Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Nearest neighbour imputation under single index models

Jun Shao & Lei Wang

STATISTICAL THEORY AND RELATED FIELDS
2019, VOL. 3, NO. 2, 208–212
https://doi.org/10.1080/24754269.2019.1675409

Nearest neighbour imputation under single index models

Jun Shao[a,b] and Lei Wang[c]

[a]School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]Department of Statistics, University of Wisconsin, Madison, WI, USA; [c]School of Statistics and Data Science & LPMC, Nankai University, Tianjin, People's Republic of China

ABSTRACT
A popular imputation method used to compensate for item nonresponse in sample surveys is the nearest neighbour imputation (NNI) method utilising a covariate to defined neighbours. When the covariate is multivariate, however, NNI suffers the well-known curse of dimensionality and gives unstable results. As a remedy, we propose a single-index NNI when the conditional mean of response given covariates follows a single index model. For estimating the population mean or quantiles, we establish the consistency and asymptotic normality of the single-index NNI estimators. Some limited simulation results are presented to examine the finite-sample performance of the proposed estimator of population mean.

ARTICLE HISTORY
Received 14 September 2019
Revised 29 September 2019
Accepted 30 September 2019

KEYWORDS
Asymptotic normality; curse of dimensionality; imputation; mean; quantiles; SAVE

1. Introduction

Let $\mathcal{P}$ be a finite population containing N units indexed by i, y_i be a univariate outcome or response of interest from unit $i \in \mathcal{P}$, x_i be a covariate vector associated with y_i, and let $\mathcal{S} \subset \mathcal{P}$ be a sample of size n taken from $\mathcal{P}$ according to some sampling design. We consider the situation where x_i is always observed if $i \in \mathcal{S}$ but y_i is subject to nonresponse, i.e., y_i is observed if and only if $i \in \mathcal{R} \subset \mathcal{S}$. In sample surveys, imputation is commonly applied to compensate for nonresponse (Kalton & Kasprzyk, 1986; Rubin, 1987; Sedransk, 1985). The nearest neighbour imputation (NNI) method imputes a missing y_j by y_l, where $l \in \mathcal{R}$ is the nearest neighbour of j in the sense that $d(x_j, x_l) = \min_{i\in\mathcal{R}} d(x_i, x_j)$ and $d(x_i, x_j)$ is a distance between x_i and x_j, e.g., the Euclidean distance. It is a popular method in many survey agencies and 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 & 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 & Ponikowski, 1993), and the Unified Enterprise Survey, the Survey of Household Spending, and the Financial Farm Survey conducted by Statistics Canada (Rancourt, 1999).

The NNI method has some nice features. First, imputed values are actually occurring y-values, not constructed values; they may not be perfect substitutes, but are unlikely to be nonsensical values. Second, the NNI method may be more efficient than imputation not using x-values, such as mean imputation or random imputation, when x 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 ratio or regression imputation based on a linear regression model. Finally, under some conditions NNI estimators (i.e., estimators calculated using standard formulas and treating nearest neighbour imputed values as observed data) are asymptotically valid not only for moments of y_i but also for the distribution and quantiles of y_i, which is a superiority over other non-random imputation methods (such as mean, ratio or regression imputation) that lead to valid moment estimators only.

For a univariate covariate x_i, some asymptotic properties of NNI are established in Chen and Shao (2000, 2001) and Shao and Wang (2008). When x_i is multivariate, however, NNI runs into the curse of dimensionality problem. The purpose of this paper is to propose a single-index NNI method for multivariate x_i and derive its asymptotic properties, under the following single index model assumption:

(A1) The population $\mathcal{P}$ can be partitioned into K subpopulations, $\mathcal{P}_1, \ldots, \mathcal{P}_K$, such that for within each $\mathcal{P}_k$, (x_i, y_i)'s are independent and identically distributed (i.i.d.) from a superpopulation with

$$E(y_i \mid x_i) = \mu_k(\beta_k' x_i),$$

where β_k' is the transpose of an unknown parameter vector β_k with the same dimension as x_i and $\mu_k(\cdot)$ is an unspecified function, $k = 1, \ldots, K$.

