1-12-2018

Predictive mean matching imputation in survey sampling

Shu Yang

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Predictive mean matching imputation in survey sampling

Abstract
Predictive mean matching imputation is popular for handling item nonresponse in survey sampling. In this article, we study the asymptotic properties of the predictive mean matching estimator of the pop- ulation mean. For variance estimation, the conventional bootstrap inference for matching estimators with fixed 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 repli- cates of the estimator directly based on linear terms, instead of individual records of variables. Extension to nearest neighbor imputation is also discussed. A simulation study confirms that the new procedure provides valid variance estimation.

Keywords
Bootstrap, Jackknife variance estimation, Martingale central limit theorem, Missing at random

Disciplines
Design of Experiments and Sample Surveys | Statistics and Probability

Comments
This is a pre-print made available through arxiv: https://arxiv.org/abs/1703.10256.

Predictive mean matching imputation in survey sampling

Shu Yang   Jae Kwang Kim

Abstract

Predictive mean matching imputation is popular for handling item nonresponse in survey sampling. In this article, we study the asymptotic properties of the predictive mean matching estimator of the population mean. For variance estimation, the conventional bootstrap inference for matching estimators with fixed 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. Extension to nearest neighbor imputation is also discussed. A simulation study confirms that the new procedure provides valid variance estimation.

Key Words: Bootstrap; Jackknife variance estimation; Martingale central limit theorem; Missing at random.

1. Introduction

Predictive mean matching imputation (Rubin; 1986; Little; 1988) is popular for handling item nonresponse in survey sampling. Hot deck imputation within imputation cells is a special case, where the predictive mean function is constant within cells. On the other hand, predictive mean matching is a version of nearest neighbor imputation. In nearest neighbor imputation, the vector of the auxiliary variables x is directly used in determining the nearest neighbor, while in predictive mean matching imputation, a scalar predictive mean function is 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. Kim et al. (2011) presented an application of nearest neighbor imputation for the US census long form data. Vink et al. (2014) and Morris et al. (2014) investigated using predictive mean matching as a tool for multiple imputation via simulation studies. Chen and Shao (2000, 2001) have developed a nice set of asymptotic theories for the nearest neighbor imputation estimator. In econometrics, Abadie and Imbens (2006, 2008, 2011, 2016) studied the matching estimator for causal effect estimation from observational studies. Up to our best knowledge, there is no literature on theoretical investigation of estimated predictive mean matching for mean estimation in survey sampling, which motivates this article.

Predictive mean matching is implemented in two steps. First, the predictive mean function is estimated. Second, for each nonrespondent, the nearest neighbor is identified among the respondents based on the predictive mean function, and then the observed outcome value of the nearest neighbor is used for imputation. Because the predictive mean function is estimated prior to matching, it is necessary to account for the uncertainty due to parameter estimation. Because of the non-smooth nature of matching, our derivation is based on the technique developed by Andreou and Werker (2012), which offers a general approach for deriving the limiting distribution of statistics that involve estimated nuisance parameters. This technique has been successfully used in Abadie and Imbens (2016) for the matching estimators of the average causal effects based on the estimated propensity score. We extend their results to the matching estimator in the survey sampling context. In addition, we establish robustness of the predictive mean matching estimator which is consistent if the mean function satisfies a certain Lipschitz continuity condition.

Lack of smoothness also makes the conventional replication methods invalid for variance estimation for the predictive mean matching estimator. Abadie and Imbens (2008) demonstrated the failure of the bootstrap for matching estimators with a fixed number of matches. We propose new replication variance estimation for the predictive mean matching estimator in survey sampling. Based on the martingale representation of the predictive mean matching estimator, we construct replicates of the estimator directly based on its linear terms. In this way, the distribution of the number of times that each unit is used as a match can be preserved, which leads to a valid variance estimation. Furthermore, our replication variance method is flexible and can accommodate bootstrap, jackknife, among others.

The rest of this paper is organized as follows. In Section 2, we introduce the basic set-up in the context of survey data and the predictive mean matching procedure. In Section 3, we establish and compare the asymptotic distributions of the predictive mean matching estimator when the predictive mean function is known or is estimated. In Section 4, we propose the new replication variance estimators and establish their consistency. In Section 5, we evaluate the finite sample performance of the proposed estimators via a simulation study. We end with a brief discussion in Section 6. All proofs are deferred to the Appendix.

2. Basic Set-up

Let $\mathcal{F}_N = \{(x_i, y_i, \delta_i) : i = 1, \ldots, N\}$ denote a finite population, where x_i is always observed, y_i has missing values, and δ_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 δ_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 ζ, and N is known. Our objective is to estimate the finite population mean $\mu = N^{-1}\sum_{i=1}^{N} y_i$. Let A denote an index set of the sample selected by a probability sampling design. Let I_i be the sampling indicator, i.e., $I_i = 1$ if unit i is selected into the sample, and $I_i = 0$ otherwise. Suppose that π_i, the probability of selection of i, is positive and known throughout the sample. 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* $\epsilon > 0$.

In order to construct the imputed values, we assume that

$$E(y_i \mid x_i) = m(x_i; \beta^*), \tag{1}$$

holds for every unit in the population, where $m(\cdot)$ is a function of x known up to β^*. Under Assumption 1, let the normalized estimating equation for β be

$$S_N(\beta) = \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} \delta_i g(x_i; \beta)\{y_i - m(x_i; \beta)\} = 0, \tag{2}$$

where $g(x; \beta)$ is any function with which the solution to (2) exists uniquely. To simply the presentation, let $g(x; \beta)$ be $\dot{m}(x; \beta) = \partial m(x; \beta)/\partial \beta$. General functions $g(x; \beta)$ can be considered at the expense of heavier notation. Under certain regularity conditions (e.g. Fuller; 2009, Ch. 2), the solution $\hat{\beta}$ converges to β^* in probability. Here, the probability distribution is the joint distribution of the sampling distribution and the superpopulation model (1). The sampling weight π_i^{-1} is used to obtain a consistent estimator of β^* even under informative sampling (Berg et al.; 2016).

Under the model (1), the predictive mean matching method can be described as follows:

Step 1. Obtain a consistent estimator of β, denoted by $\hat{\beta}$, by solving (2). For each unit i with $\delta_i = 0$, obtain a predicted value of y_i as $\hat{m}_i = m(x_i; \hat{\beta})$. Find the nearest neighbor of unit i from the respondents with the minimum distance between $\hat{m}_j$ and $\hat{m}_i$. Let $i(1)$ be the index of the nearest neighbor of unit i, which satisfies $d(\hat{m}_{i(1)}, \hat{m}_i) \leq d(\hat{m}_j, \hat{m}_i)$, for any $j \in A_R = \{i \in A : \delta_i = 1\}$, where $d(\cdot, \cdot)$ denotes a generic distance function, e.g., $d(m_i, m_j) = |m_i - m_j|$ for scalar m_i and m_j.

Step 2. The imputation estimator based on predictive mean matching is computed by

$$\hat{\mu}_{\text{PMM}} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left\{ \delta_i y_i + (1 - \delta_i) y_{i(1)} \right\}. \tag{3}$$

In (3), the imputed values are real observations. The imputation model is used only for identifying the nearest neighbor, but not for creating the imputed values. Variance estimation of $\hat{\mu}_{\text{PMM}}$ is challenging because of the nonsmoothness of the matching mechanism in Step 1. In the next section, we formally discuss the asymptotic properties of the predictive mean matching estimator.

