Semiparametric fractional imputation using empirical likelihood in survey sampling

Sixia Chen and **Jae Kwang Kim**

Abstract

The empirical likelihood method is a powerful tool for incorporating moment conditions in statistical inference. We propose a novel application of the empirical likelihood for handling item nonresponse in survey sampling. The proposed method takes the form of fractional imputation (Kim, 2011) but it does not require parametric model assumptions. Instead, only the first moment condition based on a regression model is assumed and the empirical likelihood method is applied to the observed residuals to get the fractional weights. The resulting semiparametric fractional imputation provides $\sqrt{n}$-consistent estimates for various parameters. Variance estimation is implemented using a jackknife method. Two limited simulation studies are presented to compare several imputation estimators.

1 Introduction

Missing data are frequently encountered in many areas, such as survey sampling, epidemiology and other fields. Simply ignoring missing values can potentially lead to biased estimation (Little and Rubin 2002, Kim and Shao 2013). Two statistical approaches for handling missing data have been used in practice: propensity score weighting and imputation. Propensity score weighting is used mainly to correct for unit non-response, while imputation is mainly used to handle item nonresponse. Haziza (2009) provides a comprehensive overview of the imputation methods in survey sampling.

Multiple imputation (MI), proposed by Rubin (1987), is a popular approach of imputation for general-purpose estimation due to its practical simplicity. However, the Rubin's variance estimator may be biased under certain situation (Fay 1992; Wang and Robins 1998; Kim, et. al. 2006; Yang and Kim, 2016) and its validity requires the congeniality condition of Meng (1994), which may not hold for general-purpose estimation.

Fractional imputation (FI), first proposed by Kalton and Kish (1984), provides an alternative method for handling item nonresponse. Fay (1996), Kim and Fuller (2004), Fuller and Kim (2005), Durrant (2005), and Durrant and Skinner (2006) discussed fractional hot deck imputation. Kim (2011) and Kim and Yang (2014) discussed a fully parametric approach to fractional imputation. The parametric fractional imputation provides a powerful tool for handling missing data for various situations. However, it relies on a strong parametric model

assumption and making such an assumption is not usually preferred in survey sampling. Balanced random imputation of Chauvet et al (2011) is also an attractive imputation technique, but it still requires parametric model assumptions for multipurpose estimation.

The empirical likelihood (EL) method, considered by Owen (2001) and Qin and Lawless (1994), is a useful tool for semiparametric inference in statistics. It involves a likelihood-based inference without making a parametric distributional assumption about the observed data. Qin (1993) addressed the missing survey data problem by using a biased sampling argument of Vardi (1985). Wang and Rao (2002) brought regression-type imputation approaches to empirical likelihood inference. Wang and Chen (2009) used a nonparametric regression imputation approach to handle missing data in the empirical likelihood inference. Müller (2009) considered a novel application of empirical likelihood method to handle missing data under a regression model assumption. In Müller (2009), the moment condition of the error term in the regression model is used to construct a fully imputed estimator.

In this paper, motivated by the fully imputed estimator of Müller (2009), we propose a semiparametric fractional imputation (SFI) method using empirical likelihood that can be used to handle item nonresponse in survey sampling. Because the proposed SFI uses only moment conditions in the semiparametric regression model, it is more robust than the PFI method or parametric MI method. By using a regression model assumptions, the proposed SFI method is more efficient than the nonparametric regression imputation method of Wang and Chen (2009). The proposed method takes the form of fractional imputation, so the actual implementation is very attractive in practice. The proposed SFI method can be used to estimate various parameters, including nonsmooth parameters such as population quantiles.

The paper is organized as follows. The basic setup is introduced and the proposed method is presented in Section 2. The asymptotic properties of the SFI estimators are presented in Section 3. Extensions to non-smooth statistics as well as random imputations are covered in Section 4. In Section 5, variance estimation is discussed. Some numerical results are given in Section 6. Some concluding remarks are made in Section 7.

2 Basic Setup

Consider a finite population $\mathscr{F}_N = \{(\mathbf{x}_i, y_i); i = 1, 2, \ldots, N\}$, where $\mathbf{x}_i$ is the vector of auxiliary variables that are always observed and y_i is the study variable that is subject to missingness. We assume $(\mathbf{x}_i, y_i)$ are realizations from a regression model

$$Y = m(\mathbf{X}; \boldsymbol{\beta}_0) + \varepsilon, \quad (1)$$

where $m(\mathbf{X}; \boldsymbol{\beta}_0)$ is assumed to be known with unknown parameter $\boldsymbol{\beta}_0$ and ε satisfies $E(\varepsilon|\mathbf{X}) = \mathbf{0}$. No parametric distributional assumption on $\mathbf{X}$ is made.

Let δ_i be the response indicator such that $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ otherwise. We assume missing at random (MAR) in the sense that

$$Pr(\delta=1|\mathbf{x},y)=Pr(\delta=1|\mathbf{x}). \quad (2)$$

Even though we observe δ_i only in the sample, we can conceptually assume that δ_i's are defined throughout the population. Such extended definition of δ_i has been adopted in Fay (1992), Shao and Steel (1999), Kim, Navarro, and Fuller (2006).

Given the finite population, suppose that sample A of size n is selected from the finite population by a probability sampling mechanism. Let $\pi_i, i = 1, 2,\ldots, N$, be the first order inclusion probability of unit i in the population. We are interested in estimating η_0, defined as a solution to the estimating equation $E\{U(\eta, \mathbf{x}, y)\} = 0$ where $U(\eta, \mathbf{x}, y)$ is a known function with parameter η. To avoid unnecessary details, we assume that the solution to $E\{U(\eta, \mathbf{x}, y)\} = 0$ is unique and the dimensions of η and $U(\eta, \mathbf{x}, y)$ are r. Thus, the parameter η is just-identified Under complete response, a consistent estimator of η_0 is obtained by solving

$$\sum_{i \in A} \frac{1}{\pi_i} U(\eta;\mathbf{x}_i, y_i)=0$$

for η. If some of y_i are missing, under the MAR assumption, a consistent estimator of η_0 can be obtained by solving the following expected estimating equation

$$\sum_{i \in A} \frac{1}{\pi_i} [\delta_i U(\eta;\mathbf{x}_i, y_i)+(1 - \delta_i)E\{U(\eta;\mathbf{x}_i, Y)|\mathbf{x}_i\}]=0 \quad (3)$$

for η. The conditional expectation in (3) is with respect to $f(y\,|\,\mathbf{x})$, which is unknown as we only assume (1).

In fractional imputation, our goal is to approximate the conditional expectation in (3) by the weighted mean of the fractionally imputed estimating functions. That is, we wish to achieve

$$E\{U(\eta;\mathbf{x}_i, Y)|\mathbf{x}_i\} \cong \sum_{j=1}^{m} w_{ij}^* U(\eta;\mathbf{x}_i, y_i^{*(j)}) \quad (4)$$

as closely as possible for some $(w_{ij}^*, y_i^{*(j)})$ satisfying $\sum_{j=1}^{m} w_{ij}^*=1$, where w_{ij}^*'s are desired fractional weights and $y_i^{*(j)}$'s are m imputed values for subject i. Kim (2011) and Kim and Yang (2014) developed a fractional imputation satisfying (4) using a parametric model assumption on $f(y\,|\,\mathbf{x})$.

In our proposed method, we use the empirical likelihood approach to achieve the approximation in (4). To explain the idea, assume for now that the true parameter β_0 in (1) is

known. In this case, $\varepsilon_i = y_i - m(\mathbf{x}_i; \boldsymbol{\beta}_0)$ are available among $\delta_i = 1$. Because $E(\varepsilon \mid \mathbf{x}) = 0$ holds, we can compute

$$E\{U(\eta;\mathbf{x}_i, y)|\mathbf{x}_i\} = \int U(\eta;\mathbf{x}_i, y) f(y|\mathbf{x}_i) dy = \int U(\eta;\mathbf{x}_i m(\mathbf{x}_i, \beta_0) + \varepsilon) f_\varepsilon(\varepsilon|\mathbf{x}_i) d\varepsilon,$$

where $f_\varepsilon(\varepsilon \mid \mathbf{x})$ is the (unknown) conditional density of ε given $\mathbf{x}$. To apply the empirical likelihood method, we assume that the conditional distribution of ε given $\mathbf{x}$ can be approximated by

$$F_\varepsilon(\varepsilon|\mathbf{x}) = \sum_{i \in A} \delta_i w_i I(\varepsilon_i \leq \varepsilon) \qquad (5)$$

such that $w_i \geq 0$ with $\Sigma \delta_i w_i = 1$ are the point mass assigned to the observed ε_i by assuming that the support of ε_i is equal to the set of observed ε_i. Using the approximation in (5), we can obtain

$$E\{U(\eta;\mathbf{x}_i, Y)|\mathbf{x}_i\} \cong \sum_{j \in A} \delta_j w_j U(\eta;\mathbf{x}_i; m(\mathbf{x}_i;\boldsymbol{\beta}_0) + \varepsilon_j),$$

which can be written in the fractional imputation form in (4). To determine w_j uniquely, we can use the idea of pseudo empirical likelihood method of Wu and Rao (2006) to maximize

$$l(w) = \sum_{i \in A} \delta_i \pi_i^{-1} log(w_i) \qquad (6)$$

subject to

$$\sum_{i \in A} \delta_i w_i = 1 \text{ and } \sum_{i \in A} \delta_i w_i \varepsilon_i = 0. \qquad (7)$$

In practice, we do not know $\boldsymbol{\beta}_0$ and, hence, we do not observe $\varepsilon_i = y_i - m(\boldsymbol{x}_i; \boldsymbol{\beta}_0)$. We can use $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$ to obtain $\hat{\varepsilon}_i = y_i - m(\mathbf{x}_i; \hat{\beta})$ and apply the above empirical likelihood method to the observed residuals. In general, one can use

$$\hat{U}_\beta(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i \in A} \frac{\delta_i}{\pi_i} \{y_i - m(\mathbf{x}_i; \boldsymbol{\beta})\} \mathbf{h}(\mathbf{x}_i; \boldsymbol{\beta}) = \mathbf{0} \qquad (8)$$

to obtain a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$, where $\mathbf{h}(\mathbf{x}_i; \boldsymbol{\beta})$ is an arbitrary function that enables the above equation to have a solution. If the variance function $V(y|x) = \sigma^2 q(\mathbf{x}_i; \boldsymbol{\beta}_0)$ for a known function q, then one can choose $\mathbf{h}(\mathbf{x}_i; \boldsymbol{\beta}) = \dot{m}(\mathbf{x}_i; \boldsymbol{\beta})/q(\mathbf{x}_i; \boldsymbol{\beta})$, where $\dot{m}(\mathbf{x}_i; \boldsymbol{\beta}) = \partial m(\mathbf{x}_i; \boldsymbol{\beta})/\partial \boldsymbol{\beta}$. This choice is motivated by the quasilikelihood euations for generalized linear models (McCullagh and Nelder, 1989, Ch. 9). The solution to (8) can be called complete-case (CC)

method. The CC estimator is not efficient in general, but it is efficient for estimating $\boldsymbol{\beta}$ under MAR. Thus, the resulting SFI estimator can be constructed as follows:

[Step 1] Obtain $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$ and compute $\hat{\varepsilon}_i = y_i - m(\mathbf{x}_i; \hat{\beta})$ among the respondents.

