse.

The following result is fundamental for any inference method based on normal approximation.

**Theorem 1.** Assume Assumptions A-B and, within each imputation class \( k \), conditions (4)-(5). Assume further that \( E(y_i^{yl}) < \infty \). Then

\[
\sqrt{n}(\hat{\mu} - \mu) / \sigma \rightarrow_d N(0, 1) \tag{11}
\]

for some \( \sigma > 0 \), where \( \rightarrow_d \) is convergence in distribution unconditionally with respect to the superpopulation model in Assumption A and sampling.

**Proof.** Let \( \hat{\mu}_k = \sum_{i \in R_k} \tilde{w}_{k,i} y_i + \sum_{i \in \tilde{R}_k} \tilde{w}_{k,i} \tilde{y}_i \). Then \( \hat{\mu} = \sum_k \frac{M_k}{M} \hat{\mu}_k \). Since variables are independent across imputation classes and imputation is carried out within each imputation class, \( \hat{\mu}_k \)'s are independent. Hence, it suffices to show result (11) for each \( \hat{\mu}_k \). We now drop the subscript \( k \) with the understanding that the rest of the proof is for a single fixed imputation class. Let \( S, R, \) and \( X \) be defined in Section 2 and let \( Y = \{ y_i : i \in R \} \). Then
$E(\tilde{y}_i|Y, \mathcal{X}, R, S) = \sum_{i \in R} q_i y_i$. Define $\tilde{e}_i = \tilde{y}_i - E(\tilde{y}_i|Y, \mathcal{X}, R, S)$. Consider the following decomposition:

$$\tilde{\mu} - \mu = Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6,$$

where $Q_1 = \sum_{i \in R} \tilde{w}_i \tilde{e}_i$, $Q_2 = \sum_{i \in R} \tilde{w}_i [y_i - \mu(x_i)] + (1 - p) \sum_{i \in S} q_i [y_i - \mu(x_i)]$, $Q_3 = \sum_{i \in R} \tilde{w}_i \mu(x_i) - \mu_1$, $Q_4 = (\mu_1 - \mu_0) \sum_{i \in S} \tilde{w}_i (a_i - p)$, $Q_5 = \mu (\sum_{i \in S} \tilde{w}_i - 1)$; $Q_6 = [\sum_{i \in R} \tilde{w}_i - (1 - p)] \sum_{i \in R} q_i [y_i - \mu(x_i)] + \sum_{i \in R} \tilde{w}_i [\sum_{i \in R} q_i \mu(x_i) - \mu_0]$. By repeatedly applying Lemma 1 in Schenker and Welsh (1988), Result (11) follows from

$$P(\sqrt{n}Q_1 \leq \sigma_1 t|Y, \mathcal{X}, R, S) \rightarrow \Phi(t) \text{ a.s.,}$$

$$P(\sqrt{n}Q_2 \leq \sigma_2 t|Y, \mathcal{X}, R, S) \rightarrow \Phi(t) \text{ a.s.,}$$

$$P(\sqrt{n}Q_3 \leq \sigma_3 t|R, S) \rightarrow \Phi(t) \text{ a.s.,}$$

$$P(\sqrt{n}Q_4 \leq \sigma_4 t|S) \rightarrow \Phi(t) \text{ a.s.,}$$

$$P(\sqrt{n}Q_5 \leq \sigma_5 t) \rightarrow \Phi(t),$$

$$\sqrt{n}Q_6 \rightarrow_p 0,$$

for any real $t$, where $P(\cdot|\mathcal{A})$ denotes the conditional probability given $\mathcal{A}$, $\Phi$ is the standard normal distribution function, $\sigma_i$'s are some nonnegative parameters, and $\sigma^2 = \sigma_1^2 + \cdots + \sigma_5^2$.

Conditional on $(Y, \mathcal{X}, R, S)$, $\tilde{e}_i$'s are i.i.d. with mean 0 and variance

$$\sum_{i \in R} q_i y_i^2 - \left(\sum_{i \in R} q_i y_i\right)^2 \rightarrow_p 0$$

(by Lemma 1(iii) under the condition $E(y_i^2) < \infty$), where $v_0 = E(y_i^2|a_i = 0) - [E(y_i|a_i = 0)]^2$.

Since $a_i$'s are i.i.d. with $E(a_i) = p$, it follows from condition (4) that

$$n \sum_{i \in R} \tilde{w}_i^2 - (1 - p)c_w = n \sum_{i \in S} (1 - a_i) \tilde{w}_i^2 - (1 - p)c_w \rightarrow_p 0,$$

where $c_w = \frac{n}{\sum_{i \in R} w_i}$. By Lindeberg's central limit theorem,

$$\left(\sum_{i \in R} \tilde{w}_i^2 \right)^{-1/2} \left[ \sum_{i \in R} q_i y_i^2 - \left(\sum_{i \in R} q_i y_i\right)^2 \right]^{-1/2} \sum_{i \in R} \tilde{w}_i \tilde{e}_i \rightarrow_d N(0, 1),$$

which, together with results (18) and (19), implies result (12) with $\sigma_1^2 = (1 - p)c_w v_0$. 

10
Conditional on \((\mathcal{X}, \mathcal{R}, \mathcal{S})\), \(y_i - \mu(x_i)\), \(i = 1, \ldots, r\), are independent random variables with mean 0 and variances \(\text{Var}(y_i|\mathcal{X}, \mathcal{R}, \mathcal{S}) = \text{Var}(y_i|x_i) = v(x_i)\). Similar to (19), \(n \sum_{i \in \mathcal{R}} \bar{w}_i^2 v(x_i) + pc_w E[v(x_i)|a_i = 1] \to_p 0\), since \(v(x_i), i = 1, \ldots, r\), are i.i.d. By Lemma 1,
\[
n \sum_{i \in \mathcal{R}} q_i^2 v(x_i) - \frac{3}{2p} E \left[ \frac{v(x_i)f_0(x_i)}{f_1(x_i)} \right]_{a_i = 0} \to_p 0 \tag{20}
\]
\((r/n \to p)\) and
\[
E[r q_i v(x_i)|\mathcal{R}] - E[v(x_i)|a_i = 0] \to_p 0. \tag{21}
\]
From condition (4) and Lemma 1, \(n^2 \sum_{i \in \mathcal{R}} \bar{w}_i^2 q_i^2 [v(x_i)]^2 = o_p(1)\). This result and the law of large numbers imply that
\[
n \sum_{i \in \mathcal{R}} \bar{w}_i q_i v(x_i) = n \sum_{i \in \mathcal{R}} \bar{w}_i E[q_i v(x_i)|\mathcal{R}] + o_p(1). \tag{22}
\]
It follows from (21), (22), \(r/n \to_p p\), and \(\sum_{i \in \mathcal{R}} \bar{w}_i \to_p p\) that \(n \sum_{i \in \mathcal{R}} \bar{w}_i q_i v(x_i) - E[v(x_i)|a_i = 0] \to_p 0\). This result and result (20) imply
\[
n \text{Var}(Q_2|\mathcal{X}, \mathcal{R}, \mathcal{S}) = n \sum_{i \in \mathcal{R}} \left[ \bar{w}_i + (1-p)q_i \right]^2 \text{Var}(y_i|\mathcal{X}, \mathcal{R}, \mathcal{S})
= n \sum_{i \in \mathcal{R}} \left[ \bar{w}_i^2 + (1-p)^2 q_i^2 + 2(1-p) \bar{w}_i q_i v(x_i) \right]
= \sigma^2 + o_p(1),
\]
where
\[
\sigma^2 = pc_w E[v(x_i)|a_i = 1] + \frac{3(1-p)^2}{2p} E \left[ \frac{v(x_i)f_0(x_i)}{f_1(x_i)} \right]_{a_i = 0} + 2(1-p)E[v(x_i)|a_i = 0].
\]
Let \(h(x_i) = E([y_i - \mu(x_i)]^4|\mathcal{X}, \mathcal{R}, \mathcal{S}) = E([y_i - \mu(x_i)]^4|x_i)\) (Assumption A). It follows from Lemma 1 that \(n^2 \sum_{i \in \mathcal{R}} q_i^4 h(x_i) \to_p 0\). By condition (4), \(n^2 \sum_{i \in \mathcal{R}} \bar{w}_i^4 h(x_i) \to_p 0\). Hence,
\[
n^2 \sum_{i \in \mathcal{R}} [\bar{w}_i + (1-p)q_i]^4 E\{[y_i - \mu(x_i)]^4|x_i, \mathcal{X}, \mathcal{R}, \mathcal{S}\} \to_p 0,
\]
and, therefore, it follows from Liapunov’s central limit theorem that (13) holds.

