Statistica Sinica **32** (2022), 1611-1631
doi:https://doi.org/10.5705/ss.202020.0053

ESTIMATION FOR NONIGNORABLE MISSING RESPONSE OR COVARIATE USING SEMI-PARAMETRIC QUANTILE REGRESSION IMPUTATION AND A PARAMETRIC RESPONSE PROBABILITY MODEL

Emily Berg and Cindy Yu

Iowa State University

Abstract: We address the problem of imputation when a response or covariate may be subject to a nonignorable (or, equivalently, missing not at random) nonresponse, meaning the response probability may depend on a variable that is not always observed. We discuss model identification and develop a novel estimator of the parameters of the response probability. We use a propensity score adjustment to incorporate a subset for which both the response and the covariate are missing. We derive an approximation for the large-sample variance and assess the finite-sample properties of the variance estimator using simulations. The simulation results also show that a quantile regression offers a compromise between fully parametric and nonparametric alternatives. In an application to data from a 2011 survey of pet owners, a quantile regression allows us to model complex relations between two types of veterinary expenditures, where we find evidence of a nonignorable nonresponse.

Key words and phrases: B-spline, missing not at random, survey sampling.

1. Introduction

A widely adopted remedy for missing data is to replace each missing value with one or more imputed values (Kim and Shao (2014); Rubin (1987)). An imputation model defines (1) a relationship between a response (y) and a covariate (x), and (2) the nature of the dependence between the event of responding and (x, y). A common simplifying assumption is that the data are missing at random (MAR), meaning that the probability of responding is independent of the missing variable, after conditioning on the fully observed variables. Under the MAR assumption, Kim and Yu (2011) and Wang and Chen (2009) develop fully parametric and nonparametric imputation procedures, respectively. Chen and Yu (2016) and Berg and Yu (2019) construct imputed values under the assumptions of a semiparametric quantile regression model, assuming an MAR nonresponse.

Corresponding author: Emily Berg, Department of Statistics, Iowa State University, Ames, IA 50011, USA. E-mail: emilyb@iastate.edu.

When the event of responding is not independent of missing values, given observed values, the response mechanism is called missing not at random (MNAR) or nonignorable. (Hereafter, we use MNAR and nonignorable interchangeably.) We extend Chen and Yu (2016) to a nonignorable nonresponse for a data structure in which neither the response nor the covariate is fully observed.

A condition for model identification in the presence of an MNAR nonresponse is the existence of a nonresponse instrument, a variable that is correlated with the response y, but conditionally independent of the event of responding given y (Wang, Shao and Kim (2014)). Tang, Little and Raghunathan (2003) estimates a fully parametric model for y given x, using x as an instrument, without requiring a specific form for the response probability. Zhao (2015) extends the framework of Tang, Little and Raghunathan (2003) to include an additional instrument. Other approaches, such as those of Wang, Shao and Kim (2014) and Chang and Kott (2008), use an instrumental variable to estimate a parametrized propensity score model that depends on y, but not on the instrument. Shao and Wang (2016) generalize the propensity score model of Wang, Shao and Kim (2014) to include a nonparametric component. Riddles, Kim and Im (2016) use likelihood-based methods to improve upon the efficiency of the calibration estimation. Zhao and Ma (2019) use an instrumental variable, but avoid estimating the response probability directly. Miao and Tchetgen Tchetgen (2016) develops a doubly robust estimator under the assumption that an instrumental variable (called a "shadow variable" in their work) exists. Fang, Zhao and Shao (2018) uses an instrumental variable assumption to estimate the coefficient associated with a missing covariate when the response probability depends on the covariate. However, it is well known that identifying an instrumental variable in a given data set is nontrivial. Morikawa and Kim (2017) generalize the instrumental variable condition of Wang, Shao and Kim (2014) by deriving a necessary and sufficient condition for model identification under an MNAR nonresponse. They develop an efficient propensity score estimator, assuming a univariate response variable is missing and a univariate covariate is fully observed. We extend the identification condition of Morikawa and Kim (2017) to accommodate missing covariates and construct a completed data set through imputation.

We propose generating imputed values from a semiparametric quantile regression model, and then using estimates of the response probabilities to approximate the required expectations for nonrespondents. We augment the imputation procedure with a propensity score adjustment to incorporate a subset for which both the response and the covariate are missing. In our application, x and y represent two types of veterinary expenditures, neither of which is fully observed, and ei-

Table 1. Structure of Missing Data.

Covariate x	Response y	Response Indicator δ
✓	✓	1
✓	?	2
?	✓	3

ther of which may influence the probability of responding. The semiparametric quantile regression provides the flexibility needed to model the nonlinear associations between the two types of veterinary expenditures. We define parametric and nonparametric alternatives for the purpose of comparison in the simulation study. Because our data set has a univariate covariate, we focus on that case, and briefly discuss an extension to multivariate covariates in Section 6.

We validate our proposed procedure by means of theory and simulation, and then apply the method to data from a survey of pet owners. In Section 2, we define the model assumptions and imputation and estimation procedures. In Section 3, we define a variance estimator based on a linear approximation. In Section 4, we conduct simulation studies to compare alternative imputation models and assess the finite-sample properties of the variance estimator. We apply the method to impute veterinary expenditures in Section 5. We summarize and discuss future work in Section 6.

2. Model Assumptions and Imputation and Estimation Procedures

Let x_i and y_i denote a continuous covariate and a continuous response variable, respectively, with a compact support on the box $[M_{1x}, M_{2x}] \times [M_{1y}, M_{2y}]$, where $i = 1, \ldots, n$. Let δ_i denote a response indicator variable such that $\delta_i = 1$ if both x_i and y_i are observed, $\delta_i = 2$ if x_i is observed and y_i is missing, and $\delta_i = 3$ if y_i is observed and x_i is missing. We also use $\delta_{ki} = I[\delta_i = k]$, for $k = 1, 2, 3$. Table 1 shows the data structure.

Assume that (x_i, y_i, δ_i), for $i = 1, \ldots, n$, are independent and identically distributed (iid) realizations of the random variable (X, Y, Δ) with joint cumulative distribution function (CDF) $F(x, y, \delta)$. Further, assume X and Y are absolutely continuous, and denote their corresponding conditional probability density functions by $f(y|x, \delta)$ and $f(x|y, \delta)$, respectively. Assume Δ has parametric conditional probability mass function given by

$$P(\Delta = k \mid X = x, Y = y) = \frac{\exp(\phi_{k0} + \phi_{k1}x + \phi_{k2}y)}{\sum_{k=1}^{3} \exp(\phi_{k0} + \phi_{k1}x + \phi_{k2}y)}, \qquad (2.1)$$

for $k = 1, 2, 3$, where $(\phi_{10}, \phi_{11}, \phi_{12}) = (0, 0, 0)$.

To identify the parameters of (2.1), we require an additional assumption. By a direct extension of Theorem 3.1 of Morikawa and Kim (2017) to missing covariates, the additional assumption is that $F(x, y, \delta)$ is a joint CDF such that the condition

$$\begin{aligned} &E[\exp(-\phi_{20} - \phi_{21}x_i - \phi_{22}Y) \mid x_i, \delta_i = 1] \\ &= E[\exp(-\phi'_{20} - \phi'_{21}x_i - \phi'_{22}Y) \mid x_i, \delta_i = 1] \end{aligned} \tag{2.2}$$

almost everywhere implies $(\phi_{20}, \phi_{21}, \phi_{22}) = (\phi'_{20}, \phi'_{21}, \phi'_{22})$, and the condition

$$\begin{aligned} &E[\exp(-\phi_{30} - \phi_{31}X - \phi_{32}y_i) \mid y_i, \delta_i = 1] \\ &= E[\exp(-\phi'_{30} - \phi'_{31}X - \phi'_{32}y_i) \mid y_i, \delta_i = 1] \end{aligned} \tag{2.3}$$

almost everywhere implies $(\phi_{30}, \phi_{31}, \phi_{32}) = (\phi'_{30}, \phi'_{31}, \phi'_{32})$. If $\phi_{31} = \phi_{22} = 0$, then MAR holds and the model is automatically identified.

Sufficient conditions for (2.2) and (2.3) are that

$$h_y(\phi_{22}, x) = -\log(E[\exp\{-\phi_{22}Y\} \mid x, \delta = 1]) \tag{2.4}$$

is not in the column space of x, and that

$$h_x(\phi_{31}, y) = -\log(E[\exp\{-\phi_{31}X\} \mid y, \delta = 1]) \tag{2.5}$$

is not in the column space of y. If $h_y(\phi_{22}, x)$ is in the column space of x, then ϕ_{21} is confounded with ϕ_{22}. Similarly, we require that $h_x(\phi_{31}, y)$ not be in the column space of y to prevent ϕ_{32} from being confounded with ϕ_{31}. Note that $-h_y(\phi_{22}, x)$ is the cumulant generating function of $f(y \mid x, \delta = 1)$, and likewise for $-h_x(\phi_{31}, y)$. An aspect of (2.4) and (2.5) that is of practical importance is that one can check these conditions using $\{(x_i, y_i) : \delta_i = 1\}$, as we illustrate in the data analysis of Section 5.