[Step 2] Find $\hat{w}_i$ that maximizes (6) subject to

$$\sum_{i \in A} \delta_i w_i = 1 \text{ and } \sum_{i \in A} \delta_i w_i \hat{\varepsilon}_i = 0. \tag{9}$$

The solution can be written as

$$\hat{w}_i = \frac{\pi_i^{-1}}{\sum_{k \in A} \delta_k \pi_k^{-1}} \frac{1}{1 + \hat{\lambda} \hat{\varepsilon}_i}, \tag{10}$$

where $\hat{\lambda}$ is obtained by solving the second constraint of (9).

[Step 3] Use $\hat{w}_j$ in Step 2 to approximate

$$E\{U(\eta; \mathbf{x}_i, Y) | \mathbf{x}_i\} \cong \sum_{j \in A} \delta_j w_{ij}^* U(\eta; \mathbf{x}_i, y_i^{*(j)}),$$

where $y_i^{*(j)} = \hat{y}_i + \hat{\varepsilon}_j$ and $w_{ij}^* = \hat{w}_j$.

[Step 4] The SFI estimator $\hat{\eta}_{SFI}$ of $\boldsymbol{\eta}$ is computed by solving

$$\hat{U}_\pi(\eta, \hat{\beta}, \hat{\lambda}) = \frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \left\{ \delta_i U(\eta; \mathbf{x}_i, y_i) + 1 - \delta_j) \sum_{j \in A} \delta_j w_{ij}^* U(\eta; \mathbf{x}_i, y_i^{*(j)}) \right\} = 0 \tag{11}$$

for $\boldsymbol{\eta}$.

Instead of (11), one can also consider a fully imputed estimating equation based on

$$\sum_{i \in A} \frac{1}{\pi_i} E\{U(\eta; \mathbf{x}_i, Y_i) | \mathbf{x}_i\} = 0,$$

which was considered by Müller (2009) under the independently and identically distributed (I.I.D.) setup. The fully imputed estimating equation may lead to a more efficient estimator of $\boldsymbol{\eta}$ (Matloff, 1981) but such over-imputation does not appeal to survey practice since we usually do not want to replace the true values of respondents with some imputed values. In the following section, we present the asymptotic properties of $\hat{\eta}_{SFI}$ under complex survey designs.

3 Asymptotic Properties

To discuss the asymptotic properties of the proposed SFI estimator of η, we first assume a sequence of finite populations and samples with finite fourth moments as in Fuller (2009, Ch.1). The following theorem presents the asymptotic normality of the proposed SFI estimator. The sketched proof of Theorem 1 is provided in Appendix A.

Theorem 1

Under the regularity conditions (C1)–(C13) in Appendix A, the SFI estimator defined in (11) is a $\sqrt{n}$-consistent estimator of η_0, that is

$$\sqrt{n}(\hat{\eta}_{SFI} - \eta_0) \xrightarrow{\mathscr{L}} N(0, B\textstyle\sum_{u2}B'),$$

where $B=[\,E\,\{\partial U(\eta;\mathbf{x}, y)/\partial\eta\}]^{-1}, \sum_{u2}=V(N^{-1}\sum_{i\in A}\pi_i^{-1}\zeta_i)$, and

$$\zeta_i=\delta_i U(\eta_0;\mathbf{x}_i, y_i)+(1-\delta_i)E\,\{U(\eta_0;\mathbf{x}_i, Y)|\mathbf{x}_i\}+\frac{\delta_i}{E(\delta)}C_i+D_W\left[E\left\{\delta\mathbf{h}(\mathbf{x};\boldsymbol{\beta}_0)\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\right\}\right]^{-1}\delta_i\varepsilon_i\mathbf{h}(\mathbf{x}_i;\boldsymbol{\beta}_0),$$

$$(12)$$

and

$$C_i=\overline{U}_m(\varepsilon_i) - E\left\{\overline{U}_m(\varepsilon_i)\right\} - \sigma^{-2}E\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}\varepsilon_i,$$

$$D_W=D+E\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}\sigma^{-2}E\left\{\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}|\delta=1\right\},$$

$$D=E\left((1-\delta)U(\eta_0;\mathbf{x}, y)\left[\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} - E\left\{\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}|\delta=1\right\}\right]l(\varepsilon)\right),$$

with $\sigma^2=E(\varepsilon^2), \overline{U}_m(\varepsilon)=E\,\{(1-\delta)U(\eta_0;\mathbf{x}, y)|\varepsilon\}$, and $l(\varepsilon)=-\,\partial log f_\varepsilon(\varepsilon|\mathbf{x})/\partial\varepsilon$.

Remark 1

In (12), ζ_i can be written as the sum of four terms. The first two terms is the conditional expectation of $U(\eta, \mathbf{x}, y)$, the third term is the additional term due to approximating $f(y\,|\,\mathbf{x})$ by the empirical likelihood method, and the fourth term is the additional term due to estimating $\boldsymbol{\beta}$.

According to Theorem 1, a consistent variance estimator of $\hat{\eta}_{SFI}$ can be written as

$$\hat{V}(\hat{\eta}_{SFI}) = \left\{ \hat{E}\left(\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right) \right\}^{-1} \hat{V}\left(\frac{1}{N}\sum_{i\in A}\pi_i^{-1}\zeta_i\right) \left[\left\{ \hat{E}\left(\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right) \right\}^{-1} \right]^T, \tag{13}$$

where

$$\hat{E}\left(\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right) = \frac{1}{\hat{N}}\sum_{i\in A}\pi_i^{-1}\left\{ \delta_i\frac{\partial U(\hat{\eta};\mathbf{x}_i,y_i)}{\partial \eta} + (1-\delta_i)\sum_{j\in A}\delta_j w_{ij}^*\frac{\partial U(\hat{\eta};\mathbf{x}_i y_i^{*(j)})}{\partial \eta} \right\},$$

with $\hat{N}=\sum_{i\in A}\pi_i^{-1}, \hat{\eta}=\hat{\eta}_{SFI}$ and

$$\hat{V}\left(\frac{1}{N}\sum_{i\in A}\pi_i^{-1}\zeta_i\right) = \frac{1}{\hat{N}^2}\sum_{i\in A}\sum_{j\in A}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\frac{\hat{\zeta}_i\hat{\zeta}_j}{\pi_i\pi_j} + \frac{1}{\hat{N}^2}\sum_{i\in A}\frac{(\hat{\zeta}_i-\hat{\zeta}_N)^2}{\pi_i}, \tag{14}$$

where $\hat{\zeta}_N=\hat{N}^{-1}\sum_{i\in A}\pi_i^{-1}\hat{\zeta}_i$ and $\hat{\zeta}_i$ is a plug-in estimator of ζ_i in (12). One can use

$$\begin{aligned}
\hat{\zeta}_i = &\, \delta_i U(\hat{\eta};\mathbf{x}_i,y_i) \\
&+ (1-\delta_i)\hat{\mu}(\mathbf{x}_i;\hat{\boldsymbol{\beta}},\hat{\eta}) \\
&+ \delta_i\left\{\hat{E}(\delta)\right\}^{-1}\left[\widehat{\overline{U}}_m(\hat{\varepsilon}_i) - \hat{E}\left\{\overline{U}_m(\varepsilon_i)\right\} - \hat{\sigma}^{-2}\hat{E}\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}\hat{\varepsilon}_i\right] \\
&+ \hat{D}_W\left[\hat{E}\left\{\delta\mathbf{h}(\mathbf{x};\boldsymbol{\beta}_0)\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right\}\right]^{-1}\delta_i\hat{\varepsilon}_i\mathbf{h}(\mathbf{x}_i;\hat{\boldsymbol{\beta}}),
\end{aligned}$$

with

$$\hat{\mu}(\mathbf{x}_i;\hat{\boldsymbol{\beta}},\hat{\eta}) = \frac{\sum_{j\in A}\pi_j^{-1}\delta_j U(\hat{\eta};\mathbf{x}_i,y_i^{*(j)})}{\sum_{j\in A}\pi_j^{-1}\delta_j}, \quad \hat{E}(\delta) = \frac{1}{\hat{N}}\sum_{j\in A}\pi_j^{-1}\delta_j,$$

$$\hat{\sigma}^2 = \frac{\sum_{j\in A}\pi_j^{-1}\delta_j\hat{\varepsilon}_j^2}{\sum_{j\in A}\pi_j^{-1}\delta_j}, \quad \hat{E}\left\{\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}|\delta=1\right\} = \frac{\sum_{i\in A}\pi_j^{-1}\delta_i\partial m(\mathbf{x}_i;\hat{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}}{\sum_{i\in A}\pi_i^{-1}\delta_i},$$

$$\hat{D}_W = \hat{D} + \hat{E}\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}\hat{\sigma}^{-2}\hat{E}\left\{\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}|\delta=1\right\},$$

$$\widehat{\overline{U}}_m(\hat{\varepsilon}_i) = \frac{1}{\hat{N}} \sum_{j \in A} \pi_j^{-1} (1 - \delta_j) U(\hat{\eta}; \mathbf{x}_j, y_j^{*(i)}),$$

$$\hat{E}\left\{\overline{U}_m(\varepsilon_i)\right\} = \frac{1}{\hat{N}} \sum_{i \in A} \pi_i^{-1} (1-\delta_i) \sum_{j \in A} \delta_j \hat{w}_j U(\hat{\eta}; \mathbf{x}_i, y_i^{*(j)}), \ \hat{E}\left\{\varepsilon \overline{U}_m(\varepsilon)\right\} = \frac{\sum_{j \in A} \pi_j^{-1} \delta_j \hat{\varepsilon}_j \widehat{\overline{U}}_m(\hat{\varepsilon}_j)}{\sum_{j \in A} \delta_j \pi_j^{-1}},$$

$$\hat{D} = \frac{1}{\hat{N}} \sum_{i \in A} \pi_i^{-1} (1 - \delta_i) \frac{\sum_{j \in A} \delta_j \pi_j^{-1} \partial U(\hat{\eta}; \mathbf{x}_i, y_i^{*(j)}) / \partial y \left\{\partial m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) / \partial \boldsymbol{\beta}\right\}}{\sum_{j \in A_r} \pi_j^{-1}}.$$

When $nN^{-1} = o(1)$, the second term of (14) is of smaller order and can be safely ignored.