Conditional on \((\mathcal{R}, \mathcal{S})\), \(\mu(x_i) - \mu_1\), \(i = 1, \ldots, r\), are i.i.d. with mean 0 and finite variance \(\text{Var}[\mu(x_i)|a_i = 1]\). Therefore, condition (4) and result (19) imply (14) with \(\sigma^2_3 = pc_w \text{Var}[\mu(x_i)|a_i = 1]\). Conditional on \(\mathcal{S}\), \(a_i\)'s are i.i.d. with mean \(p\) and variance \(p(1-p)\).
Then condition (4) and Lindeberg’s central limit theorem imply result (15) with \( \sigma^2_i = p(1-p)c_w(\mu_0 - \mu_1)^2 \). Result (16) follows from result (6), where \( \sigma^2_5 = \mu^2 n \text{Var} (\sum_{i \in S} \bar{w}_i) \).

Finally, it follows from the proofs of (15) and (16) that \( \sum_{i \in \mathcal{R}} \bar{w}_i - (1-p) = O_p(n^{-1/2}) \) and \( \sum_{i \in \mathcal{R}} \bar{w}_i = O_p(1) \). From Lemma 1, \( \sum_{i \in \mathcal{R}} q_i[y_i - \mu(x_i)] = o_p(1) \). From Theorem 1 in Chen and Shao (2000), \( \sum_{i \in \mathcal{R}} q_i \mu(x_i) - \mu_0 = o_p(n^{-1/2}) \). Hence, (17) holds. This completes the proof.

We now consider a variance estimator for \( \hat{\mu} \) that is a simplified version of the partially adjusted jackknife variance estimator in Chen and Shao (2001):

\[
v_n = \frac{1}{\sum_{k=1}^{K} m_k(m_k-1)M^2} \sum_{j \in \mathcal{S}_k} (m_k w_j \bar{y}_j - \bar{y}_k)^2,
\]

where \( m_k \) is the size of \( \mathcal{S}_k \), \( \bar{y}_k = \sum_{i \in \mathcal{R}_k} (1 + d_{i}^{(k)}) w_i y_i, \), \( d_{i}^{(k)} = \sum_{j \in \mathcal{R}_k} w_{ij} d_{ij}, d_{ij} = 1 \) if \( i \) is the nearest neighbor of \( j \) and \( d_{ij} = 0 \) otherwise, \( \bar{y}_j = y_j + \frac{g_{jk}(y_j - \bar{y}_{jk})}{\sqrt{6(d_{jk})^2 + 6d_{jk}^2} + 4 - 2/3d_{jk}^2} \) if \( j \in \mathcal{R}_k \) and \( \bar{y}_j = \text{the imputed value of } y_j \) if \( j \notin \mathcal{R}_k \), \( g_{jk}(k) = [\sqrt{6(d_{jk})^2 + 6d_{jk}^2} + 4 - 2/3d_{jk}^2] \) (if \( d_{jk}^2 = 0 \), and \( j_{k1} \) and \( j_{k2} \) are the two nearest neighbors of \( j \) in \( \mathcal{R}_k \). It is shown in Chen and Shao (2001) that \( v_n/V_n \to_p 1 \), where

\[
V_n = E \left[ \sum_{i \in \mathcal{R}} (1 + d_{i})^2 w_i^2 \text{Var}(y_i|x_i) \right] + \text{Var} \left[ \sum_{i \in \mathcal{S}} \bar{w}_i \mu(x_i) \right].
\]

For the purpose of showing that confidence interval \( \hat{\mu} - z_\alpha \sqrt{v_n}, \hat{\mu} + z_\alpha \sqrt{v_n} \) is asymptotically valid for \( \mu \), we need to show that \( n v_n/\sigma^2 \to_p 1 \), because result (11) in Theorem 1 shows that \( \sigma^2/n \) is the variance of the limiting distribution of \( \hat{\mu} - \mu \). Since \( v_n/V_n \to_p 1 \), this can be achieved by showing \( n V_n/\sigma^2 \to 1 \), which is the first part of the following theorem. The following theorem also shows that \( n \text{Var}(\hat{\mu})/\sigma^2 \to 1 \) and, hence, \( v_n \) is also consistent for \( n \text{Var}(\hat{\mu}) \).

**Theorem 2.** Assume the conditions in Theorem 1. Then,

(i) \( n v_n/\sigma^2 \to_p 1 \),

(ii) \( v_n/\text{Var}(\hat{\mu}) \to_p 1 \), and

(iii) \( P(\hat{\mu} - z_\alpha \sqrt{v_n} \leq \mu \leq \hat{\mu} + z_\alpha \sqrt{v_n}) \to 1 - \alpha \).

**Proof.** Without loss of generality, we assume \( K = 1 \) and drop the subscript \( k \). Write \( \hat{\mu} - \mu = H_1 + H_2 + H_3 + H_4 \), where \( H_1 = \sum_{i \in \mathcal{R}} \bar{w}_i (1 + d_{i}) [y_i - \mu(x_i)], H_2 = \sum_{i \in \mathcal{R}} \bar{w}_i [\mu(x_i) - \mu_1], \)
\( H_3 = \sum_{i \in \mathcal{R}} \bar{w}_i d_i \mu(x_i) - \sum_{j \in \mathcal{R}} \bar{w}_j \mu_0 = \sum_{j \in \mathcal{R}} \bar{w}_j (\tau_j - \mu_0), \tau_j = \sum_{i \in \mathcal{R}} d_{ij} \mu(x_i) \) if \( j \in \mathcal{R} \) and \( \tau_j = 0 \) if \( j \notin \mathcal{R} \), and \( H_4 = \sum_{i \in \mathcal{R}} \bar{w}_i \mu_1 + \sum_{j \in \mathcal{R}} \bar{w}_j \mu_0 - \mu \). Since the mean of \( y_i - \mu(x_i) \) conditional on all \( x_i \)'s, \( \mathcal{R} \), and \( S \) is 0,

\[
EH_1^4 = E \left\{ \sum_{i \in \mathcal{R}} \bar{w}_i^4 (1 + d_i)^4 | y_i - \mu(x_i) |^4 \right\}
+ \sum_{i,j \in \mathcal{R}, i \neq j} \bar{w}_i^2 \bar{w}_j^2 (1 + d_i)^2 (1 + d_j)^2 | y_i - \mu(x_i) |^2 | y_j - \mu(x_j) |^2
\]
\[
= O(n) \left\{ \sum_{i \in \mathcal{R}} \bar{w}_i^4 (1 + d_i)^4 | y_i - \mu(x_i) |^4 \right\}.
\]

It then follows from condition (4) and \( E[\mu(x_i)]^4 \leq E \bar{y}_i^4 < \infty \) that \( EH_2^4 = O(n^{-2}) \). Similarly, the mean of \( a_i [\mu(x_i) - \mu] \) conditional on \( S \) is 0, \( EH_2^4 = O(n^{-2}) \). Since

\[
H_4 = \left( \sum_{i \in \mathcal{S}} \bar{w}_i a_i - p \right) \mu_1 + \left( \sum_{i \in \mathcal{S}} \bar{w}_i (1 - a_i) - (1 - p) \right) \mu_0,
\]
we can similarly show that \( EH_4^4 = O(n^{-2}) \). By the definition of \( d_{ij} \) and \( q_i \),

\[
E \left( \sum_{j \in \mathcal{R}} \bar{w}_j \tau_j \right) = E \left( \sum_{j \in \mathcal{R}} \bar{w}_j \sum_{i \in \mathcal{R}} q_i \mu(x_i) \right).
\]

Hence,

\[
3^{-3}EH_3^4 \leq E \left[ \sum_{j \in \mathcal{R}} \bar{w}_j \tau_j - E \left( \sum_{j \in \mathcal{R}} \bar{w}_j \tau_j \right) \right]^4 + \mu_0^4 \left( \sum_{j \in \mathcal{R}} \bar{w}_j - \sum_{j \in \mathcal{S}} \bar{w}_j \right)^4
+ \left\{ E \left( \sum_{j \in \mathcal{R}} \bar{w}_j \left( \sum_{i \in \mathcal{R}} q_i \mu(x_i) - \mu_0 \right) \right) \right\}^4.
\]

Using the same argument as before, we can show that the first two terms on the right hand side of the previous equation are \( O(n^{-2}) \). By condition (4),

\[
E \left( \sum_{j \in \mathcal{R}} \bar{w}_j \left( \sum_{i \in \mathcal{R}} q_i \mu(x_i) - \mu_0 \right) \right) \leq O(1) \left| \sum_{i \in \mathcal{R}} q_i \mu(x_i) - \mu_0 \right|,
\]
which is \( o(n^{-1/2}) \) (Chen and Shao 2000). This shows that \( EH_4^4 = O(n^{-2}) \).