3. Main Result

3.1 Predictive mean matching

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 $\hat{\mu}_{\text{PMM}} = \hat{\mu}_{\text{PMM}}(\hat{\beta})$, where

$$\begin{aligned} \hat{\mu}_{\text{PMM}}(\beta) &= \frac{1}{N}\sum_{i\in A}\frac{1}{\pi_i}\{\delta_i y_i + (1-\delta_i)y_{i(1)}\} \\ &= \frac{1}{N}\left(\sum_{i\in A}\frac{1}{\pi_i}\delta_i y_i + \sum_{j\in A}\frac{1-\delta_j}{\pi_j}\sum_{i\in A}\delta_i d_{ij} y_i\right) \\ &= \frac{1}{N}\sum_{i\in A}\frac{\delta_i}{\pi_i}(1+k_{\beta,i})y_i, \end{aligned} \tag{4}$$

with

$$k_{\beta,i} = \sum_{j\in A}\frac{\pi_i}{\pi_j}(1-\delta_j)d_{ij}. \tag{5}$$

Under simple random sampling, $k_{\beta,i} = \sum_{j\in A}(1-\delta_j)d_{ij}$ is the number of times that unit i is used as the nearest neighbor for nonrespondents, where determination of the nearest neighbor is based on the predictive mean function $m(x_i;\beta)$.

We first consider the case when β^*, and hence $m(x_i) = m(x_i;\beta^*)$, is known. Suppose that the superpopulation model satisfies the following assumption.

Assumption 2 *(i) The matching variable $m(x)$ has a compact and convex support, with its density bounded and bounded away from zero. Denote $m_i = m(x_i)$. Let $g_1(m_i)$ and $g_0(m_i)$ be the conditional density of m_i given $\delta_i = 1$ and $\delta_i = 0$, respectively. Suppose that there exist constants C_{1L} and C_{1U} such that $C_{1L} \le g_1(m_i)/g_0(m_i) \le C_{1U}$; (ii) there exists $\delta > 0$ such that $E(|y|^{2+\delta} \mid x)$ is uniformly bounded for any x.*

Assumption 2 (i) is a convenient regularity condition (Abadie and Imbens; 2006). Assumption 2 (ii) is a moment condition for establishing the central limit theorem.

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

Assumption 3 is a widely accepted assumption in survey sampling (Fuller; 2009, Ch. 1).

To study the asymptotic properties of the predictive mean matching estimator, we use the following decomposition:

$$n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta) - \mu\} = D_N(\beta) + B_N(\beta), \tag{6}$$

where

$$D_N(\beta) = \frac{n^{1/2}}{N}\left(\sum_{i\in A}\frac{1}{\pi_i}\left[m(x_i;\beta) + \delta_i(1+k_{\beta,i})\{y_i - m(x_i;\beta\}\right] - \mu\right), \tag{7}$$

and

$$B_N(\beta) = \frac{n^{1/2}}{N}\sum_{i\in A}\frac{1}{\pi_i}(1-\delta_i)\{m(x_{i(1)};\beta) - m(x_i;\beta)\}. \tag{8}$$

The difference $m(x_{i(1)};\beta^*) - m(x_i;\beta^*)$ accounts for the matching discrepancy, and $B_N(\beta^*)$ contributes to the asymptotic bias of the matching estimator. In general, if the matching variable 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 $B_N(\beta^*) = O_p(n^{1/2-1/p}) \neq o_p(1)$ is not negligible; whereas, for predictive mean matching, the matching variable is a scalar function $m(x)$, and hence $B_N(\beta^*) = O_p(n^{-1/2}) = o_p(1)$. We establish the asymptotic distribution of $\hat{\mu}_{\mathrm{PMM}}(\beta^*)$.

Theorem 1 *Under Assumptions 1–3, suppose that* $m(x) = E(y \mid x) = m(x;\beta^*)$ *and* $\sigma^2(x) = \mathrm{var}(y \mid x)$. *Then,* $n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^*) - \mu\} \to \mathcal{N}(0, V_1)$ *in distribution, as* $n \to \infty$, *where*

$$V_1 = V^m + V^e \tag{9}$$

with

$$\begin{aligned} V^m &= \lim_{n\to\infty} nN^{-2}E[\mathrm{var}_p\{\sum_{i\in A}\pi_i^{-1}m(x_i)\}], \\ V^e &= \lim_{n\to\infty} nN^{-2}E\{\sum_{i=1}^{N}\pi_i^{-1}(1-\pi_i)\delta_i(1+k_{\beta^*,i})^2\sigma^2(x_i)\}, \end{aligned}$$

and $k_{\beta,i}$ *is defined in (5).*

In practice, β^* is unknown and therefore has to be estimated prior to matching. Following Abadie and Imbens (2016), the following theorem presents the approximate asymptotic distribution of $\hat{\mu}_{\mathrm{PMM}}(\hat{\beta})$.

Theorem 2 *Under Assumptions 1–3 and certain regularity conditions specified in the Appendix,* $n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\hat{\beta}) - \mu\} \to \mathcal{N}(0, V_2)$ *in distribution, as* $n \to \infty$, *where* $\hat{\beta}$ *is the solution to the estimating equation (2) and*

$$V_2 = V_1 - \gamma_1^{\mathrm{T}}V_s^{-1}\gamma_1 + \gamma_2^{\mathrm{T}}\left(\tau_{\beta^*}^{-1}V_s\tau_{\beta^*}^{-1}\right)\gamma_2, \tag{10}$$

$\gamma_1 = \lim_{n\to\infty} nN^{-2}E\{\sum_{i=1}^{N}\pi_i^{-1}(1-\pi_i)\delta_i(1+k_{\beta^*,i})g(x_i;\beta^*)\sigma^2(x_i)\}$, $\gamma_2 = E\{\dot{m}(x;\beta^*)\}$, V_1 *is defined in (9),* $V_s = \mathrm{var}\{S_N(\beta^*)\}$, $\tau_\beta = E\{p(x)\dot{m}(x;\beta)\ \dot{m}(x;\beta)^{\mathrm{T}}\}$, *and* $p(x) = \mathrm{pr}(\delta = 1 \mid x)$.

The difference between V_2 and V_1, $-\gamma_1^{\mathrm{T}} V_s^{-1} \gamma_1 + \gamma_2^{\mathrm{T}} (\tau_{\beta^*}^{-1} V_s \tau_{\beta^*}^{-1}) \gamma_2$, can be positive or negative. Thus, the estimation error in the predictive mean function should not be ignored. This is different from the result in Abadie and Imbens (2016) that matching on the estimated propensity score always improves the estimation efficiency when matching on the true propensity score. To explain the difference, we note that the propensity score is auxiliary for estimating the population mean of outcome; whereas the predictive mean function is not.

3.2 Nearest neighbor imputation

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 $j \in A_R$.

Step 2. The nearest neighbor imputation estimator of μ is computed by

$$\hat{\mu}_{\mathrm{NNI}} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left\{ \delta_i y_i + (1 - \delta_i) y_{i(1)} \right\} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \delta_i (1 + k_i) y_i, \tag{11}$$

where k_i is defined similarly as in (5), but with the matching variable x.

Following (6), write $n^{1/2}(\hat{\mu}_{\mathrm{NNI}} - \mu) = D_N + B_N$, where

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

and

$$B_N = \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} (1 - \delta_i) \{ m(x_{i(1)}) - m(x_i) \}. \tag{12}$$

Because the matching is based on a p-vector matching variable, the bias term $B_N = O_p(n^{1/2 - 1/p})$ with $p \geq 2$ is not negligible. For bias correction, let $\hat{m}(x)$ be a consistent estimator of $m(x) = E(y \mid x)$. Then, we can estimate B_N by $\hat{B}_N = n^{-1/2} N \sum_{i \in A} \pi_i^{-1} (1 - \delta_i) \{ \hat{m}(x_{i(1)}) - \hat{m}(x_i) \}$. A bias-corrected nearest neighbor imputation estimator of μ is

$$\tilde{\mu}_{\mathrm{NNI}} = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \{ \delta_i y_i + (1 - \delta_i) y_i^* \}, \tag{13}$$

where $y_i^* = \hat{m}(x_i) + y_{i(1)} - \hat{m}(x_{i(1)})$. Under certain regularity conditions imposed on the nonparametric estimator $\hat{m}(x)$, $\hat{B}_N$ is consistent for B_N, i.e., $\hat{B}_N - B_N = o_p(1)$. Then, the bias-corrected nearest neighbor imputation estimator has the same limiting distribution as the predictive mean matching estimator with known β^* has.