4 Extensions

In this section, we discuss two extensions of the proposed method. In Section 4.1, our proposed method is extended to handle non-smooth statistics including distribution functions and percentiles. In Section 4.2, an extension to stochastic imputation is discussed.

4.1 Inference for non-smooth statistics

Suppose that we are interested in estimating parameter η_0, the solution of $E\{U(\eta, \mathbf{x}, y)\} = 0$ with non-smooth function $U(\eta, \mathbf{x}, y)$, where the non-smoothness can be with respect to either η or y. For generality, we assume the non-smoothness is with respect to both η and y. Wang and Opsomer (2011) discussed asymptotic results for nondifferentiable survey estimators. Define $\theta = (\eta, \boldsymbol{\beta}), \theta_0 = (\eta_0 = \boldsymbol{\beta}_0)$ Let $\tilde{U}_n(\theta) = N^{-1} \sum_{i \in A} \pi_i^{-1} \tilde{U}(\theta; \delta_i, \mathbf{x}_i, y_i)$ and $\tilde{U}(\theta) = E\left\{\tilde{U}(\theta; \delta_i, \mathbf{x}_i, y_i)\right\}$, where

$$\tilde{U}(\theta; \delta_i, \mathbf{x}_i, y_i) = \delta_i U(\eta; \mathbf{x}_i, y_i) + (1 - \delta_i) \int U\{\eta; \mathbf{x}_i, m(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i f_\varepsilon(\varepsilon_i | \mathbf{x}_i) d\varepsilon_i.$$

Denote $\hat{\theta} = (\hat{\eta}, \hat{\beta})$ as the solution of estimating equation $\tilde{U}_n(\theta) = 0$. To discuss asymptotic properties, we replace regularity conditions (C7)–(C10) in Appendix A with the regularity conditions (C14)–(C17) in Appendix B. The following theorem presents the asymptotic expansion of $\hat{\eta}_{SFI}$ under this scenario and the sketched proof is presented in Appendix B.

Theorem 2—Under regularity conditions (C1)–(C3), and (C11)–(C17) in Appendix A and Appendix B, $\hat{\eta}_{SFI}$ has the following asymptotic expansion

$$\hat{\eta}_{SFI} - \eta_0 = -\left[\frac{\partial E\{U(\eta_0; \mathbf{x}, y)\}}{\partial \eta}\right]^{-1} \left(\frac{1}{N} \sum_{i \in A} \frac{1}{\pi_i} \zeta_{2i}\right) + o_p(n^{-1/2}),$$

where

$$
\begin{aligned}
\zeta_{2i} = &\delta_i U(\eta_0; \mathbf{x}_i, y_i) \\
&+ (1 - \delta_i)\mu(\mathbf{x}_i; \boldsymbol{\beta}_0, \eta_0) \\
&+ \frac{\delta_i}{E(\delta)}\left[\overline{U}_m(\varepsilon_i) - E\left\{\overline{U}_m(\varepsilon_i)\right\} - \frac{E\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}}{\sigma^2}\varepsilon_i \right] \\
&+ D_W^*\left[E\left\{\delta\mathbf{h}(\mathbf{x};\boldsymbol{\beta}_0)\frac{\partial m(\mathbf{x};\beta_0)}{\partial\boldsymbol{\beta}}\right\} \right]^{-1}\delta_i\varepsilon_i\mathbf{h}(\mathbf{x}_i;\boldsymbol{\beta}_0),
\end{aligned}
$$

where

$$
D_W^* = D* + E\left\{\varepsilon\overline{U}_m(\varepsilon)\right\}\sigma^{-2}E\left\{\frac{\partial m(\mathbf{x};\beta_0)}{\partial\boldsymbol{\beta}}|\delta=1\right\},
$$

and

$$
D^* = \{E(\delta)\}^{-1}\frac{\partial E\left[(1-\delta_i)\delta_j U\left\{\eta_0; \mathbf{x}_i, y_i^{*(j)}(\boldsymbol{\beta})\right\}\right]}{\partial\boldsymbol{\beta}}
$$

evaluated at $\boldsymbol{\beta}_0$ and other terms are the same as those in Theorem 1.

By Theorem 2, we can obtain

$$
\sqrt{n}(\hat{\eta}_{SFI} - \eta_0) \xrightarrow{\mathscr{L}} N(0, B\textstyle\sum_{u2}B'),
$$

where $B = [E\{\ U(\eta; \mathbf{x}, y)/\ \eta\}]^{-1}$ and $\sum_{u2} = V\left\{N^{-1}\sum_{i\in A}\pi_i^{-1}\zeta_{2i}\right\}$. If we are interested in estimating the cumulative density function of y, which is $Pr(y < t)$, then we can choose $U(\eta, \mathbf{x}, y) = I(y < t) - \eta$ and

$$
E\left[(1-\delta_i)\delta_j U\left\{\eta_0; \mathbf{x}_i, y_i^{*(j)}(\boldsymbol{\beta})\right\}\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{1 - p(\mathbf{x}_i)\}p(\mathbf{x}_j)\int_{-\infty}^{t+m(\mathbf{x}_j;\boldsymbol{\beta})-m(\mathbf{x}_i;\boldsymbol{\beta})}f_{y_j|\mathbf{x}_j}(y_j)dy_j\,d\mathbf{x}_i\,d\mathbf{x}_j,
$$

where $p(\mathbf{x}) = \Pr(\delta = 1|\mathbf{x})$. Therefore, we have

$$
D* = \{E(\delta)\}^{-1}E\left[(1-\delta_i)\delta_j\left\{\frac{\partial m(\mathbf{x}_j;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\right\}f_{y_j|\mathbf{x}_j}\{t+m(\mathbf{x}_j;\boldsymbol{\beta}_0)-m(\mathbf{x}_i;\boldsymbol{\beta}_0)\}\right].
$$

A consistent estimators of $D*$ can be written as

$$
\hat{D}* = \sum_{i\in A}\frac{1-\delta_i}{\pi_i}\frac{\sum_{j\in A}\delta_j\pi_j^{-1}\left\{\partial m(\mathbf{x}_j;\hat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta} - \partial m(\mathbf{x}_i;\hat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}\right\}\hat{f}_{y_j|\mathbf{x}_j}\left\{t+m(\mathbf{x}_j;\hat{\boldsymbol{\beta}}) - m(\mathbf{x}_i;\hat{\boldsymbol{\beta}})\right\}}{\hat{N}\sum_{j\in A_r}\pi_j^{-1}}
$$

with

$$\hat{f}_{y|\mathbf{x}}(y|\mathbf{x}) = \frac{(h_{\mathbf{x}}h_y)^{-1}\sum_{i \in A} \pi_i^{-1}\delta_i K_{\mathbf{x}}\left\{h_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{x}_i)\right\} K_y\left\{h_y^{-1}(y - y_i)\right\}}{(h_{\mathbf{x}})^{-1}\sum_{i \in A}\pi_i^{-1}\delta_i K_{\mathbf{x}}\left\{h_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{x}_i)\right\}},$$

where $K_{\mathbf{x}}$ and K_y are kernel functions for $\mathbf{x}$ and y with bandwidth $h_{\mathbf{x}}$ and h_y. Thus, a consistent variance estimator of $\hat{\eta}_{SFI}$ here can be obtained similarly to (13).

If the parameter of interest is the τ-th percentile of Y, given by $\eta = F_Y^{-1}(\tau)$, the SFI estimator $\hat{\eta}_{\tau,SFI}$ of η can be obtained by solving the estimating equation (11) with $U(\eta; \mathbf{x}, y) = I(y < \eta) - \tau$. Since $E\{I(Y < \eta)\} = F_Y(\eta)$, it can be shown that $\hat{\eta}_{\tau,SFI}$ has the asymptotic expansion in Theorem 2 with

$$\frac{\partial E\left\{U(\eta_0; \mathbf{x}, y\right\}}{\partial \eta} = f_y(\eta_0) = E\left\{f_{y|\mathbf{x}}(\eta_0|\mathbf{x})\right\},$$

where f_y is the density function for y. A consistent estimator of $E\{U(\eta_0; \mathbf{x}, y)\} / \eta$ can be written as

$$\frac{\partial \hat{E}\left\{U(\eta_0; \mathbf{x}, y)\right\}}{\partial \eta} = \frac{1}{\hat{N}}\sum_{i \in A}\pi_i^{-1}\hat{f}_{y|\mathbf{x}}(\hat{\eta}|\mathbf{x}_i),$$

and a consistent estimator of $D*$ can be written as

$$\hat{D}* = \sum_{i \in A}\frac{1 - \delta_i}{\pi_i}\frac{\sum_{j \in A}\delta_j \pi_j^{-1}\left\{\partial m(\mathbf{x}_j; \hat{\boldsymbol{\beta}})/\partial \boldsymbol{\beta}\right\}\hat{f}_{y_j|\mathbf{x}_j}\left\{\hat{\eta} + m(\mathbf{x}_j; \hat{\boldsymbol{\beta}}) - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})\right\}}{\hat{N}\sum_{j \in A_r}\pi_j^{-1}},$$

with $\hat{\eta} = \hat{\eta}_{\tau,SFI}$.