From the proof of Theorem 1, \( \sqrt{n} \sum_{i \in \mathcal{S}} \bar{w}_i \mu(x_i) - \mu \) and \( \sqrt{n} \sum_{i \in \mathcal{R}} \bar{w}_i (1 + d_i) \mu(x_i) - \mu \) = \( \sqrt{n} (H_2 + H_3 + H_4) \) have the same limiting distribution. From the previous proof, \( E(H_2 + H_3 + H_4)^4 = O(n^{-2}) \). Hence, \( \{n(H_2 + H_3 + H_4)^2\} \) is uniformly integrable and, thus,

\[
\text{Var} \left[ \sum_{i \in \mathcal{R}} \bar{w}_i (1 + d_i) \mu(x_i) \right] / \text{Var} \left[ \sum_{i \in \mathcal{S}} \bar{w}_i \mu(x_i) \right] \rightarrow 1.
\]

13
From (24) and the fact that
\[
\text{Var}(\hat{\mu}) = E \left[ \sum_{i \in R} (1 + d_i)^2 \bar{w}_i^2 \text{Var}(y_i | x_i) \right] + \text{Var} \left[ \sum_{i \in R} \bar{w}_i (1 + d_i) \mu(x_i) \right],
\]
\text{Var}(\hat{\mu})/V_n \rightarrow 1. \text{ Then, result (ii) follows from } v_n/V_n \rightarrow_p 1. \text{ Since } EH_i^4 = O(n^{-2}), l = 1, \ldots, 4, \{n(\hat{\mu} - \mu)\} \text{ is uniformly integrable, which and Theorem 1 imply } n\text{Var}(\hat{\mu})/\sigma^2 \rightarrow 1. \text{ Hence, result (i) holds. Finally, result (iii) follows from result (i) and Theorem 1.}

\section{4 Confidence Intervals for Quantiles}

Population quantiles are typically estimated by sample quantiles (Rao, Kovar and Mantel, 1990; Francisco and Fuller, 1991). In what follows we use \( F \) and \( f \) to denote the distribution and density of \( y \) with respect to the superpopulation. Note that \( f_{k,a} \) or \( f_a \) is used in the previous section for the density of \( x \) given \( a \).

Let \( I_y(t) = 1 \) if \( y \leq t \) and \( I_y(t) = 0 \) if \( y > t \), and let
\[
\hat{F}(t) = \frac{1}{M} \sum_{k} \left( \sum_{i \in R_k} w_i I_{y_i}(t) + \sum_{i \in R_k} w_i I_{\bar{y}_i}(t) \right)
\]
(25)
be the empirical distribution, a survey estimator of \( F(t) = P(y \leq t) \). Even if \( M \) is known, \( \hat{M} \) in (25) cannot be replaced by \( M \) unless \( \hat{M} = M \), since \( \hat{F}(\infty) \) has to be 1. If sampling is stratified simple random sampling, then \( \hat{M} = M \). Otherwise, \( \hat{M} \) and \( M \) may be different. For any fixed \( q \in (0, 1) \), the \( q \)th sample quantile is \( \hat{\theta} = \hat{F}^{-1}(q) = \inf \{ t : \hat{F}(t) \geq q \} \), a survey of the population quantile \( \theta = F^{-1}(q) \).

Replacing \( y_i \) by \( I_{y_i}(t) \) in Theorem 1, we obtain that, for any fixed \( t \), \( \sqrt{n}[\hat{F}(t) - F(t)] \) is asymptotically normal with mean 0. The following is a Bahadur-type representation for \( \hat{\theta} \), a result similar to that in Francisco and Fuller (1991). This result together with the asymptotic normality of \( \sqrt{n}[\hat{F}(t) - F(t)] \) imply that \( \sqrt{n}(\hat{\theta} - \theta) \) is asymptotically normal.

**Theorem 3.** Assume the conditions in Theorem 1 with \( y_i \) replaced by \( I_{y_i}(t) \) for any \( t \). Assume further that \( F \) is differentiable at \( \theta \) and \( F'(\theta) = f(\theta) > 0 \). Then,
\[
\hat{\theta} - \theta = \frac{F'(\theta) - \hat{F}'(\theta)}{f(\theta)} + o_p(n^{-1/2}).
\]
Proof. Without loss of generality, assume that $K = 1$. Since $\hat{F}$ is a ratio estimator and we can apply the delta-method, we may also assume $\hat{M} = \sum_{i \in S} w_i = M$. For any real $t$, let $\theta_{nt} = \theta + tn^{-1/2}$, $Z_{nt} = \sqrt{n}[F(\theta_{nt}) - \hat{F}(\theta)]/f(\theta)$, and $U_{nt} = \sqrt{n}[F(\theta_{nt}) - \hat{F}(\hat{\theta})]/f(\theta)$. Under condition (4), $|F(\theta) - \hat{F}(\hat{\theta})| = O(n^{-1})$. Hence,

$$U_{nt} = \sqrt{n}[F(\theta_{nt}) - F(\theta)]/f(\theta) + O(n^{-1/2}) \rightarrow t.$$

From Chen and Shao (2000), $E(Z_{nt} - Z_{n0}) = o(1)$. Also,

$$\text{Var}(Z_{nt} - Z_{n0}) = n\text{Var}[\hat{F}(\theta_{nt}) - \hat{F}(\theta)]$$

$$\leq O(1)|E[\hat{F}(\theta_{nt}) - \hat{F}(\theta)]|$$

$$= O(1)|F(\theta_{nt}) - F(\theta) + o(n^{-1/2})|$$

$$= o(1).$$

Hence, $Z_{nt} - Z_{n0} = o_p(1)$. Then, for any $t$ and $\epsilon > 0$,

$$P(\sqrt{n}(\hat{\theta} - \theta) \leq t, Z_{n0} \geq t + \epsilon) = P(Z_{nt} \leq U_{nt}, Z_{n0} \geq t + \epsilon)$$

$$\leq P(|Z_{nt} - Z_{n0}| \geq \epsilon/2) + P(|U_{nt} - t| \geq \epsilon/2)$$

$$= o(1).$$

Similarly, $P(\sqrt{n}(\hat{\theta} - \theta) \geq t + \epsilon, Z_{n0} \leq t) = o(1)$, Hence, $\sqrt{n}(\hat{\theta} - \theta) = Z_{n0} + o_p(1)$, which is the desired result.

It follows from Theorems 1–3 and the delta-method that the asymptotic variance of $\hat{\theta} - \theta$ is $V_n(\theta)/f^2(\theta)$, where, for any fixed $t$,

$$V_n(t) = F^2(t)\text{Var}(\hat{M}/M) - 2F(t)\text{Cov}(\hat{M}/M, \hat{G}(t)) + \text{Var}(\hat{G}(t)),$$  \hspace{1cm} (26)

and $\hat{G}(t)$ is defined by (25) with $\hat{M}$ replaced by $M$. In the case where $\hat{M} = M$, $V_n(t) = \text{Var}(\hat{G}(t)) = \text{Var}(\hat{F}(t))$.

A consistent estimator of $V_n(\theta)/f^2(\theta)$ and a confidence interval for $\theta$ satisfying (2) can be constructed in two steps. First, we construct a consistent estimator of $V_n(\theta)$. Second, we use the Bahadur representation.

For any fixed $t$, let $\bar{I}_k(t) = \sum_{i \in R_k} w_i (1 + d_{j}^{(k)})I_{y_j}(t)$, $\xi_j(t) = \bar{I}_{\bar{y}_j}(t)$ if $j \in R_k$ and $\xi_j(t) = I_{y_j}(t) + d_{j}^{(k)}g_{j}^{(k)}\{I_{y_j}(t) - \frac{1}{2}[I_{y_{jk_1}}(t) + I_{y_{jk_2}}(t)]\}$ if $j \in R_k$, where $d_{j}^{(k)}, g_{j}^{(k)}, y_{j_1}$ and $y_{j_2}$ are the
same as those in (23). Define

$$
\hat{V}_n(t) = \hat{F}_n(t) \sum_{k=1}^K \frac{1}{m_k(m_k-1)M^2} \sum_{j \in S_k} (m_kw_j - \hat{M}_k)^2 
- 2\hat{F}(t) \sum_{k=1}^K \frac{1}{m_k(m_k-1)M^2} \sum_{j \in S_k} [m_kw_j \xi_j(t) - \bar{I}_k(t)] (m_kw_j - \hat{M}_k) 
+ \sum_{k=1}^K \frac{1}{m_k(m_k-1)M^2} \sum_{j \in S_k} [m_kw_j \xi_j(t) - \bar{I}_k(t)]^2,
$$

where $\hat{M}_k = \sum_{i \in S_k} w_i$. From the proof of Theorem 2, $\hat{V}_n(\theta)/V_n(\theta) \to_p 1$. However, $\hat{V}_n(\theta)$ is not an estimator since $\theta$ is unknown. The following lemma, whose proof is given in the Appendix, shows that $\hat{V}_n(\theta)$ is consistent for $V_n(\theta)$.

