Received: 13 January 2020 | Revised: 12 May 2020 | Accepted: 24 June 2020

DOI: 10.1002/pst.2052

MAIN PAPER

WILEY

Augmented inverse probability weighted fractional imputation in quantile regression

Hao Cheng[1,2,3,4]

[1]National Academy of Innovation Strategy, China Association for Science and Technology, Beijing, China

[2]School of Statistics, Renmin University of China, Beijing, China

[3]The Needham Research Institute, University of Cambridge, Cambridge, UK

[4]Mailman School of Public Health, Columbia University, New York, USA

Correspondence
Hao Cheng, National Academy of Innovation Strategy, China Association for Science and Technology, Fuxing Road 3, Haidian District, Beijing, 100863, China.
Email: chenghao0524@yeah.net

Funding information
The Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China, Grant/Award Number: 16XNH102

Summary

By employing all the observed information and the optimal augmentation term, we propose an augmented inverse probability weighted fractional imputation method (AFI) to handle covariates missing at random in quantile regression. Compared with the existing completely case analysis, inverse probability weighting, multiple imputation and fractional imputation based on quantile regression model with missing covarites, we carry out simulation study to investigate its performance in estimation accuracy and efficiency, computational efficiency and estimation robustness. We also talk about the influence of imputation replicates in our AFI. Finally, we apply our methodology to part of the National Health and Nutrition Examination Survey data.

KEYWORDS

augmented inverse probability weighting, fractional imputation, quantile regression

1 | INTRODUCTION

As a promising and flexible modeling tool, quantile regression (Koenker and Bassett 1978; Koenker 2005) has the ability to capture the complex associations between a response variable and its covariates at different quantiles.[1,2] Because of this most attractive feature, quantile regression estimates are more robust to outliers than traditional least-squares regression. In this paper, we consider the following linear quantile regression model,

$$Q_Y(\tau) = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta}_{1,\tau} + \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{\beta}_{2,\tau}, \forall \tau \in (0,1) \tag{1}$$

where $Q_Y(\tau)$ stands for the τth quantile of a response variable Y, and $(\boldsymbol{x}, \boldsymbol{z})$ are both covariate vectors. We assume that the conditional quantile function of Y given $(\boldsymbol{x}, \boldsymbol{z})$ is a linear function of $(\boldsymbol{x}, \boldsymbol{z})$ with quantile specific coefficients $(\boldsymbol{\beta}_{1,\tau}, \boldsymbol{\beta}_{2,\tau})$, $\boldsymbol{z}$ contains the constant 1 and hence the intercept term is not written out separately.

Quantile regression has gained so much attention in both theoretical investigations and applications especially when the distribution of the response is heavily tailed. One of the common challenges in quantile regression is the presence of missing data. That's mainly because there is no likelihood function for quantile regression and hence most existing likelihood-based methods cannot be applied directly. In this article, we investigate missing covariates problems in

 © 2020 John Wiley & Sons Ltd

quantile regression and assume that the data are missing at random (MAR; Little and Rubin 2002).[3] That is, we assume that the covariate $\boldsymbol{x}$ may be missing at random, but $\boldsymbol{z}$ is always observed in model (1).

Existing imputation methods or algorithms dealing with missing covarites in quantile regression can be roughly classified into the following categories:

1. Complete-data-based procedures, including weighting adjustments. The popularity of complete case (CC) analysis is due to its simplicity to discard incompletely observed samples and analysis the complete samples only. It may be satisfactory with small amounts of missing data. However, it can lead to biased and inefficient estimators especially when drawing inferences for subpopulations (Little and Rubin 2002) and the rate of missingness is high (Han et al 2019).[4] Another popular and simple method is inverse probability weighting (IPW) (Horvitz and Thompson 1952).[5] Lipsitz et al (1997) reweighted the estimating equations by IPW on quantile longitudinal studies with attrition and independently identically distributed (i.i.d.) error terms, which extends the well-known conditional mean modeling to quantile modeling.[6] Sherwood et al (2013) considered an IPW quantile regression approach when the covariates are missing at random.[7] Sun et al (2012) considered quantile regression for competing risk data when the failure type was missing.[8] Chen et al (2015) proposed a kind of nonparametric IPW approach based on Wang et al (1997), which developed originally for conditional mean modeling.[9,10] Compared with IPW, augmented inverse probability weighting (AIPW) has the ability to get more efficient and doubly robust estimators such as Chen et al (2015)'s nonparametric AIPW method.
2. Imputation-based and Model-based procedures. These procedures create a predictive distribution for the imputation based on the observed data or define a model for the observed data and base inferences on distribution under that model with estimated parameters. Mean imputations and regression imputations are very popular. However, mean imputations do not add extra observed information. Yoon (2010) imputed the missing values from the conditional quantile function, but the method is valid only under i.i.d. errors.[11] Wei et al (2012) developed a multiple imputation (MI) procedure for missing covariates based on a linear quantile regression model that brings relatively expensive computational burden.[12] Cheng et al (2018) proposed a new fast imputation (FI) algorithm in quantile regression and its corresponding IPW-modified algorithm to deal with two kinds of missing covariates problems (the missingness is related with the response or not).[13] However, its performance in robustness need to be discussed further when the regression function is misspecified. Wei et al (2014) proposed an iterative EM-type algorithm to solve their unbiased estimating equations that simultaneously hold at all the quantile levels.[14] But it is computationally undesirable. Chen et al (2015)'s nonparametric estimating equations projection method was developed by Zhou et al (2008) under the conventional setup of conditional mean modeling with i.i.d. errors. This method was motivated by the fact that direct replacements of the unobserved data by their imputed values generally yield estimating equations that are biased.[9,15]

In this paper, we consider missing covariates problems in quantile regression and propose a new parametric augmented inverse probability weighted fractional imputation method (AFI). This method is based on parametric modeling of the propensity score and conditional expectation of the estimating functions. The rest of the paper is organized as follows. We described our imputation methods in Section 2, and conducted a simulation study in Section 3. Finally, we applied the proposed methods to part of the National Health and Nutrition Examination Survey study in Section 4. Some final discussions are placed in Section 5.

2 | ESTIMATION WITH AFI IMPUTATION

2.1 | Notation

Suppose (x_i, z_i, y_i), $i = 1, \dots, n$ is an i.i.d. random sample following the linear quantile model (1). The sample size equals n. We denote δ_i as the binary indicator for the existence of x_i. That is, x_i is observed when $\delta_i = 1$ for $i = 1, \dots, n_1$ and is missing when $\delta_i = 0$ for $i = n_1 + 1, \dots, n$. Let $n_0 = n - n_1$ be the number of incomplete cases, we further assume that $0 < \lim_{n \to \infty}(n_0/n_1) = \lambda < \infty$, so that the proportion of the missing observations are non-negligible and non-dominating. Thus, the missing mechanism in our paper can be represented as $P(\delta_i = 1 \mid y_i, x_i, z_i) = P(\delta_i = 1 \mid z_i)$.

2.2 | Method

Robins et al (1994) defined a class of AIPW estimators by solving the following augmented estimating equation.[16]

$$S_n^*(\beta)=\sum_{i=1}^{n}\frac{\delta_i}{p_i}\varphi(y_i-x_i^T\beta_1-z_i^T\beta_2)+\sum_{i=1}^{n}\left(1-\frac{\delta_i}{p_i}\right)E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)|y_i,z_i\} \tag{2}$$

We can easily get that the expectation of $S_n^*(\beta)$ approximates zero when $\beta=\beta_{true}$ and $Q(y\,|\,x,\,z)=(x,\,z)^T\beta_{true}$. In fact, $S_n^*(\beta)$ can be written as $A_1+A_2+A_3$, where we denote A_1 as $\sum_{i=1}^{n_1}\frac{1}{p_i}\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)$, A_2 as $\sum_{i=n_1+1}^{n}E_x\{\varphi_\tau(y_i-x_i^T\beta_1\ -z_i^T\beta_2)|y_i,z_i\}$ and A_3 as $\sum_{i=1}^{n_1}\left(1-\frac{1}{p_i}\right)E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)|y_i,z_i\}$.