4.2 Stochastic imputation

For a multi-purpose survey, stochastic imputation is often preferred to deterministic imputation since it can preserve distributional relationships better. In stochastic imputation, imputed values are generated from a stochastic imputation mechanism and with additional variability due to the imputation. For simplicity, we only consider the case where $U(\eta; \mathbf{x}, y_i^{*(j)})$ is a smooth function of η and $\boldsymbol{\beta}$. The results can be naturally extended to non-smooth statistics. The stochastic imputation estimator $\hat{\eta}_{SFI2}$ can be obtained by solving the following estimating equation

$$\hat{U}_\eta^*(\eta|\hat{\beta}, \hat{\lambda}) = \frac{1}{N}\sum_{i \in A}\frac{1}{\pi_i}\left\{\delta_i U(\eta; \mathbf{x}_i, y_i) + (1 - \delta_i)\frac{1}{M}\sum_{s=1}^{M}U(\eta; \mathbf{x}_i, y_i^{*(s)})\right\} = 0,$$

where $y_i^{*(s)}$ are randomly selected from $\left\{\hat{y}_{ij}=\hat{y}_i+\hat{\varepsilon}_j:j \in A_r\right\}$ with the selection probability, $P(y_i^{*(s)}=\hat{y}_{ij})=w_{ij}^*$ where w_{ij}^* are the fractional weights in (11). Since

$$p \lim_{M \to \infty} \frac{1}{M}\sum_{s=1}^{M}U(\eta;\mathbf{x}_i, y_i^{*(s)})=E(\hat{U}_\eta^*(\eta|\hat{\boldsymbol{\beta}}, \hat{\lambda})|I, \mathbf{x}, y, \delta)=\sum_{j \in A}\delta_j w_{ij}^* U(\eta;\mathbf{x}_i, y_i^{*(j)}),$$

where the conditional expectation is with respect to the stochastic imputation mechanism, we have

$$
\begin{aligned}
V &\left\{\hat{U}_\eta^*(\eta|\hat{\boldsymbol{\beta}}, \hat{\lambda})\right\} \\
&=V\left\{\hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})\right\} \\
&+V\left\{\hat{U}_\eta^*(\eta_0|\hat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})\right\} \\
&=V\left\{\hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})\right\} \\
&+E\left[V\left\{\hat{U}_\eta^*(\eta_0|\hat{\boldsymbol{\beta}}, \hat{\lambda})|I, \mathbf{x}, y, \delta\right\}\right].
\end{aligned}
$$

Thus, using an argument similar to Theorem 1, we can obtain

$$V(\hat{\eta}_{SFI2}) \approx \left\{E\left(\frac{\partial U(\eta_0;\mathbf{x}, y)}{\partial \eta}\right)\right\}^{-1}V_M\left\{E\left(\frac{\partial U(\eta_0;\mathbf{x}, y)}{\partial \eta'}\right)\right\}^{-1}, \quad (15)$$

where $V_M=V\left\{\hat{U}_\eta^*(\eta|\hat{\boldsymbol{\beta}}, \hat{\lambda})\right\}$ Therefore, a consistent variance estimator can be written as

$$\hat{V}(\hat{\eta}RI)=\left\{\hat{E}\left(\frac{\partial U(\eta_0;\mathbf{x}, y)}{\partial \eta}\right)\right\}^{-1}\hat{V}_M\left\{\frac{\partial U(\eta_0;\mathbf{x}, y)}{\partial \eta'}\right\}^{-1},$$

where

$$\hat{V}_M=\hat{V}\left\{\hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda}\right\} +\hat{V}\left\{\hat{U}_\eta^*(\eta_0|\hat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})|I, \mathbf{x}, y, \delta\right\}, \quad (16)$$

and $\hat{E}(\partial U(\eta_0;\mathbf{x}, y)/\partial \eta)$, $\hat{V}\left\{\hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})\right\}$ can be obtained similarly to (13) and

$$\hat{V}\left\{\hat{U}_\eta^*(\eta_0|\hat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{U}_\eta(\eta_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})|I, \mathbf{x}, y, \delta\right\} =\frac{1}{M\hat{N}^2}\sum_{i \in A}\pi_i^{-2}(1 - \delta_i)\sum_{j \in A}\delta_j w_{ij}^*\left\{U(\hat{\eta};\mathbf{x}_i, y_i^{*(j)}) - \sum_{j \in A}\delta_j w_{ij}^* U(\hat{\eta};\mathbf{x}_i, y_i^{*(j)})\right\}^2.$$

The second term of (16) estimates the additional variance due to stochastic imputation. If M is large, the second term is negligible.

5 Replication variance estimation

Estimating the variance of the estimator $\hat{\eta}_{SFI}$ can be done through the linearization formulas presented in Section 3 for smooth statistics and the formulas in Section 4 for non-smooth statistics, respectively. However, it requires tedious algebra to compute all the terms. In this section, we consider an alternative approach using replication methods. Shao and Tu (1995) considered the theoretical aspects of replication methods such as Jackknife and Bootstrap. Wolter (2007) gives a comprehensive overview of replication variance estimation methods in survey sampling.

Suppose we are interested in estimating $T = \sum_{i=1}^{N} y_i$. Define the design weight as $d_i = \pi_i^{-1}$. The design unbiased estimator of T is $\hat{T} = \sum_{i \in A} d_i y_i$ and the consistent replication variance estimator of $\hat{T}$ is given by

$$\hat{V}_R(\hat{T}) = \sum_{k=1}^{L} c_k (\hat{T}^{(k)} - \hat{T})^2,$$

where there are L replication weights, c_k is the replication factor associated with the k-th replication and $\hat{T}^{(k)} = \sum_{i \in A} d_i^{(k)} y_i$ with $d_i^{(k)}$ being the k-th replicate of d_i. For example, $c_k = (L-1)/L$ for deleting one group jackknife method. For details of corresponding c_k with different variance estimation approaches, see Wolter (2007).

To obtain replication variance estimator of our proposed SFI estimator, we apply the same SFI method to each of the replicates. In the first step, we obtain the k-th replicate of $\hat{\beta}^{(k)}$ by solving

$$\sum_{i \in A} d_i^{(k)} \delta_i \{y_i - m(\mathbf{x}_i; \boldsymbol{\beta})\} \mathbf{h}(\mathbf{x}_i; \boldsymbol{\beta}) = 0.$$

In the second step, the replicated EL weights are computed by maximizing

$$l^{(k)}(w) = \sum_{j \in A} \delta_j d_j^{(k)} log(w_j)$$

subject to constraints

$$\sum_{j \in A} \delta_j (d_j^{(k)}/d_j) w_j = 1 \text{ and } \sum_{j \in A} \delta_j (d_j^{(k)}/d_j) w_j \hat{\varepsilon}_j^{(k)} = 0,$$

with $\hat{\varepsilon}_j^{(k)} = y_j - m(\mathbf{x}_j; \hat{\boldsymbol{\beta}}^{(k)})$. In the final step, the replicated SFI estimator is computed using the replicated EL weights. For smooth statistics, the k-th replicate of $\hat{\eta}_{SFI}$, denoted by $\hat{\eta}_{SFI}^{(k)}$, is obtained by the solution to the following estimating equation

$$\sum_{i \in A} d_i^{(k)} \left\{ \delta_i U(\eta; \mathbf{x}, y) + (1 - \delta_i) \sum_{j \in A} \delta_j w_{ij}^{*(k)} U(\eta; \mathbf{x}_i, y_i^{*(j)}(\hat{\boldsymbol{\beta}}^{(k)})) \right\} = 0,$$

where $w_{ij}^{*(k)} = \hat{w}_j^{*(k)}$ *and* $y_i^{*(j)}(\hat{\boldsymbol{\beta}}^{(k)}) = m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}^{(k)}) + y_j - m(\mathbf{x}_j; \hat{\boldsymbol{\beta}}^{(k)})$. The final replication variance estimator of $\hat{\eta}_{SFI}$ is given by

$$\hat{V}_R(\hat{\eta}_{SFI}) = \sum_{k=1}^{L} c_k \left(\hat{\eta}_{SFI}^{(k)} - \hat{\eta}_{SFI} \right)^2.$$

For non-smooth statistics, our estimator is similar to that of Wang and Opsomer (2011). Define

$$\hat{u}^{(k)} = \hat{U}_\eta^{(k)}(\hat{\eta}, \hat{\boldsymbol{\beta}}, \hat{\lambda}) + \left[-\hat{E}\left\{ \varepsilon \overline{U}_m(\varepsilon) \right\} \right] (\hat{\lambda}^{(k)} - \hat{\lambda}) + \hat{D} * (\hat{\boldsymbol{\beta}}^{(k)} - \hat{\boldsymbol{\beta}}),$$

where $\hat{E}\{\varepsilon_m(\varepsilon)\}$ and $\hat{D}_*$ are defined in Section 4.1, $\hat{U}_\eta^{(k)}(\hat{\eta}, \hat{\boldsymbol{\beta}}, \hat{\lambda})$ is defined in (11) with design weight replaced by replication weight $d_i^{(k)}$ and fractional weights replaced by replication fractional weights $w_{ij}^{*(k)}$. Then the replication variance estimator can be written as:

$$\hat{V}_R(\hat{\eta}_{SFI}) = \left\{ \frac{\partial \hat{E}\{U(\hat{\eta}; \mathbf{x}, y)\}}{\partial \eta} \right\}^{-1} \sum_{k=1}^{L} c_k \left\{ \hat{u}^{(k)} - \hat{U}_\eta(\hat{\eta}, \hat{\boldsymbol{\beta}}, \hat{\lambda}) \right\}^2 \left[\left\{ \frac{\partial \hat{E}\{U(\hat{\eta}; \mathbf{x}, y\}}{\partial \eta} \right\}^{-1} \right]^T,$$

with $\hat{E}\{U(\eta; \mathbf{x}, y)\} / \eta$ defined in Section 4.1.

6 Simulation studies

In this Section, we conduct two limited simulation studies. The first one is generated from an artificial data set and the second one is based on the real data treated as a finite population.

6.1 Simulation One

We repeatedly generate $B = 2,000$ finite populations of (x_i, y_i, δ_i) of size $N = 10,000$ from a super-population model

$$y_i = 0.5 x_i + \varepsilon_i,$$

with $x_i \sim \exp(1)$ and $E(\varepsilon_i \mid x_i) = 0$. Two error distributions are considered: (E1) $\varepsilon_i \sim N(0, 1)$ and (E2) $\varepsilon \sim \{\chi^2(2) - 2\}/2$. Given (x, y), the response indicator δ has a Bernoulli distribution with $\Pr(\delta = 1|x) = \{1 + \exp(1 - x)\}^{-1}$. The overall response rate is about 50%. Given each finite population (x, y, δ), we draw a sample by using a Poisson sampling design

with the first-order inclusion probability $\pi_i = n z_i / \sum_{i=1}^N z_i$, where $n = 200$ and $z_i = \max\{0.5 y_i + 2, 1\} + u_i$, with $u_i \sim \chi^2(1)$ and $\chi^2(1)$ corresponding to the chi-squared distribution with degrees of freedom equal to one. In this simulation, we are interested in estimating three parameters:

1.
$\theta_1 = N^{-1} \sum_{i=1}^N y_i$, the population mean of y.

2.
$\theta_2 = N^{-1} \sum_{i=1}^N I(y_i < 1)$, the proportion of y less than 1.

3. $\theta_3 = F^{-1}(0.5)$, the population median of y.