Let the parameter of interest be $\theta_0 = Eg(X, Y) = \sum_{\delta=1}^{3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, dF(x, y, \delta)$. In the absence of any nonresponse, an estimator of $Eg(X, Y)$ is $\hat{\theta}_{full} = n^{-1} \sum_{i=1}^{n} g(x_i, y_i)$. The estimator $\hat{\theta}_{full}$ is not directly applicable because of the nonresponse. By Cheng (1994), a consistent estimator of θ_0 is

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} \{\delta_{1i} g(x_i, y_i) + \delta_{2i} E[g(x_i, Y) | x_i, \delta_i = 2] + \delta_{3i} E[g(X, y_i) | y_i, \delta_i = 3]\}. \tag{2.6}$$

We convert the expectations given $\delta = 2$ or $\delta = 3$ in (2.6) to expectations given $\delta = 1$ using an "exponential tilting" relationship (Kim and Yu (2011)). Under (2.1), it is straightforward to show that

$$f(y|x,\delta=2) = \frac{f(y|x,\delta=1)\exp(\phi_{22}y)}{E[\exp(\phi_{22}Y) \mid x,\delta=1]} \tag{2.7}$$

and

$$f(x|y,\delta=3) = \frac{f(x|y,\delta=1)\exp(\phi_{31}x)}{E[\exp(\phi_{31}X) \mid y,\delta=1]}, \tag{2.8}$$

where ϕ_{22} and ϕ_{31} are the tilting parameters. The equality in (2.7) allows us to express the conditional expectation for the group with $\delta = 2$ in (2.6) as a function of different expectations given $\delta = 1$ by

$$E[g(x,Y)|x,\delta=2] = \frac{E[g(x,Y)\exp(\phi_{22}Y)|x,\delta=1]}{E[\exp(\phi_{22}Y)|x,\delta=1]}. \tag{2.9}$$

Similarly, the third conditional expectation for the group with $\delta = 3$ in (2.6) converts to a ratio of two expectations given $\delta = 1$ as

$$E[g(X,y)|y,\delta=3] = \frac{E[g(X,y)\exp(\phi_{31}X)|y,\delta=1]}{E[\exp(\phi_{31}X)|y,\delta=1]}. \tag{2.10}$$

The expressions (2.9) and (2.10) show that we can estimate θ using (1) estimates of $f(y \mid x, \delta = 1)$ and $f(x \mid y, \delta = 1)$, and (2) estimates of ϕ_{22} and ϕ_{31}. We focus on using a semiparametric quantile regression to estimate $f(y \mid x, \delta = 1)$ and $f(x \mid y, \delta = 1)$. We compare the results to those from nonparametric and fully parametric alternatives in the simulations. We first define our estimation method for known (ϕ_{22}, ϕ_{31}) in Section 2.1, and then explain how to estimate unknown (ϕ_{22}, ϕ_{31}) in Section 2.2.

2.1. Approximating expectations using estimated quantiles

We approximate $f(y|x, \delta = 1)$ and $f(x|y, \delta = 1)$ through their conditional quantile regression functions, denoted by $q_\tau(x)$ and $q_\tau(y)$, respectively, for $\tau \in (0, 1)$. By definition, the quantile regression functions satisfy $\tau = P(Y \leq q_\tau(x)|x, \delta = 1)$ and $\tau = P(X \leq q_\tau(y)|y, \delta = 1)$. Assume $q_\tau(x)$ and $q_\tau(y)$ are one-to-one functions of x and y, respectively, for every τ. A well-known fact is that $q_\tau(x)$ and $q_\tau(y)$ satisfy $q_\tau(x) = \operatorname{argmin}_a \int \rho_\tau(y - a) f(y|x, \delta = 1)dy$ and $q_\tau(y) = \operatorname{argmin}_a \int \rho_\tau(x - a) f(x|y, \delta = 1)dx$, where $\rho_\tau(u)$ is the "check function" defined by $\rho_\tau(u) = u(\tau - I[u < 0])$ (Koenker (2005)). We approximate $q_\tau(x)$ and $q_\tau(y)$ using a B-spline, allowing flexibility and computational efficiency. Let $\boldsymbol{B}(x)$ be a B-spline of degree $p_{y|x}$ and with $K_{n_1,y}$ interior knots, where n_1 is the sample size for $\delta = 1$. For any $\tau \in (0,1)$, we estimate $q_\tau(x)$ by $\hat{q}_\tau(x) = \boldsymbol{B}(x)'\hat{\boldsymbol{\beta}}_{y|x}(\tau)$, where

$$\hat{\boldsymbol{\beta}}_{y|x}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \delta_{1i}\rho_\tau(y_i - \boldsymbol{B}(x_i)'\boldsymbol{\beta}) + \frac{\lambda_{n_1,y}}{2}\boldsymbol{\beta}'\boldsymbol{D}_m'\boldsymbol{D}_m\boldsymbol{\beta} \right\}, \tag{2.11}$$

$\boldsymbol{D}_m$ is a difference matrix of order m, and $\lambda_{n_1,y} > 0$ is the smoothing parameter. See Chen and Yu (2016) and Berg and Yu (2019) for a precise definition of the B-spline and the difference matrix $\boldsymbol{D}_m$. In an analogous fashion, define the estimate of $q_\tau(y)$ by $\hat{q}_\tau(y) = \boldsymbol{B}(y)'\hat{\boldsymbol{\beta}}_{x|y}(\tau)$, where

$$\hat{\boldsymbol{\beta}}_{x|y}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \delta_{1i}\rho_\tau(x_i - \boldsymbol{B}(y_i)'\boldsymbol{\beta}) + \frac{\lambda_{n_1,x}}{2}\boldsymbol{\beta}'\boldsymbol{D}_m'\boldsymbol{D}_m\boldsymbol{\beta} \right\} \text{ for a given } \tau.$$

To approximate the full distributions of $f(y_i \mid x_i, \delta_i = 1)$ and $f(x_i \mid y_i, \delta_i = 1)$, we obtain estimates $\hat{\boldsymbol{\beta}}_{x|y}(\tau)$ and $\hat{\boldsymbol{\beta}}_{y|x}(\tau)$ for a grid of τ_j defined by $\tau_j = \tau_1 + (j-1)/J$, for $j = 2, \ldots, J$, where $\tau_1 \sim Unif(0, 1/J)$. The resulting estimated quantiles, defined as $\boldsymbol{y}_i^* = \{y_{ij}^* = \hat{q}_{\tau_j}(x_i) : j = 1, \ldots, J\}$, serve as imputed values for element i with $\delta_i = 2$. Likewise, $\boldsymbol{x}_i^* = \{x_{ij}^* = \hat{q}_{\tau_j}(y_i) : j = 1, \ldots, J\}$ serve as imputed values for element i with $\delta_i = 3$.

The sequence of estimated quantiles permits us to approximate the expectations defining $\tilde{\theta}$. For any arbitrary function $m(x,y)$, a variable transformation implies

$$\begin{aligned} E[m(x,Y)|x,\delta=1] &= \int_0^1 m(x, F^{-1}_{y|x,\delta=1}(\tau)) \frac{f_{y|x,\delta=1}(F^{-1}_{y|x}(\tau) \mid x)}{f_{y|x,\delta=1}(F^{-1}_{y|x}(\tau) \mid x)} d\tau \\ &= \int_0^1 m(x, q_\tau(x)) d\tau. \end{aligned}$$

We approximate $E[m(x,Y)|x,\delta=1]$ by $\hat{E}[m(x,Y)|x,\delta=1] = J^{-1}\sum_{j=1}^{J} m(x, \hat{q}_{\tau_j}(x))$.

We approximate the numerator and denominator of (2.9) by replacing $m(x,Y)$ with $g(x,Y)\exp(\phi_{22}Y)$ and $\exp(\phi_{22}Y)$, respectively. Specifically, $\hat{E}[g(x,Y)\exp(\phi_{22}Y)|x,\delta=1] = J^{-1}\sum_{j=1}^{J} g(x,\hat{q}_{\tau_j}(x))\exp(\phi_{22}\hat{q}_{\tau_j}(x))$, and $\hat{E}[\exp(\phi_{22}Y)|x,\delta = 1] = J^{-1}\sum_{j=1}^{J}\exp(\phi_{22}\hat{q}_{\tau_j}(x))$. Then, an approximation for (2.9) is

$$\hat{E}[g(x_i,Y)|x_i,\delta_i=2] = \sum_{j=1}^{J} w_{2ij}(\boldsymbol{\phi}_2, \boldsymbol{y}_i^*) g(x_i, y_{ij}^*), \tag{2.12}$$

where $\boldsymbol{\phi}_2 = (\phi_{20}, \phi_{21}, \phi_{22})'$ and

$$w_{2ij}(\boldsymbol{\phi}_2, \boldsymbol{y}_i^*) = \frac{\exp(\phi_{22} y_{ij}^*)}{\sum_{j=1}^{J} \exp(\phi_{22} y_{ij}^*)}. \tag{2.13}$$