A_1 is just the same as inverse probability weighting method (IPW). That is, IPW weights each completely observed data by $1/\text{prob}(\delta_i=1|\ y_i,\,z_i)$.[17,18] We use monte-carlo integrations to approximate the conditional expectation $E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)\,|\,y_i,z_i\}$ in A_2. Based on the bayesian theory, $f(x\,|\,z,\,y)$ can be written as $\frac{f(y_i|x_i,z_i)f(x_i|z_i)}{\int_x f(y_i|x_i,z_i)f(x_i|z_i)dx}$. Consequently, one can rewrite the conditional expectation $E_{\mathrm{x}}\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)\,|\,y_i,z_i\}$ in A_2 by $\frac{\int_x\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)f(y_i|x_i,z_i)f(x_i|z_i)dx}{\int_x f(y_i|x_i,z_i)f(x_i|z_i)dx}$. Both of the numerator and denominator can be approximated by Monte-carlo integrations,[13,19]

$$\int_x\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)f(y_i|x_i,z_i)f(x_i|z_i)dx\approx\frac{1}{M}\sum_{k=1}^{M}\varphi_\tau(y_i-\tilde{x}_{i,k}^{*T}\beta_1-z_i^T\beta_2)f(y_i\,|\,\tilde{x}_{i,k}^*,z_i)$$

and

$$\int_x f(y_i|x_i,z_i)f(x_i|z_i)dx\approx\frac{1}{M}\sum_{k=1}^{M}f(y_i\,|\,\tilde{x}_{i,k}^{*T},z_i)$$

where $\tilde{x}_{i,k}^{*T}$ is randomly drawn from $f(x_i|\,z_i)$. Therefore,

$$\begin{aligned}A_2&=\sum_{i=n_1+1}^{n}E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)\,|\,y_i,z_i\}\\&=\sum_{i=n_1+1}^{n}\frac{\int_x\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)f(y_i|x_i,z_i)f(x_i|z_i)dx}{\int_x f(y_i|x_i,z_i)f(x_i|z_i)dx}\\&=\sum_{i=n_1+1}^{n}\frac{\frac{1}{M1}\sum_{k=1}^{M1}\varphi_\tau(y_i-\tilde{x}_{i,k}^{*T}\beta_1-z_i^T\beta_2)f(y_i\,|\,\tilde{x}_{i,k}^*,z_i)}{\frac{1}{M1}\sum_{k=1}^{M1}f(y_i\,|\,\tilde{x}_{i,k}^{*T},z_i)}\\&=\sum_{i=n_1+1}^{n}\sum_{k=1}^{M1}\frac{f(y_i\,|\,\tilde{x}_{i,k}^*,z_i)}{\sum_{k=1}^{M1}f(y_i\,|\,\tilde{x}_{i,k}^*,z_i)}\varphi_\tau(y_i-\tilde{x}_{i,k}^{*T}\beta_1-z_i^T\beta_2)\end{aligned} \tag{3}$$

where $\tilde{x}_{i,k}$ is randomly drawn from $f(x_i|\,z_i)$. More specifically, we first run a linear regression (x as the response and z as the covariate) only using completely observed data. Then we use the estimated coefficients as mean $\hat{\mu}$ and calculate the variance of residual errors as variance $\hat{\eta}$. Finally we get the values of missing $\tilde{x}_{i,k}$ from normal distribution $N(\hat{\mu},\hat{\eta})$. Thus, A_2 equals $\sum_{i=n_1+1}^{n}\sum_{k=1}^{M1}w_{i,k}\varphi_\tau\{y_i-\tilde{x}_{i,k}^{*T}\beta_1-z_i^T\beta_2\}$, where $w_{i,k}=\frac{f(y_i|\tilde{x}_{i,k}^*,z_i)}{\sum_{k=1}^{M1}f(y_i|\tilde{x}_{i,k}^*,z_i)}$.

A_3 is different from A_2 in two points. Firstly, there exists an negative weight $1-\frac{1}{p_i}$ rather than one. Secondly, the covariates x here is observed. To calculate the conditional expectation $E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)\,|\,y_i,z_i\}$ in A_3, we also use monte-carlo integrations.

$$\begin{aligned}
E_x\{\varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2)\,|\,y_i,z_i\} &= \int_x \varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2) f(x_i|y_i,z_i)dx \\
&= \int_x \varphi_\tau(y_i-x_i^T\beta_1-z_i^T\beta_2) f(x_i|y_i,z_i)dx \\
&= \frac{\frac{1}{M2}\sum_{l=1}^{M2}\varphi_\tau(y_i-\tilde{x}_{i,l}^{*T}\beta_1-z_i^T\beta_2) f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}{\frac{1}{M2}\sum_{l=1}^{M2} f(y_i\,|\,\tilde{x}_{i,l}^{*T},z_i)} \\
&= \sum_{l=1}^{M2}\frac{f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}{\sum_{l=1}^{M2} f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}\varphi_\tau(y_i-\tilde{x}_{i,l}^{*T}\beta_1-z_i^T\beta_2)
\end{aligned} \tag{4}$$

Where $\tilde{x}_{i,l}^{*T}$ is randomly drawn from $f(x_i|\,y_i,z_i)$. Because all the x, y and z in A_3 are completely observed, the density $f(x_i|\,y_i,z_i)$ can be easily estimated by maximizing a parametric likelihood over the observed (x, y, z) under the conditional independence assumption. We model x given y and z parametrically as $f(x|y,z,\eta_2)$. The missing-at-random assumption facilitates the estimation of η_2 based on the complete data. We write the estimate as $\hat{\eta}_2$, and the estimated conditional density of x given y and z as $f(x|y,z,\hat{\eta}_2)$. Thus, we can get A_3.

$$\begin{aligned}
A_3 &= \sum_{i=1}^{n1}\left(1-\frac{1}{p_i}\right)\frac{1}{M2}\sum_{l=1}^{M2}\frac{f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}{\sum_{l=1}^{M2} f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}\varphi_\tau(y_i-\tilde{x}_{i,l}^{*T}\beta_1-z_i^T\beta_2) \\
&= \sum_{i=1}^{n1}\sum_{l=1}^{M2}\left(1-\frac{1}{p_i}\right)\frac{1}{M2}\frac{f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}{\sum_{l=1}^{M2} f(y_i\,|\,\tilde{x}_{i,l}^{*},z_i)}\varphi_\tau(y_i-\tilde{x}_{i,l}^{*T}\beta_1-z_i^T\beta_2) \\
&= \sum_{i=1}^{n1}\sum_{l=1}^{M2} w_{i,l}\varphi_\tau(y_i-\tilde{x}_{i,l}^{*T}\beta_1-z_i^T\beta_2)
\end{aligned} \tag{5}$$

where $w_{i,l}=\left(1-\frac{1}{p_i}\right)\frac{1}{M2}\frac{f(y_i|\tilde{x}_{i,l}^{*},z_i)}{\sum_{l=1}^{M2} f(y_i|\tilde{x}_{i,l}^{*},z_i)}$.

In both $w_{i,k}$ and $w_{i,l}$, we need to estimate the density function $f(y\,|\,x,\ z)$ by using the first derivative of the conditional quantile function $Q_y(\tau\,|\,\boldsymbol{x},\boldsymbol{z})$ at τ_y, where τ_y is the quantile level of y, that is, $\text{pr}(Y\le y\,|\,\boldsymbol{x},\boldsymbol{z})=\tau_y$ (Wei et al, 2012). Following this direction, we model the entire conditional quantile process $Q_y(\tau\,|\,\boldsymbol{x},\boldsymbol{z})=(\mathbf{x}^{\mathbf{T}},\mathbf{z}^{\mathbf{T}})\,\boldsymbol{\beta}(\boldsymbol{\tau})$ on a fine grid of $0<\tau_1<\cdots<\tau_k<\cdots<\tau_{K_n}<1$ and approximate the density $f(y\,|\,\boldsymbol{x},\boldsymbol{z})$ by

$$\hat{f}\left\{y\,|\,\mathbf{x},\mathbf{z},\hat{\boldsymbol{\beta}}_{n1}(\tau)\right\}=\sum_{k=1}^{K_n}\frac{(\tau_{k+1}-\tau_k)I\left\{(\mathbf{x}^{\mathbf{T}},\mathbf{z}^{\mathbf{T}})\hat{\boldsymbol{\beta}}_{n_1,\tau_k}\le y<(\mathbf{x}^{\mathbf{T}},\mathbf{z}^{\mathbf{T}})\hat{\boldsymbol{\beta}}_{n_1,\tau_{k+1}}\right\}}{(\mathbf{x}^{\mathbf{T}},\mathbf{z}^{\mathbf{T}})\hat{\boldsymbol{\beta}}_{n_1,\tau_{k+1}}-(\mathbf{x}^{\mathbf{T}},\mathbf{z}^{\mathbf{T}})\hat{\boldsymbol{\beta}}_{n_1,\tau_k}}. \tag{6}$$

In addition, there exist $M1$ and $M2$ in A_2 and A_3 respectively. The selection of $(M1, M2)$ may effect the properties of estimator and computing time. Therefore, we carry out a set of simulations to investigate $(M1, M2)$ selection. Algorithm 1 shows the procedure of AFI algorithm.