From each sample, we compute the following four estimators:

1.   The complete-case (CC) estimator only based on the complete cases only. The

CC estimator is the solution to $\sum_{i \in A} \delta_i \pi_i^{-1} U(\eta; \mathbf{x}_i, y_i) = 0$, where $U(\eta; \mathbf{x}, y)$ is the corresponding estimating equation for each parameter.

2.   Full sample estimator based on the original sampling without missing data and pseudo empirical likelihood method (Full). Specifically, we maximize

$l = \sum_{i \in A} \pi_i^{-1} \log(\omega_i)$, subject to the following constraints

$$\sum_{i \in A} \omega_i = 1, \quad \sum_{i \in A} \omega_i \hat{\varepsilon}_i = 0,$$

where $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ and $(\hat{\beta}_0, \hat{\beta}_1)$ is obtained by solving the following estimating equation:

$$\sum_{i \in A} \pi_i^{-1}(y_i - \beta_0 - \beta_1 x_i)(1, x_i)^T = 0.$$

The full sample estimator serves as a benchmark for comparison.

3.   The parametric fractional imputation (PFI) estimator of Kim (2011) assuming $y_i \mid x_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ with imputation size $M = 100$.

4.   The nonparametric fractional imputation (NFI) estimator that uses the following nonparametric fractional weights:

$$\omega_{ij}^* = \frac{K_x\{h_x^{-1}(x_i - x_j)\}}{\sum_{j \in A} \delta_j K_x\{h_x^{-1}(x_i - x_j)\}}$$

for each unit $i \in A$ with $\delta_i = 0$ and $j \in A$ with $\delta_j = 1$. We use the reference bandwidth $h_x = 1.06 \hat{N}^{-1/5} \hat{\sigma}_x$ with

$$\hat{\sigma}_x = \left\{ (\hat{N} - 1)^{-1} \sum_{i \in A} \pi_i^{-1} (x_i - \hat{\mu}_x)^2 \right\}^{1/2}, \ \hat{\mu}_x = \hat{N}^{-1} \sum_{i \in A} \pi_i^{-1} x_i \text{ and}$$

$\hat{N} = \sum_{i \in A} \pi_i^{-1}$. A Gaussian kernel density function $K_x(t) = (2\pi)^{-1/2} \exp(-t^2/2)$ has also been used.

5.   The stochastic regression imputation (SRI) estimator assuming the following model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $E(\varepsilon_i) = 0$ and $V(\varepsilon_i) = \sigma^2$.

6.   The proposed semiparametric fractional imputation (SFI) estimator $\hat{\theta}_{SFI}$.

From the Monte Carlo sample of size $B = 2,000$, Monte Carlo bias, standard error and root mean squared error are computed for each point estimator. The results are presented in Table 1. Under (E1) and (E2), the CC estimators perform worst since the response mechanism is not missing completely at random (MCAR). Unless the response mechanism is MCAR, the CC estimator is biased. The FULL estimators always perform best since they assume no missing values and use moment condition (1). Under distribution (E1), the SFI and PFI estimators have similar performances. Among the three imputation estimators, the NFI and SFI estimator performs worst in terms of RMSE for all scenarios since they used less information.

Under model (E2), the SFI estimator shows negligible bias for all parameters, but the PFI estimator has non-negligible bias for estimating proportion and quantile which is due to the misspecification of the error distribution. The NFI and SRI estimators are not as efficient as the SFI estimator in terms of bias and variance. The SFI estimator outperforms PFI, NFI and SRI estimators in terms of RMSE. The overall results indicate the robustness of SFI. For variance estimation, we computed the relative bias based on the Taylor linearization and replication methods, respectively. All the relative bias are below 7%. In addition, we calculate the Monte Carlo coverage rate for the 95% confidence intervals. Under model (E1), the coverage rates are 94.8%, 93.4% and 95.0% for estimating mean, proportion and quantile by using Taylor method and 94.9%, 93.6% and 95.1% by using Replication method. The results under model (E2) are similar and the coverage rates are close to the nominal rate.

6.2 Simulation Two

In the second simulation study, we use 2013–2014 U.S. National Health Examination and Nutrition Survey (NHANES) data as a pseudo finite population. Suppose the study variable is Systolic blood pressure (BPXSY1) and the covariate variable is body mass index (BMXBMI). Keeping only the cases where both BPXSY1 and BMXBMI are greater than zero, the pseudo finite population eventually contains 7104 cases. The scatter plot of BPXSY1 versus BMXBMI is presented in Figure 1. We assume BPXSY1 is roughly linear with respect to BMXBMI. After performing linear regression of BPXSY1 versus BMXBMI, the QQ plot of residuals and residuals vs fitted values plot are presented in Figure 2. The residual plots suggest deviation from normality. The p-value from Anderson-Darling test for

normality is less than 2.2×10^{-16}. We first generate response indicators δ_i, $i = 1, 2,\dots, 7104$ from the following logistic regression model:

$$Pr(\delta_i=1|BMXBMI_i)=\frac{exp\{1-0.1log(BMXBMI_i)\}}{1+exp\{1-0.1log(BMXBMI_i)\}}.$$

The response rate is around 60%. Then given ($BPXSY1_i$, $BMXBMI$, δ_i), $B = 2000$ Monte Carlo samples are generated from simple random sampling with sample size $n = 200$. Assume the parameters of interest are:

(Mean). Finite population mean of BPXSY1, which is $\theta_m = 118.056$.

(Prop1). Finite population proportion one of BPXSY1:

$$\theta_{p1}=\frac{1}{N}\sum_{i=1}^{N}I(BPXSY1_i<80)=0.0008.$$

(Prop2). Finite population proportion two of BPXSY1:

$$\theta_{p2}=\frac{1}{N}\sum_{i=1}^{N}I(BPXSY1_i<120)=0.6017.$$

(Prop3). Finite population proportion three of BPXSY1:

$$\theta_{p3}=\frac{1}{N}\sum_{i=1}^{N}I(BPXSY1_i<160)=0.9711.$$

We consider the same PFI, NFI, SRI and SFI estimators as discussed in Simulation One. The Monte Carlo Bias, Standard Error and Root Mean Squared Error (RMSE) are presented in Table 2. For the population mean, PFI and SFI performs similarly and the NFI estimator has slightly larger bias and standard error. SRI has comparable bias as PFI and SFI, but it has larger SE, as expected. For population proportions, the PFI estimator has substantially larger bias than NFI, SRI and SFI which may be due to the misspecification of error distributions. The NFI and SRI estimators have larger standard errors than PFI and SFI estimators since the nonparametric methods are not as efficient as parametric or semiparametric methods and stochastic imputation will produce larger variance. Overall, SFI estimator performs the best in terms of both bias and variance.

7 Conclusions

Regression imputation is often used to handle item nonresponse in survey sampling. Unlike the usual regression imputation, the proposed semiparametric fractional imputation offers valid inference for a wide set of parameters such as population proportions and quantiles. Besides, only the first moment assumption is needed to obtain a consistent SFI estimator of the parameter, which leads to robust parameter estimation. The proposed SFI method shows good performances in the limited simulation studies.

The proposed method has several possible future research topics. First, instead of assuming ignorable response mechanism, we can consider an extension to nonignorable nonresponse (Kim and Yu, 2011) using an exponential tilting response model. Also, extension of the SFI for handling multivariate missing data will be an important future research topic.

Appendix

A: Proof of Theorem 1

We first assume the following regularity conditions:

(C1) The finite population is a random sample from the semiparametric regression model in (1). The regression function $m(\mathbf{x}; \boldsymbol{\beta})$ in (1) has a continuous first derivative $\partial m(\mathbf{x}; \boldsymbol{\beta})/\partial\boldsymbol{\beta}$ in the neighborhood of the true value $\boldsymbol{\beta}_0$ and $E\{m^2(\mathbf{x}; \boldsymbol{\beta})\}$ and $E\{\partial m(\mathbf{x}; \boldsymbol{\beta})/\partial\boldsymbol{\beta}\}$ are bounded in this neighborhood.

(C2) Function $\mathbf{h}(\mathbf{x}; \boldsymbol{\beta})$ in the estimating function $\hat{U}_\beta(\boldsymbol{\beta})$ in (8) has continuous first derivative $\partial\mathbf{h}(\mathbf{x}; \boldsymbol{\beta})/\partial\boldsymbol{\beta}$ in the neighborhood of the true value $\boldsymbol{\beta}_0$ and $\|\mathbf{h}(\mathbf{x}; \boldsymbol{\beta})\|^2$ and $\|\partial\mathbf{h}(\mathbf{x}; \boldsymbol{\beta})/\partial\boldsymbol{\beta}\|$ are bounded by some integrable function $G_1(\mathbf{x})$ in the neighborhood.

(C3) The model error term in (1) satisfies $E(\varepsilon^2) < \infty$ and $\max\{\|\varepsilon_i\|: i \in A\} = o_p(n^{1/2})$.

(C4) Let $U_\beta(\boldsymbol{\beta}) = E[\delta\{y - m(\mathbf{x}; \boldsymbol{\beta})\}\,\mathbf{h}(\mathbf{x}; \boldsymbol{\beta})]$, assume $\hat{U}_\beta(\boldsymbol{\beta})$ converges to $U_\beta(\boldsymbol{\beta})$ in probability uniformly in the neighborhood of the true value $\boldsymbol{\beta}_0$. For every $a > 0$, $inf_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq a}\|U_\beta(\boldsymbol{\beta})\|>0=\|U_\beta(\boldsymbol{\beta}_0)\|$.

(C5) $\partial\hat{U}_\beta(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ converges to continuous nonsingular derivative $\partial\hat{U}_\beta(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ in probability uniformly in the neighborhood of the true value $\boldsymbol{\beta}_0$.

(C6) $\sqrt{n}\hat{U}_\beta(\boldsymbol{\beta}_0) \xrightarrow{\mathscr{L}} N(0, \sum_\beta)$, as $n, N \to \infty$, where $\sum_\beta = V\left\{\sqrt{n}\hat{U}_\beta(\boldsymbol{\beta}_0)\right\}$ denotes the design model variance, the variance under the joint distribution of the superpopulation model and the sampling mechanism.

(C7) Function $U(\eta, \mathbf{x}, y)$ has continuous partial derivatives $\partial U(\eta, \mathbf{x}, y)/\partial\eta$ and $\partial U(\eta, \mathbf{x}, y)/\partial y$ in the neighborhood of the true value η_0 and $\|U(\eta, \mathbf{x}, y)\|^2$, $\|\partial U(\eta, \mathbf{x}, y)/\partial\eta\|$ and $\|\partial U(\eta, \mathbf{x}, y)/\partial y\|$ are bounded by some integrable function $G_2(\mathbf{x}, y)$ in the neighborhood.