Analogously, we estimate the expectation in (2.10) as

$$\hat{E}[g(X, y_i) | y_i, \delta_i = 3] = \sum_{j=1}^{J} w_{3ij}(\boldsymbol{\phi}_3, \boldsymbol{x}_i^*) g(x_{ij}^*, y_i), \tag{2.14}$$

where $\boldsymbol{\phi}_3 = (\phi_{30}, \phi_{31}, \phi_{32})'$ and

$$w_{3ij}(\boldsymbol{\phi}_3, \boldsymbol{x}_i^*) = \frac{\exp(\phi_{31} x_{ij}^*)}{\sum_{j=1}^{J} \exp(\phi_{31} x_{ij}^*)}. \tag{2.15}$$

2.2. Estimation of response probability

The estimated expectations in (2.12) and (2.14) require estimators of ϕ_{22} and ϕ_{31}, the two tilting parameters. We estimate $\boldsymbol{\phi} = (\boldsymbol{\phi}_2', \boldsymbol{\phi}_3')'$ using conditional probabilities. Define for $k = 2, 3$,

$$\begin{aligned} \pi_{k1}(x_i, y_i, \boldsymbol{\phi}_k) &:= P(\delta_i = k \mid x_i, y_i, \boldsymbol{\phi}_k, \delta_i \in \{1, k\}) \\ &= \frac{\exp(\phi_{k0} + \phi_{k1} x_i + \phi_{k2} y_i)}{1 + \exp(\phi_{k0} + \phi_{k1} x_i + \phi_{k2} y_i)}, \end{aligned}$$

and let $\pi_{1k\infty}(v) := P(\delta = 1 | v, \delta \in \{1, k\})$ for $v = x$ if $k = 2$, and $v = y$ if $k = 3$. Based on a result of Morikawa and Kim (2017), we can show that

$$\begin{aligned} \pi_{12\infty}(x) &= E[1 - \pi_{21}(x, Y, \boldsymbol{\phi}_2) \mid x, \delta \in \{1, 2\}] \\ &= \frac{\exp(-\phi_{20} - \phi_{21} x + h_y(-\phi_{22}, x))}{1 + \exp(-\phi_{20} - \phi_{21} x + h_y(-\phi_{22}, x))} \end{aligned} \tag{2.16}$$

and

$$\begin{aligned} \pi_{13\infty}(y) &= E[1 - \pi_{31}(X, y, \boldsymbol{\phi}_3) \mid y, \delta \in \{1, 3\}] \\ &= \frac{\exp(-\phi_{30} + h_x(-\phi_{31}, y) - \phi_{32} y)}{1 + \exp(-\phi_{30} + h_x(-\phi_{31}, y) - \phi_{32} y)}, \end{aligned} \tag{2.17}$$

where $h_y(\phi_{22}, x_i)$ and $h_x(\phi_{31}, y_i)$ are defined in (2.4) and (2.5), respectively. Note that $\pi_{12\infty}(x)$ depends only on x and $\pi_{13\infty}(y)$ depends only on y. Thus, equation

(2.16) suggests an estimator of $\boldsymbol{\phi}_2$ defined as

$$\hat{\boldsymbol{\phi}}_2 = \underset{\boldsymbol{\phi}_2}{\text{argmax}} \sum_{i:\delta_i=1} \log\left[\frac{\exp(-\phi_{20}-\phi_{21}x_i+\hat{h}_y(-\phi_{22},\hat{\boldsymbol{q}}_{yi}))}{1+\exp(-\phi_{20}-\phi_{21}x_i+\hat{h}_y(-\phi_{22},\hat{\boldsymbol{q}}_{yi}))}\right] + \sum_{i:\delta_i=2} \log\left[1-\frac{\exp(-\phi_{20}-\phi_{21}x_i+\hat{h}_y(-\phi_{22},\hat{\boldsymbol{q}}_{yi}))}{1+\exp(-\phi_{20}-\phi_{21}x+\hat{h}_y(-\phi_{22},\hat{\boldsymbol{q}}_{yi}))}\right], \tag{2.18}$$

where $\hat{h}_y(\phi_{22},\hat{\boldsymbol{q}}_{yi}) = -\log(J^{-1}\sum_{j=1}^{J}\exp\{-\phi_{22}y_{ij}^*\})$. Likewise, we estimate $\boldsymbol{\phi}_3$ as

$$\hat{\boldsymbol{\phi}}_3 = \underset{\boldsymbol{\phi}_3}{\text{argmax}} \sum_{i:\delta_i=1} \log\left[\frac{\exp(-\phi_{30}+\hat{h}_x(-\phi_{31},\hat{\boldsymbol{q}}_{xi})-\phi_{32}y_i)}{1+\exp(-\phi_{30}+\hat{h}_x(-\phi_{31},\hat{\boldsymbol{q}}_{xi})-\phi_{32}y_i)}\right] + \sum_{i:\delta_i=3} \log\left[1-\frac{\exp(-\phi_{30}+\hat{h}_x(-\phi_{31},\hat{\boldsymbol{q}}_{xi})-\phi_{32}y_i)}{1+\exp(-\phi_{30}+\hat{h}_x(-\phi_{31},\hat{\boldsymbol{q}}_{xi})-\phi_{32}y_i)}\right], \tag{2.19}$$

where $\hat{h}_x(\phi_{31},\hat{\boldsymbol{q}}_{xi}) = -\log(J^{-1}\sum_{j=1}^{J}\exp\{-\phi_{31}x_{ij}^*\})$. Note that $\hat{h}_x$ and $\hat{h}_y$ are estimates of h_x and h_y, respectively, using the imputed values y_{ij}^* and x_{ij}^*. In operation, we use the `R` function `optim` to find the maximum, where the initial value for $\boldsymbol{\phi}_2$ is from the logistic regression of $1-\delta_{2i}$ on $(1, x_i, \boldsymbol{B}(x_i)'J^{-1}\sum_{j=1}^{J}\hat{\boldsymbol{\beta}}_{y|x}(\tau_j))'$ for the set with $\delta_{3i}=0$. We define the initial value for $\boldsymbol{\phi}_3$ from the logistic regression of $1-\delta_{3i}$ on $(1, \boldsymbol{B}(y_i)'J^{-1}\sum_{j=1}^{J}\hat{\boldsymbol{\beta}}_{x|y}(\tau_j), y_i)'$ for the set with $\delta_{2i}=0$.

In summary, we define the basic steps of the estimation procedure as follows:

1. Use $\{(x_i,y_i):\delta_i=1\}$ to estimate the quantile regression model, and define imputed values y_{ij}^* and x_{ij}^*, as discussed in Section 2.1.
2. Estimate $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$, as discussed in Section 2.2.
3. Define the imputed estimator $\hat{\theta}$ by

$$\hat{\theta} = n^{-1}\sum_{i=1}^{n}\left\{\delta_{1i}g(x_i,y_i)+\delta_{2i}\sum_{j=1}^{J}w_{2ij}(\hat{\boldsymbol{\phi}}_2,\boldsymbol{y}_i^*)g(x_i,y_{ij}^*) + \delta_{3i}\sum_{j=1}^{J}w_{3ij}(\hat{\boldsymbol{\phi}}_3,\boldsymbol{x}_i^*)g(x_{ij}^*,y_i)\right\}. \tag{2.20}$$

This completes the description of our imputation and estimation procedures.

3. Large-Sample Theories and Variance Estimation

As a precursor to the statement of the large-sample distributions of $\hat{\boldsymbol{\phi}}_2$ and $\hat{\boldsymbol{\phi}}_3$, we give the large-sample distributions of the the estimates of the quantile regression coefficients as Lemma 1. We state Lemma 1 without proof because it is essentially an application of Yoshida (2013) to the set with $\delta_i = 1$. We use the linear approximation in the lemma in the subsequent derivation of the asymptotic properties of $(\hat{\boldsymbol{\phi}}_2', \hat{\boldsymbol{\phi}}_3')'$ and $\hat{\theta}$.

Lemma 1 uses the following property of Barrow and Smith (1978). The result is that the best L_∞ approximation to $q_\tau(x)$ (as a function of x), denoted $\boldsymbol{B}(x)'\boldsymbol{\beta}^*_{y|x}(\tau)$, satisfies $sup_{x\in[M_{1x},M_{2x}]} \mid q_\tau(x) + b^a_\tau(x) - \boldsymbol{B}(x)'\boldsymbol{\beta}^*_{y|x}(\tau) \mid = o(K_{n_1,y}^{-(p_{y|x}+1)})$, where $b^a_\tau(x)$ is the bias due to using a B-spline to approximate the true function $q_\tau(x)$, and is defined as in Yoshida (2013).