3.3 Robustness against the predictive mean function specification

To discuss the robustness of the predictive mean matching estimator against the predictive mean function specification, let $m(x;\beta)$ be a working model for $E(y \mid x)$, $\hat{\beta}$ be the estimator of β solving (2), and β^* be its probability limit. We also use $m = m(x;\beta^*)$ for shorthand. We require the following assumption hold for the working model.

Assumption 4 *$E(y \mid m)$ is is Lipschitz continuous in m; i.e., there exists a constant C_3 such that $|E(y \mid m_i) - E(y \mid m_j)| \leq C_3|m_i - m_j|$, for any i, j.*

Assumption 4 is trivial when $m(x;\beta)$ is correctly specified for $E(y \mid x)$, because in this case $E(y \mid m) = m$.

Theorem 3 *Under Assumptions 1–4, the predictive mean matching estimator based on the working model $m(x;\beta^*)$ is consistent for μ.*

The result can be obtained directly from the decomposition (6) by replacing $m(x;\beta)$ in $D_N(\beta)$ and $B_N(\beta)$ with $E\{y \mid m(x;\beta)\}$. The new term $D_N(\beta^*)$ is still consistent for zero; by Assumption 4, the new bias term becomes

$$\begin{aligned} |B_N(\beta^*)| &= \left|\frac{n^{1/2}}{N}\sum_{i\in A}\frac{1}{\pi_i}(1-\delta_i)\left[E\{y \mid m(x_{i(1)};\beta^*)\} - E\{y \mid m(x_{i(1)};\beta^*)\}\right]\right| \\ &\leq \frac{n^{1/2}}{N}C_3\sum_{i\in A}\frac{1}{\pi_i}(1-\delta_i)|m(x_{i(1)};\beta^*) - m(x_i;\beta^*)| = O_p(n^{-1/2}). \end{aligned}$$

4. Replication Variance Estimation

We consider replication variance estimation (Rust and Rao; 1996; Wolter; 2007) for the predictive mean matching estimator. Let $\hat{\mu}$ be the Horvitz-Thompson estimator of μ. The replication variance estimator of $\hat{\mu}$ takes the form of

$$\hat{V}_{\text{rep}}(\hat{\mu}) = \sum_{k=1}^{L} c_k(\hat{\mu}^{(k)} - \hat{\mu})^2, \tag{14}$$

where L is the number of replicates, c_k is the kth replication factor, and $\hat{\mu}^{(k)}$ is the kth replicate of $\hat{\mu}$. When $\hat{\mu} = \sum_{i\in A}\omega_i y_i$, we can write the replicate of $\hat{\mu}$ as $\hat{\mu}^{(k)} = \sum_{i\in A}\omega_i^{(k)}y_i$ with some $\omega_i^{(k)}$ for $i \in A$. The replications are constructed such that $E\{\hat{V}_{\text{rep}}(\hat{\mu})\} = \text{var}(\hat{\mu})\{1 + o(1)\}$. For example, in delete-1 jackknife under probability proportional to size sampling with $\omega_i = N^{-1}\pi_i^{-1}$, we have $L = n$, $c_k = (n-1)/n$, and $\omega_i^{(k)} = n\omega_i/(n-1)$ if $i \neq k$, and $\omega_k^{(k)} = 0$.

We propose a new replication variance estimation for the predictive mean matching estimator. We first consider $\hat{\mu}_{\text{PMM}}(\beta^*)$ with a known β^* given in (4). For simplicity, we suppress the dependence of quantities on β^*. Write $\hat{\mu}_{\text{PMM}} - \mu = (\hat{\mu}_{\text{PMM}} - \hat{\psi}_{\text{HT}}) + (\hat{\psi}_{\text{HT}} - \mu_\psi) + (\mu_\psi - \mu)$, where $\hat{\psi}_{\text{HT}} = \sum_{i \in A} \omega_i \psi_i$, $\psi_i = m(x_i) + \delta_i(1 + k_i)\{y_i - m(x_i)\}$, $\mu_\psi = N^{-1} \sum_{i=1}^N \psi_i$. By Theorem 1, $\mu_{\text{PMM}} - \hat{\psi}_{\text{HT}} = o_p(n^{-1/2})$. Together with the fact that $\mu_\psi - \mu = O_p(N^{-1/2})$ and $nN^{-1} = o(1)$, $\hat{\mu}_{\text{PMM}} - \mu = \hat{\psi}_{\text{HT}} - \mu_\psi + o_p(n^{-1/2})$. Therefore, with negligible sampling fractions, it is sufficient to estimate the variance of $\hat{\psi}_{\text{HT}} - \mu_\psi$. Because $E_p(\hat{\psi}_{\text{HT}} - \mu_\psi) = 0$, we have $\text{var}(\hat{\psi}_{\text{HT}} - \mu_\psi) = E\{\text{var}_p(\hat{\psi}_{\text{HT}} - \mu_\psi)\}$, 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 the 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 (14) for the above replicates $\hat{\psi}_{\text{HT}}^{(k)}$. It follows that $E\{\hat{V}_{\text{rep}}(\hat{\psi}_{\text{HT}})\} = \text{var}(\hat{\psi}_{\text{HT}} - \mu_\psi)\{1 + o(1)\} = \text{var}(\hat{\mu}_{\text{PMM}} - \mu)\{1 + o(1)\}$.

We now consider $\hat{\mu}_{\text{PMM}}(\hat{\beta})$, which can be expressed as $\hat{\mu}_{\text{PMM}}(\hat{\beta}) = \sum_{i \in A} \omega_i [m(x_i; \hat{\beta}) + \delta_i(1 + k_{\hat{\beta},i})\{y_i - m(x_i; \hat{\beta})\}] + o_p(n^{-1/2})$. To compute the replicates of $\hat{\mu}_{\text{PMM}}(\hat{\beta})$, we propose two steps:

Step 1. Obtain the kth replicate of $\hat{\beta}$, denoted as $\hat{\beta}^{(k)}$, by solving $S_N^{(k)}(\beta) = \sum_{i \in A} \omega_i^{(k)} \delta_i \times g(x_i; \beta)\{y_i - m(x_i; \beta)\} = 0$.

Step 2. Obtain the kth replicate as

$$\hat{\mu}_{\text{PMM}}^{(k)}(\hat{\beta}^{(k)}) = \sum_{i \in A} \omega_i^{(k)} [m(x_i; \hat{\beta}^{(k)}) + \delta_i(1 + k_{\hat{\beta}^{(k)},i})\{y_i - m(x_i; \hat{\beta}^{(k)})\}]. \tag{15}$$

If β^* is known, we do not need to reflect the effect of estimating β^*, and the above procedure with two steps reduces to the one we proposed for the case when β^* is known. On the other hand, when β^* is estimated, Step 1 is necessary, because as shown in Theorem 2, the predictive mean matching estimators by matching on the true and estimated predictive mean function may have different asymptotic distributions.

The consistency of the replication variance estimator is presented in the following theorem.

