

2019

# Nearest Neighbor Imputation for General Parameter Estimation in Survey Sampling

Shu Yang

*North Carolina State University*

Jae Kwang Kim

*Iowa State University, jkim@iastate.edu*

Follow this and additional works at: [https://lib.dr.iastate.edu/stat\\_las\\_pubs](https://lib.dr.iastate.edu/stat_las_pubs)

 Part of the [Design of Experiments and Sample Surveys Commons](#), and the [Statistical Methodology Commons](#)

The complete bibliographic information for this item can be found at [https://lib.dr.iastate.edu/stat\\_las\\_pubs/261](https://lib.dr.iastate.edu/stat_las_pubs/261). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

---

This Book Chapter is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Nearest Neighbor Imputation for General Parameter Estimation in Survey Sampling

## **Abstract**

Nearest neighbor imputation has a long tradition for handling item nonresponse in survey sampling. In this article, we study the asymptotic properties of the nearest neighbor imputation estimator for general population parameters, including population means, proportions and quantiles. For variance estimation, we propose novel replication variance estimation, which is asymptotically valid and straightforward to implement. The main idea is to construct replicates of the estimator directly based on its asymptotically linear terms, instead of individual records of variables. The simulation results show that nearest neighbor imputation and the proposed variance estimation provide valid inferences for general population parameters.

## **Keywords**

Bahadur representation, Bootstrap, Jackknife variance estimation, Matching, Missing at random, Quantile estimation

## **Disciplines**

Design of Experiments and Sample Surveys | Statistical Methodology

## **Comments**

This is a manuscript of a chapter from Yang, S. and Kim, J. (2019), "Nearest Neighbor Imputation for General Parameter Estimation in Survey Sampling", *The Econometrics of Complex Survey Data (Advances in Econometrics, Vol. 39)*, Emerald Publishing Limited, pp. 209-234. doi: [10.1108/S0731-905320190000039012](https://doi.org/10.1108/S0731-905320190000039012). Posted with permission.

# Nearest neighbor imputation for general parameter estimation in survey sampling

Shu Yang\*      Jae Kwang Kim<sup>†</sup>

## Abstract

Nearest neighbor imputation is popular for handling item nonresponse in survey sampling. In this article, we study the asymptotic properties of the nearest neighbor imputation estimator for general population parameters, including population means, proportions and quantiles. For variance estimation, the conventional bootstrap inference for matching estimators with fixed number of matches has been shown to be invalid due to the nonsmoothness nature of the matching estimator. We propose asymptotically valid replication variance estimation. The key strategy is to construct replicates of the estimator directly based on linear terms, instead of individual records of variables. A simulation study confirms that the new procedure provides valid variance estimation.

*Key Words:* Bahadur representation; Bootstrap; Hot deck; Jackknife variance estimation; Missing at random; Quantile estimation.

---

\*Department of Statistics, North Carolina State University, North Carolina 27695, U.S.A.

<sup>†</sup>Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

# 1 Introduction

Nearest neighbor imputation is popular for handling item nonresponse in survey sampling. In nearest neighbor imputation, the vector of the auxiliary variables is directly used in determining the nearest neighbor. The nearest neighbor is then used as a donor for hot deck imputation. Although these imputation methods have a long history of application, there are relatively few papers on investigating their asymptotic properties. Sande (1979) discussed nearest neighbor rules in statistical estimation with hot-deck imputation. Lee and Särndal (1994) studied methods of nearest neighbor imputation. Chen and Shao (2000, 2001) have developed a nice set of asymptotic theories for the nearest neighbor imputation estimator. Abadie and Imbens (2006) studied the matching estimator to estimate the average treatment effect from observational studies. Shao and Wang (2008) proposed methods for constructing confidence intervals for population means and quantiles with nearest neighbor imputation. Kim et al. (2011) presented an application of nearest neighbor imputation for the US Census long form data. However, most of these studies discussed either with a 1-dimensional covariate or only for mean estimation, which is restrictive both theoretically and practically.

Survey statisticians are often interested in various finite population quantities, such as the population means, proportions and quantiles (Francisco and Fuller; 1991; Wu and Sitter; 2001; Berger and Skinner; 2003), to name a few. Some corresponding sample estimators should be treated differently than others. For example, estimators of population quantiles involve nondifferentiable functions of estimated quantities. Moreover, there often are more than one auxiliary covariates available to facilitate nearest neighbor imputation. The

current framework of nearest neighbor imputation can not cover inferences in these settings.

In this article, we provide a framework of nearest neighbor imputation for general parameter estimation in survey sampling. In general, the matching estimators are not root- $n$  consistent (Abadie and Imbens; 2006), where  $n$  is the sample size. Based on a scalar matching variable  $m$  summarizing all auxiliary information, we show that nearest neighbor imputation can provide consistent estimators for a fairly general class of parameters. If the matching variable is chosen to be the mean function of the study variable, our method resembles prediction mean matching imputation. However, the validity of predictive mean matching requires the mean function to be correctly specified. Here, we show that the consistency of the nearest neighbor imputation estimator only requires the matching variable to satisfy certain Lipschitz continuity condition. For inference, intrinsically the nearest neighbor imputation estimator with fixed number of matches is not smooth. The lack of smoothness makes the conventional replication methods invalid for variance estimation, mainly because the naive replication method distorts the distribution of the number of times each unit is used as a match. We propose new replication variance estimation. Based on the linear representation of the nearest neighbor imputation estimator, we construct replicates of the estimator directly based on its linear terms. In this way, the distribution of the number of times each unit is used as a match can be preserved, which leads to a valid variance estimation. Furthermore, our replication variance method is flexible, which can accommodate bootstrap and jackknife, among others.

## 2 Basic Setup

Let  $\mathcal{F}_N = \{(x_i, y_i, \delta_i) : i = 1, \dots, N\}$  denote a finite population, where  $x_i$  is a  $p$ -dimensional vector of covariates, which is always observed,  $y_i$  has missing values, and  $\delta_i$  is the response indicator of  $y_i$ , i.e.,  $\delta_i = 1$  if  $y_i$  is observed and 0 if it is missing. The  $\delta_i$ 's are defined throughout the finite population, as in Fay (1992), Shao and Steel (1999), and Kim et al. (2006). We assume that  $\mathcal{F}_N$  is a random sample from a superpopulation model  $\zeta$ , and  $N$  is known. Our objective is to estimate the finite population parameter defined through  $\mu_g = N^{-1} \sum_{i=1}^N g(y_i)$  for some known  $g(\cdot)$ , or  $\xi_N = \inf\{\xi : S_N(\xi) \geq 0\}$ , where  $S_N(\xi) = N^{-1} \sum_{i=1}^N s(y_i - \xi)$  and  $s(\cdot)$  is a univariate real function. These parameters are fairly general, which cover many parameters of interest in survey sampling. For example, let  $g(y) = y$ ,  $\mu_g$  is the population mean of  $y$ ,  $N^{-1} \sum_{i=1}^N y_i$ . Let  $g(y) = I(y < c)$  for some constant  $c$ ,  $\mu_g$  is the population proportion of  $y$  less than  $c$ ,  $N^{-1} \sum_{i=1}^N I(y_i < c)$ . Let  $s(y_i - \xi) = I(y_i \leq \xi) - \alpha$ ,  $\xi_N$  is the population  $\alpha$ th quantile.