Algorithm 1

AFI Algorithm

Step 1: ***Quantile regression with complete data on a fine grid of quantile levels*** $0<\tau_1<\cdots<\tau_k<\cdots<\tau_{k_n}<1$.
Step 2: ***Calculate the inverse probability weights based on complete data***.
Step 3: ***Model the conditional density*** $f(x|z)$ ***and*** $f(x \mid y, z)$ ***parametrically as*** $f(x|z, \eta_1)$ ***and*** $f(x \mid y, z, \eta_2)$***, and estimate*** η_1 ***and*** η_2 ***based on the complete data***.
Step 4: ***Simulate*** $M1$ x ***from the estimated*** $f(x|z,\hat{\eta}_1)$ ***for each missing*** x_i, $1\le i\le n_1$.
Step 5: ***Simulate*** $M2$ x ***from the estimated*** $f(x \mid y,z,\hat{\eta}_2)$ ***for each observed*** x_i, $1\le i\le n_1$.
Step 6: ***Calculate the weights using the model induced density from Step 1, and assemble the weighted estimating function as in*** $S_n^*(\beta)$ ***to get the final estimator***.

3 | SIMULATION STUDY

3.1 | Models and settings

We use Monte-Carlo simulations to investigate the performance of our estimators. We consider the following location-scale model to complete the sampling process,

$$y_i=\beta_0+\beta_1x_i+\beta_2z_i+(\xi_1x_i+\xi_2z_i)e_i \tag{7}$$

where the coefficients $(\beta_0, \beta_1, \beta_2) = (1,1,1)$ and the covariates (x_i, z_i) are jointly normal with mean vector $(4, 4)^T$, variances $(1, 1)^T$ and correlation 0.5. We consider **Case 1** ($\xi_1 = \xi_2 = 1/(x_i + z_i)$) and **Case 2** ($\xi_1 = \xi_2 = 0.5$). In **Case 1**, model (7) can be written as $y_i = \beta_0 + \beta_1x_i + \beta_2z_i + e_i$ and the error is homoscedastic. In this case, the true intercept equals 1 + $Q_\tau(e_i)$ at quantile level τ, but the two slope coefficients equal 1 at every quantile level. In **Case 2**, model (7) can be written as $y_i = \beta_0 + \beta_1x_i + \beta_2z_i + (0.5x_i + 0.5z_i)e_i$ and the error is heteroscedastic. In this case, the true intercept equals 1 at every quantile level, but the two slope coefficients equal to $1 + 0.5Q_\tau(e_i)$ at quantile level τ. Furthermore, we consider two distributions for the random errors e_i, either standard normal $N(0, 1)$ or chi-square $\chi^2(1)$. We define the missing mechanism as $p(\delta_i| z_i) = \max[0, \{(z_i - 3)/10\}^{1/20}]$, such that approximately 25% observations miss x_i's, and the missingness is independent with Y. In what follow, we denote **Setting S1-1** as **Case 1** with normal e_i, and **Setting S1-2** is that with chi-square e_i. Likewise, we denote **Setting S2-1** and **Setting S2-2** as the **Case 2** with normal and chi-square random errors.

To investigate the potential bias that could be induced from misspecified $f(x| z)$ (**Setting S3**) or propensity score functions (**Setting S4**), we also examine the following settings based on model (7) with heteroscedastic errors (Case 2). In **S3**, we simulate covariates (x_i, z_i) from $x_i = (0.18 * u_{i,1} + 0.68 * u_{i,2}) + 3.14$, and $z_i = (0.68 * u_{i,1} + 0.18 * u_{i,2}) + 3.14$, where both $u_{i,1}$ and $u_{i,2}$ are independent $\chi^2(1)$ random variables (Wei et al, 2012). The constants (0.18, 0.68 and 3.14) make sure that (x_i, z_i) has the same mean, variance and correlation as the assumptions in model (7). Once simulating the non-normal covariates, we generate the responses from model (7) in **Case 2** when $\xi_1 = \xi_2 = 0.5$. For each generated sample, we allow x_i to be missing completely at random with probability 0.25. We apply the same missing mechanism and estimation procedures as above, pretending that (x_i, z_i) is jointly normal. In **S4**, we consider one kind of misspecified propensity score function based on model (7) in **Case 2** when $\xi_1 = \xi_2 = 0.5$: $p(\delta_i| z_i) = exp(2.5 * z^2)/(1 + exp(2.5 * z^2))$ for all (z, y). Based on the simulation data under the four settings with sample size 500, we can see that missing data is evenly distributed in the plot S1-1 and S2-1, while the missing data tends to locate in the bottom-left corner in S1-2 and S2-2. For brevity, we do not show the figures in our paper. Based on all the above settings, we want to investigate the following problems by conducting numerical investigations. Here we choose the sample size as 500 and the Monte-Carlo sample size as 200.

1. We compared the estimation accuracy and efficiency of AFI and its competitors CC, IPW, MI, FI estimates under all different settings and cases. We choose $M1 = M2 = 10$ in AFI, m = 10 in MI and $M = 20$ in FI.

TABLE 1 Mean biases (M.B.), standard errors (S.E.) and mean squared errors (M.S.E.) of the estimated coefficients using quantile regression at quantile levels 0.1 and 0.5 under fully observed data with sample size 500 and 200 Monte-Carlo replicates

		Case 1 with normal e_i		Case 1 with chi-square e_i		Case 2 with normal e_i		Case 2 with chi-square e_i	
τ		0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
M.B.	$\hat{\beta}_1$	0.01	0.01	0.00	0.00	0.05	0.03	0.00	−0.02
	$\hat{\beta}_2$	−0.01	−0.01	0.00	0.01	−0.03	−0.02	0.00	0.03
S.E.	$\hat{\beta}_1$	0.09	0.07	0.01	0.06	0.33	0.26	0.02	0.20
	$\hat{\beta}_2$	0.09	0.07	0.01	0.05	0.34	0.25	0.02	0.20
M.S.E.	$\hat{\beta}_1$	0.01	0.01	0.00	0.00	0.11	0.07	0.00	0.04
	$\hat{\beta}_2$	0.01	0.01	0.00	0.00	0.12	0.06	0.00	0.04

Note: Case 1, model (7) can be written as $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$; Case 2, model (7) can be written as $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$.

2. To understand the impact of the number of imputations ($M1$ and $M2$) in AFI under different settings and cases, we considered various number of imputation replicates ($M1 = M2 = 10, 20, 50$) in AFI algorithms.
3. In Setting 3 and 4 under **Case 2**, to investigate the robustness property of misspecification of either the propensity score $p(\delta| z)$ or the regression function $f(x| z)$, we compared the estimation accuracy and efficiency of the estimates from AFI and its competitors CC, IPW, MI and FI.
4. To assess the level of uncertainty brought in by estimated weights, we also compared AFI with its counterpart using true weight calculated from true density $f(y \mid x, z)$, which we denote as AFIP.
5. Under all the settings, we compared the computing times of all the imputation methods with various number of imputation replicates ($M1 = M2 = 10, 20, 50$). In addition, we choose m = 10 in MI and M = 20 in FI.

Therefore, we need to carry out the following numerical investigations by comparing AFI with CC, IPW, MI and FI, in which we choose m = 10 in MI, M = 20 in FI, sample size 500 and Monte-Carlo sample size 200. We denote S1-1-1 as standard normal e_i and $M1 = M2 = 10$ in AFI under Case 1, S1-1-2 as standard normal e_i and $M1 = M2 = 20$ in AFI under Case 1, S1-1-3 as standard normal e_i and $M1 = M2 = 50$ in AFI under Case 1, S1-2-1 as chi-square e_i and $M1 = M2 = 10$ in AFI under Case 1, S1-2-2 as chi-square e_i and $M1 = M2 = 20$ in AFI under Case 1, S1-2-3 as chi-square e_i and $M1 = M2 = 50$ in AFI under Case 1. We denote S2-1-1, S2-1-2, S2-1-3, S2-2-1, S2-2-2 and S2-2-3 as the counterparts of above settings under Case 2. In S3, we compare misspecified regression function with correct function with standard normal or chi-square e_i, $M1 = M2 = 10$ in AFI under Case 2. In S4, we compare misspecified propensity score function and correct function with standard normal or chi-square e_i, $M1 = M2 = 10$ in AFI under Case 2.

Our inference comparisons are in terms of magnitude of mean biases in estimators (M.B.), standard errors of the estimators (S.E.) and mean squared errors (M.S.E.) across 200 replications based on the bootstrap method. The benchmark results are based on the fully observed data with no missing values under the above missing data settings. We display the simulation results at quantile levels $\tau = 0.1$ and 0.5 in our paper for brevity.