(C8) Let $\hat{U}_n(\eta) = N^{-1}\sum_{i\in A}\pi_i^{-1}U(\eta; \mathbf{x}_i, y_i)$ and $U(\eta) = E\{U(\eta; \mathbf{x}_i, y_i)\}$, then $\hat{U}_n(\eta)$ converges to $U(\eta)$ in probability uniformly in the neighborhood of the true value η_0. For every $a > 0$, $inf_{\eta:\|\eta-\eta_0\|\geq a}\|U(\eta)\|>0=\|U(\eta_0)\|$

(C9) $\partial\hat{U}_n(\eta)/\delta\eta$ converges to continuous nonsingular derivative $\partial U(\eta)/\partial\eta$ in probability uniformly in the neighborhood of the true value η_0.

(C10) $\sqrt{n}\hat{U}_n(\eta_0) \xrightarrow{\mathscr{L}} N(0, \sum_\eta)$, as $n, N \to \infty$, where $\sum_\eta = V\left\{\sqrt{n}\hat{U}_n(\eta_0)\right\}$ denotes the design model variance.

(C11) The first order inclusion probabilities satisfy $K_L \leq N n^{-1} \pi_i \leq K_U$ for all i, where K_L and K_U are positive constants.

(C12) $\max_{i,j} |\pi_{ij} \pi_i^{-1} \pi_j^{-1} - 1| = o(1)$ for any $i, j = 1, 2, \ldots, N$ and $i \neq j$, where π_{ij} are the second order inclusion probability of unit i and unit j in the population.

(C13) The response probability satisfies (2) and $a < \Pr(\delta_i = 1 | \mathbf{x}_i) \leq 1$ for $i = 1, 2, \ldots, N$ for some fixed $a > 0$

Conditions (C1)–(C2) are the model assumptions about the finite population. Condition (C3) is used to control the asymptotic order of $\hat{\lambda}$ in (10). Chen and Sitter (1999, Appendix 2) argued that (C3) holds for common unequal probability sampling designs. Conditions (C4) and (C8) ensure the consistency of $\hat{\beta}$ and $\hat{\eta}$, respectively. Conditions (C5), (C6), (C9) and (C10) are the regularity conditions that ensure asymptotic normality of $\hat{\beta}$ and $\hat{\eta}$. Van der Vaart (1998, Ch. 5) used similar regularity conditions. Specifically, Conditions (C6) and (C10) have been used in many existing literature such as Wu and Rao (2006), Wang and Opsomer (2011), among others. Hajek (1960, 1964) established the asymptotic normality condition under simple random sampling and rejective sampling with unequal selection probabilities. Visek (1979) established the asymptotic normality for the Horvitz-Thompson estimator under Rao-Sampford sampling designs. Condition (C7) controls the smoothness and asymptotic behavior of estimating function $U(\eta, \mathbf{x}, \mathbf{y})$. Conditions (C11) and (C12) are the standard assumptions for the sampling designs. Similar conditions have been used in Isaki and Fuller (1982) and Wang and Opsomer (2011). Condition (C13) controls the behavior of the individual response probability. According to assumption (C3) and by using similar techniques as Wu and Rao (2006), we can show that $\hat{\lambda} = O_p(n^{-1/2})$. Assumption (C4) and Taylor linearization can establish

$$0 = \hat{U}_\beta(\hat{\beta}) = \hat{U}_\beta(\beta_0) + \frac{\partial \hat{U}_\beta(\beta_0)}{\partial \beta}\left(\hat{\beta} - \beta_0\right) + o_p(n^{-1/2}).$$

Therefore,

$$
\begin{aligned}
\hat{\beta} - \beta_0 = & -\left[E\left\{\frac{\partial \hat{U}_\beta\left(\hat{\beta}_0\right)}{\partial \beta}\right\}\right]^{-1} \hat{U}_\beta\left(\hat{\beta}_0\right) \\
& + o_p\left(n^{-1/2}\right) = \left[E\left\{\delta \mathbf{h}\left(\mathbf{x};\beta_0\right)\frac{\partial m\left(\mathbf{x};\beta_0\right)}{\partial \beta}\right\}\right]^{-1}\frac{1}{N}\sum_{i\in A}\pi_i^{-1}\delta_i\varepsilon_i \mathbf{h}\left(\mathbf{x}_i;\beta_0\right) \\
& + o_p\left(n^{-1/2}\right).
\end{aligned}
\tag{A.1}
$$

We know that $\hat{\lambda}$ is the solution of the following estimating equation

$$\hat{U}_\lambda\left(\lambda, \hat{\beta}\right) = \frac{1}{N}\sum_{i\in A}\delta_i \pi_i^{-1}\frac{\hat{\varepsilon}_i}{1 + \lambda\hat{\varepsilon}_i} = 0.$$

In addition, we have

$$\frac{\partial \hat{U}_\lambda (0, \boldsymbol{\beta}_0)}{\partial \lambda} = -\frac{1}{N} \sum_{i \in A} \delta_i \pi_i^{-1} \varepsilon_i^2 = -E(\delta) \sigma^2 + o_p(1), \tag{A.2}$$

and

$$\frac{\partial \hat{U}_\lambda (0, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} = -\frac{1}{N} \sum_{i \in A} \frac{\delta_i}{\pi_i} \frac{\partial m(\mathbf{x}_i; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} = -E\left\{ \frac{\partial m\{\mathbf{x}; \boldsymbol{\beta}_0\}}{\partial \boldsymbol{\beta}} |\delta = 1 \right\} E(\delta) \tag{A.3}$$

Based on (A.2), (A.3), by using Taylor linearization, we have

$$0 = \hat{U}_\lambda \left(\hat{\lambda}, \hat{\boldsymbol{\beta}} \right) = \hat{U}_\lambda (0, \boldsymbol{\beta}_0) + \frac{\partial \hat{U}_\lambda (0, \boldsymbol{\beta}_0)}{\partial \lambda} \hat{\lambda} + \frac{\partial \hat{U}_\lambda (0, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) + o_p \left(n^{-1/2} \right). \tag{A.4}$$

According to (A.1)–(A.4) and after some algebra, it can be shown that

$$\begin{aligned}
\hat{\lambda} &= \frac{1}{\sigma^2} \frac{1}{N} \sum_{i \in A} \pi_i^{-1} \frac{\delta_i}{E(\delta)} \varepsilon_i \\
&\quad - \frac{1}{\sigma^2} E\left\{ \frac{\partial m(\mathbf{x}; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} |\delta = 1 \right\} \\
&\quad \times \left[E\left\{ \mathbf{h} \left(\mathbf{x}; \boldsymbol{\beta}_0 \frac{\partial m(\mathbf{x}; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} |\delta = 1 \right) \right\} \right]^{-1} \frac{1}{N} \sum_{i \in A} \pi_i^{-1} \frac{\delta_i}{E(\delta)} \varepsilon_i \mathbf{h}(\mathbf{x}_i; \boldsymbol{\beta}_0) \\
&\quad + o_p \left(n^{-1/2} \right),
\end{aligned} \tag{A.5}$$

where σ^2 is the variance for the residuals. With condition (C6), it can be shown that $\hat{\eta} = \eta_0 + o_p(1)$. In addition, we have

$$\begin{aligned}
&\frac{\partial \hat{U}_\eta (\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \lambda} \\
&= \frac{1}{N} \sum_{i \in A} \pi_i^{-1} (1 \\
&\quad - \delta_i) \sum_{j \in A} \frac{\delta_j \pi_j^{-1}}{\sum_{k \in A} \delta_k \pi_k^{-1}} \varepsilon_j U(\eta; \mathbf{x}_i, y_i^{*(j)}(\boldsymbol{\beta}_0)) \\
&= -\{E(\delta)\}^{-1} E\left\{ (1 - \delta_i) \delta_j \varepsilon_j U(\eta_0; \mathbf{x}_i, y_i^{*(j)}(\boldsymbol{\beta}_0)) \right\} \\
&\quad + o_p(1) = -E\left\{ \varepsilon \overline{U}_m(\varepsilon) \right\} + o_p(1),
\end{aligned} \tag{A.6}$$

$$\frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \boldsymbol{\beta}}$$

$$=\frac{1}{N}\sum_{i\in A}\frac{(1-\delta_i)}{\pi_i}\sum_{j\in A}\frac{\delta_j\pi_j^{-1}}{\sum_{k\in A}\delta_k\pi_k^{-1}}\frac{\partial U(\eta_0;\mathbf{x}_i,y_i)}{\partial y}\left\{\frac{\partial m(\mathbf{x}_i,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}-\frac{\partial m(\mathbf{x}_j;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right\}$$

$$=\{E(\delta)\}^{-1}E\left[(1-\delta_i)\delta_j\frac{\partial U(\eta_0;\mathbf{x}_i,y_i)}{\partial y}\left\{\frac{\partial m(\mathbf{x}_i;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}-\frac{\partial m(\mathbf{x}_j;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right\}\right]$$

$$+o_p(1)=D+o_p(1), \tag{A.7}$$

and

$$\frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \eta}$$

$$=\frac{1}{N}\sum_{i\in A}\frac{\delta_i}{\pi_i}\frac{\partial U(\eta_0;\mathbf{x}_i,y_i)}{\partial \eta}$$

$$+\frac{1}{N}\sum_{i\in A}\frac{(1-\delta_i)}{\pi_i}\sum_{j\in A}\frac{\delta_j\pi_j^{-1}}{\sum_{k\in A}\delta_k\pi_k^{-1}}\frac{\partial U(\eta_0;\mathbf{x}_i,y_i^{*(j)}))}{\partial \eta}$$

$$=E\left\{\delta\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right\}$$

$$+E\left\{(1-\delta)\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right\}$$

$$+o_p(1)=E\left\{\frac{\partial U(\eta_0;\mathbf{x},y)}{\partial \eta}\right\}+o_p(1), \tag{A.8}$$

where $m(\varepsilon) = E\{(1-\delta)\ U(\eta_0;\ \mathbf{x},\ y)|\varepsilon\}$ and

$$D=E\left((1-\delta)U(\eta_0;\mathbf{x},y)\left[\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}-E\left\{\frac{\partial m(\mathbf{x};\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}|\delta=1\right\}\right]l(\varepsilon)\right),$$

with $l(\varepsilon) = -f'(\varepsilon)/^{-1}(\varepsilon)$. Define

$$S=\frac{1}{N(N-1)E(\delta)}\sum_{i\in A}\omega_i(1-\delta_i)\sum_{j\in A, j\neq i}\omega_j\delta_j U\left(\eta_0;\mathbf{x}_i,y_i^{*(j)}(\boldsymbol{\beta}_0)\right),$$