**Lemma 2.** Assume the conditions in Theorem 1 with $y_i$ replaced by $I_{y_i}(t)$ for any $t$. Assume further that $F$ is continuous at $\theta$. Then $\hat{V}_n(\theta)/V_n(\theta) \to_p 1$.

From the Bahadur representation, we propose the following Woodruff's confidence interval for $\theta$:

$$CI = [\hat{F}^{-1}(q - z_\alpha s_n), \hat{F}^{-1}(q + z_\alpha s_n)],$$

where $s_n = [\hat{V}_n(\theta)]^{1/2}$, and the following variance estimator for $\theta$:

$$v_n = [\hat{F}^{-1}(q + z_\alpha s_n) - \hat{F}^{-1}(q - z_\alpha s_n)]^2/(4z_\alpha^2).$$

The next theorem establishes the asymptotic validity of $CI$ and $v_n$.

**Theorem 4.** Assume the conditions in Theorem 3. Assume further that $F$ is differentiable in a neighborhood of $\theta$ and $f = F'$ is continuous at $\theta$. Then,

(i) $v_n/[V_n(\theta)/f^2(\theta)] \to_p 1$ and

(ii) $P(\theta \in CI) \to 1 - \alpha$.

**Proof.** Let $S_n(\theta) = [V_n(\theta)]^{1/2}$. Without loss of generality, assume that $\sqrt{n}S_n(\theta)$ has a nonzero limit. Define $c_n = cS_n(\theta)$ for a constant $c \neq 0$ and $\theta_n = F^{-1}(q + c_n)$. From the proof of Theorem 3, we have the following Bahadur representation:

$$\hat{F}^{-1}(q + c_n) - F^{-1}(q + c_n) = \frac{F(\theta_n) - \hat{F}(\theta_n)}{f(\theta_n)} + o_p(n^{-1/2}).$$

Under the condition on $F$, $F^{-1}(q + c_n) = \theta + c_n/f(\theta) + o(n^{-1/2})$ and $f(\theta_n) = f(\theta) + o(1)$. 

16
From the proof of Theorem 3, \( F(\theta_n) - \hat{F}(\theta_n) = F(\theta) - \hat{F}(\theta) + o_p(n^{-1/2}) \). Hence,

\[
\hat{F}^{-1}(q + c_n) = \hat{\theta} + \frac{c_n}{f(\hat{\theta})} + \frac{F(\theta) - \hat{F}(\theta)}{f(\theta)} + o_p(n^{-1/2}).
\]

This result and Theorem 3 imply that

\[
\hat{F}^{-1}(q + cS_n(\theta)) = \hat{\theta} + \frac{cS_n(\theta)}{f(\theta)} + o_p(n^{-1/2}).
\] (29)

From Lemma 2, \( s_n / S_n(\theta) \to_p 1 \). For any fixed \( \epsilon > 0 \), \( P((1 - \epsilon)S_n(\theta) \leq s_n \leq (1 + \epsilon)S_n(\theta)) \to 1 \) and, therefore, we may assume \( (1 - \epsilon)S_n(\theta) \leq s_n \leq (1 + \epsilon)S_n(\theta) \) in the rest of the proof.

Since \( \hat{F}^{-1} \) is a monotone function,

\[
\hat{F}^{-1}(q + (1 - \epsilon)z_\alpha S_n(\theta)) \leq \hat{F}^{-1}(q + z_\alpha s_n) \leq \hat{F}^{-1}(q + (1 + \epsilon)z_\alpha S_n(\theta)).
\]

This and result (29) imply that

\[
\frac{(1 - \epsilon)z_\alpha S_n(\theta)}{f(\theta)} + o_p(n^{-1/2}) \leq \hat{F}^{-1}(q + z_\alpha s_n) - \hat{\theta} \leq \frac{(1 + \epsilon)z_\alpha S_n(\theta)}{f(\theta)} + o_p(n^{-1/2}).
\]

Taking the limit and using the fact that \( \epsilon \) is arbitrary, we obtain that

\[
\hat{F}^{-1}(q + z_\alpha s_n) = \hat{\theta} + \frac{z_\alpha S_n(\theta)}{f(\theta)} + o_p(n^{-1/2}).
\]

Similarly, the same result holds if \( z_\alpha \) is replaced by \( -z_\alpha \). Thus,

\[
\hat{F}^{-1}(q + z_\alpha s_n) - \hat{F}^{-1}(q - z_\alpha s_n) = \frac{2z_\alpha S_n(\theta)}{f(\theta)} + o_p(n^{-1/2}),
\]

i.e., result (i) follows. Result (ii) also follows, since

\[
P(\theta \in CI) = P\left( \hat{F}^{-1}(q - z_\alpha s_n) \leq \theta \leq \hat{F}^{-1}(q + z_\alpha s_n) \right)
= P\left( \hat{\theta} - z_\alpha S_n(\theta) / f(\theta) + o_p(n^{-1/2}) \leq \theta \leq \hat{\theta} + z_\alpha S_n(\theta) / f(\theta) + o_p(n^{-1/2}) \right)
= P\left( -z_\alpha + o_p(1) \leq \frac{\hat{\theta} - \theta}{S_n(\theta) / f(\theta)} \leq z_\alpha + o_p(1) \right)
\rightarrow \Phi(z_\alpha) - \Phi(-z_\alpha) = 1 - \alpha.
\]

5 Simulation Results

A simulation study was performed to examine the finite sample performance of the proposed variance estimators and confidence intervals. Stratified simple random samples were generated from a population that matches the main characteristics of an aggregated dataset from
the 1998 Financial Farm Survey (FFS) published by Statistics Canada (Rancourt 1999). The FFS is a biannual survey collecting information on agriculture operations in Canada. The survey collects information on revenues, expenses, assets, investments, and liabilities for the reference year. Nonrespondents in the survey are imputed by NNI for some variables (Rancourt 1999). We focus on dairy farms and two variables, the net assets \((x)\) and the cash income \((y)\). Strata in the FFS are constructed using the size of farm and province (5 provinces and ALT, a group of small provinces, and 3 size classes in each province). These 18 strata are also used as imputation classes and, hence, imputation does not cut across strata. Information about population size, sample size, number of respondents, mean and standard deviation of \(x\) and \(y\), and the correlation coefficient between \(x\) and \(y\) in each stratum are listed in Table 1. The overall sampling fraction \(n/M\) is about 7%. Note that the stratum mean and standard deviations in Table 1 are the same as those in Chen and Shao (2001), but the correlation coefficients between \(x\) and \(y\) are changed in order to study whether NNI is more efficient than the random hot deck that does not use covariate information.

For each pair \((x, y)\), a \(y\)-respondent is generated according to the response probability function

\[
P(a = 1|x) = \frac{\exp(\gamma_1 + \gamma_2(x - \mu_x)\sigma_x^{-1})}{1 + \exp(\gamma_1 + \gamma_2(x - \mu_x)\sigma_x^{-1})}
\]

with some \(\gamma_1\) and \(\gamma_2\). For each pair \((\gamma_1, \gamma_2)\), we define a model as follows:

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma_1)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>(\infty)</td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(p)</td>
<td>0.607</td>
<td>0.610</td>
<td>0.628</td>
<td>0.700</td>
<td>0.703</td>
<td>0.735</td>
<td>0.846</td>
<td>0.848</td>
<td>0.883</td>
<td>1.000</td>
</tr>
</tbody>
</table>

where \(p = E[P(a = 1|x)]\) is the average response probability and Model 10 corresponds to the case of no nonresponse.

We considered the estimation of 6 different parameters of the distribution of \(y\), the mean, the median, the 10th, 25th, 75th, and 95th percentiles. In addition to the NNI, we considered the linear regression imputation (LRI) that assumes a linear model between \(y\) and \(x\), the random hot deck imputation (RHD), and the Bayesian bootstrap multiple imputation given by Rubin (1987) with 10 imputations (MI10).
Since \( P(a = 1|x) \) depends on \( x \) when \( \gamma_2 \neq 0 \) (models 1, 2, 4, 5, 7, and 8), the RHD and MI10 are biased. On the other hand, when \( \gamma_2 = 0 \) (models 3, 6, and 9), \( P(a = 1|x) \) is a constant and the RHD and MI10 are unbiased. The LRI is biased for percentile estimation. For the estimation of the mean, the LRI is biased when \( P(a = 1|x) \) depends on \( x \), because the relationship between \( y \) and \( x \) is nonlinear. When \( P(a = 1|x) \) is a constant (models 3, 6, and 9), the LRI is unbiased even though the relationship between \( y \) and \( x \) is not linear.

Table 2 provides empirical results based on 2,000 simulations in terms of the relative bias of the point estimator, the variance of the estimator, the relative bias (RB) of the variance estimator (given in (23) and (28) for NNI), the coverage probability of the confidence interval of the form \( \hat{\mu} \pm z_\alpha \sqrt{\hat{v}_n} \) for the case of sample mean and the Woodruff’s confidence interval given in (27) for the case of sample quantiles (\( 1 - \alpha \) is chosen to be 95%), and the average length of the confidence interval.