Lemma 1. *Assume $q_\tau^{(p_{y|x}+1)}(x)$ is continuous, where $q_\tau^{(p_{y|x}+1)}(x)$ denotes the* p+1 *derivative of $q_\tau(x)$ with respect to x, $K_{n_1,y} = O(n_1^{1/(2p_{y|x}+3)})$, and $\lambda_{n_1,y} = O(n_1^{\nu_y})$ for $\nu_y < (p_{y|x}+m+1)/(2p_{y|x}+3)$. Then,*

$$\sqrt{\frac{n_1}{K_{n_1,y}}}\left(\boldsymbol{B}(x)'\hat{\boldsymbol{\beta}}_{y|x}(\tau) - q_\tau(x) - b^a_\tau(x) - b^\lambda_\tau(x)\right) = W_{n_1} + o_p(1),$$

where

$$W_{n_1} = \sqrt{\frac{n_1}{K_{n_1,y}}}\boldsymbol{B}(x)'\boldsymbol{H}^{-1}_{n_1,y|x}(\tau)\frac{1}{n_1}\sum_{i:\delta_i=1}\boldsymbol{B}(x_i)\psi_\tau(e_{y|x,i}(\tau)),$$

$\psi_\tau(u) = \tau - I[u<0]$, $e_{y|x,i}(\tau) = y_i - q_\tau(x_i)$,

$$\boldsymbol{H}_{n_1,y|x}(\tau) = \boldsymbol{\Phi}_{y|x}(\tau) + n_1^{-1}\lambda_{n_1,y}\boldsymbol{D}_m'\boldsymbol{D}_m,$$

$$b^\lambda_\tau(x) = -\frac{\lambda_{n_1,y}}{n_1}\boldsymbol{B}(x)'\left(\boldsymbol{\Phi}_{y|x}(\tau) + \frac{\lambda_{n_1,y}}{n_1}\boldsymbol{D}_m'\boldsymbol{D}_m\right)^{-1}\boldsymbol{D}_m'\boldsymbol{D}_m\boldsymbol{\beta}^*_{y|x}(\tau), \tag{3.1}$$

and $\boldsymbol{\Phi}_{y|x}(\tau) = \lim_{n_1\to\infty} n_1^{-1}\sum_{i:\delta_i=1} f_{y|(x,\delta=1)}(x_i, q_\tau(x_i))\boldsymbol{B}(x_i)\boldsymbol{B}(x_i)'$.

Lemma 1 holds for a given τ, but the order of the approximation does not depend on τ. A result analogous to Lemma 1 holds for $\hat{\boldsymbol{\beta}}_{x|y}(\tau)$. We assume that the degree of $\boldsymbol{B}(y)$, denoted $p_{x|y}$, is such that $p_{x|y}\cdot p_{y|x}^{-1} = O(1)$. We also assume that the number of interior knots used to define $\boldsymbol{B}(y)$, denoted by $K_{n_1,x}$, satisfies $K_{n_1,y}\cdot K_{n_1,x}^{-1} = O(1)$.

3.1. Asymptotic variance of $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$

We state the large-sample distributions of $\hat{\boldsymbol{\phi}}_2$ and $\hat{\theta}$ as Theorems 1 and 2, respectively. Section S1 of the Supplementary Material contains a result for $\hat{\boldsymbol{\phi}}_3$ analogous to Theorem 1, as well as proofs.

Theorem 1. *In addition to the assumptions of Lemma* 1*, assume* $\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2 = o_p(1)$*,* $J = O(n^{0.5+\delta})$ *for some* $\delta > 0$*, and the conditions in the Supplementary Material hold. Then,* $\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2 = O_p(n^{-0.5})$*,* $\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2 = n^{-1}\sum_{i=1}^n \boldsymbol{I}_{\phi_2}^{-1}\boldsymbol{U}_{\phi_2 i} + o_p(n^{-0.5})$*, and* $\sqrt{n}\boldsymbol{V}_{\phi_2}^{-1/2}(\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2) \xrightarrow{d} N(0, \boldsymbol{I}_3)$*, where*

$$\boldsymbol{V}_{\phi_2} = \lim_{n\to\infty} n^{-1}\boldsymbol{I}_{\phi_2}^{-1}\left(\sum_{i=1}^n \boldsymbol{U}_{\phi_2 i}\boldsymbol{U}_{\phi_2 i}'\right)\boldsymbol{I}_{\phi_2}^{-1}, \tag{3.2}$$

$\boldsymbol{I}_{\phi_2} = \lim_{n\to\infty} \boldsymbol{I}_{n,\phi_2}(\boldsymbol{q}_y), \boldsymbol{I}_{n,\phi_2}(\boldsymbol{q}_y) = n^{-1}\sum_{i\in A_{12}} \pi_{12i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi})(1 - \pi_{12i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi}))$ $\boldsymbol{z}_{2i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi})\boldsymbol{z}_{2i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi})'$, $\boldsymbol{U}_{\phi_2 i} = (\delta_{1i} + \delta_{2i})\boldsymbol{S}_{i\infty}(\boldsymbol{\phi}_2) + \phi_{22}\delta_{1i}\int_{M_{1x}}^{M_{2x}} p_1^{-1}\pi_{12\infty}(x)(1 - \pi_{12\infty}(x))\boldsymbol{z}_{2\infty}(x)\boldsymbol{B}(x)'(\int_0^1 exp(\phi_{22}q_\tau(x))\boldsymbol{\ell}_i(\tau)d\tau/\int_0^1 exp(\phi_{22}q_\tau(x))d\tau)dF(x \mid \delta_1 + \delta_2 = 1)$, $p_1 = \lim_{n\to\infty} n^{-1}n_1$, $\boldsymbol{S}_{i\infty}(\boldsymbol{\phi}_2) = (\delta_{1i} - \pi_{12\infty}(x_i))\boldsymbol{z}_{2i\infty}, \boldsymbol{\ell}_i(\tau) = \boldsymbol{H}_{n_1,y|x}^{-1}(\tau)$ $\boldsymbol{B}(x_i)\psi_\tau(e_{y|x,i}(\tau)), \boldsymbol{z}_{2\infty}(x) = (-1, -x, -E_2(Y \mid x))', \boldsymbol{z}_{2i\infty} = \boldsymbol{z}_{2\infty}(x_i), \boldsymbol{z}_{2i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi})$ $= (-1, -x_i, -E_{2,J}(Y \mid x_i; \boldsymbol{\phi}_2, \boldsymbol{q}_{yi}))$, $\boldsymbol{q}_{yi} = \{q_{\tau_j}(x_i) : j = 1, \ldots, J\}$, $E_{2,J}(Y \mid x_i; \boldsymbol{\phi}_2, \boldsymbol{q}_{yi}) = \sum_{j=1}^J w_{2ij}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi})q_{\tau_j}(x_i)$,

$$\pi_{12i}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi}) = \left\{1 + \exp\left[\phi_{20} + \phi_{21}x_i + \log\left(J^{-1}\sum_{j=1}^J \exp\{\phi_{22}q_{\tau_j}(x_i)\}\right)\right]\right\}^{-1},$$

$A_{12} = \{i : \delta_{1i} + \delta_{2i} = 1\}$*,* $\boldsymbol{q}_y = \{\boldsymbol{q}_{yi} : \delta_{1i} + \delta_{2i} = 1\}$*, and* $E_2[Y \mid x] = E[Y \mid x, \delta = 2]$*.*

An estimator of the variance of $\hat{\boldsymbol{\phi}}_2$ is

$$\hat{V}\{\hat{\boldsymbol{\phi}}_2\} = n^{-2}\hat{\boldsymbol{I}}_{n,\phi_2}^{-1}\left(\sum_{i=1}^n \hat{\boldsymbol{U}}_{\phi_2 i}\hat{\boldsymbol{U}}_{\phi_2 i}'\right)\hat{\boldsymbol{I}}_{n,\phi_2}^{-1}, \tag{3.3}$$

where we substitute the unknown parameters with their corresponding estimators to define $\hat{\boldsymbol{I}}_{n,\phi_2}$ and $\hat{\boldsymbol{U}}_{\phi_2 i}$, as defined explicitly in Section S2 of the Supplementary Material.

