Parameter estimation through semiparametric quantile regression imputation

Senniang Chen and Cindy L. Yu[*]

Department of Statistics, Iowa State University, Ames, Iowa, USA
e-mail: cindyyu@iastate.edu

Abstract: In this article, we consider an imputation method to handle missing response values based on semiparametric quantile regression estimation. In the proposed method, the missing response values are generated using the semiparametrically estimated conditional quantile regression function at given values of covariates. Then the imputed values are used to estimate a parameter defined as the expected value of a function involving the response and covariate variables. We derive the asymptotic distribution of our estimator constructed with the imputed data and provide a variance estimator. In simulation, we compare our semiparametric quantile regression imputation method to fully parametric and nonparametric alternatives and evaluate the variance estimator based on the asymptotic distribution. We also discuss an extension for estimating a parameter defined through an estimation equation.

Keywords and phrases: GMM, imputation, semiparametric quantile regression.

1. Introduction

Missing data is frequently encountered in many disciplines. If the variables that explain the missing data are related to the response of interest, an inference based on ignoring missing undermines efficiency and often leads to biases and misleading conclusions. Missing data adjustments include weighting and imputation as two broad classes. In weighting, weights are derived through calibration or propensity score estimation. However, weighting is poorly suited to item nonresponse since for a given unit the weight should be the same for all items. Imputation provides a complete data set by replacing missing response variables with imputed values. In the presence of item nonresponse, imputation simplifies analyses because standard analytical tools can be applied to any imputed data set and the resulting point estimates are consistent across different users.

Many different imputation approaches have been developed in the literature and some prominent examples are included here. The pioneer work of Rubin (1987) discussed multiple imputation (MI) based on Bayesian methods to

---

[*]Corresponding author: Cindy L. Yu, Associate Professor, Department of Statistics, Iowa State University, Ames, IA 50011, USA.

generate pseudo values from the posterior predictive distribution and impute multiple data sets. Despite its simple form, however, the variance estimator of MI will have convergence problems if the congeniality and self-sufficiency conditions are not met (Meng (1994)). Fractional imputation was proposed to retain both estimation efficiency of multiple imputation and consistency of the Rao-Shao variance estimator (Rao and Shao (1992)). In fractional imputation, multiple values are imputed for each missing cell with assigned weights. Kim (2011) proposed parametric fractional imputation (PFI) with inspirations from importance sampling and calibration weighting to reduce the computation burden. Noticeably, both PFI and MI assume a parametric regression model, and therefore may suffer from model misspecification. While MI and PFI resort to the creation of artificial responses, hot-deck imputation (HDI) replaces missing units with observed data through matching methods. By using covariate information, the matching method could be classifying donors and recipients into similar categorical classes (Brick and Kalton (1996); Kim and Fuller (2004)), or creating metrics to match donors and recipients (Rubin (1987); Little (1988)). More examples are documented in Andridge and Little (2010). In a recent work by Wang and Chen (2009), multiple imputed values are independently drawn from observed respondents with probabilities proportional to kernel distances between missing cells and donors. Both HDI and Wang and Chen (2009) are purely nonparametric, so the stability and accuracy of the estimators depend on the dimensionality and the sample size concerned.

To leverage the advantages of both parametric and nonparametric methods and avoid the limitation of a pure or exclusive approach, we propose an imputation method based on semiparametric quantile regression, which has the following set up. Define $f(y|\mathbf{x})$ as the conditional density where y is the response subject to missing and $\mathbf{x}$ is the covariate always observed, and $q_\tau(\mathbf{x})$ as the τ-th conditional quantile function, which is the inverse conditional distribution function $F^{-1}(\tau|\mathbf{x})$. Instead of estimating $f(y|\mathbf{x})$ parametrically or non-parametrically, we estimate $q_\tau(\mathbf{x})$ semiparametrically using observed data under the missing at random (MAR) assumption, in the sense intended by Rubin (1976). Then multiple imputed values $y_j^*(j = 1, \cdots, J)$ are obtained via $y_j^*|\mathbf{x} = \hat{q}_{\tau_j}(\mathbf{x})$, where τ_j is independently drawn from Uniform$[0, 1]$. The semiparametric quantile regression imputation (hereafter called SQRI) is expected to have appealing features. Firstly, the entire conditional distribution function is used to draw imputed values, hence preserving the conditional density of the filled-in response values. Secondly, because different conditional quantiles instead of conditional means or actual observations are used in imputation, the method is less sensitive to outliers, as quantiles are known to be less affected by extremes. Thirdly, it does not require strong model assumptions as in a fully parametric solution, and therefore is robust against model violations. Lastly, imputed values can be easily created through random numbers generated from Uniform$[0, 1]$ once $\hat{q}_\tau(\mathbf{x})$ is estimated.