The following is a summary for the results in Table 2.

1. The relative bias. For the NNI estimator, its relative bias is smaller than 1% in absolute value in all cases (it is actually not larger than 0.5% in absolute value for most cases). The relative bias of the RHD, MI10, and LRI estimators is smaller than 0.5% in absolute value when \( P(a = 1|x) \) is constant (models 3, 6, and 9), but it is not negligible in cases where \( P(a = 1|x) \) depends on \( x \) (models 1, 2, 4, 5, 7, and 8). Although the relative bias in some cases is small (e.g., it is 0.8% for LRI in the estimation of the mean under model 7), it still leads to a low coverage probability of the associated confidence interval.

2. The variance. Under models 3, 6, and 9, the LRI, RHD, and MI10 are unbiased but the NNI is more efficient in terms of the variance because the RHD and MI10 do not use the covariate information and the LRI assumes a linear relationship between \( y \) and \( x \); the variance of the MI10 estimator is smaller than that of the RHD estimator for the estimation of the mean, but it is larger for the estimation of percentiles. In situations where the LRI, RHD, and MI10 are biased, their variances are sometimes smaller than that of the NNI, but having a small variance is not necessarily an advantage when the point estimator is biased.

3. Variance estimators. The proposed variance estimator for the NNI performs well in
the estimation of mean and median. For the estimation of other percentiles, it overestimates, especially for the estimation of extreme percentiles.

4. The performance of the confidence interval. For the NNI, the coverage probability of confidence interval is close to the nominal level of 95% for mean estimation and is between 91% and 94% for percentile estimation. For the mean estimation, the coverage probability of the confidence intervals associated with the LRI, RHD, and MI10 is comparable to that of the NNI under models 3, 6, and 9, but can be much lower than the nominal level of 95% when \( P(a = 1|x) \) depends on \( x \). For the percentile estimation, only the RHD has a comparable coverage probability with the NNI under models 3, 6, and 9; the coverage probability for the LRI is clearly low due to its bias; the coverage probability for the MI10 is much lower than that for the RHD.

The conclusion is that the empirical results are consistent with our theoretical findings. In the cases where the LRI, RHD, and MI10 are unbiased, the NNI is better than the RHD and MI10 when useful covariate information is used; the NNI is better than the LRI when the relationship between \( y \) and \( x \) is not linear. When \( P(a = 1|x) \) depends on \( x \), the NNI is still unbiased but LRI, RHD and MI10 may not.

6 Discussion

The results in Sections 3 are for the case where \( \mu \) is the parameter of interest. If the parameter of interest is the finite population mean \( \hat{Y} \) instead of \( \mu \), then we first need an asymptotic distribution for \( \sqrt{n}(\hat{\mu} - \hat{Y}) \). Note that \( \hat{Y} = \frac{1}{M} \sum_k \left( \sum_{i \in R_k} y_i + \sum_{i \notin R_k, i \in P_k} y_i \right) \) and \( \{y_i : i \in R_k\} \) and \( \{y_i : i \notin R_k, i \in P_k\} \) are independent. Hence, it follows from the central limit theorem and our Theorem 1 that \( \sqrt{n}(\hat{\mu} - \hat{Y}) = \sqrt{n}(\hat{\mu} - \mu) + \sqrt{n}(\mu - \hat{Y}) \) is asymptotically normal with mean 0 but with a variance that may be different from the one in result (11). Since \( \mu - \hat{Y} = O_p(M^{-1/2}) \), the limiting variances of \( \sqrt{n}(\hat{\mu} - \hat{Y}) \) and \( \sqrt{n}(\hat{\mu} - \mu) \) are the same if \( n/M \to 0 \). Hence, if \( n/M \to 0 \), the variance estimator \( \nu_n \) in (23) is still consistent and the confidence interval for \( \hat{Y} \) of the form \( \hat{\mu} \pm z_\alpha \sqrt{\nu_n} \) satisfies (2). A similar discussion applies to the estimation of quantiles. If \( n/M \) is not negligible, then our variance estimators may not be consistent and an extra effort is needed to derive consistent variance estimators.
The main difficulty for extending our results to cluster sampling or multistage sampling is that the independence of \((x_i, y_i, a_i)\)'s (Assumption A) does not hold, since values within a cluster are typically dependent. In a few steps in our proof, we apply some result for order statistics of i.i.d. \(x_i\)'s, which is not available for general dependent \(x_i\)'s. If the cluster sizes in cluster sampling (or the first stage cluster sizes in multistage sampling) are bounded by a fixed integer, then our proof can be modified to establish similar asymptotic results. For general cases, however, further research is needed.

NNI can be applied when the covariate \(x\) is multivariate, but the study of its properties faces the curse of dimensionality, a problem for many other nonparametric imputation methods. Because of the curse of dimensionality, the NNI with a multivariate \(x\) may not be efficient. As an alternative, we are working on how to find a linear combination of the multivariate \(x\) to define neighbors.

**Appendix: Proofs of Lemmas**

**Proof of Lemma 1.** (i) Let \(F_j\) be the cumulative distribution with respect to \(f_j\), \(j = 0, 1\), and let \(z_i, i = 1, \ldots, m\), be i.i.d. random variables from \(F_0\), \(t = (t_1, \ldots, t_m)\), \(f_0(t) = f_0(t_1) \cdots f_0(t_m)\), \(s_t = (t_i - x_i)\), \(s = (s_1, \ldots, s_m)\), and \(1 = (1, \ldots, 1)\). Then

\[
r^m E(q^m_i | x_i) = r^m E\left[ P^m(|z_i - x_i| \leq |z_j - x_j| \forall j \neq i | \mathcal{X}) | x_i \right]
\]

\[
= r^m E\left[ P(|z_i - x_i| \leq |z_l - x_j| \forall j \neq i \text{ and } 1 \leq l \leq m | \mathcal{X}) | x_i \right]
\]

\[
= r^m P\left(|z_i - x_i| \leq |z_l - x_j| \forall j \neq i \text{ and } 1 \leq l \leq m | x_i \right)
\]

\[
= r^m \int P^{l-1}\left(|t_i - x_i| \leq |t_l - x_j| \forall 1 \leq l \leq m | x_i \right) f_0(t) dt
\]

\[
= \int P^{l-1}\left(|s_l|/r \leq |s_l/r + x_i - x_j| \forall 1 \leq l \leq m | x_i \right) f_0(s/r + x_i) ds
\]

\[
= \int P^{l-1}\left(x_j \notin B(x_i + s_l/r, |s_l/r|) \forall 1 \leq l \leq m | x_i \right) f_0(s/r + x_i) ds
\]

\[
= \int \left[ 1 - P\left(x_j \in \bigcup_{l=1}^m B(x_i + s_l/r, |s_l/r|) | x_i \right) \right]^{l-1} f_0(s/r + x_i) ds
\]

\[
= \int \left[ 1 - P\left(x_j \in (x_i + 2\alpha/r, x_i + 2\beta/r) | x_i \right) \right]^{l-1} f_0(s/r + x_i) ds
\]

\[
= \int \left[ 1 - \left(F_1(x_i + 2\beta/r) - F_1(x_i + 2\alpha/r) \right) \right]^{l-1} f_0(s/r + x_i) ds,
\]
where \( \alpha = \min \{0, s_1, \ldots, s_m\} \), \( \beta = \max \{0, s_1, \ldots, s_m\} \), and \( B(x, |r|) \) denotes the interval \((x - |r|, x + |r|)\). For any fixed \( x_i \), \( f_1(x_i) > 0 \). By the continuity of \( f_1 \), there exists \( \delta > 0 \) such that \( c_{x_i} = \min_{x_i - 2\delta \leq x \leq x_i + 2\delta} f_1(x) \) > 0. When \( \beta r^{-1} > \delta \) and \( |\alpha| r^{-1} < \delta \), \( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \geq 2(\beta - \alpha)r^{-1}c_{x_i} \). When \( \beta r^{-1} \geq \delta \), \( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \geq F_1(x_i + 2\beta r^{-1}) - F_1(x_i) \geq F_1(x_i + 2\delta) - F(x_i) \geq 2\delta c_{x_i} \). Similarly, when \( |\alpha| r^{-1} \geq \delta \), \( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \geq 2\delta c_{x_i} \). Note that \( f_0 \leq c_f \) for some constant \( c_f > 0 \). Hence, when \( r \geq 2 \), the integrand in (30) is bounded by

\[
\exp \left\{ (r - 1) \log \left[ 1 - \left( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \right) \right] \right\} f_0(sr^{-1} + x_i 1) \\
\leq \exp \left\{ -(r - 1) \left( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \right) \right\} f_0(sr^{-1} + x_i 1) \\
\leq c_f^m \exp \left\{ - (\beta - \alpha) c_{x_i} \right\} + 2 \exp \left\{ - 2(r - 1) \delta c_{x_i} \right\} f_0(sr^{-1} + x_i 1).
\]