Imputation for nonrespondents are typically done within each $\mathcal{P}_k$ and, hence, $\mathcal{P}_k$'s are often referred to

CONTACT Lei Wang lwangstat@nankai.edu.cn

as imputation classes. They 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_i, y_i) within each imputation class ensures exchangeability of units within each $\mathcal{P}_k$. The single index model assumption is a semiparametric assumption, since μ_k is unspecified.

Details of the proposed method are presented in Section 2, where we also show that estimators based on single-index NNI are consistent and asymptotically normal under some limiting process as the sample size n increases to infinity. To complement the theory, some simulation results are presented in Section 3 to examine the finite sample performance of proposed estimators.

2. Method and theory

We consider one stage sampling without clusters. Let w_i be the survey weight for unit $i \in \mathcal{P}$, which is equal to the inverse of probability that unit i is selected, a known quantity from sampling design. When there is no nonresponse, a simple and popular estimator of the unknown population total $Y = \sum_{i\in\mathcal{P}} y_i$ is the Horvitz–Thompson estimator $\hat{Y} = \sum_{i\in\mathcal{S}} w_i y_i$, which has the unbiasedness property

$$E_s(\hat{Y}) = E_s\left(\sum_{i\in\mathcal{S}} w_i y_i\right) = \sum_{i\in\mathcal{P}} y_i = Y, \qquad (1)$$

where E_s is the expectation with respect to sampling. If the total number of units in $\mathcal{P}$, N, is known, then the population mean Y/N is estimated by $\hat{Y}/N$. If N is unknown, then Y/N is estimated by $\hat{Y}/\hat{N}$, where $\hat{N} = \sum_{i\in\mathcal{S}} w_i$ satisfying $E_s(\hat{N}) = N$.

The most important population parameter in a survey study concerning a variable y is the population mean. Estimation of population quantiles has also become more and more important in modern survey studies. For income variables, for example, the median income or other quantiles could be 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). Let $I(y_i \le t)$ be the indicator of $y_i \le t$ for any fixed value t. Using property (1) with y_i replaced by $I(y_i \le t)$, we obtain an approximately unbiased estimator $\sum_{i\in\mathcal{S}} w_i I(y_i \le t)/\hat{N}$ of the population cumulative distribution of y_i at t, which further leads to an approximately unbiased estimator of any quantile of the distribution of y_i.

When y_i has nonresponse, however, the previously discussed estimators cannot be used. Imputation is a popular technique to handle nonresponse. It fills in a value for every nonrespondent y_j, such that an unbiased or approximate unbiased estimator can be obtained using the formula for the situation of no nonresponse with imputed values treated as observed values. That is, if $\hat{y}_j$ is an imputed value for nonrespondent y_j, then our estimator of the population total Y is

$$\hat{Y}_I = \sum_{i\in\mathcal{R}} w_i y_i + \sum_{j\in\bar{\mathcal{R}}} w_j \hat{y}_j, \qquad (2)$$

where $\mathcal{R}$ and $\bar{\mathcal{R}}$ are the sets of respondents and nonrespondents, respectively, in the sample $\mathcal{S} = \mathcal{R} \cup \bar{\mathcal{R}}$.

Under (A1), we consider NNI within each imputation class and independently across imputation classes. For a multivariate x_i, if β_k in (A1) is known, we can apply a single-index NNI by defining the distance between x_i and x_j as $|\beta_k' x_i - \beta_k' x_j|$, to avoid the curse of dimensionality issue in multivariate NNI. As β_k is generally unknown, we can first estimate β_k by $\hat{\beta}_k$ using a nonparametric method such as the sliced inverse regression (SIR) proposed by Li and Duan (1991) or the sliced average variance estimation (SAVE) proposed by Cook and Weisberg (1991), and then apply single-index NNI using $|\hat{\beta}_k' x_i - \hat{\beta}_k' x_j|$ as the distance between x_i and x_j, i.e., a nonrespondent y_j in imputation class k is imputed by $\hat{y}_j = y_l$ with l satisfying

$$|\hat{\beta}_k' x_l - \hat{\beta}_k' x_j| = \min_{i\in\mathcal{R}\cap\mathcal{P}_k} |\hat{\beta}_k' x_i - \hat{\beta}_k' x_j|. \qquad (3)$$