then by using Taylor linearization,

$$\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)$$

$$=\frac{1}{N}\sum_{i\in A}\pi_i^{-1}\delta_i U(\eta_0;\mathbf{x}_i,y_i)+E(S)+S-E(S)-\frac{E(S)}{E(\delta)}\left\{\bar{\delta}_N-E(\delta)\right\}+o_p(n^{-1/2})$$

$$=\frac{1}{N}\sum_{i\in A}\pi_i^{-1}\delta_i U(\eta_0;\mathbf{x}_i,y_i)+S-\frac{E(S)}{E(\delta)}\left\{\bar{\delta}_N-E(\delta)\right\}+o_p(n^{-1/2}),$$

with $E(S) = E\{(1 - \delta)U(\eta_0; \mathbf{x}, y)\}$ and $\overline{\delta}_N = N^{-1}\sum_{i \in A} \delta_i \pi_i^{-1}$. According to the Hoeffding decomposition,

$$
\begin{aligned}
S &= \frac{1}{N(N-1)E(\delta)}\sum_{i \in A}\pi_i^{-1}(1-\delta_i)\sum_{j \in A, j \neq i}\pi_j^{-1}\delta_j U\left\{\eta_0; \mathbf{x}_i, y_i^{*(j)}(\boldsymbol{\beta}_0)\right\} \\
&= \frac{1}{N}\sum_{i \in A}\left[\pi_i^{-1}(1-\delta_i)\{U(\eta_0; \mathbf{x}_i, y_i)|x_i\} + \pi_i^{-1}\frac{\delta_i}{E(\delta)}E\{(1-\delta_i)U(\eta_0; \mathbf{x}_i, y_i)|\varepsilon_i\}\right] \\
&\quad - E(S) + o_p(n^{-1/2}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0) \\
&= \frac{1}{N}\sum_{i \in A}\pi_i^{-1}\delta_i U(\eta_0; \mathbf{x}_i, y_i) + \frac{1}{N}\sum_{i \in A}\left[\pi_i^{-1}(1-\delta_i)E\{U(\eta_0; \mathbf{x}_i, y_i)|x_i\} + \pi_i^{-1}\frac{\delta_i}{E(\delta)}E\{1-\delta_i)U(\eta_0; \mathbf{x}_i, y_i)|\varepsilon_i\}\right] \\
&\quad - \frac{E(S)}{E(\delta)}\overline{\delta}_N + o_p(n^{-1/2}).
\end{aligned}
$$

(A.9)

According to Taylor linearization, we have

$$
0 = \hat{U}_\eta(\hat{\eta}, \hat{\boldsymbol{\beta}}, \hat{\lambda}) = \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0) + \frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \eta}(\hat{\eta} - \eta_0) + \frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \lambda}\hat{\lambda} + o_p(n^{-1/2}).
$$

(A.10)

By (A.1),(A.5)–(A.10), after some algebra, we can show that

$$
\hat{\eta} - \eta_0 = -\left\{E\left(\frac{\partial U(\eta_0; \mathbf{x}, y)}{\partial \eta}\right)\right\}^{-1}\left(\frac{1}{N}\sum_{i \in A}\pi_i^{-1}\zeta_i\right) + o_p(n^{-1/2}),
$$

where ζ_i is defined in (12) of Theorem 1.

B: Proof of Theorem 2

We replace regularity conditions (C7)–(C10) in Appendix A with the following regularity conditions (C14)–(C17):

(C14) $\tilde{U}_n(\theta)$ converges to $\tilde{U}(\theta)$ in probability uniformly in the neighborhood of the true value θ_0. For every $a > 0$. $inf_{\theta:\|\theta - \theta_0\| \geq a}\|\tilde{U}(\theta)\| > 0 = \|\tilde{U}(\theta_0)\|$

(C15) There exists a measurable function $L(\delta, \mathbf{x}, y)$ with $E\{L^2(\delta, \mathbf{x}, y)\} < \infty$ and for every θ_1 and θ_2 in the neighborhood of the true value θ_0, $\|\tilde{U}(\theta_\rho \; \delta, \mathbf{x}, y) - \tilde{U}(\theta_2; \delta, \mathbf{x}, y)\| \quad L(\delta, \mathbf{x}, y)\|\theta_1 - \theta_2\|$.

(C16) Assume that $E\left\{\tilde{U}^2(\theta_0; \delta, \mathbf{x}, y)\right\} < \infty$ and $E\left\{\tilde{U}(\theta; \delta, \mathbf{x}, y)\right\}$ has continuous and invertible first derivatives with respect to θ and the corresponding first derivatives are bounded by some integrable function in the neighborhood of the true value θ_0.

(C17) $\sqrt{n}\tilde{U}_n(\theta_0) \xrightarrow{\mathscr{L}} N(0, \sum_\theta)$, as $n, N \to \infty$, where $\sum_\theta = V\left\{\sqrt{n}\tilde{U}_n(\theta_0)\right\}$ denotes the design model variance.

Similar as conditions (C4) and (C8), condition (C14) ensures the consistency of proposed estimator. Conditions (C15) and (C16) are required to derive asymptotic expansion of proposed estimator. See Van der Vaart (1998, Ch. 5) for more details for those conditions. Similar as conditions (C6) and (C10), Condition (C17) is used to derive the central limit theory.

The proof of the consistency of $\hat{\beta}$ and $\hat{\eta}$ is similar to the relevant proof in Theorem 1. According to the regularity conditions (C10), (C11), (C12) and by using similar techniques as that of Theorem 19.26 of Van der Vaart (1998), we can show that

$$
\begin{aligned}
0 =& \hat{U}_\eta(\hat{\eta}, \hat{\boldsymbol{\beta}}, \hat{\lambda}) \\
=& \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0) + \frac{\partial \hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)}{\partial \lambda}\hat{\lambda} + \frac{\partial E\left\{\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)\right\}}{\partial \boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \frac{\partial E\left\{\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)\right\}}{\partial \eta}(\hat{\eta} \\
& - \eta_0) + o_p(n^{-1/2}).
\end{aligned}
$$

(B.1)

In addition, we have

$$
\frac{\partial E\left\{\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)\right\}}{\partial \eta} = \frac{\partial E\left\{\delta U(\eta_0, \mathbf{x}, y)\right\}}{\partial \eta} + \frac{1}{E(\delta)}\frac{\partial E\left\{(1-\delta_i)\delta_j U_\eta(\eta_0; \mathbf{x}_i, y_i)\right\}}{\partial \eta} + o_p(1) = \frac{\partial U(\eta_0; \mathbf{x}, y)}{\partial \eta} + o_p(1),
$$

(B.2)

and

$$
\frac{\partial E\left\{\hat{U}_\eta(\eta_0, \boldsymbol{\beta}_0, 0)\right\}}{\partial \boldsymbol{\beta}} = D* + o_p(1),
$$

(B.3)

where D^* is defined in Theorem 2. According to (A.1), (A.5), (A.6), (A.9), (B.1)–(B.3), we have

$$\hat{\eta}_{SFI} - \eta_0 = - \left[\frac{\partial E\left\{ U(\eta_0; \mathbf{x}, y) \right\}}{\partial \eta} \right]^{-1} \left(\frac{1}{N} \sum_{i \in A} \pi_i^{-1} \zeta_i \right) + o_p(n^{-1/2}),$$

where ζ_i is defined in Theorem 2.

References

Chen J, Sitter R. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. Statistica Sinica. 1999; 9:385–406.

Chauvet G, Deville JC, Haziza D. On balanced random imputation in surveys. Biometrika. 2011; 98:459–471.

Durrant GB. Imputation methods for handling item-nonresponse in the social sciences: a methodological review. ESRC National Center for Research Methods and Southampton Stat Sci.s Research Institute NCRM Methods Review Papers NCRM/002. 2005

Durrant GB, Skinner C. Using missing data methods to correct for measurement error in a distribution function. Survey Methodology. 2006; 32(1):25–36.

Fay RE. When are inferences from multiple imputation valid? Proceedings of the Survey Research Methods Section of the American Statistical Association. 1992; 81:227–32.

Fay RE. Alternative paradigms for the analysis of imputed survey data. Journal of the American Statistical Association. 1996; 91(434):490–498.

Fuller WA, Kim JK. Hot deck imputation for the response model. Survey Methodology. 2005; 31:139–149.

Fuller, WA. Sampling Statistics. Wiley; Hoboken, NJ: 2009.

Haziza, D. Imputation and inference in the presence of missing data. In: Pfeffermann, D., Rao, CR., editors. Handbook of Statistics. Vol. 29, Sample Surveys: Theory, Methods and Inference. Amsterdam: Elsevier BV; 2009. p. 215-46.

Kalton G, Kish L. Some efficient random imputation methods. Communications in Statistics A. 1984; 13:1919–1939.

Kim JK, Fuller WA. Fractional hot deck imputation. Biometrika. 2004; 91(3):559–578.

Kim JK, Brick J, Fuller WA, Kalton G. On the bias of the multiple-imputation variance estimator in survey sampling. Journal of Royal Statistical Society: Series B. 2006; 68(3):509–521.

Kim JK, Navarro A, Fuller WA. Replicate variance estimation after multi-phase stratified sampling. Journal of the American Statistical Association. 2006; 101:312–320.

Kim JK. Parametric fractional imputation for missing data analysis. Biometrika. 2011; 98:119–132.

Kim JK, Yu CL. A semi-parametric estimation of mean functionals with non-ignorable missing data. Journal of the American Statistical Association. 2011; 106:157–165.

Kim, JK., Shao, J. Statistical methods for handling incomplete data. London: Chapman and Hall/CRC; 2013.

Kim JK, Yang S. Fractional hot deck imputation for robust inference under item nonresponse in survey sampling. Survey Methodology. 2014; 40:211–230.

Little, RJA., Rubin, DB. Statistical Analysis With Missing Data. 2nd. Hoboken, NJ: Wiley; 2002.

Matloff NS. Use of regression functions for improved estimation of means. Biometrika. 1981; 68:685–689.

McCullagh, P., Nelder, J. Generalized Linear Models. London: Chapman and Hall; 1989.

Meng XL. Multiple-imputation inferences with uncongenial sources of input. Statistical Science. 1994; 9:538–558.

Müller UU. Estimating linear functionals in nonlinear regression with response missing at random. Annals of Statistics. 2009; 98:2245–2277.

Owen, AB. Empirical Likelihood. Chapman and Hall/CRC; New York: 2001.

Qin J. Empirical likelihood in biased sample problems. Annals of Statistics. 1993; 21(3):1182–1196.