Let  $A$  denote an index set of the sample selected by a probability sampling design. Let  $I_i$  be the sampling indicator function, i.e.,  $I_i = 1$  if unit  $i$  is selected into the sample, and  $I_i = 0$  otherwise. Suppose that  $\pi_i$ , the first-order inclusion probability of unit  $i$ , is positive and known throughout the sample. If  $y_i$  were fully observed throughout the sample, the sample estimator of  $\mu_g$  and  $\xi_N$  are  $\hat{\mu}_g = N^{-1} \sum_{i \in A} \pi_i^{-1} g(y_i)$  and  $\hat{\xi} = \inf\{\xi : \hat{S}_N(\xi) \geq 0\}$  with  $\hat{S}_N(\xi) = N^{-1} \sum_{i \in A} \pi_i^{-1} s(y_i - \xi)$ , respectively.

We make the following assumption for the missing data process.

**Assumption 1 (Missing at random and positivity)** *The missing data*

process satisfies  $\text{pr}(\delta = 1 \mid x, y) = \text{pr}(\delta = 1 \mid x)$ , which is denoted by  $p(x)$ , and with probability 1,  $p(x) > \epsilon$  for a constant  $\epsilon > 0$ .

Our primary focus will be on the imputation estimators of  $\mu_g$  and  $\xi_N$  given by  $\hat{\mu}_{g,I} = N^{-1} \sum_{i \in A} \pi_i^{-1} \{\delta_i g(y_i) + (1 - \delta_i) g(y_i^*)\}$  and  $\hat{\xi}_I = \inf\{\xi : \hat{S}_I(\xi) \geq 0\}$ , with  $\hat{S}_I(\xi) = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} s(y_i - \xi) \{\delta_i s(y_i - \xi) + (1 - \delta_i) s(y_i^* - \xi)\}$ , where  $y_i^*$  is an imputed value of  $y_i$  for unit  $i$  with  $\delta_i = 0$ . To find suitable imputed values, the classical nearest neighbor imputation can be described in the following steps:

**Step 1.** For each unit  $i$  with  $\delta_i = 0$ , find the nearest neighbor from the respondents with the minimum distance between  $x_j$  and  $x_i$ . Let  $i(1)$  be the index set of its nearest neighbor, which satisfies  $d(x_{i(1)}, x_i) \leq d(x_j, x_i)$ , for  $j \in A_R$ , where  $d(x_i, x_j)$  is a distance function between  $x_i$  and  $x_j$ . For example,  $d(x_i, x_j) = \|x_i - x_j\|$ , where  $\|x\| = (x^T x)^{1/2}$ . Other norms of the form  $\|x\|_D = (x^T D x)^{1/2}$ , where  $D$  is a positive definite symmetric matrix  $D$ , are equivalent to the Euclidean norm, since  $\|x\|_D = \{(Qx)^T (Qx)\}^{1/2} = \|Qx\|$  with  $Q^T Q = D$ . In particular, Mahalanobis distance is commonly used, where  $D = \hat{\Sigma}^{-1}$  with  $\hat{\Sigma}$  the empirical covariance matrix of  $x$ .

**Step 2.** The nearest neighbor imputation estimators of  $\mu_g$  and  $\xi_N$  are computed by

$$\hat{\mu}_{g,\text{NNI}} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \{\delta_i g(y_i) + (1 - \delta_i) g(y_{i(1)})\}, \quad (1)$$

and  $\hat{\xi}_{\text{NNI}} = \inf\{\xi : \hat{S}_{\text{NNI}}(\xi) \geq 0\}$ , respectively, with

$$\hat{S}_{\text{NNI}}(\xi) = \frac{1}{N} \sum_{i \in A} \pi_i^{-1} \{\delta_i s(y_i - \xi) + (1 - \delta_i) s(y_{i(1)} - \xi)\}. \quad (2)$$

In (1) and (2), the imputed values are real observations.

### 3 Main result

For asymptotic inference, we follow the framework of Isaki and Fuller (1982) where the asymptotic properties of estimators are established under a fixed sequence of populations and a corresponding sequence of random samples. Denote  $E_p(\cdot)$  and  $\text{var}_p(\cdot)$  to be the expectation and the variance under the sampling design, respectively. We impose the following regularity conditions on the sampling design.

**Assumption 2** (i) *There exist positive constants  $C_1$  and  $C_2$  such that  $C_1 \leq \pi_i N n^{-1} \leq C_2$ , for  $i = 1, \dots, N$ ; (ii) the sequence of the Hotvitz-Thompson estimators  $\hat{\mu}_{g,\text{HT}} = N^{-1} \sum_{i \in A} \pi_i^{-1} g(y_i)$  satisfies  $\text{var}_p(\hat{\mu}_{g,\text{HT}}) = O(n^{-1})$  and  $\{\text{var}_p(\hat{\mu}_{g,\text{HT}})\}^{-1/2}(\hat{\mu}_{g,\text{HT}} - \mu_g) \mid \mathcal{F}_N \rightarrow \mathcal{N}(0, 1)$  in distribution, as  $n \rightarrow \infty$ .*

Assumption 2 is a widely accepted assumption in survey sampling (Fuller; 2009).

We introduce additional notation. Let  $A = A_R \cup A_M$ , where  $A_R$  and  $A_M$  are the sets of respondents and nonrespondents, respectively. Define  $d_{ij} = 1$  if  $y_{j(1)} = y_i$ , i.e., unit  $i$  is used as a donor for unit  $j \in A_M$ , and  $d_{ij} = 0$  otherwise. We write  $\hat{\mu}_{g,\text{NNI}}$  in (1) as

$$\hat{\mu}_{g,\text{NNI}} = \frac{1}{N} \left\{ \sum_{i \in A} \frac{1}{\pi_i} \delta_i g(y_i) + \sum_{j \in A} \frac{1 - \delta_j}{\pi_j} \sum_{i \in A} \delta_i d_{ij} g(y_i) \right\} = \frac{1}{N} \sum_{i \in A} \frac{\delta_i}{\pi_i} (1 + k_i) g(y_i), \quad (3)$$

with

$$k_i = \sum_{j \in A} \frac{\pi_i}{\pi_j} (1 - \delta_j) d_{ij}. \quad (4)$$



Under simple random sampling,  $k_i = \sum_{j \in A} (1 - \delta_j) d_{ij}$  is the number of times that unit  $i$  is used as the nearest neighbor for nonrespondents.