After imputation, the population total Y is estimated by $\hat{Y}_I$ in (2) with $\hat{y}_j$ defined by (3). The population cumulative distribution of y_i at any t is estimated by

$$\hat{F}_I(t) = \frac{1}{\hat{N}}\left\{\sum_{i\in\mathcal{R}} w_i I(y_i \le t) + \sum_{j\in\bar{\mathcal{R}}} w_j I(\hat{y}_j \le t)\right\},$$

regardless whether N is known or unknown (to ensure that the estimate $\to 1$ when $t \to \infty$).

To consider asymptotic properties of estimators based on singe-index NNI, we assume that the finite population $\mathcal{P}$ is a member of a sequence of finite populations indexed by ν. All limiting processes in this paper are understood to be as $\nu \to \infty$. We need the following assumptions in addition to (A1).

(A2) The size of $\mathcal{P}_k$ and sample size of $\mathcal{S} \cap \mathcal{P}_k$ increase to infinity as $\nu \to \infty$, while the number of subpopulations, K, is fixed.

(A3) There is a fixed constant $c > 0$ (not depending on ν) such that

$$\max_{i\in\mathcal{P}} \frac{nw_i}{N} \leq c \quad \text{and} \quad \frac{n}{N^2} E_s \left(\sum_{i\in\mathcal{S}} w_i \right)^2 \leq c.$$

Recall that N is the size of $\mathcal{P}$ and n is the sample size. The first condition in (A3) ensures that none of the weights w_i's is disproportionately large (see Krewski & Rao, 1981). The second condition in (A3) means that the sampling variance of $\sum_{i\in\mathcal{S}} w_i/N$ is at most of the order n^{-1}. These conditions are typically satisfied, e.g., they are satisfied under stratified simple random sampling designs.

Let a_i be the response indicator, i.e., $a_i = 1$ if y_i is observed and $a_i = 0$ if y_i is a nonrespondent.

(A4) Within each $\mathcal{P}_k$, (x_i, y_i, a_i)'s are i.i.d. from a superpopulation with $E(y_i^8) < \infty$, (x_i, y_i, a_i)'s from different imputation classes are independent, and sampling is independent of the superpopulation.

(A5) Within each $\mathcal{P}_k$, under the superpopulation, $P(a_i = 1 \mid x_i, y_i, k) = P(a_i = 1 \mid x_i, k) > 0$, which is continuous in x_i.

(A6) Within each $\mathcal{P}_k$, the conditional distribution of x_i given a_i has a bounded and continuous Lebesgue density and $\mu_k(\cdot)$ in (A1) is a differentiable function.

(A7) Within each $\mathcal{P}_k$,

$$q_{k,i}(\gamma) = P\left(|\gamma' x - \gamma' x_i| = \min_{j\in\mathcal{R}_k} |\gamma' x - \gamma' x_j| \,\Big|\, \mathcal{X}_k, \mathcal{R}_k, \mathcal{S}_k \right)$$

is differentiable with respect to γ, where P is with respect to x under superpopulation, $\mathcal{S}_k = \mathcal{S} \cap \mathcal{P}_k$, $\mathcal{R}_k = \mathcal{R} \cap \mathcal{P}_k$, and $\mathcal{X}_k = \{x_i : i \in \mathcal{R}_k\}$.

(A8) For each k, $n^{1/2}(\hat\beta_k - \beta_k) = n^{-1/2} \sum_{i\in\mathcal{S}\cap\mathcal{P}_k} \phi(x_i, y_i, a_i) + o_p(1)$, where ϕ is a function satisfying $E\{\phi(x_i, y_i, a_i)\} = 0$ and $E\{\phi(x_i, y_i, a_i)\}^2 < \infty$, and $o_p(1)$ denotes a term converging to 0 in probability.