Theorem 4 *Under the assumptions in Theorem 2, suppose that $\hat{V}_{\text{rep}}(\hat{\mu})$ in (14) is consistent for $\text{var}_p(\hat{\mu})$. Then, if $nN^{-1} = o(1)$, the replication variance estimators for $\hat{\mu}_{\text{PMM}}(\hat{\beta})$ is consistent, i.e., $n\hat{V}_{\text{rep}}\{\hat{\mu}_{\text{PMM}}(\hat{\beta})\}/V_2 \to 1$ in probability, as $n \to \infty$, where the replicates of $\hat{\mu}_{\text{PMM}}(\hat{\beta})$ are given in (15), and V_2 is given in (10).*

5. A Simulation Study

In this simulation study, we investigate the performance of the proposed replication variance estimator. 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}, 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) $y_i = -1.167 + x_{1i} + x_{2i} + (x_{1i} - 0.5)^2 + (x_{2i} - 0.5)^2 + e_i$, and (P3) $y_i = -1.5 + x_{1i} + \cdots + x_{6i} + e_i$. The covariates are fully observed, but y_i is not. The response indicator of y_i, δ_i, is generated from Bernoulli(p_i) with $\text{logit}\{p(x_i)\} = 0.2 + x_{1i} + x_{2i}$. This results in the average response rate about 75%. The parameter of interest is $\mu = N^{-1} \sum_{i=1}^{N} y_i$. To generate samples, we consider two sampling designs: (S1) simple random sampling with $n = 400$; (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 $\pi_i = 400 s_i / \sum_{i=1}^{N} s_i$. Therefore, (S2) is informative, where units with larger y_i values have larger probabilities to be selected into the sample.

For estimation, we consider predictive mean matching imputation, nearest neighbor imputation, and stochastic regression imputation. In stochastic regression imputation, for units with $\delta_i = 0$, the imputation of y_i is obtained as $y_i^* = \hat{y}_i + \hat{e}_i^*$, where $\hat{y}_i = m(x_i; \hat{\beta})$ and $\hat{e}_i^*$ is randomly selected from the observed residuals $\{\hat{e}_i = y_i - \hat{y}_i : \delta_i = 1\}$. For (P1) and (P2), we specify the predictive mean function to be $m(x; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Note that for (P1), $m(x; \beta)$ is correctly specified; whereas for (P2), $m(x; \beta)$ is misspecified. For (P3), we specify the mean function to be $m(x; \beta) = \beta_0 + \beta^{\mathrm{T}} x$, where $x = (x_1, \cdots, x_6)$. 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 point estimate and $\hat{V}_I$ is the variance estimate obtained by the proposed jackknife variance estimation. For stochastic regression imputation, the kth replicate of μ is given by $\hat{\mu}_{\text{REG}}^{(k)}(\hat{\beta}^{(k)}) = \sum_{i \in A} \omega_i^{(k)} [m(x_i; \hat{\beta}^{(k)}) + \delta_i (1 + k_i) \{y_i - m(x_i; \hat{\beta}^{(k)})\}]$, where $\hat{\beta}^{(k)}$ is obtained from the estimating equation of β based on the replication weights, and k_i is the number of times that $\hat{e}_i$ is selected to impute the missing values of y based on the original data.

Table 1 presents the simulation results based on 2,000 Monte Carlo samples. When the covariate is 2-dimensional, all three imputation estimators have small biases, even when the mean function is misspecified. In addition, the proposed jackknife method provides valid coverage of confidence intervals for the predictive mean matching and stochastic regression imputation estimators in all scenarios. This suggests that the proposed replication method can be used widely even for stochastic regression imputation. When the covariate is 6-dimensional, nearest neighbor imputation presents large biases and low coverage rates.

6. Discussion

Propensity score matching has been recently proposed for inferring causal effects of treatments in the context of survey data; however, their asymptotic properties are underdeveloped (Lenis et al.; 2017). Because causal inference is inherently a missing data problem (e.g., Ding and Li; 2017), the proposed methodology here can be easily generalized to investigate the asymptotic properties of propensity score matching estimators with survey weights.

Table 1: Simulation results: 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.

	PMM		NNI		SRI		PMM		NNI		SRI	
	Bias	S.E.	Bias	S.E.	Bias	S.E.	RB	CR	RB	CR	RB	CR
Simple Random Sampling												
(P1)	-0.15	6.46	-0.21	6.54	-0.23	6.44	4	95.2	3	95.1	5	95.8
(P2)	-0.22	6.54	-0.25	6.55	-0.37	6.46	6	95.5	3	95.3	5	95.6
(P3)	1.90	11.85	18.59	11.06	0.11	11.17	5	95.1	4	63.8	4	95.5
Probability Proportional to Size Sampling												
(P1)	0.05	6.46	0.13	6.37	0.18	6.53	3	95.3	3	94.8	2	94.9
(P2)	0.30	6.52	0.12	6.47	0.16	6.60	2	95.3	0	95.3	3	94.9
(P3)	1.33	10.99	17.53	10.70	0.40	11.10	6	95.6	3	65.5	-3	95.6

PMM: predictive mean matching; NNI: nearest neighbor imputation; SRI: stochastic regression imputation.

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 topic for future research.

Appendix

A1 Proof for Theorem 1

Based on the decomposition in (6), write

$$n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^*) - \mu\} = D_N(\beta^*) + B_N(\beta^*), \tag{A1}$$

where $D_N(\beta)$ and $B_N(\beta)$ are defined in (7) and (8), respectively. For simplicity, we introduce the following notation: $m_i = m(x_i;\beta^*)$ and $e_i = y_i - m_i$.

Under Assumption 2, for the predictive mean matching estimator, $m_{i(1)} - m_i = O_p(1)$. Together with Assumption 3, we derive the order of $B_N(\beta^*)$ as

$$B_N(\beta^*) = \frac{n^{1/2}}{N}\sum_{i\in A}\frac{1}{\pi_i}(1-\delta_i)(m_{i(1)} - m_i) = O_p(n^{-1/2}) = o_p(1).$$

Therefore, (A1) reduces to

$$n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^*) - \mu\} = D_N(\beta^*) + o_p(1).$$

Then, to study the asymptotic properties of $n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^*) - \mu\}$, we only need to study the asymptotic properties of $D_N(\beta^*)$. We express

$$D_N(\beta^*) \quad = \quad \frac{n^{1/2}}{N}\left[\sum_{i\in A}\frac{1}{\pi_i}\left\{m_i + \delta_i(1+k_{\beta^*,i})e_i\right\} - \mu\right]$$

$$= \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) m_i + \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \delta_i (1 + k_{\beta^*,i}) e_i$$

$$+ \frac{n^{1/2}}{N} \sum_{i=1}^{N} (m_i - \mu) + \frac{n^{1/2}}{N} \sum_{i=1}^{N} \delta_i (1 + k_{\beta^*,i}) e_i$$

$$= \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) m_i + \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \delta_i (1 + k_{\beta^*,i}) e_i + o_p(1), \quad \text{(A2)}$$

given $nN^{-1} = o(1)$. We can verify that the covariance of the two terms in (A2) is zero. Thus, the asymptotic variance of $D_N(\beta^*)$ is

$$\mathrm{var} \left\{ \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) m_i \right\} + \mathrm{var} \left\{ \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \delta_i (1 + k_{\beta^*,i}) e_i \right\}.$$

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

$$V^m = \lim_{n \to \infty} \frac{n}{N^2} E \left\{ \mathrm{var}_p \left(\sum_{i \in A} \frac{m_i}{\pi_i} \right) \right\},$$

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

$$V^e = \mathrm{plim} \frac{n}{N^2} \sum_{i=1}^{N} \frac{1 - \pi_i}{\pi_i} \delta_i (1 + k_{\beta^*,i})^2 \mathrm{var}(e_i \mid x_i).$$

The remaining is to show that $V^e = O(1)$. To do this, the key is to show that the moments of $k_{\beta^*,i}$ are bounded. Under Assumption 3, it is easy to verify that

$$\underline{\omega} \tilde{k}_{\beta^*,i} \le k_{\beta^*,i} \le \bar{\omega} \tilde{k}_{\beta^*,i}, \quad \text{(A3)}$$

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

Combining all results, the asymptotic variance of $n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^*) - \mu\}$ is $V^m + V^e$. By the central limit theorem, the result in Theorem 1 follows.