To study the asymptotic properties of the nearest neighbor imputation estimator  $\hat{\mu}_{g,\text{NNI}}$ , we use the following decomposition:

$$n^{1/2}(\hat{\mu}_{g,\text{NNI}} - \mu_g) = D_N + B_N, \quad (5)$$

where

$$D_N = n^{1/2} \left[ \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \{ \mu_g(x_i) + \delta_i(1 + k_i) \{ g(y_i) - \mu_g(x_i) \} - \mu_g \right], \quad (6)$$

and

$$B_N = \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} (1 - \delta_i) \{ \mu_g(x_{i(1)}) - \mu_g(x_i) \}. \quad (7)$$

The difference  $\mu_g(x_{i(1)}) - \mu_g(x_i)$  accounts for the matching discrepancy, and  $B_N$  contributes to the asymptotic bias of the matching estimator. In general, if  $x$  is  $p$ -dimensional, Abadie and Imbens (2006) showed that  $d(x_{i(1)}, x_i) = O_p(n^{-1/p})$ . Therefore, for nearest neighbor imputation with  $p \geq 2$ , the bias  $B_N = O_p(n^{1/2-1/p}) \neq o_p(1)$  is not negligible.

To address for the matching discrepancy due to a non-scalar  $x$ , we first summarize the covariate information into a scalar matching variable  $m = m(x)$ , and then apply nearest neighbor imputation based on this scalar variable. For simplicity of notation, we may suppress the dependence of  $m$  on  $x$  if there is no ambiguity. For nearest neighbor imputation with a scalar matching variable, we then have  $p = 1$  and  $B_N = O_p(n^{-1/2}) = o_p(1)$ . We assume the superpopulation model and the matching variable  $m$  satisfy the following assumption.

**Assumption 3** (i) *The matching variable  $m$  has a compact and convex support, with density bounded and bounded away from zero. Let  $f_1(m)$  and  $f_0(m)$  be the conditional density of  $m$  given  $\delta = 1$  and  $\delta = 0$ , respectively. Suppose that there exist constants  $C_{1L}$  and  $C_{1U}$  such that  $C_{1L} \leq f_1(m)/f_0(m) \leq C_{1U}$ ;* (ii)  $\mu_g(x) = E\{g(y) \mid x\}$  and  $\mu_s(\xi, x) = E\{s(y - \xi) \mid x\}$  *satisfy certain Lipschitz continuous condition; i.e., there exists a constant  $C_2$  such that  $|\mu_g(x_i) - \mu_g(x_j)| < C_2|m_i - m_j|$  and  $|\mu_s(\xi, x_i) - \mu_s(\xi, x_j)| < C_2|m_i - m_j|$  for any  $i$  and  $j$ ;* (iii) *there exists  $\delta > 0$  such that  $E(|g(y)|^{2+\delta} \mid x)$  and  $E(|s(y - \xi)|^{2+\delta} \mid x)$  are uniformly bounded for any  $x$  and  $\xi$  in the neighborhood of  $\xi_N$ .*

Assumption 3 (i) is a convenient regularity condition (Abadie and Imbens; 2006). Assumption 3 (ii) imposes a smoothness condition for  $\mu_g(x)$ ,  $\mu_s(\xi, x)$  and  $m(x)$ , which is not restrictive (Chen and Shao; 2000). Assumption 3 (iii) is a moment condition for establishing the central limit theorem.

We establish the asymptotic distribution of  $\hat{\mu}_{g,\text{NNI}}$ , with the proof deferred to the Appendix.

**Theorem 1** *Under Assumptions 1–2, suppose that  $\mu_g(x) = E\{g(y) \mid x\}$  and  $\sigma_g^2(x) = \text{var}\{g(y) \mid x\}$ . Then,  $n^{1/2}\{\hat{\mu}_{g,\text{NNI}} - \mu_g\} \rightarrow \mathcal{N}(0, V_g)$  in distribution, as  $n \rightarrow \infty$ , where*

$$V_g = V_g^\mu + V_g^e \tag{8}$$

*with  $V_g^\mu = \lim_{n \rightarrow \infty} nN^{-2}E[\text{var}_p\{\sum_{i \in A} \pi_i^{-1}\mu_g(x_i)\}]$ ,  $V_g^e = \lim_{n \rightarrow \infty} nN^{-2}E[\sum_{i \in A} \{\pi_i^{-1}\delta_i(1+k_i) - 1\}^2\sigma_g^2(x_i)]$ , and  $k_i$  is defined in (4).*

We now establish a similar result for  $\hat{\xi}_{\text{NNI}}$ , with the proof deferred to the Appendix.

**Theorem 2** *Under Assumptions 1–2, suppose the population parameter  $\xi_N$  and the population estimating function  $S_N(\cdot)$  satisfy certain regularity conditions specified in Assumptions A4 and A5. We obtain an asymptotic linearization representation of  $\hat{\xi}_{\text{NNI}}$ :*

$$n^{1/2}(\hat{\xi}_{\text{NNI}} - \xi) = -n^{1/2}\{\hat{S}_{\text{NNI}}(\xi) - S_N(\xi)\}/S'(\xi) + o_p(1). \quad (9)$$

*It follows that  $n^{1/2}(\hat{\xi}_{\text{NNI}} - \xi_N) \rightarrow \mathcal{N}(0, V_\xi)$  in distribution, as  $n \rightarrow \infty$ , where  $V_\xi = \dot{S}(\xi_N)^{-2}\text{var}\{\hat{S}_{\text{NNI}}(\xi_N)\}$ ,  $\dot{S}(\xi_N) = dS(\xi_N)/d\xi$ , and*

$$\begin{aligned} \text{var}\{\hat{S}_{\text{NNI}}(\xi_N)\} &= \lim_{n \rightarrow \infty} \frac{n}{N^2} E \left( \text{var}_p \left[ \sum_{i \in A} \frac{E\{s(y_i - \xi_N) \mid x_i\}}{\pi_i} \right] \right) \\ &+ \text{plim} \frac{n}{N^2} \sum_{i=1}^N \left\{ \frac{I_i}{\pi_i} \delta_i (1 + k_i) - 1 \right\}^2 \text{var} [s(y_i - \xi_N) - E\{s(y_i - \xi_N) \mid x_i\} \mid x_i], \end{aligned} \quad (10)$$

*and  $k_i$  is defined in (4).*

For illustration, we use quantile estimation as an example.

**Example 1 (Quantile estimation)** *The estimating function for the  $\alpha$ th quantile is  $s(y_i - \xi) = I(y_i \leq \xi) - \alpha$ , and the population estimating equation  $S_{\alpha, N}(\xi) = F_N(\xi) - \alpha$ , where  $F_N(\xi) = N^{-1} \sum_{i=1}^N I(y_i \leq \xi)$ . The nearest neighbor imputation estimator  $\hat{\xi}_{\alpha, \text{NNI}}$  is defined as*

$$\hat{\xi}_{\alpha, \text{NNI}} = \inf\{\xi : \hat{S}_{\alpha, \text{NNI}}(\xi) \geq 0\},$$