3.2 | Results

3.2.1 | Estimation under fully observed data

Table 1 displays the benchmark results where data are fully observed with no missing values. Under all the models and settings, the results show that no biases of any substantial magnitude are exhibited, and both variances and mean squared errors perform well. As expected, the benchmark results using fully observed data are the most accurate, and the accuracy of estimators across all performance dimensions getting worse when covariates are missing.

3.2.2 | Comparisons of estimation accuracy and efficiency

Tables 2 to 4 display the mean biases, standard errors and mean squared errors of the estimated coefficients under S1 and S2, using CC, IPW, MI, FI, AFI and AFIP. It should be noted that, although there do not exist differences in models,

TABLE 2 Mean biases (M.B.) and standard errors (S.E.) of the estimated coefficients using CC, IPW, MI, FI, AFI, and AFIP at quantile levels 0.1 and 0.5 under the setting S1 with sample size 500 and 200 Monte-Carlo replicates

		S1-1-1		S1-1-2		S1-1-3		S1-2-1		S1-2-2		S1-2-3	
τ		0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
M.B.	$\hat{\beta}_{1,CC}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	−0.01	0.00	−0.01	0.00	−0.01
	$\hat{\beta}_{1,IPW}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{1,MI}$	−0.01	0.01	−0.01	0.01	−0.01	0.01	0.00	−0.02	0.00	−0.02	0.00	−0.02
	$\hat{\beta}_{1,FI}$	−0.02	0.00	−0.02	0.00	−0.02	0.00	0.00	−0.03	0.00	−0.03	0.00	−0.03
	$\hat{\beta}_{1,AFI}$	−0.02	−0.01	−0.02	0.00	−0.01	0.00	0.00	−0.03	0.00	−0.02	0.00	−0.02
	$\hat{\beta}_{1,AFIP}$	−0.02	−0.01	−0.01	0.00	−0.01	0.00	0.00	−0.04	0.00	−0.02	0.00	−0.02
	$\hat{\beta}_{2,CC}$	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	0.00	0.02	0.00	0.02	0.00	0.02
	$\hat{\beta}_{2,IPW}$	−0.01	0.00	−0.01	0.00	−0.01	0.00	0.00	0.02	0.00	0.02	0.00	0.02
	$\hat{\beta}_{2,MI}$	−0.02	−0.01	−0.02	−0.01	−0.02	−0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{2,FI}$	−0.01	0.00	−0.01	0.00	−0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{2,AFI}$	0.00	0.00	−0.01	0.00	−0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	$\hat{\beta}_{2,AFIP}$	0.00	0.01	−0.01	0.00	−0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01
S.E.	$\hat{\beta}_{1,CC}$	0.10	0.08	0.10	0.08	0.10	0.08	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{1,IPW}$	0.10	0.08	0.10	0.08	0.10	0.08	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{1,MI}$	0.10	0.07	0.10	0.07	0.10	0.07	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{1,FI}$	0.10	0.08	0.10	0.08	0.10	0.08	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{1,AFI}$	0.10	0.07	0.10	0.08	0.10	0.07	0.01	0.07	0.01	0.06	0.01	0.07
	$\hat{\beta}_{1,AFIP}$	0.10	0.07	0.10	0.07	0.10	0.07	0.01	0.06	0.01	0.06	0.01	0.07
	$\hat{\beta}_{2,CC}$	0.13	0.09	0.13	0.09	0.13	0.09	0.01	0.08	0.01	0.08	0.01	0.08
	$\hat{\beta}_{2,IPW}$	0.13	0.09	0.13	0.09	0.13	0.09	0.01	0.08	0.01	0.08	0.01	0.08
	$\hat{\beta}_{2,MI}$	0.12	0.08	0.12	0.08	0.12	0.08	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{2,FI}$	0.12	0.08	0.12	0.08	0.12	0.08	0.01	0.07	0.01	0.07	0.01	0.07
	$\hat{\beta}_{2,AFI}$	0.11	0.08	0.12	0.08	0.12	0.08	0.01	0.06	0.01	0.07	0.01	0.07
	$\hat{\beta}_{2,AFIP}$	0.12	0.08	0.12	0.08	0.12	0.08	0.01	0.06	0.01	0.06	0.01	0.07

Note: All the estimates are under Case 1: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$. S1-1-1, $M1 = M2 = 10$ in AFI and AFIP with normal e_i; S1-1-2, $M1 = M2 = 20$ in AFI and AFIP with normal e_i; S1-1-3, $M1 = M2 = 50$ in AFI and AFIP with normal e_i; S1-2-1, $M1 = M2 = 10$ in AFI and AFIP with chi-square e_i; S1-2-2, $M1 = M2 = 20$ in AFI and AFIP with chi-square e_i; S1-2-3, $M1 = M2 = 50$ in AFI and AFIP with chi-square e_i; $\hat{\beta}_{1,\mathrm{METHOD}}$, the estimated X coefficients using METHOD; $\hat{\beta}_{2,METHOD}$, the estimated Z coefficients using METHOD; here METHOD stands for CC, IPW, MI, FI, AFI or AFIP.

settings and assumptions among CC, IPW, MI and FI no matter which value A equals, we want to compare the performances between different AFI and other methods under all settings. Therefore, we run CC, IPW, MI and FI under each S1-1-A, S1-2-A, S2-1-A and S2-2-A (where $A = 1$, 2 and 3, respectively) and display the results in Tables 2 to 4.

The upper half of Table 2 displays the mean biases of the estimated coefficients under S1, and the bottom half shows standard errors. All six methods are nearly unbiased and efficient. The structure of Table 3, which is under S2, is the same as Table 2. Different from S1, S2 represents model (7) with heteroscedastic errors (Case 2). Under S2-1-1, S2-1-2 and S2-1-3, MI, FI, AFI and AFIP yield relatively smaller biases than CC and IPW at quantile levels 0.1 and 0.5. Under S2-2-1, S2-2-2 and S2-2-3, the estimates from the six methods are nearly comparable in their biases at quantile level 0.1, while both CC and IPW slightly outperform the other methods for x's estimated coefficients and perform worse for z's estimated coefficients at quantile level 0.5. As expected from the theory, the variances of IPW estimators are larger than other estimators except S2-2-1, S2-2-2 and S2-2-3 at quantile level 0.1, in which settings all the six methods are fairly comparable. In addition, the variances of CC estimators are similar to IPW. In Table 4, the upper half displays the mean squared errors under S1, while the bottom half shows those under S2. Under S1, the mean squared errors of all six methods nearly equal 0.00. Under S2, MI, FI, AFI and AFIP are nearly comparable under all settings and perform better than both CC and IPW except S2-2-1, S2-2-2 and S2-2-3 at quantile level 0.1 (all equal 0.00).

TABLE 3 Mean Biases (M.B.) and Standard Errors (S.E.) of the estimated coefficients using CC, IPW, MI, FI, AFI, and AFIP at quantile levels 0.1 and 0.5 under the setting S2 with sample size 500 and 200 Monte-Carlo replicates