A2 Le Cam's third Lemma

Consider two sequences of probability measures $(Q^{(N)})_{N=1}^{\infty}$ and $(P^{(N)})_{N=1}^{\infty}$. Assume that under $P^{(N)}$, a statistic T_N and the likelihood ratios $dQ^{(N)}/dP^{(N)}$ satisfy

$$\begin{pmatrix} T_N \\ \log(dQ^{(N)}/dP^{(N)}) \end{pmatrix} \to \mathcal{N} \left\{ \begin{pmatrix} 0 \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \tau^2 & c \\ c & \sigma^2 \end{pmatrix} \right\}$$

in distribution, as $N \to \infty$. Then, under $Q^{(N)}$,

$$T_N \to \mathcal{N}(c, \tau^2)$$

in distribution, as $N \to \infty$. See Le Cam and Yang (1990), Bickel et al. (1993) and van der Vaart (2000) for textbook discussions.

A3 Proof for Theorem 2

Let P be the distribution of $(x_i, y_i, \delta_i, I_i)$, for $i = 1, \dots, N$, induced by the marginal distribution of x_i, the conditional distribution of y_i given x_i, the conditional distribution of δ_i given (x_i, y_i), and the conditional distribution of I_i given (x_i, y_i, δ_i). Consider P to be restricted by the moment condition through the predictive mean function (1) with the true parameter value β^*. We can treat the consistent estimator $\hat{\beta}$ as the solution to the normalized estimating equation

$$S_N(\beta) = \frac{n^{1/2}}{N} \sum_{i=1}^{N} \frac{I_i}{\pi_i} \delta_i g(x_i; \beta)\{y_i - m(x_i; \beta)\} = 0. \tag{A4}$$

To discuss the asymptotic properties of $\hat{\mu}_{\mathrm{PMM}}(\hat{\beta})$, we rely on Le Cam's third lemma and consider an auxiliary parametric model P^{β} defined locally around β^* with a density

$$\frac{\exp\left\{n^{1/2}(\beta - \beta^*)^{\mathrm{T}} \tau_{\beta^*} V_s^{-1} S_N(\beta^*) - 2^{-1} n (\beta - \beta^*)^{\mathrm{T}} \Lambda^{-1} (\beta - \beta^*)\right\}}{E\left[\exp\left\{n^{1/2}(\beta - \beta^*)^{\mathrm{T}} \tau_{\beta^*} V_s^{-1} S_N(\beta^*) - 2^{-1} n (\beta - \beta^*)^{\mathrm{T}} \Lambda^{-1} (\beta - \beta^*)\right\}\right]}. \tag{A5}$$

Because under P^{β^*}, $S_N(\beta^*) \to \mathcal{N}(0, V_s)$ in distribution, the normalizing constant in the denominator converges to 1 as $n \to \infty$. The Fisher information under the parametric model (A5) is $n\Lambda^{-1}$. Therefore, $\hat{\beta}$ is efficient under (A5).

We now consider sequences that are local to β^*, $\beta_N = \beta^* + n^{-1/2} h$, indexed by N. In our context, we have the population size N goes to infinity with sample size n. Consider $(x_i, y_i, \delta_i, I_i)$, for $i = 1, \dots, N$, with the local shift P^{β_N} (Bickel et al.; 1993). We make the following regularity assumptions:

Assumption A5 *(i) The superpopulation model is regular (Bickel et al.; 1993, pp 12–13); (ii) under P^{β_N}: $S_N(\beta_N) \to \mathcal{N}(0, V_s)$ in distribution, as $n \to \infty$; (iii) τ_β is nonsingular around β^*, and $n^{1/2}(\hat{\beta} - \beta_N) = \tau_{\beta^*}^{-1} S_N(\beta_N) + o_p(1)$; (iv) for all bounded continuous functions $h(x, y, \delta, I)$, the conditional expectation $E_{\beta_N}\{h(x, y, \delta, I) \mid x, \delta = 1\}$ converges in distribution to $E\{h(x, y, \delta, I) \mid x, \delta = 1\}$, where E_{β_N} is the expectation with respect to P^{β_N}.*

We now give a sketch proof for Theorem 2.

Under (A5), the likelihood ratio under P^{β_N} is

$$\begin{aligned}\log(dP^{\beta^*}/dP^{\beta_N}) &= -h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N(\beta^*) + \frac{1}{2}h^{\mathrm{T}}\Lambda^{-1}h + o_p(1) \\ &= -h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N(\beta_N) - \frac{1}{2}h^{\mathrm{T}}\Lambda^{-1}h + o_p(1),\end{aligned}$$

where the second equality follows by the Taylor expansion of $S_N(\beta^*)$ at β_N.

We can derive that under P^{β_N},

$$\begin{pmatrix} n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta_N) - \mu(\beta_N)\} \\ n^{1/2}(\hat{\beta} - \beta_N) \\ \log(dP^{\beta^*}/dP^{\beta_N}) \end{pmatrix}$$

$$\to \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \\ \frac{-1}{2}h^{\mathrm{T}}\Lambda^{-1}h \end{pmatrix}, \begin{pmatrix} V_1 & \gamma_1^{\mathrm{T}}\tau_{\beta^*}^{-1} & -\gamma_1^{\mathrm{T}}V_s^{-1}\tau_{\beta^*}h \\ \tau_{\beta^*}^{-1}\gamma_1 & \Lambda & -h \\ -h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}\gamma_1 & -h^{\mathrm{T}} & h^{\mathrm{T}}\Lambda^{-1}h \end{pmatrix} \right\} \quad \text{(A6)}$$

in distribution, as $n \to \infty$. Here, we write $\mu = \mu(\beta_N)$ to reflect its dependence on β_N. We then express $\mu(\beta_N) = \mu(\beta^*) + \gamma_2^{\mathrm{T}}(n^{-1/2}h) + o(n^{-1/2})$, and use the shorthand μ for $\mu(\beta^*)$.

By Le Cam's third lemma, under P^{β^*}, we have

$$\begin{pmatrix} n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta_N) - \mu\} \\ n^{1/2}(\hat{\beta} - \beta_N) \end{pmatrix} \to \mathcal{N}\left\{ \begin{pmatrix} -\gamma_1^{\mathrm{T}}V_s^{-1}\tau_{\beta^*}h - \gamma_2^{\mathrm{T}}h \\ -h \end{pmatrix}, \begin{pmatrix} V_1 & \gamma_1^{\mathrm{T}}\tau_{\beta^*}^{-1} \\ \tau_{\beta^*}^{-1}\gamma_1 & \Lambda \end{pmatrix} \right\}$$

in distribution, as $n \to \infty$. Replacing β_N by $\beta^* + n^{-1/2}h$ yields that under P^{β^*},

$$\begin{pmatrix} n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^* + n^{-1/2}h) - \mu\} \\ n^{1/2}(\hat{\beta} - \beta^*) \end{pmatrix} \to \mathcal{N}\left\{ \begin{pmatrix} -\gamma_1^{\mathrm{T}}V_s^{-1}\tau_{\beta^*}h - \gamma_2^{\mathrm{T}}h \\ 0 \end{pmatrix}, \begin{pmatrix} V_1 & \gamma_1^{\mathrm{T}}\tau_{\beta^*}^{-1} \\ \tau_{\beta^*}^{-1}\gamma_1 & \Lambda \end{pmatrix} \right\}$$

in distribution, as $n \to \infty$.

Heuristically, if the normal distribution was exact, then

$$n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta^* + n^{-1/2}h) - \mu\} \mid n^{1/2}(\hat{\beta} - \beta^*) = h \sim \mathcal{N}\left(-\gamma_2^{\mathrm{T}}h, V_1 - \gamma_1^{\mathrm{T}}V_s^{-1}\gamma_1\right). \quad \text{(A7)}$$

Given $n^{1/2}(\hat{\beta} - \beta^*) = h$, we have $\beta^* + n^{-1/2}h = \hat{\beta}$, and hence $\hat{\mu}_{\mathrm{PMM}}(\beta^* + n^{-1/2}h) = \hat{\mu}_{\mathrm{PMM}}(\hat{\beta})$. Integrating (A7) over the asymptotic distribution of $n^{1/2}(\hat{\beta} - \beta^*)$, we derive

$$n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\hat{\beta}) - \mu\} \sim \mathcal{N}\left(0, V_1 - \gamma_1^{\mathrm{T}}V_s^{-1}\gamma_1 + \gamma_2^{\mathrm{T}}\Lambda\gamma_2\right). \quad \text{(A8)}$$

The formal technique to derive (A8) can be find in Andreou and Werker (2012). (A8) gives the result in Theorem 2.