*where  $\hat{S}_{\alpha, \text{NNI}}(\xi) = \hat{F}_{\text{NNI}}(\xi) - \alpha$ ,  $\hat{F}_{\text{NNI}}(\xi) = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} \delta_i (1 + k_i) I(y_i \leq \xi)$ ,  $\hat{N} = \sum_{i \in A} \pi_i^{-1}$ , and  $k_i$  is defined in (4). Let  $F(\xi) = \text{pr}(y \leq \xi)$  be the cumulative distribution function of  $y$ . Then,  $\hat{F}_{\text{NNI}}(\xi)$  is a Hajek estimator*

for  $F(\xi)$ , which is asymptotically equivalent to the one using  $N$  instead of  $\hat{N}$ . Even with a known  $N$ , it is necessary to use  $\hat{N}$  because  $\hat{F}_{\text{NNI}}(\xi)$  for  $\xi = \infty$  should be 1. The limiting function of  $S_{\alpha,N}(\xi)$  is  $S_\alpha(\xi) = F(\xi) - \alpha$ . The asymptotic linearization representation of  $\hat{\xi}_{\alpha,\text{NNI}}$  is

$$\hat{\xi}_{\alpha,\text{NNI}} - \xi = -\frac{\hat{F}_{\text{NNI}}(\xi) - F_N(\xi)}{f(\xi)} + o_p(n^{-1/2}), \quad (11)$$

where  $f(\xi) = F'(\xi)$ . Expression (11) is called the Bahadur-type representation for  $\hat{\xi}_{\alpha,\text{NNI}}$  (Francisco and Fuller; 1991).

**Remark 1 (The choice of the scalar matching variable)** *By judicious choice, the scalar matching variable should ensure that Assumption 3 holds. If the conditional mean function of the outcome variable given the covariates is feasible, we can choose the matching variable to be the conditional mean function. We note that in this case the proposed nearest neighbor imputation resembles predictive mean matching imputation. However, our method is more general than predictive mean matching imputation, because the latter requires the mean function to be correctly specified.*

## 4 Replication variance estimation

We consider replication variance estimation (Rust and Rao; 1996; Wolter; 2007) for nearest neighbor imputation.

Let  $\hat{\mu}_g$  be the Horvitz-Thompson estimator of  $\mu_g$ . The replication variance estimator of  $\hat{\mu}_g$  takes the form of

$$\hat{V}_{\text{rep}}(\hat{\mu}_g) = \sum_{k=1}^L c_k (\hat{\mu}_g^{(k)} - \hat{\mu}_g)^2, \quad (12)$$

where  $L$  is the number of replicates,  $c_k$  is the  $k$ th replication factor, and  $\hat{\mu}_g^{(k)}$  is the  $k$ th replicate of  $\hat{\mu}_g$ . When  $\hat{\mu}_g = \sum_{i \in A} \omega_i g(y_i)$ , we can write the replicate of  $\hat{\mu}_g$  as  $\hat{\mu}_g^{(k)} = \sum_{i \in A} \omega_i^{(k)} g(y_i)$  with some  $\omega_i^{(k)}$  for  $i \in A$ . The replications are constructed such that  $E_p\{\hat{V}_{\text{rep}}(\hat{\mu}_g)\} = \text{var}_p(\hat{\mu}_g)\{1 + o(1)\}$ .

We propose a new replication variance estimation for  $\hat{\mu}_{g,\text{NNI}}$ . Let  $\omega_i = N^{-1}\pi_i^{-1}$ . Write  $\hat{\mu}_{g,\text{NNI}} - \mu_g = (\hat{\mu}_{g,\text{PMM}} - \hat{\psi}_{\text{HT}}) + (\hat{\psi}_{\text{HT}} - \mu_\psi) + (\mu_\psi - \mu_g)$ , where  $\hat{\psi}_{\text{HT}} = \sum_{i \in A} \omega_i \psi_i$ ,  $\psi_i = \mu_g(x_i) + \delta_i(1 + k_i)\{g(y_i) - \mu_g(x_i)\}$ ,  $\mu_\psi = N^{-1} \sum_{i=1}^N \psi_i$ . Because  $\mu_{g,\text{NNI}} - \hat{\psi}_{\text{HT}} = o_p(n^{-1/2})$  by Theorem 1 and  $\mu_\psi - \mu_g = O_p(N^{-1/2})$ , we have  $\hat{\mu}_{g,\text{NNI}} - \mu_g = \hat{\psi}_{\text{HT}} - \mu_\psi + o_p(n^{-1/2})$ , if  $nN^{-1} = o(1)$ . Therefore, with negligible sampling fractions, it is sufficient to estimate the variance of  $\hat{\psi}_{\text{HT}} - \mu_\psi$ . Because  $E(\hat{\psi}_{\text{HT}} - \mu_\psi \mid \mathcal{F}_N) = 0$ , we have  $\text{var}(\hat{\psi}_{\text{HT}} - \mu_\psi) = E\{\text{var}(\hat{\psi}_{\text{HT}} - \mu_\psi \mid \mathcal{F}_N)\}$ , which is essentially the sampling variance of  $\hat{\psi}_{\text{HT}}$ . This suggests that we can treat  $\{\psi_i : i \in A\}$  as pseudo observations in applying replication variance estimator. Otsu and Rai (2016) used a similar idea to develop a wild bootstrap technique for a matching estimator. To be specific, we construct replicates of  $\hat{\psi}_{\text{HT}}$  as follows:  $\hat{\psi}_{\text{HT}}^{(k)} = \sum_{i \in A} \omega_i^{(k)} \psi_i$ , where  $\omega_i^{(k)}$  is the replication weight that account for complex sampling design. The replication variance estimator of  $\hat{\psi}_{\text{HT}}$  is obtained by applying  $\hat{V}_{\text{rep}}(\cdot)$  in (12) for the above replicates  $\hat{\psi}_{\text{HT}}^{(k)}$ . It follows that  $E\{\hat{V}_{\text{rep}}(\hat{\psi}_{\text{HT}})\} = \text{var}(\hat{\psi}_{\text{HT}} - \mu_\psi)\{1 + o(1)\} = \text{var}(\hat{\mu}_{g,\text{NNI}} - \mu_g)\{1 + o(1)\}$ . Because  $\mu_g(x)$  is unknown, we use a plug-in kernel estimator  $\hat{\mu}_g(x)$ .

In summary, the new replication variance estimation for  $\hat{\mu}_{g,\text{NNI}}$  proceeds as follows:

**Step 1.** Obtain a consistent kernel estimator  $\hat{\mu}_g(x)$ .

**Step 2.** Construct replicates of  $\hat{\mu}_{g,\text{NNI}}$  as

$$\hat{\mu}_{g,\text{NNI}}^{(k)} = \sum_{i \in A} \omega_i^{(k)} [\hat{\mu}_g(x_i) + \delta_i(1 + k_i) \{g(y_i) - \hat{\mu}_g(x_i)\}], \quad (13)$$

where  $\omega_i^{(k)}$  is the  $k$ th replication weight for unit  $i$ .

**Step 3.** Apply  $\hat{V}_{\text{rep}}(\cdot)$  in (12) for the above replicates to obtain the replication variance estimator of  $\hat{\mu}_{g,\text{NNI}}$ .