		S2-1-1		S2-1-2		S2-1-3		S2-2-1		S2-2-2		S2-2-3	
τ		0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
M.B.	$\hat{\beta}_{1,\text{CC}}$	0.05	0.04	0.05	0.04	0.05	0.04	0.00	−0.02	0.00	−0.02	0.00	−0.02
	$\hat{\beta}_{1,\text{IPW}}$	0.04	0.04	0.04	0.04	0.04	0.04	0.00	−0.03	0.00	−0.03	0.00	−0.03
	$\hat{\beta}_{1,\text{MI}}$	0.03	0.03	0.03	0.03	0.03	0.03	−0.02	−0.07	−0.02	−0.07	−0.02	−0.07
	$\hat{\beta}_{1,\text{FI}}$	0.02	0.02	0.02	0.02	0.02	0.02	−0.01	−0.07	−0.01	−0.07	−0.01	−0.07
	$\hat{\beta}_{1,\text{AFI}}$	0.01	0.01	0.02	0.01	0.01	0.02	−0.01	−0.08	−0.01	−0.07	−0.01	−0.06
	$\hat{\beta}_{1,\text{AFIP}}$	0.01	0.01	0.02	0.02	0.02	0.03	−0.01	−0.08	−0.01	−0.07	−0.01	−0.06
	$\hat{\beta}_{2,\text{CC}}$	−0.04	−0.03	−0.04	−0.03	−0.04	−0.03	0.01	0.08	0.01	0.08	0.01	0.08
	$\hat{\beta}_{2,\text{IPW}}$	−0.06	−0.02	−0.06	−0.02	−0.06	−0.02	0.00	0.09	0.00	0.09	0.00	0.09
	$\hat{\beta}_{2,\text{MI}}$	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	0.02	0.07	0.02	0.07	0.02	0.07
	$\hat{\beta}_{2,\text{FI}}$	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	0.02	0.06	0.02	0.06	0.02	0.06
	$\hat{\beta}_{2,\text{AFI}}$	−0.02	0.00	−0.02	0.00	−0.02	−0.01	0.02	0.06	0.01	0.06	0.01	0.06
	$\hat{\beta}_{2,\text{AFIP}}$	−0.01	0.00	−0.02	−0.01	−0.02	−0.01	0.02	0.06	0.01	0.06	0.01	0.06
S.E.	$\hat{\beta}_{1,\text{CC}}$	0.42	0.32	0.42	0.32	0.42	0.32	0.02	0.27	0.02	0.27	0.02	0.27
	$\hat{\beta}_{1,\text{IPW}}$	0.40	0.31	0.40	0.31	0.40	0.31	0.02	0.27	0.02	0.27	0.02	0.27
	$\hat{\beta}_{1,\text{MI}}$	0.37	0.29	0.37	0.29	0.37	0.29	0.04	0.25	0.04	0.25	0.04	0.25
	$\hat{\beta}_{1,\text{FI}}$	0.38	0.30	0.38	0.30	0.38	0.30	0.03	0.24	0.03	0.24	0.03	0.24
	$\hat{\beta}_{1,\text{AFI}}$	0.37	0.29	0.38	0.29	0.38	0.29	0.02	0.24	0.03	0.24	0.03	0.25
	$\hat{\beta}_{1,\text{AFIP}}$	0.38	0.29	0.38	0.30	0.38	0.30	0.02	0.24	0.02	0.25	0.03	0.25
	$\hat{\beta}_{2,\text{CC}}$	0.57	0.39	0.57	0.39	0.57	0.39	0.03	0.35	0.03	0.35	0.03	0.35
	$\hat{\beta}_{2,\text{IPW}}$	0.56	0.39	0.56	0.39	0.56	0.39	0.03	0.35	0.03	0.35	0.03	0.35
	$\hat{\beta}_{2,\text{MI}}$	0.35	0.27	0.35	0.27	0.35	0.27	0.06	0.25	0.06	0.25	0.06	0.25
	$\hat{\beta}_{2,\text{FI}}$	0.35	0.27	0.35	0.27	0.35	0.27	0.04	0.24	0.04	0.24	0.04	0.24
	$\hat{\beta}_{2,\text{AFI}}$	0.36	0.27	0.36	0.27	0.36	0.27	0.04	0.23	0.04	0.24	0.03	0.24
	$\hat{\beta}_{2,\text{AFIP}}$	0.36	0.27	0.36	0.27	0.36	0.27	0.04	0.23	0.04	0.24	0.04	0.24

Note: All the estimates are under Case 2: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$. S2-1-1, $M1 = M2 = 10$ in AFI and AFIP with normal e_i; S2-1-2, $M1 = M2 = 20$ in AFI and AFIP with normal e_i; S2-1-3, $M1 = M2 = 50$ in AFI and AFIP with normal e_i; S2-2-1, $M1 = M2 = 10$ in AFI and AFIP with chi-square e_i; S2-2-2, $M1 = M2 = 20$ in AFI and AFIP with chi-square e_i; S2-2-3, $M1 = M2 = 50$ in AFI and AFIP with chi-square e_i; $\hat{\beta}_{1,\text{METHOD}}$, the estimated X coefficients using METHOD; $\hat{\beta}_{2,\text{METHOD}}$, the estimated Z coefficients using METHOD; here METHOD stands for CC, IPW, MI, FI, AFI or AFIP.

3.2.3 | The selection of *M*1 and *M*2 in AFI

In this subsection, we investigate how the numbers of imputation replicates $M1$ and $M2$ affect the estimation accuracy and computation time of our proposed AFI method. We repeat the AFI estimation in settings S1-1-A, S1-2-A, S2-1-A and S2-2-A with $M1 = M2 = 10$ when $A = 1$, $M1 = M2 = 20$ when $A = 2$, $M1 = M2 = 50$ when $A = 3$, respectively. (ie, S1-1-1 stands for $M1 = M2 = 10$ in under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with normal e_i; S1-1-2 stands for $M1 = M2 = 20$ in under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with normal e_i; S1-1-3 stands for $M1 = M2 = 50$ in under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with normal e_i; S1-2-1 stands for $M1 = M2 = 10$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with chi-square e_i; S1-2-2 stands for $M1 = M2 = 20$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with chi-square e_i; S1-2-3 stands for $M1 = M2 = 50$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with chi-square e_i; S2-1-1 stands for $M1 = M2 = 10$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with normal e_i; S2-1-2 stands for $M1 = M2 = 20$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with normal e_i; S2-1-3 stands for $M1 = M2 = 50$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with normal e_i; S2-2-1 stands for $M1 = M2 = 10$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with chi-square e_i; S2-2-2 stands for $M1 = M2 = 20$ in

TABLE 4 Mean Squared Errors (M.S.E.) of the estimated coefficients using CC, IPW, MI, FI, AFI, and AFIP at quantile levels 0.1 and 0.5 under the setting S1 and S2 with sample size 500 and 200 Monte-Carlo replicates

		1–1		1–2		1–3		2–1		2–2		2–3	
τ		0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
S1	$\hat{\beta}_{1,CC}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{1,IPW}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{1,MI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{1,FI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{1,AFI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{1,AFIP}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{2,CC}$	0.02	0.01	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{2,IPW}$	0.02	0.01	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{2,MI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
	$\hat{\beta}_{2,FI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{2,AFI}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	$\hat{\beta}_{2,AFIP}$	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01
S2	$\hat{\beta}_{1,CC}$	0.18	0.10	0.18	0.10	0.18	0.10	0.00	0.07	0.00	0.07	0.00	0.07
	$\hat{\beta}_{1,IPW}$	0.16	0.10	0.16	0.10	0.16	0.10	0.00	0.07	0.00	0.07	0.00	0.07
	$\hat{\beta}_{1,MI}$	0.14	0.09	0.14	0.09	0.14	0.09	0.00	0.07	0.00	0.07	0.00	0.07
	$\hat{\beta}_{1,FI}$	0.14	0.09	0.14	0.09	0.14	0.09	0.00	0.06	0.00	0.06	0.00	0.06
	$\hat{\beta}_{1,AFI}$	0.14	0.08	0.14	0.09	0.14	0.09	0.00	0.06	0.00	0.06	0.00	0.06
	$\hat{\beta}_{1,AFIP}$	0.14	0.09	0.15	0.09	0.14	0.09	0.00	0.06	0.00	0.07	0.00	0.07
	$\hat{\beta}_{2,CC}$	0.33	0.15	0.33	0.15	0.33	0.15	0.00	0.13	0.00	0.13	0.00	0.13
	$\hat{\beta}_{2,IPW}$	0.32	0.15	0.32	0.15	0.32	0.15	0.00	0.13	0.00	0.13	0.00	0.13
	$\hat{\beta}_{2,MI}$	0.12	0.07	0.12	0.07	0.12	0.07	0.00	0.07	0.00	0.07	0.00	0.07
	$\hat{\beta}_{2,FI}$	0.13	0.07	0.13	0.07	0.13	0.07	0.00	0.06	0.00	0.06	0.00	0.06
	$\hat{\beta}_{2,AFI}$	0.13	0.07	0.13	0.07	0.13	0.08	0.00	0.06	0.00	0.06	0.00	0.06
	$\hat{\beta}_{2,AFIP}$	0.13	0.07	0.13	0.08	0.13	0.07	0.00	0.06	0.00	0.06	0.00	0.06

Note: S1, estimates under Case 1: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$; S2, estimates under Case 2: $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$. 1-1, $M1 = M2 = 10$ in AFI and AFIP with normal e_i; 1-2, $M1 = M2 = 20$ in AFI and AFIP with normal e_i; 1-3, $M1 = M2 = 50$ in AFI and AFIP with normal e_i; 2-1, $M1 = M2 = 10$ in AFI and AFIP with chi-square e_i; 2-2, $M1 = M2 = 20$ in AFI and AFIP with chi-square e_i; 2-3, $M1 = M2 = 50$ in AFI and AFIP with chi-square e_i; $\hat{\beta}_{1,\text{METHOD}}$, the estimated X coefficients using METHOD; $\hat{\beta}_{2,\text{METHOD}}$, the estimated Z coefficients using METHOD; here METHOD stands for CC, IPW, MI, FI, AFI or AFIP.

$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with chi-square e_i; S2-2-3 stands for $M1 = M2 = 50$ in $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with chi-square e_i.)