We are not the first to use quantile regression for imputation. Papers pertaining to quantile regression imputation include Munoz and Rueda (2009), Wei, Ma, and Caroll (2012) and Yoon (2013). Our paper is distinctive from these

papers in terms of objective, type of imputation and theory. (i) For objective, Wei, Ma, and Caroll (2012) and Yoon (2013) aimed to estimate quantile regression coefficients, while we are interested in estimating more general parameters, e.g. parameters defined through a smooth estimation equation. It is also worth noting that the setting in Wei, Ma, and Caroll (2012) is different since they dealt with missing covariates, not missing responses. (ii) For type, Wei, Ma, and Caroll (2012) imputed multiple data sets, and Munoz and Rueda (2009) proposed a single and deterministic imputation. Our method utilizes fractional imputation because the variance estimator used in multiple imputation is not consistent for some estimated parameters. For examples, see Wang and Robins (1998) and Kim, Brick, Fuller, and Kalton (2006). (iii) For theory, Wei, Ma, and Caroll (2012) and Yoon (2013) assumed a linear quantile regression model, while we rely on a semiparametric approach incorporating penalty for model complexity. And the key idea used to arrive at the asymptotic normality in our proof is substantially different from that of Wei, Ma, and Caroll (2012) and Yoon (2013).

We demonstrate the theoretical validity and applicability of our SQRI estimators in this paper. The rest of paper is organized as follows. Section 2 introduces the SQRI algorithm and the estimator constructed using the imputed values. Section 3 presents large sample theories and a variance estimator. Section 4 discusses an extension for estimating a parameter defined through an estimation equation. Section 5 demonstrates the properties of SQRI estimators through simulations. Section 6 concludes with some remarks. Appendix outlines proofs of the theorems appearing in the main text.

2. Semiparametric quantile regression imputation (SQRI)

We consider $(\mathbf{x}_i, y_i)^T (i = 1, \cdots, n)$ to be a set of i.i.d. observations of random variables $(\mathbf{X}, Y)$, where Y is the response variable subject to missing, and $\mathbf{X}$ is a d_x-dimension variable always observed. Each dimension of $\mathbf{X}$ has a compact support set on $[0, 1]$. Let $q_\tau(\mathbf{x})$ be the unknown conditional $100\tau\%$ quantile of response Y given $\mathbf{X} = \mathbf{x}$. For a given $\tau \in (0, 1)$, the conditional quantile function $q_\tau(\mathbf{x})$ is defined as a function satisfying,

$$P(Y < q_\tau(\mathbf{x})|\mathbf{X} = \mathbf{x}) = \tau. \tag{2.1}$$

When $\tau = 0.5$, $q_\tau(\mathbf{x})$ is the conditional median of Y. It is easy to show that $q_\tau(\mathbf{x})$ satisfies

$$q_\tau(\mathbf{x}) = \arg \min_{h(\mathbf{x})} E\{\rho_\tau(Y - h(\mathbf{x}))|\mathbf{X} = \mathbf{x}\}, \tag{2.2}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function proposed in Koenker and Bassett (1978). The true parameter of interest, $\boldsymbol{\theta}_0$, is a d_θ-dimensional vector defined as,

$$\boldsymbol{\theta}_0 = E\{\mathbf{r}(y_i, \mathbf{x}_i)\}, \tag{2.3}$$

where $\mathbf{r}(y_i, \mathbf{x}_i)$ is a d_θ-dimensional function. If the full sample is observed, we can estimate $\boldsymbol{\theta}_0$ in (2.3) by,

$$\widehat{\boldsymbol{\theta}}_{full} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}(y_i, \mathbf{x}_i). \tag{2.4}$$

Because y_i is unobserved for missing units, $\widehat{\boldsymbol{\theta}}_{full}$ is not available. In Section 2.2, we outline a procedure to obtain an imputed version of (2.4) by using semiparametric quantile regression.