We now consider a replication variance estimation for  $\hat{\xi}_{\text{NNI}}$ . Following the previous section, we directly obtain the asymptotic variance of  $\hat{\xi}_{\text{NNI}}$  using  $\text{var}\{\hat{S}_{\text{NNI}}(\xi)\}$  and  $S'(\xi)$ . First to estimate  $\text{var}\{\hat{S}_{\text{NNI}}(\xi)\}$ , we can use the similar replication variance estimation earlier in this section. Now to estimate  $S'(\xi)$ , we follow the kernel-based derivative estimation of Deville (1999):

$$\hat{S}'(\xi) = \frac{1}{Nh} \sum_{i \in A} \frac{1}{\pi_i} \int s(y_i - x) K' \left( \frac{\xi - x}{h} \right) dx \quad (14)$$

where  $K(\cdot)$  is a kernel function in  $\mathcal{R}$ ,  $K'(x) = dK(x)/dx$ , and  $h$  is the bandwidth. Under Assumption A6 for the kernel function and bandwidth and previously stated regularity conditions on the superpopulations and sampling designs, the kernel-based estimator (14) is consistent for  $S'(\xi)$ .

In summary, the new replication variance estimation for  $\hat{\xi}_{\text{NNI}}$  proceeds as follows:

**Step 1.** Obtain a consistent kernel estimator  $\hat{\mu}_s(\hat{\xi}_{\text{NNI}}, x)$

**Step 2.** Construct replicates of  $\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}})$  as

$$\hat{S}_{\text{NNI}}^{(k)}(\hat{\xi}_{\text{NNI}}) = \sum_{i \in A} \omega_i^{(k)} [\hat{\mu}_s(\hat{\xi}_{\text{NNI}}, x_i) + \delta_i(1 + k_i) \{s(y_i - \hat{\xi}_{\text{NNI}}) - \hat{\mu}_s(\hat{\xi}_{\text{NNI}}, x_i)\}]. \quad (15)$$

**Step 3.** Apply  $\hat{V}_{\text{rep}}(\cdot)$  in (12) for the above replicates to obtain the replication variance estimator of  $\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}})$ , denoted as  $\hat{V}_{\text{rep}}\{\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}})\}$ .

**Step 4.** Obtain the kernel-based derivative estimator  $\hat{S}'(\hat{\xi}_{\text{NNI}})$ , where  $\hat{S}'(\xi)$  is defined in (14).

**Step 5.** Calculate the variance estimator of  $\hat{\xi}_{\text{NNI}}$  as  $\hat{S}'(\hat{\xi}_{\text{NNI}})^{-2}\hat{V}_{\text{rep}}\{\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}})\}$ .

For illustration, we continue with Example 1.

**Example 2 (Quantile estimation (Cont.))** Obtain kernel-based estimators for  $F(\xi) = \text{pr}(y \leq \xi)$  and  $f(\xi)$ , denoted as  $\hat{F}(\xi)$  and  $\hat{f}(\xi)$ , respectively. Construct replicates of  $\hat{F}_{\text{NNI}}(\hat{\xi}_{\alpha, \text{NNI}})$  as  $\hat{F}_{\text{NNI}}^{(k)}(\hat{\xi}_{\alpha, \text{NNI}}) = \sum_{i \in A} \omega_i^{(k)} [\hat{F}(\hat{\xi}_{\alpha, \text{NNI}}) + \delta_i(1 + k_i)\{I(y_i \leq \hat{\xi}_{\alpha, \text{NNI}}) - \hat{F}(\hat{\xi}_{\alpha, \text{NNI}})\}]$ . Apply  $\hat{V}_{\text{rep}}(\cdot)$  in (12) for the above replicates to obtain the replication variance estimator of  $\hat{F}_{\text{NNI}}(\hat{\xi}_{\alpha, \text{NNI}})$ , denoted as  $\hat{V}_{\text{rep}}\{\hat{F}_{\text{NNI}}(\hat{\xi}_{\alpha, \text{NNI}})\}$ . Calculate the variance estimator of  $\hat{\xi}_{\alpha, \text{NNI}}$  as  $\hat{f}(\hat{\xi}_{\alpha, \text{NNI}})^{-2}\hat{V}_{\text{rep}}\{\hat{F}_{\text{NNI}}(\hat{\xi}_{\alpha, \text{NNI}})\}$ .

**Theorem 3** Under the assumptions in Theorem 2, suppose that  $\hat{V}_{\text{rep}}(\hat{\mu}_g)$  in (12) is consistent for  $\text{var}_p(\hat{\mu}_g)$ . Then, if  $nN^{-1} = o(1)$ , the replication variance estimators for  $\hat{\mu}_{g, \text{NNI}}$  is consistent, i.e.,  $n\hat{V}_{\text{rep}}\{\hat{\mu}_{g, \text{NNI}}\}/V_g \rightarrow 1$  in probability, as  $n \rightarrow \infty$ , where  $\hat{V}_{\text{rep}}(\cdot)$  is given in (12), the replicates of  $\hat{\mu}_{g, \text{NNI}}$  are given in (13), and  $V_g$  is given in (8).

Given that the kernel-based estimator  $\hat{S}'(\xi)$  in (14) is consistent for  $S'(\xi)$ , the replication variance estimators for  $\hat{\xi}_{\text{NNI}}$  is consistent, i.e.,  $n\hat{V}_{\text{rep}}\{\hat{\xi}_{\text{NNI}}\}/V_\xi \rightarrow 1$  in probability, as  $n \rightarrow \infty$ , where  $\hat{V}_{\text{rep}}(\cdot)$  is given in (12), the replicates of  $\hat{S}_{\text{NNI}}^{(k)}(\hat{\xi}_{\text{NNI}})$  are given in (15), and  $V_\xi$  is given in (10).

The formal proof follows by straightforward asymptotic bounding arguments from the assumptions and therefore is omitted.

## 5 Simulation study

In this simulation study, we investigate the performance of the proposed replication variance estimation. For generating finite populations of size  $N = 50,000$ : first, let  $x_{1i}$ ,  $x_{2i}$  and  $x_{3i}$  be generated independently from Uniform $[0, 1]$ , and  $x_{4i}$ ,  $x_{5i}$  and  $x_{6i}$  and  $e_i$  be generated independently from  $\mathcal{N}(0, 1)$ ; then, let  $y_i$  be generated as (P1)  $y_i = -1 + x_{1i} + x_{2i} + e_i$ , (P2)  $y_i = -1.5 + x_{1i} + x_{2i} + x_{3i} + x_{4i} + e_i$ , (P3)  $y_i = -1.5 + x_{1i} + \dots + x_{6i} + e_i$ , (P4)  $y_i = -1 + x_{1i} + x_{2i} + x_{1i}^2 + x_{2i}^2 - 2/3 + e_i$ , (P5)  $y_i = -1.5 + x_{1i} + x_{2i} + x_{3i} + x_{4i} + x_{1i}^2 + x_{2i}^2 - 2/3 + e_i$  and (P6)  $y_i = -1.5 + x_{1i} + \dots + x_{6i} + x_{1i}^2 + x_{2i}^2 - 2/3 + e_i$ . The covariates are fully observed, but  $y_i$  is not. The response indicator of  $y_i$ ,  $\delta_i$ , is generated from Bernoulli( $p_i$ ) with  $\text{logit}\{p(x_i)\} = x_i^T \mathbf{1}$ , where  $x_i$  includes all corresponding covariates under each data generating mechanism and  $\mathbf{1}$  is a vector of 1 with a compatible length. This results in the average response rate about 75%. The parameters of interest are  $\mu = N^{-1} \sum_{i=1}^N y_i$ ,  $\eta = N^{-1} \sum_{i=1}^N I(y_i < c)$ , where  $c$  is the 80th quantile such that the true value of  $\eta$  is 0.8, and the median  $\xi$ . To generate samples, we consider two sampling designs: (S1) simple random sampling with  $n = 800$ ; (S2) probability proportional to size sampling. In (S2), for each unit in the population, we generate a size variable  $s_i$  as  $\log(|y_i + \nu_i| + 4)$ , where  $\nu_i \sim \mathcal{N}(0, 1)$ . The selection probability is specified as  $\pi_i = 400s_i / \sum_{i=1}^N s_i$ . Therefore, (S2) is informative, where units with larger  $y_i$  values have larger probabilities to be selected into the sample.