Because of (A4), NNI is carried out within each $\mathcal{S} \cap \mathcal{P}_k$. (A5) assumes that, within an imputation class, the nonresponse mechanism is covariate-dependent (Little, 1995) or unconfounded (Lee, Rancourt, & Särndal, 1994), an assumption made for the validity of many other popular imputation methods. This actually is the main reason to construct imputation classes, in addition to the exchangeability of (x_i, y_i)'s. Although (x_i, y_i, a_i)'s within an imputation class are i.i.d., the nonresponse mechanism is still not completely at random, since $P(a_i = 1 \mid x_i, k)$ depends on the covariate x_i. Finally, (A8) is satisfied if $\hat\beta_k$ is obtained using SIR (Li & Duan, 1991) or SAVE (Cook & Weisberg, 1991).

The following is our main theoretical result.

Theorem: *Assume* (A1)–(A8). *Let* $\hat Y_I$ *by defined by* (2) *with imputed* $\hat y_j$ *based on single-index NNI. Then*

$$\sqrt{n}\left(\frac{\hat Y_I}{N} - \frac{Y}{N} \right) \Big/ \sigma \to_d N(0, 1)$$

for some $\sigma > 0$, *where* $\to_d$ *is convergence in distribution unconditionally with respect to the superpopulation and sampling.*

Similar results can be obtained for $\hat F_I(t)$ with any t and quantiles related with $\hat F_I$.

Proof of Theorem: The proof follows the same argument in Shao and Wang (2008). Since variables are independent across imputation classes and imputation is carried out within each imputation class, it suffices to show the result within each imputation class or, equivalently, the result when $K = 1$. We now drop the subscript k in this proof. Let $\mathcal{S}$, $\mathcal{R}$ and $\mathcal{X}$ be defined as before with subscript k dropped. Then

$$E(\hat y_i \mid \mathcal{X}, \mathcal{R}, \mathcal{S}) = \sum_{i\in\mathcal{R}} q_i(\hat\beta) y_i,$$

where $q_i(\hat\beta)$ is the probability that $i \in \mathcal{R}$ is selected as the nearest neighbour of a nonrespondent and $q_i(\beta)$ is defined in (A7) with subscript k dropped. Define $\hat\mu_I = \hat Y_I/N$, $\mu = Y/N$, $\mu_1 = E(y_i \mid a_i = 1)$, $\mu_0 = E(y_i \mid a_i = 0)$, $p = P(a_i = 1)$, $\bar w_i = w_i/N$, $\hat e_i = \hat y_i - \sum_{i\in\mathcal{R}} q_i(\hat\beta) y_i$, $Q_1 = \sum_{i\in\bar{\mathcal{R}}} \bar w_i \hat e_i$, $\hat Q_2 = \sum_{i\in\mathcal{R}} \bar w_i [y_i - \mu(\hat\beta' x_i)] + (1 - p) \sum_{i\in\mathcal{R}} q_i(\hat\beta)[y_i - \mu(\hat\beta' x_i)]$, $\hat Q_3 = \sum_{i\in\mathcal{R}} \bar w_i [\mu(\hat\beta' x_i) - \mu_1]$, $\hat Q_4 = \sum_{i\in\bar{\mathcal{R}}} [\bar w_i - (1 - p)] \sum_{i\in\mathcal{R}} q_i(\hat\beta)[y_i - \mu(\hat\beta' x_i)] + \sum_{i\in\bar{\mathcal{R}}} \bar w_i [\sum_{i\in\mathcal{R}} q_i(\hat\beta)\mu(\hat\beta' x_i) - \mu_0]$, $Q_5 = (\mu_1 - \mu_0) \sum_{i\in\mathcal{S}} \bar w_i (a_i - p)$ and $Q_6 = \mu(\sum_{i\in\mathcal{S}} \bar w_i - 1)$. Also, for $l = 2, 3, 4$, define Q_l to be $\hat Q_l$ with $\hat\beta$ replaced by β. Then

$$\begin{aligned}\hat\mu_I - \mu &= Q_1 + \hat Q_2 + \hat Q_3 + \hat Q_4 + Q_5 + Q_6 \\ &= Q_1 + Q_2 + Q_3 + Q_4 + Q_5 + Q_6 \\ &\quad + (\hat Q_2 - Q_2) + (\hat Q_2 - Q_3) + (\hat Q_4 - Q_4).\end{aligned}$$