Note that

\[
\int \exp \left\{ - 2(r - 1) \delta c_{x_i} \right\} f_0(sr^{-1} + x_i 1) ds = r^m \exp \left\{ - 2(r - 1) \delta c_{x_i} \right\} \int \tilde{f}_0(t) dt \to 0.
\]

For any \( c > 0 \),

\[
\int \exp \left\{ - c(\beta - \alpha) \right\} ds = \sum_{i=0}^{m} \binom{m}{i} \int_{s_1 > 0, \ldots, s_i > 0} e^{-c \max\{s_{i+1}, \ldots, s_m\}} ds_1 \cdots ds_i \\
= \int_{s_{i+1} < 0, \ldots, s_m < 0} e^{-c \max\{s_{i+1}, \ldots, s_m\}} ds_{i+1} \cdots ds_m \\
= c^{-m} \sum_{i=0}^{m} \binom{m}{i} i!(m-i)! \\
= \frac{(m+1)!}{c^m}.
\]

Hence, \( \int \exp \left\{ - (\beta - \alpha) c_{x_i} \right\} ds < \infty \). By the dominated convergence theorem,

\[
\lim \left[ r^m E(q_i^m | x_i) \right] \\
= \int \lim \left\{ \left[ 1 - \left( F_1(x_i + 2\beta r^{-1}) - F_1(x_i + 2\alpha r^{-1}) \right) \right]^{r^{-1}} \tilde{f}_0(sr^{-1} + x_i 1) \right\} ds \\
= \int \exp \left\{ - 2f_1(x_i)(\beta - \alpha) \right\} f_0^m(x_i) ds \\
= \frac{(m+1)!}{2^m} \frac{f_0^m(x_i)}{f_1^m(x_i)}.
\]

Note that

\[
E \left[ r^{m-1} \sum_{i \in \mathcal{R}} q_i^m g(x_i) \right] = E[r^m q_i^m g(x_i)] = E[r^m g(x_i) E(q_i^m | x_i)].
\]

22
To show that (8) holds, it suffices to show that $r^m g(x_i) E(q_i^m | x_i)$ is uniformly integrable, which is implied by $E[r^m g(x_i) E(q_i^m | x_i)]^{6/5} = O(1)$. From Hölder’s inequality,

$$E[r^m g(x_i) E(q_i^m | x_i)]^{6/5} \leq \left\{ E[r^m E(q_i^m | x_i)]^3 \right\}^{2/5} \left\{ E[g(x_i)]^2 \right\}^{3/5} \leq \left\{ E(r^{3m} q_i^{3m}) \right\}^{2/5} \left\{ E[g(x_i)]^2 \right\}^{3/5}.$$  

Hence, it suffices to show that

$$E(r^{3m} q_i^{3m}) = E \left( r^{3m-1} \sum_{i \in \mathcal{R}} q_i^{3m} \right) = O(1). \quad (31)$$

Let $u_i, i \in \mathcal{R},$ be i.i.d. random variables having the uniform distribution on $[0, 1]$. Since $F_i^{-1}(u_i), i \in \mathcal{R},$ has the same distribution as $x_i, i \in \mathcal{R},$ and the quantities in (31) are expectations, without loss of generality we may assume that $x_{(i)} = F_i^{-1}(u_{(i)})$, where $x_{(i)}$ is the $i$th ordered value of $x_i$’s and $u_{(i)}$ is the $i$th ordered value of $u_i$’s. Let $x_{(0)} = x_-$, $x_{(r+1)} = x_+$, and $q_{(i)}$ be the $q$-value corresponding to $x_{(i)}$. Note that

$$q_{(i)} = F_0 \left( \frac{x_{(i+1)} + x_{(i)}}{2} \right) - F_0 \left( \frac{x_{(i)} + x_{(i-1)}}{2} \right) \leq F_0 \left( x_{(i+1)} \right) - F_0 \left( x_{(i-1)} \right) = F_0 \left( x_{(i+1)} - F_0 \left( x_{(i)} \right) + F_0 \left( x_{(i)} \right) - F_0 \left( x_{(i-1)} \right).$$

Hence, (31) follows from

$$E \left\{ r^{3m-1} \sum_{i=1}^{r} \left[ F_0 \left( x_{(i+1)} \right) - F_0 \left( x_{(i)} \right) \right]^{3m} \right\} = O(1). \quad (32)$$

Let $G(t) = F_0(F_i^{-1}(t))$. Then $G'(t) = f_0(F_i^{-1}(t))/f_1(F_i^{-1}(t))$. By the mean-value theorem, for each $i$, there exists $\zeta_i \in (x_{(i)}, x_{(i+1)})$ such that

$$F_0 \left( x_{(i+1)} \right) - F_0 \left( x_{(i)} \right) = F_0 \left( F_i^{-1}(u_{(i+1)}) \right) - F_0 \left( F_i^{-1}(u_{(i)}) \right) = \frac{f_0(\zeta_i)}{f_1(\zeta_i)} (u_{(i+1)} - u_{(i)}).$$

Under Assumption B, $f_0(x)/f_1(x) \leq c$ for $x \in [x_-, x_+]$ and a constant $c > 0$. Hence,

$$E \left[ F_0 \left( F_i^{-1}(u_{(i+1)}) \right) - F_0 \left( F_i^{-1}(u_{(i)}) \right) \right]^{3m} \leq c^{3m} E(u_{(i+1)} - u_{(i)})^{3m}.$$

Then, result (32) follows from

$$E \left[ r^{3m-1} \sum_{i=1}^{r} (u_{(i+1)} - u_{(i)})^{3m} \right] = O(1). \quad (33)$$
Note that \( u_{(i+1)} - u_{(i)} \) has the beta distribution \( \text{Beta}(1, r) \) (Arnold, Balakrishnan, and Nagaraja, 1992, p. 32). Therefore, the left hand side of (33) is equal to

\[
E \left\{ r^{3m+1} \int_0^1 t^{3m} (1 - t)^{r-1} \, dt \right\} = E \left\{ \frac{r^{3m+1} \Gamma(3m + 1) \Gamma(r)}{\Gamma(3m + 1 + r)} \right\} \\
= E \left\{ \frac{\Gamma(3m + 1) r^{3m+1}}{(3m + r)(3m - 1 + r) \cdots r} \right\} \\
\leq \Gamma(3m + 1).
\]

This proves (33) and, thus, (8).

(ii) Similarly,

\[
r^{2m} E(q_i^m q_j^m | x_i, x_j) \\
= \int \int [1 - P(x_i \in (x_i + 2\alpha_i r^{-1}, x_i + 2\beta_i r^{-1}) \cup (x_j + 2\alpha_j r^{-1}, x_j + 2\beta_j r^{-1}))]^{r-1} \\
\times \tilde{f}_0(s_i r^{-1} + x_i 1) \tilde{f}_0(s_j r^{-1} + x_j 1) \, ds_i ds_j,
\]

where \( s_i = (s_{i1}, ..., s_{im}) \), \( \alpha_i = \min \{0, s_{i1}, ..., s_{im} \} \), and \( \beta_i = \max \{0, s_{i1}, ..., s_{im} \} \). Using a similar argument as before, we can also show that it is valid to exchange the order of limitation and expectation. As \( r \to \infty \), the two intervals \( (x_i + 2\alpha_i r^{-1}, x_i + 2\beta_i r^{-1}) \) and \( (x_j + 2\alpha_j r^{-1}, x_j + 2\beta_j r^{-1}) \) are separated from each other. Hence

\[
P(x_i \in (x_i + 2\alpha_i r^{-1}, x_i + 2\beta_i r^{-1}) \cup (x_j + 2\alpha_j r^{-1}, x_j + 2\beta_j r^{-1})) \\
= P(x_i \in (x_i + 2\alpha_i r^{-1}, x_i + 2\beta_i r^{-1})) + P(x_i \in (x_j + 2\alpha_j r^{-1}, x_j + 2\beta_j r^{-1})) \\
= 2r^{-1} [f_1(x_i)(\beta_i - \alpha_i) + f_1(x_j)(\beta_j - \alpha_j)] + o(r^{-1}).
\]