Tables 2 to 4 display the resulting mean biases, standard errors and mean squared errors of the estimated AFI and AFIP coefficients with different $M1$ and $M2$ at quantile levels 0.1 and 0.5. We found that, when $M1$ and $M2$ increases, the mean biases, standard errors and mean squared errors remain nearly unchanged. Small values of $M1$ and $M2$ (equal 10) are sufficient to stabilize the estimated coefficients. Bigger $M1$ and $M2$ do not further improve the accuracy in our simulations.

3.2.4 | Comparison of computing time

Table 5 displays the average computing time from 200 Monte-Carlo simulations under all settings (Seconds), using CC, IPW, MI, FI and AFI. It should be noted that, there exists only one variable to represent the number of imputation replicates in both MI (ie, m) and FI (ie, M), while there exist two in AFI (ie, $M1$ and $M2$).

TABLE 5 Average computing time from 200 Monte-Carlo simulations under all settings (Seconds)

	S1-1	S1-2	S2-1	S2-2
CC	0.01	0.01	0.01	0.01
IPW	0.02	0.02	0.01	0.01
MI(m = 10)	7.35	6.79	5.44	5.36
FI(M = 20)	0.74	0.69	0.50	0.49
AFI($M1 = M2 = 10$)	1.61(S1-1-1)	1.39(S1-2-1)	1.04(S2-1-1)	1.03(S2-2-1)
AFI($M1 = M2 = 20$)	3.50(S1-1-2)	3.91(S1-2-2)	2.14(S2-1-2)	2.17(S2-2-2)
AFI($M1 = M2 = 50$)	9.26(S1-1-3)	8.08(S1-2-3)	7.21(S2-1-3)	7.34(S2-2-3)

Note: S1-1, estimates under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with normal e_i; S1-2, estimates under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$ with chi-square e_i; S2-1, estimates under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with normal e_i; S2-2, estimates under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ with chi-square e_i; S1-1-1, $M1 = M2 = 10$ in S1-1; S1-1-2, $M1 = M2 = 20$ in S1-1; S1-1-3, $M1 = M2 = 50$ in S1-1; S1-2-1, $M1 = M2 = 10$ in S1-2; S1-2-2, $M1 = M2 = 20$ in S1-2; S1-2-3, $M1 = M2 = 50$ in S1-2; S2-1-1, $M1 = M2 = 10$ in S2-1; S2-1-2, $M1 = M2 = 20$ in S2-1; S2-1-3, $M1 = M2 = 50$ in S2-1; S2-2-1, $M1M1 = M2 = 10$ in S2-2; S2-2-2, $M1 = M2 = 20$ in S2-2; S2-2-3, $M1 = M2 = 50$ in S2-2.

Under S1-1, the average computing times of AFI ($M1 = M2 = 10$), AFI ($M1 = M2 = 20$) and AFI ($M1 = M2 = 50$) are 1.61, 3.50 and 9.26 seconds, respectively. With the $M1$ and $M2$ increased from 10 to 20 and from 20 to 50, the latter AFI costs nearly double computing time of the former. The average computing time of MI is 7.35 seconds, which cost more than 10 times of FIs (M = 20) computing time (Cheng et al 2018). MI costs more than twice as much as AFI ($M1 = M2 = 20$) and nearly five times as much as AFI ($M1 = M2 = 10$). AFI ($M1 = M2 = 10$) costs nearly two times of FI. These conclusions can also be confirmed in S1-2, S2-1 and S2-2. Based on all these settings, the proposed AFI ($M1 = M2 = 10$) almost costs only one fifth of MI's average computing time. It concludes that AFI ($M1 = M2 = 10$ or 20) is able to greatly relieve the computation burden in MI algorithm.

3.2.5 | Estimation robustness

In this subsection, we investigate the robustness property of our proposed AFI when the regression function $f(x| z)$ (Table 6) or propensity score function $p(\delta_i| z_i)$ (Table 7) is misspecified.

On the basis of Table 6, the standard errors and mean squared errors from the AFI and AFIP estimators with the misspecified $f(x| z)$ are nearly the smallest, especially when the e_i is normal for all the estimated coefficients and e_i is chi-square for x's estimated coefficients. More exactly, our AFI and AFIP are comparable with the existing MI and FI when $f(x| z)$ is misspecified, but better than CC and IPW in both standard errors and mean squared errors except z's estimated coefficients when e_i is chi-square. Compared with S2-1-1 and S2-2-1 in Tables 3 and 4 when $f(x| z)$ is correct, the difference between the AFI estimators using correct and misspecified densities are small relative to their standard errors, indicating that the AFI estimator is also fairly robust against the misspecification of $f(x| z)$.

Table 7 displays the results of S4 when the propensity score function $p(\delta_i| z_i)$ (Table 7) is misspecified. As expected, AFI performs better than IPW in mean biases, standard errors and mean squared errors when the e_i is normal. When e_i is chi-square, both standard errors and mean squared errors of AFI are smaller than IPW at quantile level 0.5. Compared with S2-1-1 and S2-2-1 in Tables 3and 4 when $p(\delta_i| z_i)$ is correct, the difference between the AFI estimators using correct and misspecified propensity score function are small relative to their standard errors, indicating that the AFI estimator is also fairly robust against the misspecification of $p(\delta_i| z_i)$.

4 | APPLICATION TO REAL DATA STUDY

In this section, we illustrate the performance of our AFI method using part of the Examination Data from National Health and Nutrition Examination Survey (NHANES) 2015-2016, which is a program of studies designed to assess the health and nutritional status of adults and children in the United States. In these data, we find that Upper Leg Length

TABLE 6 Mean biases (M.B.), standard errors (S.E.) and mean squared errors (M.S.E.) of the estimated coefficients using CC, IPW, MI, FI, AFI, and AFIP at quantile levels 0.1 and 0.5 under S3 with sample size 500 and 200 Monte-Carlo replicates

	Normal e_i						**Chi-square e_i**					
	M.B.		**S.E.**		**M.S.E.**		**M.B.**		**S.E.**		**M.S.E.**	
τ	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**
$\hat{\beta}_{1,CC}$	−0.03	−0.03	0.50	0.39	0.25	0.15	0.00	0.00	0.03	0.31	0.00	0.10
$\hat{\beta}_{1,IPW}$	−0.04	−0.03	0.52	0.38	0.27	0.15	0.00	0.00	0.03	0.31	0.00	0.10
$\hat{\beta}_{1,MI}$	−0.06	−0.04	0.48	0.36	0.24	0.13	−0.01	−0.04	0.03	0.29	0.00	0.09
$\hat{\beta}_{1,FI}$	−0.04	−0.05	0.49	0.36	0.25	0.13	−0.01	−0.05	0.03	0.29	0.00	0.09
$\hat{\beta}_{1,AFI}$	−0.05	−0.04	0.48	0.36	0.23	0.13	−0.01	−0.05	0.03	0.29	0.00	0.09
$\hat{\beta}_{1,AFIP}$	−0.05	−0.04	0.48	0.36	0.24	0.13	−0.01	−0.05	0.03	0.28	0.00	0.08
$\hat{\beta}_{2,CC}$	0.00	−0.01	0.53	0.36	0.28	0.13	0.01	0.00	0.03	0.27	0.00	0.08
$\hat{\beta}_{2,IPW}$	0.00	−0.02	0.53	0.35	0.28	0.13	0.01	0.00	0.03	0.27	0.00	0.07
$\hat{\beta}_{2,MI}$	0.00	0.00	0.50	0.35	0.25	0.12	0.02	0.04	0.03	0.28	0.00	0.08
$\hat{\beta}_{2,FI}$	0.00	0.01	0.51	0.34	0.26	0.12	0.02	0.04	0.03	0.28	0.00	0.08
$\hat{\beta}_{2,AFI}$	−0.01	0.01	0.50	0.34	0.25	0.12	0.02	0.04	0.03	0.27	0.00	0.08
$\hat{\beta}_{2,AFIP}$	0.00	0.01	0.50	0.34	0.25	0.12	0.02	0.04	0.03	0.28	0.00	0.08

Note: S3, misspecified regression function $f(x| z)$ under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ and $M1 = M2 = 10$ in AFI; $\hat{\boldsymbol{\beta}}_{1,METHOD}$, the estimated X coefficients using METHOD; $\hat{\boldsymbol{\beta}}_{2,METHOD}$, the estimated Z coefficients using METHOD; here METHOD stands for CC, IPW, MI, FI, AFI, or AFIP.