For nearest neighbor imputation, the matching scalar variable  $m$  is set to be the conditional mean function of  $y$  given  $x$ ,  $m(x)$ , approximated by power series estimation. For investigating the effect of the matching variable, we



consider the power series including all first and second order terms under (P1)–(P3) and only first order terms under (P4)–(P6), so that  $m(x)$  is accurate for the mean function under (P1)–(P3) but inaccurate under (P4)–(P6). We construct 95% confidence intervals using  $(\hat{\mu}_I - z_{0.975}\hat{V}_I^{1/2}, \hat{\mu}_I + z_{0.975}\hat{V}_I^{1/2})$ , where  $\hat{\mu}_I$  is the joint estimate and  $\hat{V}_I$  is the variance estimate obtained by the proposed jackknife variance estimation and a naive jackknife variance estimation that calculates a sample estimator for each replicate. For the jackknife replication method under (S2), in the  $k$ th replicate, the replication weights are  $\omega_i^{*(k)} = n\omega_i/(n-1)$  for all  $i \neq k$ , and  $\omega_k^{*(k)} = 0$ . In the proposed jackknife variance estimation, the  $k$ th replicates of  $\hat{\mu}_{\text{NNI}}$ ,  $\hat{\eta}_{\text{NNI}}$  and  $\hat{\xi}_{\text{NNI}}$  are given by

$$\begin{aligned}\hat{\mu}_{\text{NNI}}^{(k)} &= \sum_{i=1}^n \omega_i^{(k)} [\hat{\mu}(x_i) + \delta_i(1+k_i)\{I(y_i < c) - \hat{\mu}(x_i)\}], \\ \hat{\eta}_{\text{NNI}}^{(k)} &= \sum_{i=1}^n \omega_i^{(k)} [\hat{\mu}_\eta(x_i) + \delta_i(1+k_i)\{I(y_i < c) - \hat{\mu}_\eta(x_i)\}], \\ \hat{\xi}_{\text{NNI}}^{(k)}(\hat{\xi}_{\text{NNI}}) &= \hat{f}(\hat{\xi}_{\text{NNI}})^{-2} \sum_{i=1}^n \omega_i^{(k)} [\hat{\mu}_s(\hat{\xi}_{\text{NNI}}, x_i) + \delta_i(1+k_i)\{I(y_i \leq \hat{\xi}_{\text{NNI}}) - \hat{\mu}_s(\hat{\xi}_{\text{NNI}}, x_i)\}],\end{aligned}$$

where  $\hat{\mu}_\eta(x)$ ,  $\hat{\mu}_s(\xi, x)$  and  $\hat{f}(x)$  are nonparametric estimators of  $\mu_\eta(x) = \text{pr}(y < c \mid x)$ ,  $\mu_s(\xi, x) = \text{pr}(y < \xi \mid x)$  and  $f(\xi)$ , respectively, and  $k_i$  is the number of times that  $y_i$  is selected to impute the missing values of  $y$  based on the original data. These are obtained by kernel regression using a Gaussian kernel with bandwidth  $h = 1.5n^{-1/5}$ . The variance estimators are compared in terms of empirical coverage rate and relative bias,  $\{E(\hat{V}_I) - V\}/V$ , where  $V$  is the true variance simulated by Monte Carlo.

Tables 1 and 2 present the simulation results under simple random sampling and probability proportional to size sampling, respectively, based on

2,000 Monte Carlo samples. Under both sampling designs, the nearest neighbor imputation estimator has small biases for all parameters  $\mu$ ,  $\eta$  and  $\xi$ , under (P1)–(P3) with  $m(x)$  accurate approximation for the mean function and (P4)–(P6) with  $m(x)$  inaccurate approximation of the mean function. For variance estimation, as expected, the naive jackknife variance estimator is severely biased, indicating that the lack of smoothness of the matching estimator needs to be taken into account in variance estimation. In contrast, the proposed jackknife variance estimators provide satisfactory results under both sampling designs and for all parameters. The relative biases are small and the empirical coverage rates are close to the nominal coverage. Overall, the simulation results suggest that the proposed variance estimator works reasonably well under the settings we considered.

## 6 Discussion

Instead of choosing the nearest neighbor as a donor for missing items, we can consider fractional imputation (Kim and Fuller; 2004; Yang and Kim; 2016) using  $K$  ( $K > 1$ ) nearest neighbors. Such extension remains an interesting avenue for future research.

## Appendix

The Appendix includes proofs of Theorems 1 and 2 and additional assumptions.

Table 1: Simulation results for the population mean  $\mu$ , the population proportion  $\eta = 0.8$  and the population median  $\xi$  under simple random sampling: Bias ( $\times 10^2$ ) and S.E. ( $\times 10^2$ ) of the point estimator, Relative Bias of jackknife variance estimates ( $\times 10^2$ ) and Coverage Rate (%) of 95% confidence intervals.

Simple Random Sampling								
	$m(x)$	Bias	S.E.	Prop JK		Naive JK		
				RB	CR	RB	CR	
$\mu$	(P1)	a	0.00	4.87	0.1	94.9	>1000	100
	(P2)	a	0.12	6.08	0.5	95.3	>1000	100
	(P3)	a	1.09	8.42	2.2	95.3	>1000	100
	(P4)	i	-0.10	5.41	3.6	96.0	>1000	100
	(P5)	i	0.20	6.59	0.1	95.4	>1000	100
	(P6)	i	1.17	8.81	0.3	94.8	>1000	100
$\eta$	(P1)	a	0.00	1.77	0.4	95.0	>1000	100
	(P2)	a	0.00	1.53	-0.1	94.9	>1000	100
	(P3)	a	-0.01	1.50	-5.1	94.7	>1000	100
	(P4)	i	0.03	1.63	6.1	95.4	>1000	100
	(P5)	i	0.05	1.48	4.3	95.5	>1000	100
	(P6)	i	-0.01	1.47	-0.7	94.9	>1000	100
$\xi$	(P1)	a	-0.25	6.15	2.7	94.8	>1000	100
	(P2)	a	-0.40	7.60	2.5	94.7	>1000	100
	(P3)	a	-0.37	10.19	4.0	94.6	>1000	100
	(P4)	i	-0.25	7.09	3.2	94.6	>1000	100
	(P5)	i	-0.35	8.17	7.2	96.0	>1000	100
	(P6)	i	-0.54	10.78	1.8	94.1	>1000	100

Prop JK: proposed jackknife variance estimation; Naive JK: naive jackknife variance estimation. a: accurate and i: inaccurate.