Using this result, we obtain that

\[
r^{2m} E[q_i^m g(x_i) q_j^m g(x_j)] = \left[ \frac{(m + 1)!}{2^m} \right]^2 E \left[ g(x_i) g(x_j) \frac{f_0^m(x_i)}{f_1^m(x_i)} \frac{f_0^m(x_j)}{f_1^m(x_j)} \right] + o(1) \\
= \left[ \frac{(m + 1)!}{2^m} \right]^2 E \left[ g(x_i) \frac{f_0^m(x_i)}{f_1^m(x_i)} \right] E \left[ g(x_j) \frac{f_0^m(x_j)}{f_1^m(x_j)} \right] + o(1) \\
= E[r^{2m} q_i^m g(x_i)] E[r^{2m} q_j^m g(x_j)] + o(1),
\]

where second equality follows from the independence of \( x_i \) and \( x_j \) and the last equality follows from result (8). From this result and result (8),

\[
\text{Cov} \left( r^{2m} q_i^m g(x_i), r^{2m} q_j^m g(x_j) \right) \to 0.
\]
Note that
\[
\text{Var} \left( \frac{r^{m-1}}{n} \sum_{i \in \mathcal{R}} q_i^m g(x_i) \right) = \frac{1}{r} \text{Var} \left( \frac{r^{m-1}}{n} q_i^m g(x_i) \right) + \frac{r - 1}{r} \text{Cov} \left( \frac{r^{m-1}}{n} q_i^m g(x_i), \frac{r^{m-1}}{n} q_j^m g(x_j) \right).
\]
Using result (8) with \( g \) replaced by \( g^2 \), we obtain that \( \text{Var} \left( \frac{r^{m-1}}{n} q_i^m g(x_i) \right) \leq E \left[ \frac{r^{m-1}}{n} q_i^m g(x_i) \right]^2 \to 0 \), which, together with (34), implies that \( \text{Var} \left( \frac{r^{m-1}}{n} \sum_{i \in \mathcal{R}} q_i^m g(x_i) \right) \to 0. \)
This proves (9).

(iii) Note that
\[
E \left( \sum_{i \in \mathcal{R}} q_i y_i^t \mid \mathcal{X} \right) = \sum_{i \in \mathcal{R}} q_i E(y_i^t \mid \mathcal{X}) = \sum_{i \in \mathcal{R}} q_i E(y_i^t \mid x_i) = E(\hat{y}_i^t \mid a_i = 0) + o_p(1),
\]
where the last equality follows from (9) with \( m = 1 \) and \( g(x_i) = E(\hat{y}_i^t \mid x_i) \) and the second equality follows from \( E(\hat{y}_i^t \mid x_i, a_i = 1) = E(\hat{y}_i^t \mid x_i) \) (Assumption A). Also,
\[
E \left[ \text{Var} \left( \sum_{i \in \mathcal{R}} q_i y_i^t \mid \mathcal{X} \right) \right] = E \left[ \sum_{i \in \mathcal{R}} q_i^2 \text{Var}(y_i^t \mid \mathcal{X}) \right] = E \left[ \sum_{i \in \mathcal{R}} q_i^2 \text{Var}(y_i^t \mid x_i) \right] = o(1),
\]
where the last equality follows from (8) with \( g(x_i) = \text{Var}(y_i^t \mid x_i) \). This proves (10).

**Proof of Lemma 2.** Without loss of generality, we assume \( K = 1 \) and that \( n V_n(\theta) \) has a positive limit. From Theorem 2, \( n [\hat{V}_n(\theta) - V_n(\theta)] \to_p 0 \). Hence, it remains to show that \( n[\hat{V}_n(\hat{\theta}) - \hat{V}_n(\theta)] \to_p 0 \). From (28) for the case of \( K = 1 \),
\[
n \hat{V}_n(t) = \frac{1}{n-1} \sum_{l=1}^{3} \sum_{k=1}^{3} \sum_{j \in \mathcal{R}} c_{jl}(t) c_{jk}(t) + \frac{1}{n-1} \sum_{l=3}^{4} \sum_{k=3}^{4} \sum_{j \in \mathcal{R}} c_{jl}(t) c_{jk}(t),
\]
where \( c_{jl}(t) = n \bar{w}_j (1 + d_j g_j) I_{y_j}(t) \), \( c_{j2}(t) = -2^{-1} n \bar{w}_j d_j g_j [I_{y_j}(t) + I_{y_j}(t)] \), \( c_{j3}(t) = - \sum_{j \in \mathcal{R}} \bar{w}_j (1 + d_j) I_{y_j}(t) \), and \( c_{4j}(t) = n \bar{w}_j I_{y_j}(t) \). Then, it suffices to show that for each \( (l, k) \) and \( \mathcal{A} = \mathcal{R} \) or \( \tilde{\mathcal{R}} \), \((n-1)^{-1} \sum_{j \in \mathcal{A}} c_{jl}(\hat{\theta}) c_{jk}(\hat{\theta}) - \sum_{j \in \mathcal{A}} c_{jl}(\theta) c_{jk}(\theta) = o_p(1) \). We now prove this for \( l = k = 1 \) and \( \mathcal{A} = \mathcal{R} \). Other cases can be similarly treated. Let \( H_n(t) = (n-1)^{-1} \sum_{j \in \mathcal{R}} c_{j1}^2(t) \). We need to show
\[
H_n(\hat{\theta}) - H_n(\theta) = o_p(1).
\]
Since \( \hat{\theta} \to_p \theta \), we can assume that, for any \( \epsilon > 0 \), \( \theta - \epsilon \leq \hat{\theta} \leq \theta + \epsilon \). Note that \( c_{jk}(t) \) is monotone in \( t \). Hence,
\[
H_n(\theta - \epsilon) - H_n(\theta) \leq H_n(\hat{\theta}) - H_n(\theta) \leq H_n(\theta + \epsilon) - H_n(\theta).
\]
From Theorem 2 and its proof,

\[ H_n(\theta \pm \epsilon) - H_n(\theta) = E[H_n(\theta \pm \epsilon) - H_n(\theta)] + o_p(1). \]

Also,

\[
E[H_n(\theta + \epsilon) - H_n(\theta)] = E \left\{ \frac{1}{n-1} \sum_{j \in R} [n\bar{w}_j(1 + d_jg_j)]^2 [I_{y_j}(\theta + \epsilon) - I_{y_j}(\theta)] \right\} \\
\leq O(1) E \left\{ \sum_{j \in R} \bar{w}_j(1 + d_j)[I_{y_j}(\theta + \epsilon) - I_{y_j}(\theta)] \right\} \\
\leq O(1)[F(\theta + \epsilon) - F(\theta)],
\]

where the two inequalities follow from Lemma 1 and its proof. Since \(\epsilon\) is arbitrary and \(F\) is continuous at \(\theta\), we conclude that (35) holds, which completes the proof.

References


Table 1. Population Characteristics in Each Stratum

<table>
<thead>
<tr>
<th>Province</th>
<th>Farm Size</th>
<th>$M$</th>
<th>$n$</th>
<th>$\mu_x$</th>
<th>$\sigma_x$</th>
<th>$\mu_y$</th>
<th>$\sigma_y$</th>
<th>corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT</td>
<td>1</td>
<td>557</td>
<td>101</td>
<td>1060451</td>
<td>218560</td>
<td>42269</td>
<td>8029</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>602</td>
<td>148</td>
<td>1311046</td>
<td>395854</td>
<td>43754</td>
<td>8476</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>181</td>
<td>86</td>
<td>1673880</td>
<td>376296</td>
<td>45679</td>
<td>9928</td>
<td>0.658</td>
</tr>
<tr>
<td>QUE</td>
<td>1</td>
<td>2589</td>
<td>44</td>
<td>694503</td>
<td>127627</td>
<td>41082</td>
<td>7565</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3401</td>
<td>51</td>
<td>830875</td>
<td>170994</td>
<td>43996</td>
<td>7662</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2565</td>
<td>55</td>
<td>1151957</td>
<td>221437</td>
<td>47714</td>
<td>9772</td>
<td>0.597</td>
</tr>
<tr>
<td>ONT</td>
<td>1</td>
<td>2840</td>
<td>105</td>
<td>1175844</td>
<td>296864</td>
<td>53273</td>
<td>9453</td>
<td>0.608</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>950</td>
<td>124</td>
<td>1433480</td>
<td>361390</td>
<td>55395</td>
<td>10114</td>
<td>0.624</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>621</td>
<td>111</td>
<td>1810432</td>
<td>484186</td>
<td>53520</td>
<td>10816</td>
<td>0.595</td>
</tr>
<tr>
<td>MAN</td>
<td>1</td>
<td>271</td>
<td>20</td>
<td>1107272</td>
<td>359383</td>
<td>43580</td>
<td>5578</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>270</td>
<td>41</td>
<td>1175414</td>
<td>154871</td>
<td>45916</td>
<td>9176</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>194</td>
<td>31</td>
<td>1559805</td>
<td>169610</td>
<td>46782</td>
<td>9501</td>
<td>0.637</td>
</tr>
<tr>
<td>ALB</td>
<td>1</td>
<td>406</td>
<td>44</td>
<td>1833741</td>
<td>392292</td>
<td>109162</td>
<td>10452</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>406</td>
<td>46</td>
<td>2170270</td>
<td>340039</td>
<td>116319</td>
<td>15952</td>
<td>0.639</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>274</td>
<td>49</td>
<td>2846832</td>
<td>612906</td>
<td>115036</td>
<td>14385</td>
<td>0.628</td>
</tr>
<tr>
<td>B.C.</td>
<td>1</td>
<td>315</td>
<td>45</td>
<td>1776791</td>
<td>247117</td>
<td>67906</td>
<td>10240</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>326</td>
<td>47</td>
<td>2216035</td>
<td>392315</td>
<td>67258</td>
<td>14411</td>
<td>0.577</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>221</td>
<td>45</td>
<td>2801296</td>
<td>366018</td>
<td>69825</td>
<td>17473</td>
<td>0.446</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>16989</td>
<td>1193</td>
<td>1195710</td>
<td>545773</td>
<td>52315</td>
<td>19621</td>
<td>0.732</td>
</tr>
</tbody>
</table>