TABLE 7 Mean Biases (M.B.), Standard Errors (S.E.) and Mean squared Errors (M.S.E.) of the estimated coefficients using CC, IPW, MI, FI, AFI, and AFIP at quantile levels 0.1 and 0.5 under S4 with sample size 500 and 200 Monte-Carlo replicates

	Normal e_i						**Chi-square e_i**					
	M.B.		**S.E.**		**M.S.E.**		**M.B.**		**S.E.**		**M.S.E.**	
τ	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**	**0.1**	**0.5**
$\hat{\beta}_{1,IPW}$	0.05	0.04	0.42	0.32	0.18	0.10	0.00	−0.02	0.02	0.27	0.00	0.07
$\hat{\beta}_{1,AFI}$	0.01	0.00	0.39	0.29	0.15	0.08	−0.02	−0.09	0.03	0.24	0.00	0.06
$\hat{\beta}_{2,IPW}$	−0.04	−0.03	0.57	0.39	0.33	0.15	0.01	0.08	0.03	0.35	0.00	0.13
$\hat{\beta}_{2,AFI}$	0.00	0.01	0.35	0.27	0.12	0.07	0.02	0.06	0.04	0.23	0.00	0.06

Note: S4, misspecified propensity score function $p(\delta_i| z_i)$ under $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + (0.5x_i + 0.5z_i)e_i$ and $M1 = M2 = 10$ in AFI; $\hat{\boldsymbol{\beta}}_{1,METHOD}$, the estimated X coefficients using METHOD; $\hat{\boldsymbol{\beta}}_{2,METHOD}$, the estimated Z coefficients using METHOD; here METHOD stands for IPW or AFI.

(BMXLEG), Upper Arm Length (BMXARML), Arm Circumference (BMXARMC), Waist Circumference (BMXWAIST) and Average Sagittal Abdominal Diameter (BMDAVSAD) are relatively highly correlated with Standing Height (BMXHT). The correlation coefficients are 0.805, 0.891, 0.545, 0.453 and 0.410 respectively. Thus we build a model with y_i being BMXHT for the i^{th} person, $x_{i,1}$ being BMDAVSAD, $x_{i,2}$ being BMXLEG, $x_{i,3}$ being BMXARML, $x_{i,4}$ being BMXARMC and $x_{i,\ 5}$ being BMXWAIST. Considering the the distributions of the variables are commonly skewed, we use quantile regression and the model can be written as

$$y_i = \beta_{0,\tau} + \beta_{1,\tau} x_{i,1} + \beta_{2,\tau} x_{i,2} + \beta_{3,\tau} x_{i,3} + \beta_{4,\tau} x_{i,4} + \beta_{5,\tau} x_{i,5} + e_i. \tag{8}$$

We use 200 bootstraps among 721 subjects. The covariate $x_{i,1}$ (BMDAVSAD) is missing (the missing rate is about 20.83%), while other covariates are completely observed. Here we apply CC, IPW, MI, FI, AFI to obtain the estimated coefficients $\beta_{i,\ \tau}$, $i = 1, \ldots, n$, with $\boldsymbol{x}$ as BMDAVSAD, and $\boldsymbol{z}$ as BMXLEG, BMXARML, BMXARMC and BMXWAIST. We

TABLE 8 Raw estimation before bootstrap (Raw) and standard errors (S.E.), relative efficiencies (R.E.), *p* value from 200 bootstraps of the estimated coefficients in model (8)

		BMDAVSAD		BMXLEG		BMXARML		BMXARMC		BMXWAIST	
		0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5
CC	Raw	−0.31	−0.24	0.99	1.26	2.64	2.24	−0.01	−0.13	0.02	0.09
	S.E.	0.33	0.22	0.16	0.10	0.19	0.15	0.14	0.12	0.08	0.06
	P	0.34	0.29	0.00	0.00	0.00	0.00	0.92	0.28	0.79	0.11
IPW	Raw	−0.19	−0.29	1.03	1.24	2.52	2.21	0.01	−0.20	0.01	0.15
	S.E.	0.32	0.21	0.15	0.10	0.18	0.16	0.16	0.12	0.08	0.05
	P	0.56	0.17	0.00	0.00	0.00	0.00	0.93	0.10	0.93	0.00
	R.E.(%)	102	104	106	103	108	96	92	99	102	118
MI	Raw	−0.27	−0.32	1.07	1.28	2.54	2.19	−0.13	−0.24	0.06	0.15
	S.E.	0.32	0.22	0.14	0.08	0.19	0.12	0.13	0.09	0.08	0.05
	P	0.39	0.15	0.00	0.00	0.00	0.00	0.31	0.01	0.48	0.00
	R.E.(%)	104	102	115	128	99	128	109	131	106	112
FI	Raw	−0.31	−0.32	1.07	1.28	2.50	2.19	−0.08	−0.25	0.06	0.16
	S.E.	0.31	0.21	0.14	0.08	0.19	0.12	0.13	0.09	0.08	0.05
	P	0.33	0.13	0.00	0.00	0.00	0.00	0.54	0.01	0.47	0.00
	R.E.(%)	106	104	116	126	100	125	107	131	107	113
AFI	Raw	−0.27	−0.34	1.08	1.26	2.45	2.17	−0.10	−0.25	0.07	0.18
	S.E.	0.31	0.21	0.14	0.08	0.18	0.13	0.15	0.10	0.08	0.05
	P	0.39	0.10	0.00	0.00	0.00	0.00	0.49	0.01	0.39	0.00
	R.E.(%)	106	108	116	129	105	122	98	126	103	125

Note: R.E., the ratio between the estimated variances of the CC estimator and the other estimators.

set $m = 10$ in MI, $M = 20$ in FI and $M1 = M2 = 10$ in AFI. To illustrate the improved efficiency from AFI, we calculated the relative efficiencies of different methods relative to CC. The estimated coefficients, standard errors and relative efficiencies from different approaches at quantile levels 0.1 and 0.5 are listed in Table 8.

Table 8 consists of five layers. Each layer represents one estimator, including the estimated coefficients (raw estimation) before bootstrap, standard errors from 200 bootstraps and P-value. In addition, we also list the relative efficiency comparing to the CC estimates for the latter four layers. Based on Table 8, we find that almost all the imputation methods (IPW, MI, FI and AFI) have smaller estimated standard errors than the CC and IPW estimates. These are expected as both of CC and IPW only use the completely observed data. Figure 1 displays raw estimation and standard errors of the estimated coefficients.

We also calculate average computing times (ACT) of five imputation approaches from 200 bootstraps based on 721 subjects (Seconds). We find that AFI (1.51 seconds) is much more faster than MI (4.62 seconds). On average, CC, IPW and FI cost less than 1 second (0.01, 0.01, and 0.53 second) for each estimation process. Thus in our real data, AFI's average computing time is about 32.68% of MI's.

5 | DISCUSSION

In many applications, some observations could be missing for various reasons.[20,21] Ignoring the missing data will undermine study efficiency, and sometimes introduce substantial bias. Based on model (2), there already existed many related investigation works.[22-26] In our paper, we propose an augmented inverse probability weighted fractional imputation method (AFI) to handle missing covariates in quantile regression. The proposed AFI has the following advantages compared with the existing methods: 1) More efficient because making fully use of information not only from complete observations such as CC and IPW; 2) More robust to the misspecification of either the propensity score or the regression

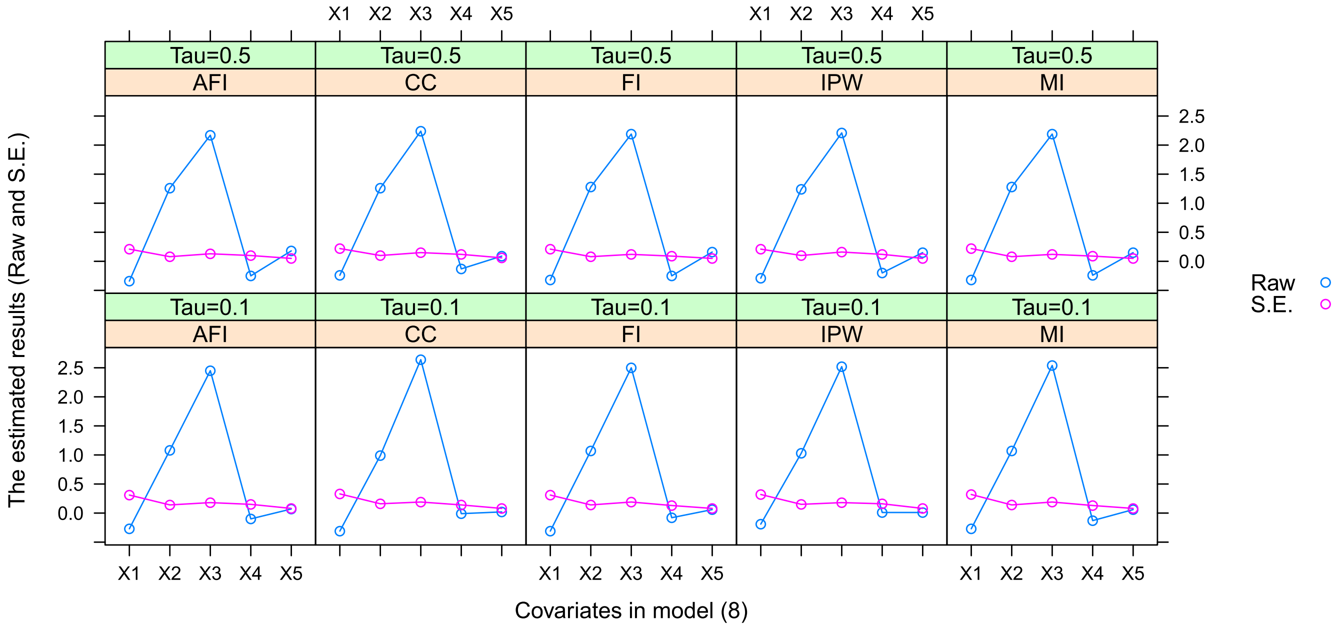

FIGURE 1 Raw estimation and standard errors (S.E.) of the estimated coefficients in Model (8). X1, BMDAVSAD; X2, BMXLEG; X3, BMXARML; X4, BMXARMC; X5, BMXWAIST