For each Q_l, it is shown in Shao and Wang (2008) that each $n^{1/2} Q_l$ is an approximately linear function of random variables converging in distribution to a normal distribution with mean 0. Under (A6)–(A8) and Taylor expansions, we can show that each $\hat Q_l - Q_l$, $l = 2, 3, 4$, can be approximated by a linear function of random variables converging in distribution to a normal distribution with mean 0. Hence, the result follows by repeatedly applying Lemma 1 in Schenker and Welsh (1988). ■

Table 1. Simulation bias and standard deviation (SD) in estimating μ (1000 runs).

				$n = 200$			$n = 500$		
Model	μ	ε dist		$\hat{\mu}$	$\tilde{\mu}_I$	$\hat{\mu}_I$	$\hat{\mu}$	$\tilde{\mu}_I$	$\hat{\mu}_I$
Linear	3	(a)	bias	−0.008	0.003	0.001	−0.003	0.001	−0.002
			SD	0.218	0.252	0.258	0.138	0.164	0.167
		(b)	bias	−0.002	0.000	0.003	0.005	0.009	0.005
			SD	0.208	0.242	0.246	0.130	0.152	0.151
		(c)	bias	0.001	0.009	0.013	−0.003	−0.002	0.012
			SD	0.314	0.388	0.394	0.202	0.252	0.259
Nonlinear	7.25	(a)	bias	−0.004	−0.035	−0.029	0.003	−0.003	0.000
			SD	0.579	0.597	0.600	0.377	0.383	0.387
		(b)	bias	0.003	−0.020	−0.017	−0.001	−0.009	−0.009
			SD	0.580	0.590	0.599	0.377	0.390	0.389
		(c)	bias	0.034	0.010	0.016	0.001	−0.014	−0.016
			SD	0.638	0.681	0.686	0.398	0.420	0.417

3. Simulation results

A simulation study is performed to examine the finite sample performance of $\hat{\mu}_I = \hat{Y}_I/N$ with $\hat{Y}_I$ defined in (2) and $w_i = N/n$. With sample of size $n = 200$ or 500, data $(x_1, y_1, a_1), \ldots, (x_n, y_n, a_n)$ are i.i.d. generated as follows. First, a three-dimensional covariate vector x_i is generated from the multivariate normal distribution with mean vector (1, 1, 1) and covariance matrix

$$\begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}$$

Conditioned on x_i, y_i is generated according to a linear model: $y_i = \beta' x_i + \varepsilon_i$, or a nonlinear model: $y_i = 0.5(\beta' x_i)^2 + \varepsilon_i$, where $\beta' = (1, 1, 1)$ and ε_i is generated from one of the following three distributions:

(a) normal distribution $N(0, 4)$,
(b) mixture normal distribution $0.4N(0, 1) + 0.6N(0, 9)$,
(c) heteroscedastic normal distribution $N(0, x_{i1}^2 + 1)$, where x_{i1} is the first component of x_i.

Conditioned on x_i, the response indicator a_i is generated from the Bernoulli distribution with probability

$$\pi(x_i) = 1/[1 + \exp(-0.4 - 0.1\beta' x_i)],$$

where the coefficients in $\pi(x_i)$ are chosen so that the unconditional rates of missing data are between 20% and 40%. For each i, x_i is observed and y_i is observed if and only if $a_i = 1$.