In the following, we provide the proof to (A6). Asymptotic normality of $n^{1/2}\{\hat{\mu}_{\mathrm{PMM}}(\beta_N) - \mu\}$ under P^{β_N} follows from Theorem 1. Asymptotic joint normality of $n^{1/2}(\hat{\beta} - \beta_N)$ and $\log(dP^{\beta^*}/dP^{\beta_N})$ follows from Assumption A5. Therefore, the remaining is to show that, under P^{β_N}:

$$\begin{pmatrix} D_N(\beta_N) \\ S_N(\beta_N) \end{pmatrix} \to \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_1 & \gamma_1^T \\ \gamma_1 & V_s \end{pmatrix} \right\} \quad \text{(A9)}$$

in distribution, as $n \to \infty$. To prove (A9), consider the linear combination $c_1 D_N(\beta_N) + c_2^{\mathrm{T}} S_N(\beta_N)$, which has the same limiting distribution as

$$
\begin{aligned}
C_N \quad = \quad & c_1 \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) m(x_i; \beta_N) \\
& + c_1 \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \delta_i (1 + k_{\beta_N, i}) \{ y_i - m(x_i; \beta_N) \} \\
& + c_2^{\mathrm{T}} \frac{n^{1/2}}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \delta_i g(x_i; \beta_N) \{ y_i - m(x_i; \beta_N) \},
\end{aligned}
$$

given $nN^{-1} = o(1)$.

We analyze C_N using the martingale theory. First, we rewrite $C_N = \sum_{k=1}^{N} \xi_{N,k}$, where

$$
\xi_{N,k} = c_1 \frac{n^{1/2}}{N} \left(\frac{I_k}{\pi_k} - 1 \right) m(x_k; \beta_N)
$$

$$
\begin{aligned}
& + c_1 \frac{n^{1/2}}{N} \left(\frac{I_k}{\pi_k} - 1 \right) \delta_k (1 + k_{\beta_N, k}) \{ y_k - m(x_k; \beta_N) \} \\
& + c_2^{\mathrm{T}} \frac{n^{1/2}}{N} \left(\frac{I_k}{\pi_k} - 1 \right) \delta_k g(x_k; \beta_N) \{ y_k - m(x_k; \beta_N) \}.
\end{aligned}
$$

Consider the σ-fields $\mathcal{F}_{N,k} = \sigma\{x_1, \ldots, x_N, \delta_1, \ldots, \delta_N, y_1, \ldots, y_k, I_1, \ldots, I_k\}$ for $1 \leq k \leq N$. Then, $\{\sum_{k=1}^{i} \xi_{N,k}, \mathcal{F}_{N,i}, 1 \leq i \leq N\}$ is a martingale for each $N \geq 1$. Therefore, the limiting distribution of C_N can be studied using the martingale central limit theorem (Theorem 35.12, Billingsley; 1995). Under Assumption 2, and the fact that $k_{\beta_N,k}$ has uniformly bounded moments, it follows that $\sum_{k=1}^{N} E_{\beta_N}(|\xi_{N,k}|^{2+\delta}) \to 0$ for some $\delta > 0$. It then follows that Lindeberg's condition in Billingsley's theorem holds. As a result, we obtain that under P^{β_N}, $C_N \to \mathcal{N}(0, \sigma^2)$ in distribution, as $n \to \infty$, where $\sigma^2 = \operatorname{plim} \sum_{k=1}^{N} E_{\beta_N}(\xi_{N,k}^2 \mid \mathcal{F}_{N,k-1})$. Assumption A5 further implies the following expressions:

$$
\begin{aligned}
\sigma^2 \quad = \quad & \operatorname{plim} \sum_{k=1}^{N} E_{\beta_N}(\xi_{N,k}^2 \mid \mathcal{F}_{N,k-1}) \\
= \quad & c_1^2 \operatorname{plim} \frac{n}{N^2} \sum_{k=1}^{N} E_{\beta_N} \left[\left\{ \left(\frac{I_k}{\pi_k} - 1 \right) m(x_k; \beta_N) \right\}^2 \mid \mathcal{F}_{N,k-1} \right] \\
& + c_1^2 \operatorname{plim} \frac{n}{N^2} \sum_{k=1}^{N} E_{\beta_N} \left(\left[\left(\frac{I_k}{\pi_k} - 1 \right) \delta_k (1 + k_{\beta_N,k}) \{ y_k - m(x_k; \beta_N) \} \right]^2 \mid \mathcal{F}_{N,k-1} \right) \\
& + 2 c_2^{\mathrm{T}} \operatorname{plim} \frac{n}{N^2} \sum_{k=1}^{N} E_{\beta_N} \left[\left(\frac{I_k}{\pi_k} - 1 \right)^2 \delta_k (1 + k_{\beta_N,k}) g(x_k; \beta_N) \{ y_k - m(x_k; \beta_N) \}^2 \mid \mathcal{F}_{N,k-1} \right] c_1 \\
& + c_2^{\mathrm{T}} \operatorname{plim} \frac{n}{N^2} \sum_{k=1}^{N} E_{\beta_N} \left[\left(\frac{I_k}{\pi_k} - 1 \right)^2 \delta_k g(x_k; \beta_N) g(x_k; \beta_N)^{\mathrm{T}} \{ y_k - m(x_k; \beta_N) \}^2 \mid \mathcal{F}_{N,k-1} \right] c_2
\end{aligned}
$$

$$
\begin{aligned}
&= c_1^2 \text{plim} \frac{n}{N^2} \text{var}_p \left(\sum_{k \in A} \frac{m_k}{\pi_k} \right) + c_1^2 \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k (1 + k_{\beta^*,k})^2 \sigma^2(x_k) \\
&\quad + 2c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k (1 + k_{\beta^*,k}) g(x_k; \beta^*) \sigma^2(x_k) c_1 \\
&\quad + c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k g(x_k; \beta^*) g(x_k; \beta^*)^{\mathrm{T}} \sigma^2(x_k) c_2 \\
&= c_1^2 V^m + c_1^2 V^e + 2c_2^{\mathrm{T}} \gamma_1 c_1 + c_2^{\mathrm{T}} V_s c_2.
\end{aligned}
$$

By the martingale central limit theorem, under P^{β_N}, (A9) follows.

A.4 Proof for Theorem 4

The replication method implicitly induces replication weights ω_i^* and random variables u_i such that $E^*(\omega_i^* u_i) = N^{-1}\pi_i^{-1}$ and $\text{var}^*(\omega_i^* u_i) = N^{-2}(1-\pi_i)\pi_i^{-2}$, for $i = 1, \ldots, N$, where $E^*(\cdot)$ and $\text{var}^*(\cdot)$ denote the expectation and variance for the resampling given the observed data. For example, in delete-1 jackknife under probability proportional to size sampling with $nN^{-1} = o(1)$, we have $\omega_i^{(k)} = (n-1)^{-1} n\omega_i$ if $i \neq k$, and $\omega_k^{(k)} = 0$. Then, the induced random variables u_i follows a two-point mass distribution as

$$
u_i = \begin{cases} 1, & \text{with probability } \frac{n-1}{n}, \\ 0, & \text{with probability } \frac{1}{n}, \end{cases}
$$

and weights $\omega_i^* = (n-1)^{-1} n\omega_i$. It is straightforward to verify that $E^*(\omega_i^* u_i) = \omega_i = N^{-1}\pi_i^{-1}$ and $\text{var}^*\{(\omega_i^* u_i)^2\} = (n-1)^{-1}\omega_i^2 \approx n^{-1}N^{-2}(1-\pi_i)\pi_i^{-2}$.