function; 3) Less computation burden than MI estimator, for example, the proposed AFI ($M1 = M2 = 10$) almost costs only one-third to one-fifth of MI's average computing time.

To propose AFI, we investigate its estimating function and divide it into three parts: $\sum_{i=1}^{n1} \frac{1}{p_i} \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right)$, $\sum_{i=n_1+1}^{n} E_x \left\{ \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right) | y_i, z_i \right\}$ and $\sum_{i=1}^{n1} \left(1 - \frac{1}{p_i}\right) E_x \left\{ \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right) | y_i, z_i \right\}$. Considering that conditional density of y given the covariates is unspecified under a typical quantile regression setting and classical likelihood-based approaches cannot be applied directly, solving missing covariate problem in the quantile regression is challenging. Here we use IPW for $\sum_{i=1}^{n1} \frac{1}{p_i} \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right)$, bayesian theory and monte-carlo integrations for $\sum_{i=n_1+1}^{n} E_x \left\{ \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right) | y_i, z_i \right\}$ and monte-carlo integrations for $\sum_{i=1}^{n1} \left(1 - \frac{1}{p_i}\right) E_x \left\{ \varphi_\tau \left(y_i - x_i^T \beta - z_i^T \beta\right) | y_i, z_i \right\}$. In addition, we estimate the density function $f(x|z_i)$ by maximizing a parametric likelihood over the observed (x, z) and $f(y \mid x, z)$ by using the first derivative of the the conditional quantile function $Q_y(\tau \mid x, z)$ (Wei et al, 2012).

Based on all the above, we will carry out the following investigation works in the future: (a) We will consider arbitrary nonlinear quantile functions instead of only focusing on linear quantile regressions; (b) Until now, all the methods that we talked about or proposed are parametric. We will consider semiparametric or nonparametric functions[27]; (c) It is possible to investigate missing data problems based on other kinds of basic model such as varying-coefficient quantile regression rather than linear quantile regression in our paper.[28-30] Therefore, we can carry out longitudinal data researches and solve missing covariates problems in that context[31,32]; (d) Last but not least, we will apply our methods to more real data applications.[33-35]

ACKNOWLEDGEMENTS

The author is very grateful to the reviewers and editors for their many helpful comments and suggestions that have led to better presentation and improvements in the paper. Hao Cheng's work is supported by The Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (16XNH102). The author wants to thank his parents' care and his wife Yujie Liu's love and support.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from National Health and Nutrition Examination Survey Database. And the website is https://www.cdc.gov/nchs/nhanes/index.htm.

ORCID

Hao Cheng https://orcid.org/0000-0003-1143-9601

REFERENCES

1. Koenker R, Bassett GJ. Regression quantiles. *Econometrica*. 1978;46(1):33-50.
2. Koenker R. *Quantile regression*. Cambridge: Cambridge University Press; 2005.
3. Little RJA, Rubin DB. *Statistical analysis with missing data*. New York, NY: Wiley; 2002.
4. Han PS, Kong LL, Zhao JW, Zhou XC. A general framework for quantile estimation with incomplete data. *J R Stat Soc Series B Stat Methodology*. 2019;81(2):305-333.
5. Horvitz DG, Tompson DJ. A generalization of sampling without replacement from a finite population. *J Am Stat Assoc*. 1952;47(260): 663-685.
6. Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Ann Appl Stat*. 1997;46(4):463-476.
7. Sherwood B, Wang L, Zhou X. Weighted quantile regression for analyzing health care cost data with missing covariates. *Stat Med*. 2013; 32(28):4967-4979.
8. Sun YQ, Wang H, Gilbert P. Quantile regression for competing risks data with missing cause of failure. *Stat Sin*. 2012;22(2):703-728.
9. Chen XR, WAN T, A K, Zhou Y. Efficient quantile regression analysis with missing observations. *J Am Stat Assoc*. 2015;110(510): 723-741.
10. Wang CY, Wang SJ, Zhao L, Ou ST. Weighted semiparametric estimation in regression analysis with missing covariate data. *J Am Stat Assoc*. 1997;92(438):512-525.
11. Yoon, J. Quantile regression analysis with missing response with applications to inequality measures and data combination (working paper). 2010. Retrieved from http://www.cmc.edu/pages/faculty/jyoon/.
12. Wei Y, Ma Y, Carroll RJ. Multiple imputation in quantile regression. *Biometrika* 2012. 2012;99(2):423-438.
13. Cheng H, Wei Y. A fast imputation algorithm in quantile regression. *Computational Statistics*. 2018;33(4):1017-1036.
14. Wei Y, Yang YK. Quantile regression with covariates missing at random. *Statistica Sinica*. 2014;24(3):1277-1299.
15. Zhou Y, Wan ATK, Wang X. Estimating equation inference with missing data. *J Am Stat Assoc*. 2008;103(483):1187-1199.
16. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coeffcients when some regressors are not always observed. *J Am Stat Assoc*. 1994;89(427):846-866.
17. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2011;22(3):278-295.
18. Wooldridge JM. Inverse probability weighted estimation for general missing data problems. *J Econometrics*. 2007;141(2):1281-1301.
19. Kim JK. Parametric fractional imputation for missing data analysis. *Biometrika*. 2011;98(1):119-132.
20. Little RJA. Regression with missing X's: A review. *J Am Stat Assoc*. 1992;87(420):1227-1237.
21. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002;7(2):147-177.
22. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*. 2005;61(4):962-972.
23. Cao W, Tstiatis AA, Daviadian M. Improving efficiency and robustness of doubly robust estimator for a population mean with incomplete data. *Biometrika*. 2009;96(3):723-734.
24. Graham BS, Pinto C, Egel D. Inverse probability tilting for moment condition models with missing data. *Rev Econ Stud*. 2012;79(3):1052-1079.
25. Qin J, Leung DHY, Zhang B. Efficient augmented inverse probability weighted estimation in missing data problems. *J Bus Econ Stat*. 2017;35(1):86-97.
26. Tan ZQ. Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*. 2010;97(3):661-682.
27. Tstiatis AA. *Semiparametric theory and missing data*. New York, NY: Springer Series in Statistics; 2006.
28. Hastie T, Tibshirani R. Varying-coefficient models. *J R Stat Soc Series B Stat Methodology*. 1993;55(4):757-796.
29. Honda T. Quantile regression in varying coefficient models. *Journal of Statistical Planning & Inference*. 2004;121(1):113-125.
30. Kim MO. Quantile regression with varying coefficients. *The Annals of Statistics*. 2007;35(1):92-108.
31. Zhao MT, Xu XL. Asymptotic estimation for longitudinal additive partial linear EV models. *Statistics and Information Forum*. 2019;34 (11):3-11.
32. Zhao MT, Xu XL, Gao W. Estimation of single-index longitudinal data models. *Statistics and Information Forum*. 2019;34(1):13-19.
33. Chen WD, Liu YX. An study on the methods for longitudinal data with missing values for emergency statistic under public emergency. *Statistics and Information Forum*. 2009;24(11):3-8.
34. Cheng H. An application research of inverse probability weighted multiple imputation method on factors of residents income in China. *Statistics and Information Forum*. 2019;34(7):26-34.
35. Yang GJ, Sun LL, Li L. Response propensity score matching imputation. *Statistics and Information Forum*. 2018;33(8):3-11.

How to cite this article: Cheng H. Augmented inverse probability weighted fractional imputation in quantile regression. *Pharmaceutical Statistics*. 2020;1–14. https://doi.org/10.1002/pst.2052