Theorem 2. *Continue to assume the conditions of Theorem* 1*. In addition, assume* $g(X, Y)$ *has bounded* $2 + c$ *moments for* $c > 0$*, and has bounded second derivatives with respect to both* x *and* y*. Let* $K_{n_1} = \max\{K_{n_1,y}, K_{n_1,x}\}$*. Then,*

$\sqrt{n}V_g^{-0.5}(\hat{\theta} - E[g(X,Y)]) \xrightarrow{d} N(0,1)$, *where* $V_g = \lim_{n\to\infty}(n-1)^{-1}\sum_{i=1}^{n}(r_i - \bar{r})^2$, $\bar{r} = n^{-1}\sum_{i=1}^{n} r_i$,

$$\begin{aligned} r_i = {} & g(x_i, y_i) - Eg(X,Y) + \delta_{2i}(E_2[g(x_i, Y) \mid x_i] \\ & -g(x_i, y_i)) + \delta_{3i}(E_3[g(X, y_i) \mid y_i] - g(x_i, y_i)) \\ & +(\delta_{1i} + \delta_{2i}) \left\{\bar{C}_{2\infty}\right\} \boldsymbol{e}_2' \boldsymbol{I}_{\phi_2}^{-1} \boldsymbol{U}_{\phi_2,i} + (\delta_{1i} + \delta_{3i}) \left\{\bar{C}_{3\infty}\right\} \boldsymbol{e}_3' \boldsymbol{I}_{\phi_3}^{-1} \boldsymbol{U}_{\phi_3,i} \\ & +\delta_{1i}\Bigg(\int_0^1 \int_{M_{1x}}^{M_{2x}} \boldsymbol{C}_y(x,\tau)' \boldsymbol{\ell}_i(\tau) dF(x \mid \delta = 2) d\tau \\ & + \int_0^1 \int_{M_{1y}}^{M_{2y}} \boldsymbol{C}_x(y,\tau)' \boldsymbol{m}_i(\tau) dF(y \mid \delta = 3) d\tau \Bigg), \end{aligned} \tag{3.4}$$

$\bar{C}_{2\infty} = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \delta_{2k} Cov_2(g(x_k, Y), Y \mid x_k)$, $\bar{C}_{3\infty} = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \delta_{3k} Cov_3(g(X, y_k), X \mid y_k)$, $Cov_2(g(x,Y), Y \mid x) = Cov(g(x,Y), Y \mid X = x, \delta = 2), Cov_3(g(X,y), X \mid y) = Cov(g(X,y), X \mid Y = y, \delta = 3), E_2[g(x,Y) \mid x] = E[g(x,Y) \mid \delta = 2, X = x], E_3[g(X,y) \mid y] = E[g(X,y) \mid \delta = 3, Y = y], \boldsymbol{e}_2 = (0,1,0)', \boldsymbol{e}_3 = (0,0,1)'$, $\boldsymbol{C}_y(x,\tau) = \tilde{c}_y(x,\tau)\boldsymbol{B}(x)$, $\boldsymbol{C}_x(y,\tau) = \tilde{c}_x(y,\tau)\boldsymbol{B}(y)$,

$$\begin{aligned} \tilde{c}_y(x,\tau) &= \frac{c_y(x,\tau)}{\int_0^1 \exp(\phi_{22} q_\tau(x))} - E_2[g(x,Y) \mid x] \frac{\phi_{22}\exp(\phi_{22} q_\tau(x))}{\int_0^1 \exp(\phi_{22} q_\tau(x))} \\ \tilde{c}_x(y,\tau) &= \frac{c_x(y,\tau)}{\int_0^1 \exp(\phi_{31} q_\tau(y)) d\tau} - E_3[g(X,y) \mid y] \frac{\phi_{31}\exp(\phi_{31} q_\tau(y))\phi_{31}}{\int_0^1 \exp(\phi_{31} q_\tau(y)) d\tau}, \end{aligned}$$

$c_y(x,\tau) = \exp(\phi_{22} q_\tau(x)) g_y'(x, q_\tau(x)) + g(x, q_\tau(x))\exp(\phi_{22} q_\tau(x))\phi_{22}$, $c_x(y,\tau) = \exp(\phi_{31} q_\tau(y)) g_x'(q_\tau(y), y) + g(q_\tau(y), y)\exp(\phi_{31} q_\tau(y))\phi_{31}$, *and* $\boldsymbol{m}_i(\tau)$, $\boldsymbol{I}_{\phi_3}$, *and* $\boldsymbol{U}_{\boldsymbol{\phi}_3}$ *are defined in the Supplementary Material for the linear approximation for* $\hat{\boldsymbol{\phi}}_3$.

A proof of Theorem 2 is presented in Section S1 of the Supplementary Material. An estimator of the variance of the imputed estimator is

$$\hat{V}\{\hat{\theta}\} = (n(n-1))^{-1}\sum_{i=1}^{n}(\hat{r}_i - \bar{\hat{r}})^2, \tag{3.5}$$

where $\hat{r}_i$ is a plug-in estimator of r_i defined in Section S2.3 of the Supplementary Material, and $\bar{\hat{r}} = n^{-1}\sum_{i=1}^{n}\hat{r}_i$. In the Supplementary Material Section S2.4, we define how to use a Taylor linearization to estimate the variance of "composite" estimators of the form $\hat{\theta} = h(\hat{\theta}_1, \ldots, \hat{\theta}_K)$ of a parameter $\theta = h(\theta_1, \ldots, \theta_K)$, where each θ_k is of the form $Eg_k(X,Y)$, for some function $g_k(X,Y)$.

3.2. Propensity score-adjusted imputed estimator

The data set may contain a fourth group for which both x_i and y_i are missing. Let $\delta_{4i} = 1$ if both x_i and y_i are missing. In this context, we interpret the probabilities (2.1) as conditional probabilities, given that $\delta_{4i} = 0$. We apply the imputation procedure to $\{i : \delta_{4i} = 0\}$, as described in Section 2. We then apply a propensity score adjustment using a p-dimensional covariate $\boldsymbol{v}_i$, known for all $i = 1, \ldots, n$. Assume

$$P(\delta_{4i} = 0) = \exp(\phi_{40} + \boldsymbol{\phi}'_{41}\boldsymbol{v}_i)[1 + \exp(\phi_{40} + \boldsymbol{\phi}'_{41}\boldsymbol{v}_i)]^{-1} := p_{4i}(\boldsymbol{\phi}_4). \tag{3.6}$$

Estimate the $(p+1)$-dimensional parameter $\boldsymbol{\phi}'_4 = (\phi_{40}, \boldsymbol{\phi}'_{41})'$ as $\hat{\boldsymbol{\phi}}_4 = (\hat{\phi}_{40}, \hat{\boldsymbol{\phi}}'_{41})'$ satisfying $S_4(\hat{\boldsymbol{\phi}}_4) = \mathbf{0}$, where $S_4(\boldsymbol{\phi}_4) = \sum_{i=1}^n (1, \boldsymbol{v}_i)'(1 - \delta_{4i} - p_{4i}(\boldsymbol{\phi}_4))$. Then, let $\hat{p}_{4i} = p_{4i}(\hat{\boldsymbol{\phi}}_4)$. Assumption (3.6) justifies the propensity score-adjusted imputed estimator defined by

$$\hat{\theta}_{PSA-IMP} = \frac{1}{n}\Bigg\{ \sum_{i=1}^n \delta_{1i}\frac{g(y_i, x_i)}{\hat{p}_{4i}} + \delta_{2i}\frac{\sum_{j=1}^J w_{2ij}(\hat{\boldsymbol{\phi}}_2, \boldsymbol{y}_i^*)g(x_i, y_{ij}^*)}{\hat{p}_{4i}} + \delta_{3i}\frac{\sum_{j=1}^J w_{3ij}(\hat{\boldsymbol{\phi}}_3, \boldsymbol{x}_i^*)g(x_{ij}^*, y_i)}{\hat{p}_{4i}} \Bigg\}.$$

The propensity weights $\hat{p}_{4i}^{-1}$ extrapolate the set $\{i : \delta_{1i} + \delta_{2i} + \delta_{3i} = 1\}$ onto the full sample $\{i = 1, \ldots, n\}$. In the Supplementary Material Section S3, we define an estimator of the variance of $\hat{\theta}_{PSA-Imp}$ as a straightforward extension of (3.5), and we verify through simulation that $\hat{\theta}_{PSA-Imp}$ and the corresponding variance estimator are approximately unbiased.

4. Simulation Study

We assess the finite-sample properties of the proposed estimator. We first compare the estimator of Section 2 to competitive alternatives. We then assess the properties of the variance estimator proposed in Section 3.

4.1. Comparison of alternative imputation estimators

We consider two distributions for $F(y, x, \delta)$. For both, the parameter of interest is $\boldsymbol{\theta} = (EY, EX, V(Y), V(X), C(X, Y))'$, where $V(Y)$ (or $V(X)$) and $C(X, Y)$ denote the variance of Y (or X) and the correlation between X and Y, respectively. We compute the estimators for a Monte Carlo (MC) sample size of 500 and define $\boldsymbol{\theta}$ based on a separate simulation of size 500,000.