Table 2: Simulation results for the population mean  $\mu$ , the population proportion  $\eta = 0.8$  and the population median  $\xi$  under probability proportional to size sampling: Bias ( $\times 10^2$ ) and S.E. ( $\times 10^2$ ) of the point estimator, Relative Bias of jackknife variance estimates ( $\times 10^2$ ) and Coverage Rate (%) of 95% confidence intervals.

		Probability Proportional to Size						
		$m(x)$	Bias	S.E.	Prop JK		Naive JK	
					RB	CR	RB	CR
$\mu$	(P1)	a	0.07	4.71	1.8	95.4	>1000	100
	(P2)	a	0.20	5.71	6.1	95.9	>1000	100
	(P3)	a	0.73	7.71	6.0	96.1	>1000	100
	(P4)	i	-0.06	5.29	2.4	95.5	>1000	100
	(P5)	i	0.22	6.08	7.0	95.9	>1000	100
	(P6)	i	0.99	8.23	5.4	95.1	>1000	100
$\eta$	(P1)	a	-0.01	1.89	-6.0	94.5	>1000	100
	(P2)	a	0.02	1.63	-1.9	95.3	>1000	100
	(P3)	a	0.08	1.66	-5.5	94.4	>1000	100
	(P4)	i	0.02	1.79	-4.0	95.2	>1000	100
	(P5)	i	0.03	1.60	1.8	95.2	>1000	100
	(P6)	i	0.08	1.67	-8.7	93.7	>1000	100
$\xi$	(P1)	a	-0.31	6.34	6.2	94.8	>1000	100
	(P2)	a	-0.06	8.30	0.8	94.5	>1000	100
	(P3)	a	-0.42	11.36	5.4	94.6	>1000	100
	(P4)	i	-0.32	7.57	4.1	94.0	>1000	100
	(P5)	i	-0.34	8.91	7.0	94.8	>1000	100
	(P6)	i	-0.49	12.22	2.2	94.4	>1000	100

Prop JK: proposed jackknife variance estimation; Naive JK: naive jackknife variance estimation. a: accurate and i: inaccurate.

## A7 Proof for Theorem 1

With a scalar matching variable  $m$ , we have

$$\begin{aligned} B_N &= \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} (1 - \delta_i) \{ \mu_g(x_{i(1)}) - \mu_g(x_i) \} \\ &\leq \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} (1 - \delta_i) | m_{i(1)} - m_i | = o_p(1), \end{aligned}$$

where  $\leq$  in the second line follows by Assumption 3 (ii). Based on the decomposition in (5), we can write

$$n^{1/2}(\hat{\mu}_{g,\text{NNI}} - \mu_g) = D_N + o_p(1), \quad (\text{A1})$$

where  $D_N$  is defined in (6). Then, to study the asymptotic properties of  $n^{1/2}(\hat{\mu}_{g,\text{NNI}} - \mu_g)$ , we only need to study the asymptotic properties of  $D_N$ . For simplicity, we introduce the following notation:  $\mu_{g,i} = \mu_g(x_i) \equiv E\{g(y) \mid x_i\}$  and  $e_i = g(y_i) - \mu_{g,i}$ . We express

$$\begin{aligned} D_N &= \frac{n^{1/2}}{N} \left[ \sum_{i \in A} \frac{1}{\pi_i} \{ \mu_{g,i} + \delta_i(1 + k_i)e_i \} - \sum_{i=1}^N g(y_i) \right] \\ &= \frac{n^{1/2}}{N} \sum_{i=1}^N \left( \frac{I_i}{\pi_i} - 1 \right) \mu_{g,i} + \frac{n^{1/2}}{N} \sum_{i=1}^N \left\{ \frac{I_i}{\pi_i} \delta_i(1 + k_i) - 1 \right\} e_i, \quad (\text{A2}) \end{aligned}$$

and we can verify that the covariance of the two terms in (A2) is zero. Thus,

$$\text{var}(D_N) = \text{var} \left\{ \frac{n^{1/2}}{N} \sum_{i=1}^N \left( \frac{I_i}{\pi_i} - 1 \right) \mu_{g,i} \right\} + \text{var} \left[ \frac{n^{1/2}}{N} \sum_{i=1}^N \left\{ \frac{I_i}{\pi_i} \delta_i(1 + k_i) - 1 \right\} e_i \right].$$

The first term, as  $n \rightarrow \infty$ , becomes

$$V_g^\mu = \lim_{n \rightarrow \infty} \frac{n}{N^2} E \left\{ \text{var}_p \left( \sum_{i \in A} \frac{\mu_{g,i}}{\pi_i} \right) \right\},$$

and the second term, as  $n \rightarrow \infty$ , becomes

$$V_g^e = \text{plim} \frac{n}{N^2} \sum_{i=1}^N \left\{ \frac{I_i}{\pi_i} \delta_i (1 + k_i) - 1 \right\}^2 \text{var}(e_i | x_i).$$

The remaining is to show that  $V_g^e = O(1)$ . To do this, the key is to show that the moments of  $k_i$  are bounded. Under Assumption 2, it is easy to verify that

$$\underline{\omega} \tilde{k}_i \leq k_i \leq \bar{\omega} \tilde{k}_i, \quad (\text{A3})$$

for some constants  $\underline{\omega}$  and  $\bar{\omega}$ , where  $\tilde{k}_i = \sum_{j=1}^n (1 - \delta_j) d_{ij}$  is the number of unit  $i$  used as a match for the nonrespondents. Under Assumption 3,  $\tilde{k}_i = O_p(1)$  and  $E(\tilde{k}_i)$  and  $E(\tilde{k}_i^2)$  are uniformly bounded over  $n$  (Abadie and Imbens; 2006, Lemma 3); therefore, together with (A3), we have  $k_i = O_p(1)$  and  $E(k_i)$  and  $E(k_i^2)$  are uniformly bounded over  $n$ . Therefore, a simple algebra yields  $V_g^e = O(1)$ .

Combining all results, the asymptotic variance of  $n^{1/2}(\hat{\mu}_{g,\text{NNI}} - \mu_g)$  is  $V_g^\mu + V_g^e$ . By the central limit theorem, the result in Theorem 1 follows.

## A8 Proof for Theorem 2

We impose the following assumptions for the population parameter  $\xi_N$  and the population estimating function  $S_N(\cdot)$ ; see also Wang et al. (2011).