For simplicity, we consider $K = 1$ in (A1) and $N = n$. Then, $\hat{\mu}_I = \hat{Y}_I/n$ is considered as an estimator of the super-population mean $\mu = E(y_i)$, which is $\mu = 3$ under linear model and $\mu = 7.25$ under nonlinear model. To apply single-index NNI in (2), SAVE (Cook & Weisberg, 1991) is used to obtain estimator $\hat{\beta}$.

To evaluate the performance, we add two oracle estimators, in addition to $\hat{\mu}_I$. The first oracle estimator is $\hat{\mu} = \sum_{i=1}^n y_i/n$, the sample mean without nonresponse, assuming we observe all y_i's. The second oracle estimator is $\tilde{\mu}_I$, which is the same as $\hat{\mu}_I$ except that the true β, instead of $\hat{\beta}$, is used in finding the nearest neighbour.

Table 1 provides simulation bias and standard error (SD) of $\hat{\mu}$, $\tilde{\mu}_I$ and $\hat{\mu}_I$ based on 1000 runs. It can be seen from Table 1 that all biases are negligible. In terms of the SD, $\hat{\mu}_I$ based on single-index NNI is just slightly worse than the oracle estimator $\tilde{\mu}_I$ using the true β instead of $\hat{\beta}$.

The empirical results are consistent with our theoretical findings.

Acknowledgements

We would like to thank two referees for their comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by the National Natural Science Foundation of China grants 11831008 and 11871287, the U.S. National Science Foundation grants DMS-1612873 and DMS-1914411, the Natural Science Foundation of Tianjin (18JCYBJC41100) and the Fundamental Research Funds for the Central Universities.

Notes on contributors

Dr Jun Shao holds a PhD in statistics from the University of Wisconsin-Madison. He is a professor of statistics at the University of Wisconsin-Madison and East China Normal University. His research interests include variable selection and inference with high dimensional data,sample surveys, and missing data problems.

Dr Lei Wang holds a PhD in statistics from East China Normal University. He is an assistant professor of statistics at Nankai University. His research interests include empirical likelihood and missing data problems.

References

Chen, J., & Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics*, *16*, 113–132.

Chen, J., & Shao, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association*, *96*, 260–269.

Cook, R. D., & Weisberg, S. (1991). Discussion of 'Sliced inverse regression for dimension reduction'. *Journal of the American Statistical Association*, *86*, 28–33.

Farber, J. E., & Griffin, R. (1998). A comparison of alternative estimation methodologies for Census 2000. *Proceedings of the section on survey research methods* (pp. 629–634). American Statistical Association.

Fay, R. E. (1999). Theory and application of nearest neighbor imputation in Census 2000. *Proceedings of the section on survey research methods* (pp. 112–121). American Statistical Association.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing data. *Survey Methodology*, *12*, 1–16.

Kosorok, M. R. (1999). Two-sample quantile tests under general conditions. *Biometrika*, *86*, 909–921.

Krewski, D., & Rao, J. N. K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, *9*, 1010–1019.

Lee, H., Rancourt, E., & Särndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, *10*, 231–243.

Li, K. C., & Duan, N. (1991). Regression analysis under link violation. *The Annals of Statistics*, *17*, 1009–1052.

Little, R. J. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121.

Montaquila, J. M., & Ponikowski, C. H. (1993). Comparison of methods for imputing missing responses in an establishment survey. *Proceedings of the section on survey research methods* (pp. 446–451). American Statistical Association.

Rancourt, E. (1999). Estimation with nearest neighbor imputation at statistics Canada. *Proceedings of the section on survey research methods* (pp. 131–138). American Statistical Association.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schenker, N., & Welsh, A. H. (1988). Asymptotic results for multiple imputation. *The Annals of Statistics*, *16*, 1550–1566.

Sedransk, J. (1985). The objective and practice of imputation. *Proceedings of the first annual research conference* (pp. 445–452). Washington, DC: Bureau of the Census.

Shao, J., & Wang, H. (2008). Confidence intervals based on survey data with nearest neighbor imputation. *Statistica Sinica*, *18*, 281–297.

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, *88*, 89–96.