We compare the estimator proposed in Section 2 (abbreviated “Imp”) to three alternatives. To assess the impact of accounting for an MNAR nonresponse, we consider an ignorable (Ign) estimator that is essentially that of Chen and Yu (2016), and is obtained by setting $\boldsymbol{\phi} = \mathbf{0}$ so that $w_{2ij}(\boldsymbol{\phi}_2, \boldsymbol{q}_{yi}) = w_{3ij}(\boldsymbol{\phi}_3, \boldsymbol{q}_{xi}) = J^{-1}$. We define parametric (Par) and nonparametric (NP) alternatives that involve implementing the three steps of Section 2.3, including estimating $\boldsymbol{\phi}$, but generating the imputed values differently. For Par, we assume that $y_i = \beta_{0,y} + \beta_{1,y}x_i + \beta_{2,y}x_i^2 + \beta_{3,y}x_i^3 + e_{i,y}$, where $e_{i,y} \overset{i.i.d.}{\sim} N(0, \sigma^2_{e,y})$, and likewise, $x_i = \beta_{0,x} + \beta_{1,x}y_i + \beta_{2,x}y_i^2 + \beta_{3,x}y_i^3 + e_{i,x}$, where $e_{i,x} \overset{i.i.d.}{\sim} N(0, \sigma^2_{e,x})$. The imputed values for Par are $v^*_{ij} = \hat{v}_i + e^*_{vij}$, where for $\nu = x, y$, $e^*_{\nu ij} \overset{i.i.d.}{\sim} N(0, \hat{\sigma}^2_{e,\nu})$, $\hat{v}_i$ is the predicted mean using the ordinary least squares coefficients $(\hat{\beta}_{0,v}, \hat{\beta}_{1,v}, \hat{\beta}_{2,v}, \hat{\beta}_{3,v})$, and $\hat{\sigma}^2_{e,\nu} = (n-4)^{-1}\sum_{i=1}^n(\nu_i - \hat{\nu}_i)^2$. For NP, we generate imputed values independently and with replacement from the set of observed values such that $P\{y^*_{ij} = y_k\} = K(x_k - x_i)[\sum_{\ell=1}^n \delta_{1\ell}K(x_\ell - x_i)]^{-1}$, and $P\{x^*_{ij} = x_k\} = K(y_k - y_i)[\sum_{\ell=1}^n \delta_{1\ell}K(y_\ell - y_i)]^{-1}$, where $K(\cdot)$ is a Gaussian kernel with bandwidth defined by applying the `R` function `bw.ucv` to the sets $\{x_i : \delta_{1i} = 1\}$ and $\{y_i : \delta_{1i} = 1\}$ individually. Owing to the adjustment for the MNAR nonresponse, by estimating $\boldsymbol{\phi}$, the Par and NP estimators proposed above are themselves innovations upon Kim and Yu (2011) and Wang and Chen (2009), respectively.

We define the $FlippedExp$ simulation model by

$$y_i = h(x_i) + 1.25(1 + x_i)(\epsilon_i - 0.2) \tag{4.1}$$

and $\epsilon_i \overset{i.i.d.}{\sim} Beta(1,4)$, where $h(x_i) = \{2\exp(-2) - \exp(-2(x_i - 1))\}I[x_i < 2] + \{2\exp(2) - \exp(-2(x_i - 5))\}I[2 < x_i < 4] + \exp(-2(x_i - 3))I[4 < x_i < 6]$, $x_i \overset{i.i.d.}{\sim} Unif(0, 6)$, for $i = 1, \ldots, n$, and $(\phi_{20}, \phi_{21}, \phi_{22}, \phi_{30}, \phi_{31}, \phi_{32}) = (-1, 0.033, 0.12, -0.800, 0.1, 0.033)$. We consider $n = 100, 1000$, and 5000. The penalties $(\lambda_{n_1,y}, \lambda_{n_1,x})$ are (0.2, 2), (1, 10), and (3, 30) for $n = 100$, 1000, and 5000, respectively. They are based on a rule of $(\lambda_{n_1,y}, \lambda_{n_1,x}) \approx (0.1, 0.01)n^{6/9}$, determined from an exploratory analysis of simulated data using generalized cross-validation (Chen and Yu (2016)) and the relation between $\lambda_{n_1,y}$ and n in Lemma 1. We define $J \approx n^{0.5}$, giving $J = 10$, 30, and 70 for $n = 100$, 1000, and 5000, respectively. The knots are the $k/(K+1)$ quantiles of $\{x_i : \delta_{1i} + \delta_{2i} = 1, i = 1, \ldots, n\}$ and $\{y_i : \delta_{1i} + \delta_{3i} = 1, i = 1, \ldots, n\}$, where $k = 1, \ldots, K$, and $K = 20, 30$, and 35 for n =100, 1000, and 5000, respectively. The values of K are based loosely on the rule of thumb, $K = min\{n/4, 35\}$ (Ruppert, Wand and Carroll (1987)).

Table 2. MC bias and RMSE of alternative estimators of $\boldsymbol{\theta}$ for $FlippedExp$.

		Bias				RMSE			
	True	Ign	Imp	Par	NP	Ign	Imp	Par	NP
					$n = 100$				
EY	4.412	-0.034	**0.022**	0.041	-0.125	**0.637**	0.644	0.751	0.751
EX	3.000	-0.838	**-0.001**	-0.010	-0.014	0.862	**0.180**	0.186	0.196
$V(Y)$	42.865	**-0.182**	0.619	3.469	-1.610	**5.545**	6.235	14.460	5.575
$V(X)$	3.000	0.954	-0.018	**0.011**	-0.040	1.010	0.315	**0.308**	0.312
$C(X,Y)$	0.939	-0.341	**-0.000**	-0.015	-0.039	0.352	**0.014**	0.030	0.059
					$n = 1000$				
EY	4.412	-0.039	**0.004**	-0.044	0.011	0.209	**0.208**	0.229	0.211
EX	3.000	-0.839	**-0.000**	-0.003	-0.002	0.841	**0.053**	0.056	0.053
$V(Y)$	42.865	-0.564	**0.003**	-0.114	-0.038	1.472	**1.430**	1.809	1.441
$V(X)$	3.000	1.001	**-0.002**	0.015	-0.005	1.006	**0.089**	0.095	0.090
$C(X,Y)$	0.939	-0.341	**0.000**	-0.003	-0.005	0.342	**0.004**	0.005	0.007
					$n = 5000$				
EY	4.412	-0.035	0.007	-0.037	**0.007**	0.101	0.097	0.110	**0.096**
EX	3.000	-0.838	0.001	0.001	**0.001**	0.838	**0.025**	0.025	0.025
$V(Y)$	42.865	-0.574	0.011	-0.190	**-0.010**	0.833	0.654	0.833	**0.637**
$V(X)$	3.000	1.002	-0.004	0.012	**-0.004**	1.003	**0.039**	0.042	0.039
$C(X,Y)$	0.939	-0.341	**0.000**	-0.002	-0.001	0.341	**0.002**	0.003	0.002

Tables 2 and 3 contain the MC biases and RMSEs of the estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, respectively, with the smallest absolute value among competitors shown in bold. For $n = 100$, variation from estimating additional parameters causes the RMSE of Imp to exceed those of Par and Ign, except for EX and $Cor(X,Y)$. For $n = 1000$, the Imp procedure is efficient. As n increases to 5,000, the efficiency of NP improves. The Imp estimator of $\boldsymbol{\phi}$ typically has the smallest absolute bias and RMSE.

To construct a model that better satisfies the assumptions of the Par estimator, we define the Exp configuration by (4.1) with $h(x_i) = \exp(2x_i)$, where $x_i \overset{i.i.d.}{\sim} Unif(-1,1)$, and $(\phi_{20}, \phi_{21}, \phi_{22}, \phi_{30}, \phi_{31}, \phi_{32}) = (-0.9, 0.15, 0.2, -0.8, 0.15, 0.1)$. A rule of $\lambda_{n_1,y} = \lambda_{n_1,x} \approx n^{6/9}$ gives penalties of 20 and 100 for $n = 100$ and 1000, respectively. We define the knots and τ_j in the same way as for FlippedExp.

The results for Exp in Table 4 favor Par because the assumed cubic approximates the Exp function well. An exception is for $Var(X)$, where Imp has a smaller RMSE than Par for $n = 100$ and $n = 1000$. Imp and Par are superior to NP in Table 4, owing to the small sample size. The results for $\hat{\boldsymbol{\phi}}$ and $n = 5000$ (omitted for brevity) lead to similar conclusions.

Table 3. MC bias and RMSE of alternative estimators of $\boldsymbol{\phi}$ for *FlippedExp*.

			Bias			RMSE		
	n	True	Imp	Par	NP	Imp	Par	NP
ϕ_{20}	100	-1.0000	**0.1008**	-0.2918	-0.3151	**1.0756**	1.3114	1.0886
ϕ_{21}	100	0.0333	**-0.0853**	0.1236	0.1046	**0.5482**	0.6408	0.5548
ϕ_{22}	100	0.1200	0.0277	-0.0323	**-0.0176**	**0.1475**	0.1712	0.1663
ϕ_{30}	100	-0.8000	-0.0923	**-0.0709**	-0.0852	**1.3860**	1.5112	1.7724
ϕ_{31}	100	0.1000	**-0.0023**	-0.0231	-0.0477	**0.7272**	0.7588	0.8893
ϕ_{32}	100	0.0333	**0.0061**	0.0103	0.0223	**0.1928**	0.2037	0.2286
ϕ_{20}	1,000	-1.0000	**0.0021**	-0.1877	-0.0042	0.3077	0.4324	**0.3001**
ϕ_{21}	1,000	0.0333	**-0.0049**	0.0995	-0.0059	0.1531	0.2242	**0.1510**
ϕ_{22}	1,000	0.1200	**0.0025**	-0.0248	0.0042	**0.0400**	0.0598	0.0402
ϕ_{30}	1,000	-0.8000	**0.0075**	0.0664	0.0549	**0.3613**	0.4875	0.3721
ϕ_{31}	1,000	0.1000	**-0.0045**	-0.0379	-0.0296	**0.1824**	0.2535	0.1892
ϕ_{32}	1,000	0.0333	**0.0014**	0.0097	0.0078	**0.0479**	0.0669	0.0498
ϕ_{20}	5,000	-1.0000	**0.0014**	-0.1442	-0.0016	0.1411	0.2273	**0.1402**
ϕ_{21}	5,000	0.0333	-0.0012	0.0798	**0.0000**	0.0691	0.1198	**0.0688**
ϕ_{22}	5,000	0.1200	0.0006	-0.0199	**0.0006**	0.0180	0.0311	**0.0179**
ϕ_{30}	5,000	-0.8000	**0.0057**	0.0191	0.0221	**0.1630**	0.2210	0.1654
ϕ_{31}	5,000	0.1000	**-0.0027**	-0.0100	-0.0113	**0.0829**	0.1139	0.0843
ϕ_{32}	5,000	0.0333	**0.0008**	0.0025	0.0030	**0.0214**	0.0295	0.0218

Table 4. Comparison of imputation procedures for *Exp*.