The $k$the replication of $\hat{\beta}$, $\hat{\beta}^{(k)}$, can be viewed as one realization of $\hat{\beta}^*$ which is the solution to the estimating equation

$$
S_N^*(\beta) = n^{1/2} \sum_{i \in A} \omega_i^* u_i \delta_i g(x_i; \beta)\{y_i - m(x_i; \beta)\} = 0. \tag{A10}
$$

Let P^* be the distribution of $z_i^* = (\omega_i^* u_i x_i, \omega_i^* u_i y_i, \omega_i^* u_i \delta_i, \omega_i^* u_i I_i)$, for $i = 1, \ldots, N$, given the observed data induced by bootstrap resampling satisfying

$$
\begin{aligned}
E^*\{S_N^*(\hat{\beta})\} &= n^{1/2} E^* \left[\sum_{i \in A} \omega_i^* u_i \delta_i g(x_i; \hat{\beta})\{y_i - m(x_i; \hat{\beta})\} \right] \\
&= \frac{n^{1/2}}{N} \sum_{i \in A} \frac{1}{\pi_i} \delta_i g(x_i; \hat{\beta})\{y_i - m(x_i; \hat{\beta})\} = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& E^*\left\{S_N^*(\hat{\beta})S_N^*(\hat{\beta})^{\mathrm{T}}\right\} \\
= \; & E^*\left[\left\{S_N^*(\hat{\beta}) - S_N(\hat{\beta})\right\}\left\{S_N^*(\hat{\beta}) - S_N(\hat{\beta})\right\}^{\mathrm{T}}\right] \\
= \; & nE^*\left[\sum_{i\in A}\left(\omega_i^* u_i - \frac{1}{N\pi_i}\right)^2 \delta_i g(x_i;\hat{\beta})g(x_i;\hat{\beta})^{\mathrm{T}}\{y_i - m(x_i;\hat{\beta})\}^2\right] \\
= \; & \frac{n}{N^2}\sum_{i\in A}\frac{1-\pi_i}{\pi_i^2}\delta_i g(x_i;\hat{\beta})g(x_i;\hat{\beta})^{\mathrm{T}}\{y_i - m(x_i;\hat{\beta})\}^2.
\end{aligned}
$$

We consider an auxiliary parametric model P^{β} defined locally around $\hat{\beta}$ with a density

$$
\frac{\exp\left\{n^{1/2}(\beta-\hat{\beta})^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N^*(\hat{\beta}) - 2^{-1}n(\beta-\hat{\beta})^{\mathrm{T}}\Lambda^{-1}(\beta-\hat{\beta})\right\}}{E^*\left[\exp\left\{n^{1/2}(\beta-\hat{\beta})^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N^*(\hat{\beta}) - 2^{-1}n(\beta-\hat{\beta})^{\mathrm{T}}\Lambda^{-1}(\beta-\hat{\beta})\right\}\right]}. \tag{A11}
$$

Consider sequences that are local to $\hat{\beta}$, $\beta_N^* = \hat{\beta} + n^{-1/2}h$, indexed by N, and z_i^*, for $i = 1,\ldots,N$, with the local shift $P^{\beta_N^*}$. We make the following regularity assumptions:

Assumption A6 *(i) Model (A11) is regular; (ii) under $P^{\beta_N^*}$: $S_N^*(\beta_N^*) \to \mathcal{N}(0,V_s)$ in distribution, as $n\to\infty$; (iii) $n^{1/2}(\hat{\beta}^* - \beta_N^*) = \tau_{\beta^*}^{-1}S_N^*(\beta_N^*) + o_p(1)$; (iv) for all bounded continuous functions $h(z_i^*)$, the conditional expectation $E^*_{\beta_N^*}\{h(z_i^*)\}$ converges in distribution to $E^*_{\hat{\beta}}\{h(z_i^*)\}$, where $E_{\beta_N^*}$ is the expectation with respect to $P^{\beta_N^*}$.*

Under (A11), the likelihood ratio under $P^{\beta_N^*}$ is

$$
\begin{aligned}
\log(dP^{\hat{\beta}}/dP^{\beta_N^*}) &= -h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N^*(\hat{\beta}) + \frac{1}{2}h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}\tau_{\beta^*}h + o_p(1) \\
&= -h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}S_N^*(\beta_N^*) - \frac{1}{2}h^{\mathrm{T}}\tau_{\beta^*}V_s^{-1}\tau_{\beta^*}h + o_p(1),
\end{aligned}
$$

where the second equality follows by the Taylor expansion of $S_N^*(\hat{\beta})$ at β_N^*.

The $k$the replication of $\hat{\mu}_{\mathrm{PMM}}(\hat{\beta})$, $\hat{\mu}_{\mathrm{PMM}}^{(k)}(\hat{\beta}^{(k)})$, can be viewed as one realization of

$$
\hat{\mu}^*_{\mathrm{PMM}}(\hat{\beta}^*) = \sum_{i\in A}\omega_i^* u_i[m(x_i;\hat{\beta}^*) + \delta_i(1+k_{\hat{\beta}^*,i})\{y_i - m(x_i;\hat{\beta}^*)\}]. \tag{A12}
$$

We can derive that under $P^{\beta_N^*}$, the sequence $[\; n^{1/2}\{\hat{\mu}^*_{\mathrm{PMM}}(\beta_N^*) - \hat{\mu}_{\mathrm{PMM}}(\beta_N^*)\} \;\; n^{1/2}(\hat{\beta}^* - \beta_N^*)^{\mathrm{T}} \;\; \log(dP^{\hat{\beta}}/dP^{\beta_N^*})]^{\mathrm{T}}$ has the same limiting distribution as in (A6). Then, following the same argument in the proof of Theorem 2, we can obtain that the asymptotic conditional variance of $n^{1/2}\hat{\mu}^*_{\mathrm{PMM}}(\hat{\beta}^*)$, given the observed data, is V_2.

The remaining is to show that, under $P^{\beta_N^*}$ given the observed data:

$$
\begin{pmatrix} n^{1/2}\{\hat{\mu}^*_{\mathrm{PMM}}(\beta_N^*) - \hat{\mu}_{\mathrm{PMM}}(\beta_N^*)\} \\ S_N^*(\beta_N^*) \end{pmatrix} \to \mathcal{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_1 & \gamma_1^T \\ \gamma_1 & V_s \end{pmatrix}\right\} \tag{A13}
$$

in distribution, as $n \to \infty$. To prove (A13), given the observed data, consider the linear combination $c_1 n^{1/2}\{\hat{\mu}^*_{\mathrm{PMM}}(\beta^*_N) - \hat{\mu}_{\mathrm{PMM}}(\beta^*_N)\} + c_2^{\mathrm{T}} S^*_N(\beta^*_N)$, which has the same limiting distribution as

$$\begin{aligned} C^*_N &= c_1 n^{1/2} \sum_{i=1}^{N} I_i \left(\omega^*_i u_i - \frac{1}{N\pi_i} \right) m(x_i; \beta^*_N) \\ &\quad + c_1 n^{1/2} \sum_{i=1}^{N} I_i \left(\omega^*_i u_i - \frac{1}{N\pi_i} \right) \delta_i (1 + k_{\beta^*_N, i}) \{y_i - m(x_i; \beta^*_N)\} \\ &\quad + c_2^{\mathrm{T}} n^{1/2} \sum_{i=1}^{N} I_i \left(\omega^*_i u_i - \frac{1}{N\pi_i} \right) \delta_i g(x_i; \beta^*_N) \{y_i - m(x_i; \beta^*_N)\}. \end{aligned}$$