Qin J, Lawless J. Empirical likelihood and general estimating equations. The Annals of Statistics. 1994; 22:300–325.

Rubin, DB. Multiple Imputation for Nonresponse in Surveys. Wiley; 1987.

Shao, J., Tu, D. The Jackknife and Bootstrap. Springer; 1995.

Shao J, Steel P. Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. Journal of the American Statistical Association. 1999; 94:254–265.

Vardi Y. Empirical distributions in selection bias models. Annals of Statistics. 1985; 13:178–203.

Van der Vaart, AW. Asymptotic Statistics. New York: Cambridge University Press; 1998.

Víšek, JA. Asymptotic distribution of simple estimate for rejective, Sampford and successive sampling. In: Jurecková, J., editor. Contributions to Statistics: Jaroslav Hįek Memorial. Academia, Prague & D. Reidel; Dordrecht: 1979. p. 263-275.

Wang N, Robins JM. Large-sample theory for parametric multiple imputation procedures. Biometrika. 1998; 85(4):935–948.

Wang Q, Rao JNK. Empirical likelihood-based inference under imputation for missing response data. The Annals of Statistics. 2002; 30:896–924.

Wang D, Chen SX. Empirical likelihood for estimating equations with missing values. The Annals of Statistics. 2009; 37:490–517.

Wang JQ, Opsomer JD. On asymptotic normality and variance estimation for nondifferentiable survey estimators. Biometrika. 2011; 98:91–106.

Wolter, KM. Introduction to Variance Estimation. Wiley; New York: 2007.

Wu C, Rao JNK. Pseudo empirical likelihood ratio confidence intervals for complex surveys. The Canadian Journal of Statistics. 2006; 34:359–375.

Yang S, Kim JK. A Note on Multiple Imputation for General-Purpose Estimation. Biometrika. 2016; 103:244–251.
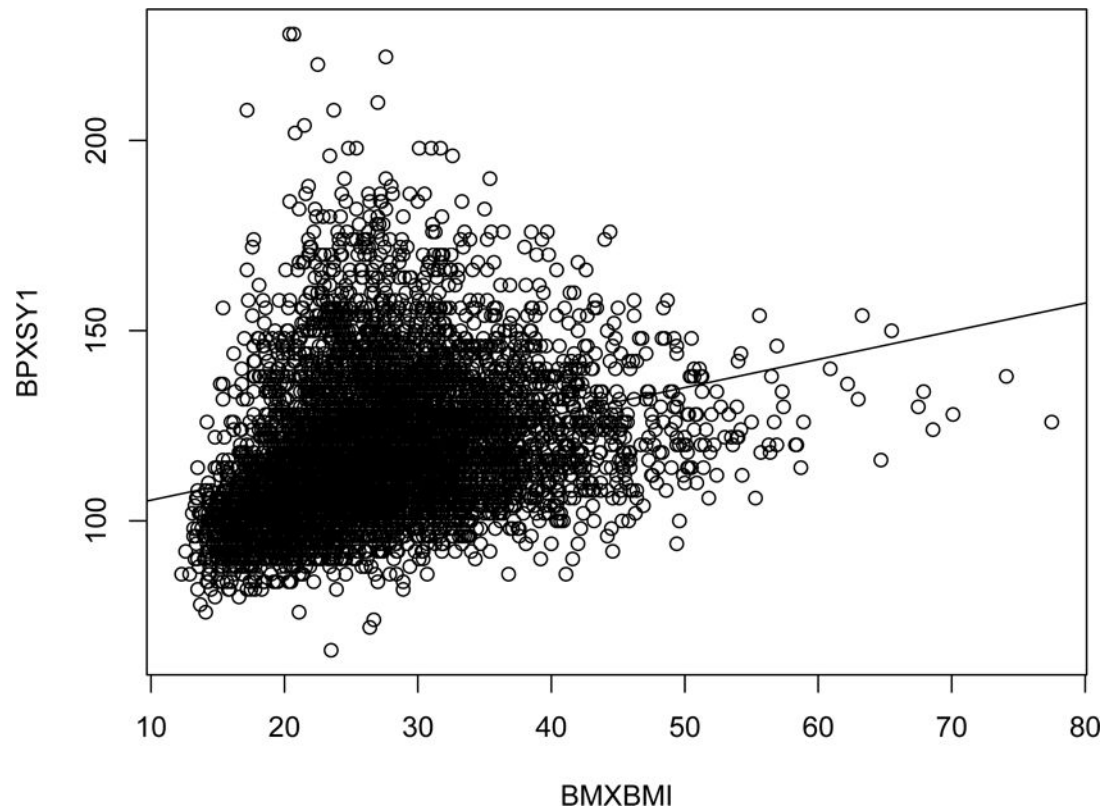
Figure 1.
Scatter plot of BPXSY1 vs BMXBMI
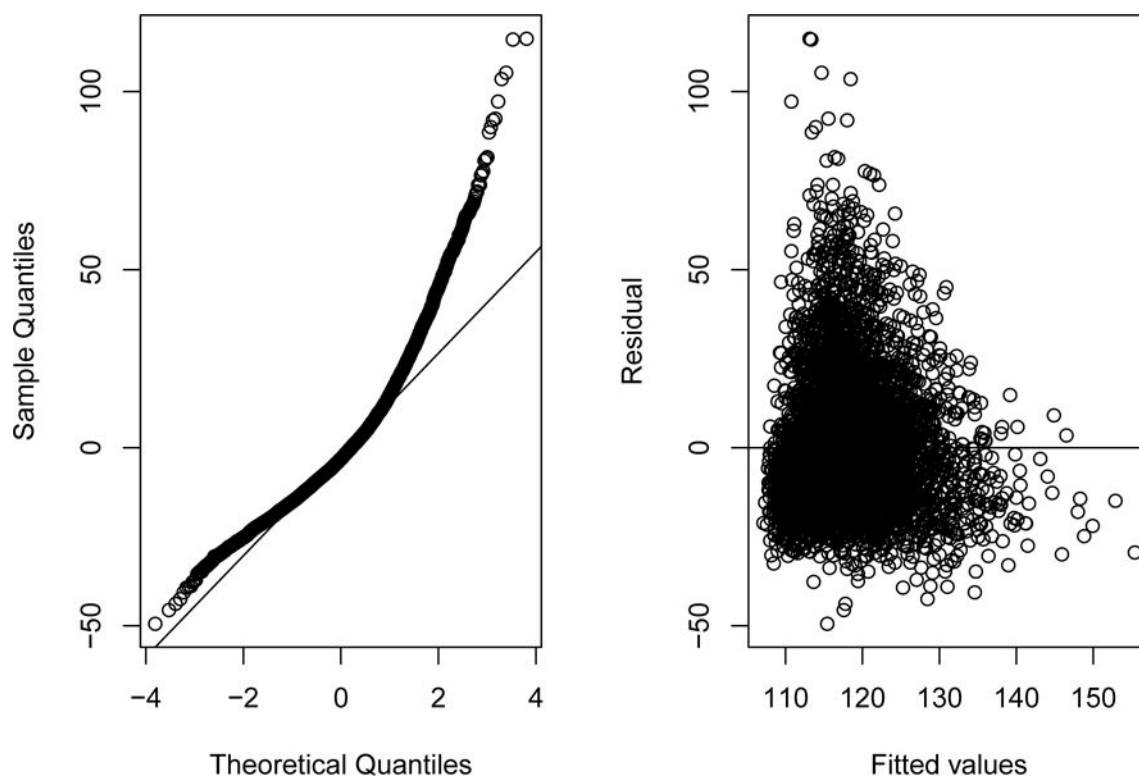
Figure 2.
QQ plot (left panel) and Residual vs fitted value plot (right panel)

Table 1

The Monte Carlo Bias $(\times 10^{-2})$, Standard Error (SE) $(\times 10^{-2})$ and Root Mean Squared Error (RMSE) $(\times 10^{-2})$ for four different methods with two error distributions in Simulation One.

Par	Method	(E1)			(E2)		
		Bias	**SE**	**RMSE**	**Bias**	**SE**	**RMSE**
$E(y)$	CC	18.6	13.0	22.7	20.0	19.8	28.1
	FULL	0.1	5.9	5.9	0.3	9.0	9.0
	PFI	−0.1	7.8	7.8	0.7	12.0	12.0
	NFI	0.7	12.9	12.9	2.9	21.3	21.5
	SRI	−0.2	13.9	13.9	1.7	23.3	23.3
	SFI	−0.2	6.9	6.9	0.3	12.4	12.4
$Pr(y < 1)$	CC	−6.3	5.3	8.2	−3.5	4.8	5.9
	FULL	0.0	3.0	3.0	0.0	2.6	2.6
	PFI	0.2	3.1	3.1	−5.4	3.1	6.2
	NFI	−0.1	5.1	5.1	−0.5	4.9	4.9
	SRI	0.2	5.0	5.0	−0.4	5.0	5.0
	SFI	0.1	3.2	3.2	0.1	3.3	3.3
Quantile	CC	15.0	13.9	20.4	21.9	22.8	31.6
	FULL	0.1	7.8	7.8	0.2	13.2	13.2
	PFI	−0.3	8.5	8.5	30.0	14.5	33.3
	NFI	−0.8	14.9	14.9	2.7	25.3	25.4
	SRI	−1.0	15.7	15.7	2.4	25.4	25.5
	SFI	−0.5	9.1	9.1	0.2	17.0	17.0

Table 2

The Monte Carlo Bias ($\times 10^{-2}$), Standard Error (SE) ($\times 10^{-2}$) and Root Mean Squared Error (RMSE) ($\times 10^{-2}$) for four different methods and four parameters.

Par	Method	Bias	SE	RMSE
Mean	COM	−2.9	124.8	124.9
	PFI	−2.3	153.2	153.2
	NFI	−5.0	153.5	153.6
	SRI	1.4	169.7	169.7
	SFI	−2.2	153.3	153.3
Prop1	COM	0.0	0.2	0.2
	PFI	0.5	0.3	0.6
	NFI	0.0	0.3	0.3
	SRI	0.1	0.3	0.3
	SFI	0.0	0.2	0.2
Prop2	COM	0.0	3.4	3.4
	PFI	−2.2	3.8	4.4
	NFI	−0.5	4.2	4.3
	SRI	0.5	4.2	4.3
	SFI	0.2	3.9	3.9
Prop3	COM	0.0	1.2	1.2
	PFI	0.7	1.1	1.3
	NFI	0.2	1.4	1.4
	SRI	−0.3	1.6	1.6
	SFI	0.1	1.4	1.4