		Bias				RMSE			
	True	Ign	Imp	Par	NP	Ign	Imp	Par	NP
					$n = 100$				
EY	1.813	-0.012	**0.000**	0.007	-0.025	0.203	0.202	**0.199**	0.223
EX	0.000	-0.008	**0.002**	0.009	-0.015	**0.053**	0.059	0.063	0.069
$V(Y)$	3.613	-0.211	-0.161	**-0.069**	-0.197	0.685	0.675	**0.648**	0.846
$V(X)$	0.333	-0.084	**0.001**	0.024	-0.005	0.089	**0.032**	0.095	0.033
$C(X,Y)$	0.888	-0.129	**0.004**	0.005	-0.075	0.139	**0.015**	0.018	0.107
					$n = 1000$				
EY	1.813	-0.015	-0.007	-0.008	**-0.006**	0.063	0.062	**0.061**	0.062
EX	0.000	-0.009	-0.001	**-0.001**	-0.003	**0.018**	0.019	0.019	0.019
$V(Y)$	3.613	-0.099	-0.069	-0.070	**-0.046**	0.216	0.208	**0.200**	0.206
$V(X)$	0.333	-0.084	**-0.001**	0.002	-0.002	0.084	**0.009**	0.010	0.010
$C(X,Y)$	0.888	-0.124	0.001	**0.001**	-0.010	0.125	0.006	**0.005**	0.014

4.2. Variance estimator for imputed estimator

Table 5 contains the MC variances ($V_{MC}(\hat{\theta})$) of the Imp estimators, the percent relative biases (RB%) of the variance estimator $(100(E_{MC}[\hat{V}] - V_{MC}(\hat{\theta}))/$

Table 5. Properties of variance estimator for Imp for *FlippedExp*.

	$n = 100$			$n = 1000$			$n = 5000$		
	$V_{MC}(\hat{\theta})$ $\times 10^3$	RB%	CR%	$V_{MC}(\hat{\theta})$ $\times 10^3$	RB%	CR%	$V_{MC}(\hat{\theta})$ $\times 10^3$	RB%	CR%
EY	517.943	-9.061	93.4	43.164	2.387	94.6	8.854	0.771	95.6
EX	37.112	-10.713	93.8	3.102	-0.216	94.0	0.633	-2.467	94.8
$V(Y)$	36,090.740	-11.852	93.0	1,962.765	-1.616	94.4	435.427	-10.830	93.6
$V(X)$	84.684	33.543	96.6	7.271	10.397	95.2	1.623	-3.466	94.2
$C(X,Y)$	0.185	29.068	96.0	0.014	0.177	92.8	0.003	12.921	95.6
ϕ_{20}	1,121.574	5.837	97.4	100.770	-3.255	94.2	19.616	-2.589	95.2
ϕ_{21}	298.448	1.864	97.0	25.131	-5.082	93.8	4.974	-6.440	94.6
ϕ_{22}	21.847	-0.004	96.0	1.759	-5.536	93.6	0.343	-6.006	93.4
ϕ_{30}	1,707.902	-5.553	97.4	131.168	-3.995	95.0	25.236	-2.336	94.2
ϕ_{31}	467.293	-8.326	96.2	33.185	-3.753	95.0	6.163	1.473	95.0
ϕ_{32}	32.245	-5.849	97.0	2.351	-5.450	95.0	0.411	5.636	95.4

Table 6. Number of records in each group for pet data.

Group	Count	Group	Count
1: X^* and Y^* observed	3,338	3: Only Y^* observed	262
2: Only X^* observed	2,461	4: X^* and Y^* missing	1,169

$V_{MC}(\hat{\theta})$, where $E_{MC}[\hat{V}]$ denotes the MC mean of the variance estimator (3.5)), and the percent of normal theory confidence intervals that contain the true parameter values (CR%). For $n = 100$, the absolute RB% can exceed 30% and the CR% can exceed 97%. For $n \in \{100, 5000\}$, the absolute RB% is below 15% and the CR% is within 2.2% of 95%.

5. Data Analysis

We analyze data from the 2011 Pet Demographic Survey (PDS), a national survey that collects information about pet ownership. The Iowa State Center for Survey Statistics and Methodology (CSSM) received the data as an agreement to plan for the 2017 survey. Variables of interest on the PDS include the number and type of pets owned, body types of those pets, and expenditures on veterinary services. We consider X^*, the sum of the most recent vet visit expenditures for a dog and cat combined, as a covariate for Y^*, the average vet visit expenditures in 2011 for dogs and cats. Table 6 has the number of observations for X^* and Y^*, with four missing data patterns. We apply the propensity-score-adjusted imputed estimator to estimate the veterinary expenditures for dogs and cats.

The nature of the relationship between X^* and Y^*, as well as extreme values, preclude us from finding a quantile regression model that fits the sample data well in the original scale. Furthermore, the 75 zeros for X^* and 64 zeros for Y^* make a log transformation problematic. After exploring several transformations, including the square root, cube root, and fifth root, we find that the cube root transformation allows us to construct a quantile regression model that appears adequate.

We apply the quantile regression procedure to first construct imputed values for $X = (X^*)^{1/3}$ and $Y = (Y^*)^{1/3}$ for groups 2 and 3. The generalized cross-validation criterion of Chen and Yu (2016) suggests $\lambda_{n_1,y} = 100$. The rule used for the *Exp* configuration of $\lambda_{n_1,y} \approx n_{123}^{6/9}$, where n_{123} is the number of observations in groups 1, 2, and 3, suggests $\lambda_{n_1,y} \approx 330$. First, we tried the approximate mid-point of $\lambda_{n_1,y} \approx 200$, and obtained negative estimated quantiles for y_i for τ_1 and small values of x_i. Increasing the penalty to $\lambda_{n_1,y} = 300$ successfully avoided negatives. We present the results for $\lambda_{n_1,y} = 300$. We use a fixed sequence of $\tau_j = j/(J+1)$, for $j = 1, \ldots, J$, with $J = 80 \approx n_{123}^{0.5}$. The fixed sequence avoids extreme quantiles and ensures that the data analysis is reproducible. (Chen and Yu (2016) compare results for fixed and random τ_j.) We define knots at the $k/(K+1)$ quantiles of $\{x_i : \delta_{1i} + \delta_{2i} = 1 : i = 1, \ldots, n\}$ and $\{y_i : \delta_{1i} + \delta_{3i} = 1 : i = 1, \ldots, n\}$, for $k = 1, \ldots, K$, where $K = 35$.

We assess the model identification conditions (2.4) and (2.5) using the estimated functions $\hat{h}_y(\hat{\phi}_{22}, x)$ and $\hat{h}_x(\hat{\phi}_{31}, y)$ plotted in Figure 1. To construct the left plot in Figure 1, we first define an estimate of $h_y(\hat{\phi}_{22}, x_i)$ in equation (2.4) as the negative logarithm of the LOWESS regression of $\exp(-\hat{\phi}_{22} y_i)$ on x_i for the $\{i : \delta_i = 1\}$, where $\hat{\phi}_{22}$ is the estimated exponential tilting parameter in (2.7) obtained using the method described in Section 2.3. The right plot is constructed analogously, interchanging the roles of x_i and y_i, and replacing $\hat{\phi}_{22}$ with $\hat{\phi}_{31}$. The nonlinearities seen in Figure 1 support the model identification conditions (2.4) and (2.5).

Table 7 gives estimates and corresponding standard errors for the propensity score model. The covariates, given in the column headings, are selected using step-wise selection, starting with a model that contains all fully observed covariates and using the BIC criterion. The gender variable is one for females and zero for males. The other covariates (defined in Section S3.3 of the Supplementary Material) are defined by ordered categories, and are treated as numeric. The response variable is the indicator that unit i is not in group 4. Therefore, a positive coefficient is associated with a higher probability of providing a response. As such, we estimate that women with higher income and education who live alone or with one other person are more likely to provide a response to at least one of the questions about veterinary expenses.