This is because under $P^{\beta^*_N}$, the extra term in C^*_N compared with $c_1 n^{1/2}\{\hat{\mu}^*_{\mathrm{PMM}}(\beta^*_N) - \hat{\mu}_{\mathrm{PMM}}(\beta^*_N)\} + c_2^{\mathrm{T}} S^*_N(\beta^*_N)$ is

$$\begin{aligned} & n^{1/2} \sum_{i=1}^{N} \frac{I_i}{N\pi_i} \delta_i g(x_i; \beta^*_N) \{y_i - m(x_i; \beta^*_N)\} \\ =\ & \frac{n^{1/2}}{N} \sum_{i=1}^{N} \frac{I_i}{\pi_i} \delta_i g(x_i; \hat{\beta}) \{y_i - m(x_i; \hat{\beta})\} + O_p(\beta^*_N - \hat{\beta}) \\ =\ & 0 + O_p(n^{-1/2}) = o_p(1). \end{aligned}$$

We analyze C^*_N using the martingale theory. First, we rewrite $C^*_N = \sum_{k=1}^{N} \xi^*_{N,k}$, where

$$\begin{aligned} \xi^*_{N,k} &= c_1 n^{1/2} I_k \left(\omega^*_k u_k - \frac{1}{N\pi_i} \right) m(x_k; \beta^*_N) \\ &\quad + c_1 n^{1/2} I_k \left(\omega^*_k u_k - \frac{1}{N\pi_i} \right) \delta_k (1 + k_{\beta^*_N, k}) \{y_k - m(x_k; \beta^*_N)\} \\ &\quad + c_2^{\mathrm{T}} n^{1/2} I_k \left(\omega^*_k u_k - \frac{1}{N\pi_i} \right) \delta_k g(x_k; \beta^*_N) \{y_k - m(x_k; \beta^*_N)\}. \end{aligned}$$

for $1 \le k \le N$. Consider the σ-fields

$$\mathcal{F}^*_{N,k} = \sigma\{x_1, \ldots, x_N, I_1, \ldots, I_N, \delta_1, \ldots, \delta_N, y_1, \ldots, y_N, \omega^*_1 u_1, \ldots, \omega^*_k u_k\}$$

for $1 \le k \le N$. Then, $\{\sum_{k=1}^{i} \xi^*_{N,k}, \mathcal{F}^*_{N,i}, 1 \le i \le N\}$ is a martingale for each $N \ge 1$. As a result, we obtain that under $P^{\beta_N^*}$, $C_N^* \to \mathcal{N}(0, \tilde{\sigma}^2)$ in distribution, as $n \to \infty$, where

$$\begin{aligned}
\tilde{\sigma}^2 &= \text{plim} \sum_{k=1}^{N} E^*_{\beta_N^*}(\xi^{*2}_{N,k} \mid \mathcal{F}_{N,k-1}) \\
&= c_1^2 \text{plim} n \sum_{k=1}^{N} E^*_{\beta_N^*} \left[\left\{ I_k \left(\omega_k^* u_k - \frac{1}{N\pi_i} \right) m(x_k; \beta_N^*) \right\}^2 \mid \mathcal{F}_{N,k-1} \right] \\
&\quad + c_1^2 \text{plim} n \sum_{k=1}^{N} E^*_{\beta_N^*} \left(\left[I_k \left(\omega_k^* u_k - \frac{1}{N\pi_i} \right) \delta_k (1 + k_{\beta_N^*,k}) \{ y_k - m(x_k; \beta_N^*) \} \right]^2 \mid \mathcal{F}_{N,k-1} \right) \\
&\quad + 2c_2^{\mathrm{T}} \text{plim} n \sum_{k=1}^{N} E^*_{\beta_N^*} \left[I_k \left(\omega_k^* u_k - \frac{1}{N\pi_i} \right)^2 \delta_k (1 + k_{\beta_N^*,k}) g(x_k; \beta_N^*) \{ y_k - m(x_k; \beta_N^*) \}^2 c_1 \mid \mathcal{F}_{N,k-1} \right] \\
&\quad + c_2^{\mathrm{T}} \text{plim} n \sum_{k=1}^{N} E^*_{\beta_N^*} \left[I_k \left(\omega_k^* u_k - \frac{1}{N\pi_i} \right)^2 \delta_k g(x_k; \beta_N^*) g(x_k; \beta_N^*)^{\mathrm{T}} \{ y_k - m(x_k; \beta_N^*) \}^2 \mid \mathcal{F}_{N,k-1} \right] c_2 \\
&= c_1^2 \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{I_k(1-\pi_k)}{\pi_k^2} m(x_k; \hat{\beta})^2 + c_1^2 \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{I_k(1-\pi_k)}{\pi_k^2} \delta_k (1 + k_{\hat{\beta},k})^2 \{ y_k - m(x_k; \hat{\beta}) \}^2 \\
&\quad + 2c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{I_k(1-\pi_k)}{\pi_k^2} \delta_k (1 + k_{\hat{\beta},k}) g(x_k; \hat{\beta}) \{ y_k - m(x_k; \hat{\beta}) \}^2 c_1 \\
&\quad + c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{I_k(1-\pi_k)}{\pi_k^2} \delta_k g(x_k; \hat{\beta}) g(x_k; \hat{\beta})^{\mathrm{T}} \{ y_k - m(x_k; \hat{\beta}) \}^2 c_2 \\
&= c_1^2 \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} m(x_k; \beta^*)^2 + c_1^2 \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k (1 + k_{\beta^*,k})^2 \sigma^2(x_k) \\
&\quad + 2c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k (1 + k_{\beta^*,k}) g(x_k; \beta^*) \sigma^2(x_k) c_1 \\
&\quad + c_2^{\mathrm{T}} \text{plim} \frac{n}{N^2} \sum_{k=1}^{N} \frac{1-\pi_k}{\pi_k} \delta_k g(x_k; \beta^*) g(x_k; \beta^*)^{\mathrm{T}} \sigma^2(x_k) c_2.
\end{aligned}$$

Therefore, by the martingale central limit theorem, conditional on the observed data under $P^{\beta_N^*}$, (A13) follows.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects, *Econometrica* **74**: 235–267.

Abadie, A. and Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators, *Econometrica* **76**: 1537–1557.

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects, *Journal of Business & Economic Statistics* **29**: 1–11.

Abadie, A. and Imbens, G. W. (2016). Matching on the estimated propensity score, *Econometrica* **84**: 781–807.

Andreou, E. and Werker, B. J. (2012). An alternative asymptotic analysis of residual-based statistics, *Rev Econ Stat* **94**: 88–99.

Berg, E., Kim, J. K. and Skinner, C. (2016). Imputation under informative sampling, *J. Surv. Statist. Methodol.* **4**: 436–462.

Bickel, P. J., Klaassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Inference in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

Billingsley, P. (1995). *Probability and Measure*, 3 edn, Wiley: New York.

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.

Ding, P. and Li, F. (2017). Causal inference: A missing data perspective, *arXiv preprint arXiv:1712.06170* .

Fuller, W. A. (2009). *Sampling Statistics*, Wiley, Hoboken.

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.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*, Springer: Berlin.

Lenis, D., Nguyen, T. Q., Dong, N. and Stuart, E. A. (2017). It's all about balance: propensity score matching in the context of complex survey data, *Biostatistics* p. kxx063.

Little, R. J. (1988). Missing-data adjustments in large surveys, *Journal of Business & Economic Statistics* **6**: 287–296.

Morris, T. P., White, I. R. and Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws, *BMC Med Res Methodol* **14**: 75.

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.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations, *Journal of Business & Economic Statistics* **4**: 87–94.

Rust, K. F. and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques, *Stat Methods Med Res* **5**: 283–310.

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.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge University Press, Cambridge, MA.

Vink, G., Frank, L. E., Pannekoek, J. and Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables, *Statistica Neerlandica* **68**: 61–90.

Wolter, K. (2007). *Introduction to Variance Estimation*, 2 edn, Springer, New York.

Yang, S. and Kim, J. K. (2016). Fractional imputation in survey sampling: A comparative review, *Statist. Sci.* **31**: 415–432.