$M =$ population size; $n =$ sample size; $\mu_x$ and $\sigma_x =$ mean and standard deviation for $x$ (net assets); $\mu_y$ and $\sigma_y =$ mean and standard deviation for $y$ (cash income); corr = correlation coefficient between $x$ and $y$
Table 2. Average Results based on 2,000 Simulations

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$M$</th>
<th>Performance of point estimator</th>
<th>Relative bias (%)</th>
<th>Variance/1000</th>
<th>Relative bias of variance estimator (%)</th>
<th>Performance of confidence interval</th>
<th>Coverage prob (%)</th>
<th>Length/1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNI</td>
<td>LRI</td>
<td>RHD</td>
<td>MI10</td>
<td>NNI</td>
<td>LRI</td>
<td>RHD</td>
<td>MI10</td>
</tr>
<tr>
<td>$\mu_y$</td>
<td>0</td>
<td>-0.01</td>
<td>1.60</td>
<td>-3.31</td>
<td>-3.33</td>
<td>229</td>
<td>258</td>
<td>355</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>1.72</td>
<td>3.89</td>
<td>3.90</td>
<td>227</td>
<td>247</td>
<td>253</td>
<td>210</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.11</td>
<td>-0.01</td>
<td>0.00</td>
<td>206</td>
<td>221</td>
<td>297</td>
<td>253</td>
</tr>
<tr>
<td>4</td>
<td>-0.04</td>
<td>1.28</td>
<td>-2.37</td>
<td>-2.36</td>
<td>204</td>
<td>226</td>
<td>294</td>
<td>237</td>
</tr>
<tr>
<td>5</td>
<td>0.08</td>
<td>1.24</td>
<td>3.18</td>
<td>3.19</td>
<td>204</td>
<td>207</td>
<td>225</td>
<td>189</td>
</tr>
<tr>
<td>6</td>
<td>0.02</td>
<td>0.07</td>
<td>0.00</td>
<td>0.01</td>
<td>189</td>
<td>192</td>
<td>258</td>
<td>211</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.80</td>
<td>-0.09</td>
<td>-0.09</td>
<td>150</td>
<td>186</td>
<td>203</td>
<td>183</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>0.57</td>
<td>1.51</td>
<td>1.82</td>
<td>172</td>
<td>165</td>
<td>178</td>
<td>151</td>
</tr>
<tr>
<td>9</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>183</td>
<td>181</td>
<td>207</td>
<td>188</td>
</tr>
<tr>
<td>10</td>
<td>-0.03</td>
<td>0.15</td>
<td>165</td>
<td>-5.86</td>
<td>94.3</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\theta$: the parameter of interest; $q_a = \text{the } a\text{th percentile of } y$

$M$: the model for simulation

NNI: nearest neighbor imputation

LRI: linear regression imputation

RHD: random hot deck imputation

MI10: Bayesian bootstrap multiple imputation with 10 imputations
<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$M$</th>
<th>Performance of point estimator</th>
<th>Performance of confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Relative bias (%) Variance/1000</td>
<td>Coverage prob (%) Length/1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NNI LRI RHD MI10</td>
<td>NNI LRI RHD MI10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NNI LRI RHD MI10</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.07 1.52 -2.60 -2.61 779 753 756 836</td>
<td>92.4 79.9 59.2 53.9 1.71 1.58 1.71 1.59</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.08 -1.44 2.42 2.49 501 408 510 623</td>
<td>93.4 75.8 56.2 47.3 1.41 1.36 1.47 1.37</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.01 -1.31 -0.02 0.00 557 475 643 713</td>
<td>93.0 79.6 92.6 88.2 1.48 1.39 1.56 1.45</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.01 1.24 -1.70 -1.81 602 561 664 652</td>
<td>93.3 82.2 72.6 66.7 1.52 1.42 1.55 1.43</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.04 -1.16 1.91 1.91 454 371 497 517</td>
<td>92.9 82.3 64.2 59.5 1.30 1.30 1.36 1.27</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.07 -0.91 0.02 0.04 499 442 559 572</td>
<td>92.8 85.9 92.2 89.5 1.35 1.29 1.42 1.32</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.08 -0.80 -0.66 -0.68 448 452 460 484</td>
<td>92.7 85.7 80.1 87.2 1.28 1.22 1.31 1.25</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.03 -0.54 1.07 1.07 378 350 401 394</td>
<td>93.6 90.7 81.6 79.1 1.19 1.20 1.23 1.18</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-0.03 -0.44 -0.02 -0.01 386 354 403 416</td>
<td>92.9 91.6 93.2 91.3 1.21 1.18 1.24 1.19</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.00 330</td>
<td>93.3</td>
</tr>
<tr>
<td>0.70</td>
<td>1</td>
<td>0.02 2.63 -5.80 -5.78 1225 1225 1703 1935</td>
<td>93.6 89.8 66.2 62.0 2.24 2.56 2.64 2.45</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.88 11.90 10.10 10.10 2106 1269 1534 1728</td>
<td>92.4 9.5 26.6 22.4 2.95 2.40 2.60 2.39</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.12 5.36 0.02 -0.02 1470 1188 1926 2124</td>
<td>93.4 66.4 93.7 89.7 2.44 2.39 2.80 2.59</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.06 2.04 -4.14 -4.11 1143 1139 1476 1626</td>
<td>93.2 90.4 70.3 73.7 2.14 2.34 2.49 2.30</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.53 9.43 8.43 8.38 1731 1188 1397 1604</td>
<td>90.9 20.2 35.0 31.9 2.65 2.20 2.43 2.26</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.19 3.99 0.08 0.10 1191 1148 1656 1693</td>
<td>93.8 74.8 92.2 89.2 2.29 2.25 2.53 2.35</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.18 1.26 -1.74 -1.74 973 1068 1189 1693</td>
<td>94.2 90.6 91.7 89.3 2.02 2.08 2.23 2.11</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.27 5.15 5.09 5.07 1260 1092 1257 1309</td>
<td>93.3 56.8 62.8 59.8 2.27 2.04 2.22 2.11</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.23 1.89 0.20 0.20 1095 1072 1245 1282</td>
<td>93.2 87.2 92.4 91.3 2.09 2.07 2.20 2.10</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.18 937</td>
<td>93.7</td>
</tr>
<tr>
<td>0.65</td>
<td>1</td>
<td>-0.31 1.45 -3.78 -3.77 5402 4477 8518 9410</td>
<td>92.7 93.0 66.1 61.3 4.73 5.15 5.37 4.98</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.04 4.44 4.43 4.44 7777 4996 5624 6119</td>
<td>91.5 38.7 52.8 40.0 5.48 5.91 4.93 4.52</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.11 1.77 -0.08 -0.10 6211 4422 8298 9459</td>
<td>92.4 83.0 91.6 89.9 4.89 4.14 5.54 5.15</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.17 1.23 -2.67 -2.65 5073 4567 7296 7738</td>
<td>92.3 92.6 77.2 72.8 4.53 4.80 5.11 4.77</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.02 3.35 3.65 3.63 6741 4503 5633 5713</td>
<td>91.3 52.5 60.8 55.4 4.99 3.64 4.64 4.27</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-0.18 1.30 -0.14 -0.10 5229 4074 6914 7089</td>
<td>93.4 87.4 93.0 89.2 4.61 4.07 5.14 4.78</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.01 0.78 -1.13 -1.12 4395 4180 5769 5914</td>
<td>92.9 91.4 90.6 87.6 4.23 4.37 4.66 4.41</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.13 1.45 2.03 2.05 5129 3863 4699 4780</td>
<td>92.1 81.1 81.9 77.9 4.48 3.61 4.35 4.14</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-0.11 0.57 -0.09 -0.12 4674 4142 5027 5357</td>
<td>92.8 91.0 92.8 90.5 4.25 4.00 4.51 4.33</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>-0.01 3864</td>
<td>93.8</td>
</tr>
</tbody>
</table>