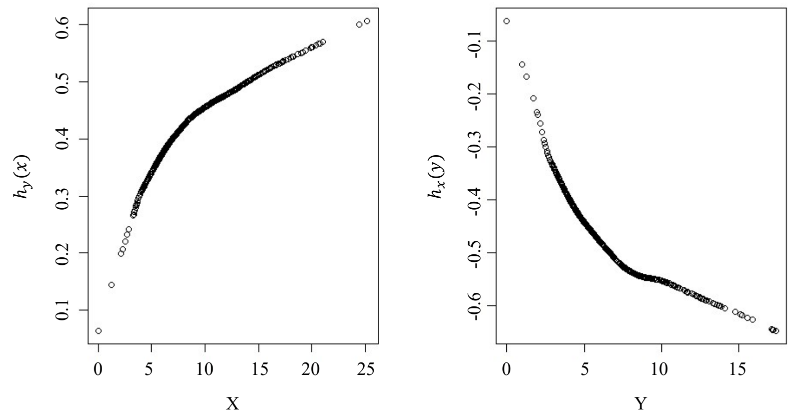

Figure 1. Estimated $\hat{h}_y(\hat{\phi}_{22}, x)$ (left) and $\hat{h}_x(\hat{\phi}_{31}, y)$ (right).

Table 7. Estimated $\hat{\boldsymbol{\phi}}_4$ and SE for propensity score model.

	Intercept	Age	Gender	Income	Education	Household Size
Est.	0.16252	0.10355	0.38652	0.38395	0.21250	-0.31212
SE	0.20897	0.02687	0.09208	0.03056	0.03671	0.05129

Table 8. Estimates and standard errors for $\boldsymbol{\phi} = (\boldsymbol{\phi}_2', \boldsymbol{\phi}_3')'$ for the pet data.

	ϕ_{2j} Est.	ϕ_{2j} SE	ϕ_{3j} Est.	ϕ_{3j} SE
$j = 0$	0.5136	0.2782	-1.0677	0.4484
$j = 1$	-0.0561	0.0271	-0.2590	0.1037
$j = 2$	-0.0810	0.0903	0.0609	0.0587

Table 9. Complete case and Imp-PSA estimators of selected parameters, along with standard errors for the Imp-PSA estimator.

	EY	EX	$Var(Y)$	$Var(X)$	$Cor(X,Y)$	EY^3	EX^3
Complete Case	5.210	7.359	3.729	7.073	0.420	208.336	575.560
SE Complete Case	0.032	0.035	0.161	0.225	0.033	5.664	11.409
Imp-PSA	5.052	7.274	3.269	6.979	0.442	185.164	566.653
SE Imp-PSA	0.077	0.035	0.184	0.218	0.016	7.297	10.912

Table 8 contains estimates of $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$ (obtained using (2.18) and (2.19)), along with associated standard errors (defined in (3.3)). The estimator of ϕ_{21} differs significantly from zero at the 5% level, but after accounting for x_i, y_i is no longer significantly associated with the response indicator, δ_{2i}. Interestingly, the estimate of ϕ_{31} is more than double the standard error. The component of the model that accounts for a nonignorable nonresponse is important for δ_{3i}.

Table 9 compares the propensity-score-adjusted imputed estimator (Imp-

PSA) to the complete case estimator, which naively ignores missing values. The parameters EY^3 and EX^3 represent the mean expenditures in the original scale, and are defined by $g(x, y) = y^3$ and $g(x, y) = x^3$, respectively. We also estimate the means and the correlation in the cube root scale. The comparison of the complete case and the imputed estimators suggests that ignoring the missing data would overstate the expenditures and understate the correlation between X and Y. As a result of the nonignorable nonresponse, the complete-case standard errors are also invalid. Imputation requires estimating additional parameters, and can therefore lead to an increase in the SE relative to the complete-case SE. The sample size for the complete-case estimator of the correlation is smaller than that used to estimate the other parameters, because the complete-case estimator of the correlation only uses pairs where both x_i and y_i are observed simultaneously.

6. Discussion

The theory, simulations, and data analysis demonstrate that the proposed semiparametric quantile regression imputation procedure is a viable method of constructing imputed values when the probability of responding may depend on the value of a missing response or covariate. We prove that the imputed estimator is asymptotically normal, and verify through simulation that an estimate of the large-sample covariance matrix has reasonable finite-sample properties. The simulations also show that failing to account for a nonignorable nonresponse can lead to severe bias. The squared bias of the ignorable predictor can account for over 90% of the MSE. In contrast, the ratio of the squared bias to the MSE for the proposed (Imp) estimator is consistently below 1%. In our simulations, the quantile regression is more robust than the fully parametric imputation, and more efficient than the nonparametric imputation at small sample sizes. We do not have theoretical support for the superiority of semiparametric quantile regression relative to the nonparametric regression, and therefore do not expect these results to hold broadly. A further advantage of the quantile regression over the nonparametric estimator of Wang and Chen (2009) is that the quantile regression permits a linearization-based variance estimator. In the application, the proposed procedure allows us to use one type of veterinary expenditure to impute the other, while allowing for a nonignorable nonresponse and modeling complex patterns in the data. Furthermore, we develop a propensity score adjustment to incorporate a set for which neither veterinary expenditure is observed.

We have used a fully parametric model for the response probability. As demonstrated in Robins and Ritov (1997), identification for a nonignorable nonresponse is elusive without any restrictions. Nonetheless, relaxing the parametric assumptions of the response probability model, along the lines of Shao and Wang (2016), is a possible avenue for future work.

In principle, our approach of modeling the conditional distribution of the covariate given a response extends to multivariate covariates. One must ensure that the quantile regression model adequately describes each full univariate conditional, and that identification conditions are satisfied. We define an identification condition for multivariate covariates in Section S4 of the Supplementary Material. An alternative approach for missing covariates is to use Bayes' rule to deduce $f(x \mid y)$ from a specification of $f(y \mid x)$ and $f(x)$ (Yang and Kim (2017)) . Our preliminary studies suggest that an extension of Yang and Kim (2017) to a nonignorable nonresponse and a quantile regression is a promising direction for future work.

7. Supplementary Material

The online Supplementary Material provides proofs of theorems, details of variance estimation, and simulation results for the propensity-score-adjusted imputed estimator.

8. Acknowledgments

We acknowledge the grant NSF MMS-1733572.

References

Barrow, D. L. and Smith, P. W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by spline with variable knots. *Quarterly of Applied Mathematics* **36**, 293–304.

Berg, E. and Yu, C. (2019). Semiparametric quantile regression imputation for a complex survey with application to the Conservation Effects Assessment Project. *Survey Methodology* **45**, 249–270.

Chen, S. and Yu, C. L. (2016). Parameter estimation through semiparametric quantile regression imputation. *Electronic Journal of Statistics* **10**, 3621–3647.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association* **89**, 81–87.

Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95**, 555–571.

Fang, F., Zhao, J. and Shao, J. (2018). Imputation-based adjusted score equations in generalized linear models with nonignorable missing covariate values. *Statistica Sinica* **28**, 1677–1701.

Kim, J. K. and Shao, J. (2014). *Statistical Methods for Handling Incomplete Data.* Chapman & Hall, Boca Raton.

Kim, J. K. and Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association* **106**, 157–165.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, New York.

Miao, W. and Tchetgen Tchetgen, E. J. (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika* **103**, 475–482.

Morikawa, K. and Kim, J. K. (2017). Semiparametric adaptive estimation with nonignorable nonresponse data. *arXiv preprint arXiv:1612.09207.*

Riddles, M. K., Kim, J. K. and Im, J. (2016). A Propensity-score-adjustment method for nonignorable nonresponse. *Journal of Survey Statistics and Methodology.* **4**, 215–245.

Robins, J. M. and Ritov, Y. A. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**, 285–319.

Rubin, D. B. (1987). *Multiple Imputation for Nonresposne in Surveys.* Wiley, New York.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression (No. 12).* Cambridge University Press, Cambridge.

Shao, J. and Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika* **103**, 175–187.

Tang, G., Little, R. J. and Raghunathan, T. E. (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika* **90**, 747–764.

Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics* **37**, 490–517.

Wang, S., Shao, J. and Kim, J. K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* **24**, 1097–1116.

Yang, S. and Kim, J. K. (2017). A semiparametric inference to regression analysis with missing covariates in survey data. *Statistica Sinica* **27**, 261–285.

Yoshida, T. (2013). Asymptotics for penalized spline estimators in quantile regression. *Communications in Statistics - Theory and Methods.* DOI: `10.1080/03610926.2013.765477`.

Zhao, J. and Ma, Y. (2019). A Versatile Estimation Procedure without Estimating the Nonignorable Missingness Mechanism. *Journal of the American Statistical Association.* DOI: `10.1080/01621459.2021.1893176`.

Zhao, J. and Shao, J. (2015). Semiparametric pseudo-likelihoods in generalized linear models with nonignorable missing data. *Journal of the American Statistical Association* **110**, 1577–1590.

Emily Berg

Department of Statistics, Iowa State University, Ames, IA 50011, USA.

E-mail: emilyb@iastate.edu

Cindy Yu

Department of Statistics, Iowa State University, Ames, IA 50011, USA.

E-mail: cindyyu@iastate.edu

(Received February 2020; accepted January 2021)