**Assumption A4 (i)** *The population parameter  $\xi_N$  lies in a closed interval*

$$\mathcal{I}_\xi \text{ on } \mathcal{R};$$

**(ii)** *the function  $s(\cdot)$  is bounded;*

- (iii) the population estimating function  $S_N(\xi)$  converges to  $S(\xi)$  uniformly on  $\mathcal{I}_\xi$  as  $N \rightarrow \infty$ , and the equation  $S(\xi) = 0$  has a unique root in the interior of  $\mathcal{I}_\xi$ ;
- (iv) the limiting function  $S(\xi)$  is strictly increasing and absolutely continuous with finite first derivative in  $\mathcal{I}_\xi$ , and the derivative  $S'(\xi)$  is bounded away from 0 for  $\xi$  in  $\mathcal{I}_\xi$ ;
- (v) the population quantities

$$\sup_{\xi \in \mathcal{I}_s} N^\alpha |S_N(\xi_N + N^{-\alpha}\xi) - S_N(\xi_N) - S(\xi_N + N^{-\alpha}\xi) + S(\xi_N)| \rightarrow 0,$$

and

$$\sup_{\xi \in \mathcal{I}_s} N^{-1} \sum_{i=1}^N |s(y_i - \xi_N - N^{-\alpha}\xi) - s(y_i - \xi_N)| = O_p(N^{-\alpha}),$$

where  $\mathcal{I}_s$  is a large enough compact set in  $\mathcal{R}$  and  $\alpha \in (1/4, 1/2]$ .

Assumption A4 (v) holds with probability one under suitable assumptions on the probability mechanism generating the  $y_i$ 's and on the function  $s(\cdot)$ , and therefore is justifiable. Under Assumption A4, by the standard arguments from the theory on M-estimators (Serfling; 1980),  $\hat{\xi}_{\text{NNI}}$  is consistent for  $\xi_N$ . We further make the following assumption.

**Assumption A5** *The nearest neighbor imputation estimator  $\hat{\xi}_{\text{NNI}}$  is root- $n$  consistent for  $\xi_N$ ,*

Now, we give proof for Theorem 2. Under Assumptions A4 and A5, we can write

$$\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}}) - S_N(\xi_N) = \{\hat{S}_{\text{NNI}}(\xi_N) - S_N(\xi_N)\} + S'(\xi_N)(\hat{\xi}_{\text{NNI}} - \xi_N) + o_p(n^{-1/2}). \quad (\text{A4})$$

By Assumption A4 (iv),  $S(\xi)$  is smooth, and therefore  $S_N(\xi_N) = O_p(N^{-1})$ ,  $\hat{S}_{\text{NNI}}(\hat{\xi}_{\text{NNI}}) = O_p(n^{-1})$ , and the left hand side of (A4) is  $o_p(n^{-1/2})$ . Therefore, we can obtain a linearization for  $\hat{\xi}_{\text{NNI}}$  as in (9).

Based on the linearization (9), the asymptotic variance  $V_\xi = \dot{S}(\xi)^{-2} \text{var}\{\hat{S}_{\text{NNI}}(\xi)\}$ . Following a similar derivation in the proof for Theorem 1, it is easy to show that

$$\begin{aligned} \text{var}\{\hat{S}_N(\xi)\} &= \lim_{n \rightarrow \infty} \frac{n}{N^2} E \left( \text{var}_p \left[ \sum_{i \in A} \frac{E\{s(y_i - \xi) \mid x_i\}}{\pi_i} \right] \right) \\ &+ \text{plim} \frac{n}{N^2} \sum_{i=1}^N \left\{ \frac{I_i}{\pi_i} \delta_i (1 + k_i) - 1 \right\}^2 \text{var} [s(y_i - \xi) - E\{s(y_i - \xi) \mid x_i\} \mid x_i]. \end{aligned}$$

## A9 Assumptions

**Assumption A6** *The following conditions hold for kernel function  $K(\cdot)$  and bandwidth  $h$ :*

- (i) *the kernel function  $K(\cdot)$  is absolutely continuous with nonzero finite derivative  $K'(\cdot)$  and  $\int K(x)dx = 1$ ;*
- (ii) *the bandwidth  $h \rightarrow 0$  and  $nh \rightarrow \infty$  as  $n \rightarrow \infty$ ;*
- (iii) *there exists a constant  $c$ , such that  $|h^{-1}K'(x_1/h) - h^{-1}K'(x_2/h)| \leq c|x_1 - x_2|$  for any  $x_1, x_2$  and  $h$  arbitrarily small.*

Assumption A6 states conditions on the smoothness and tail behavior of the kernel functions. Popular kernel functions, including Epanechnikov, Gaussian, and triangle kernels, satisfy the required conditions.



## References

- Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects, *Econometrica* **74**: 235–267.
- Berger, Y. G. and Skinner, C. J. (2003). Variance estimation for a low income proportion, *Journal of the Royal Statistical Society: series C (applied statistics)* **52**: 457–468.
- Chen, J. and Shao, J. (2000). Nearest neighbor imputation for survey data, *J. Offic. Stat.* **16**: 113–131.
- Chen, J. and Shao, J. (2001). Jackknife variance estimation for nearest-neighbor imputation, *J. Amer. Statist. Assoc.* **96**: 260–269.
- Deville, J. C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques, *Surv. Methodol.* **25**: 193–204.
- Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design, *Ann. Statist.* **19**: 454–469.
- Fuller, W. A. (2009). *Sampling Statistics*, Wiley, Hoboken.
- Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model, *J. Amer. Statist. Assoc.* **77**: 89–96.
- Kim, J. K. and Fuller, W. A. (2004). Fractional hot deck imputation, *Biometrika* **91**: 559–578.
- Kim, J. K., Fuller, W. A., Bell, W. R. et al. (2011). Variance estimation for

- nearest neighbor imputation for US Census long form data, *The Annals of Applied Statistics* **5**: 824–842.
- Kim, J. K., Navarro, A. and Fuller, W. A. (2006). Replication variance estimation for two-phase stratified sampling, *J. Amer. Statist. Assoc.* **101**: 312–320.
- Lee, H. and Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values, *J. Offic. Stat.* **10**: 231–243.
- Otsu, T. and Rai, Y. (2016). Bootstrap inference of matching estimators for average treatment effects, *J. Amer. Statist. Assoc.* p. DOI:10.1080/01621459.2016.1231613.
- Rust, K. F. and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques, *Stat Methods Med Res* **5**: 283–310.
- Sande, I. G. (1979). A personal view of hot deck imputation procedures, *Surv. Methodol.* **5**: 238–258.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Hoboken, NJ: Wiley.
- Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions, *J. Amer. Statist. Assoc.* **94**: 254–265.
- Shao, J. and Wang, H. (2008). Confidence intervals based on survey data with nearest neighbor imputation, *Statist. Sinica* **18**: 281–297.

- Wang, J. C., Opsomer, J. D. et al. (2011). On asymptotic normality and variance estimation for nondifferentiable survey estimators, *Biometrika* **98**: 91–106.
- Wolter, K. (2007). *Introduction to Variance Estimation*, 2 edn, Springer, New York.
- Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data, *J. Amer. Statist. Assoc.* **96**: 185–193.
- Yang, S. and Kim, J. K. (2016). Fractional imputation in survey sampling: A comparative review, *Statist. Sci.* **31**: 415–432.