Let $\delta_i = 1$ be the response indicator, defined by $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ if y_i is missing. Assume $\delta_i \sim Bernoulli(p_i)$, where the response probability p_i is defined as,

$$p_i = p(x_i) = P(\delta_i = 1|y_i, \mathbf{x}_i). \tag{2.5}$$

Assume that the response mechanism satisfies the condition of missing at random (MAR) by Rubin (1976). MAR asserts that the response variable y_i is conditionally independent of the indicator δ_i and can be expressed in terms of the response probability as,

$$P(\delta_i = 1|y_i, \mathbf{x}_i) = P(\delta_i = 1|\mathbf{x}_i) = p_i, \tag{2.6}$$

or in terms of the conditional distribution of y_i given $\mathbf{x}_i$ as,

$$f_{y|x}(y_i|\mathbf{x}_i, \delta_i) = f_{y|x}(y_i|\mathbf{x}_i). \tag{2.7}$$

2.1. SQRI using penalized B-splines

Many papers have studied the estimation of $q_\tau(\mathbf{x})$ based on parametric methods, and a summary of relevant literature can be found in Koenker (2005). Parametric model assumptions may not hold sometimes, giving rise to nonparametric methods. Nonparametric quantile regression, including the kernel quantile regression in Yu and Jones (1994) and the smoothing spline method in Koenker, Ng, and Portnoy (1994), has also been intensively studied. Among many findings is the well-known trade-off between computational cost and smoothness. In other words, spline smoothing methods demand massive computation, and the unpenalized spline tends to give wiggly curves despite its cheap computational cost. There are some recent work on nonparametric quantile regression (thanks to a referee for drawing our attention). Belloni, Chetverikov, and Fernandez-Val (2016) developed the nonparametric quantile regression series framework in which the conditional quantile function is approximated by a linear combination of series terms. They showed that this framework can cover many models as a special case. Chao, Volgushev, and Cheng (2016) established weak convergence for the quantile regression estimator in a general series approximation framework, and studied a partial linear model which is included as a special case of such framework and might be a solution to the curse of dimensionality in a

nonparametric approach. In this paper, we employ a semiparametric quantile regression method based on penalized B-splines, as suggested in Yoshida (2013), that features a relatively smoothed quantile function at reduced computational burden.

Without loss of generality, we consider X as an univariate variable with a distribution function $F_x(x)$ on $[0, 1]$ to simplify notations in theories. We discuss how to deal with multivariate $\mathbf{X}$ in the simulation study of Section 5. Let $K_n - 1$ be the number of knots within the range $(0, 1)$, and p be the degree of B-splines. In order to construct the p-th degree B-spline basis, we define equidistantly located knots as $\kappa_k = K_n^{-1} k, (k = -p + 1, \cdots, K_n + p)$. Note there are $K_n - 1$ knots located in $(0, 1)$. The p-th B-spline basis is

$$\mathbf{B}(x) = (B_{-p+1}^{[p]}(x), B_{-p}^{[p]}(x), \cdots, B_{K_n}^{[p]}(x))^T,$$

where $B_k^{[p]}(x)(k = -p + 1, \cdots, K_n)$ are defined recursively as,

- For $s = 0$:

$$B_k^{[0]}(x) = \left\{ \begin{array}{ll} 1, \kappa_{k-1} < x \leq \kappa_k, \\ 0, \text{ otherwise,} \end{array} \right. \quad \text{where } k = -p + 1, \cdots, K_n + p;$$

- For $s = 1, 2, \cdots, p$:

$$B_k^{[s]}(x) = \frac{x - \kappa_{k-1}}{\kappa_{k+s-1} - \kappa_{k-1}} B_k^{[s-1]}(x) + \frac{\kappa_{k+s} - x}{\kappa_{k+s} - \kappa_k} B_{k+1}^{[s-1]}(x),$$

where $k = -p + 1, ..., K_n + p - s$.

Readers can refer to de Boor (2001) for more details and properties of the B-spline functions. The estimated conditional quantile regression function is

$$\hat{q}_\tau(x) = \mathbf{B}^T(x)\hat{\mathbf{b}}(\tau), \tag{2.8}$$

where $\hat{\mathbf{b}}(\tau)$ is a $(K_n + p) \times 1$ vector obtained by,

$$\hat{\mathbf{b}}(\tau) = \arg\min_{\mathbf{b}(\tau)} \sum_{i=1}^n \delta_i \rho_\tau[y_i - \mathbf{B}^T(x_i)\mathbf{b}(\tau)] + \frac{\lambda_n}{2} \mathbf{b}^T(\tau)\mathbf{D}_m^T \mathbf{D}_m \mathbf{b}(\tau). \tag{2.9}$$

Here $\lambda_n (> 0)$ is the smoothing parameter, and $\mathbf{D}_m$ is the m-th difference matrix and is $(K_n + p - m) \times (K_n + p)$ dimensional with its element defined as

$$d_{ij} = \left\{ \begin{array}{ll} (-1)^{|i-j|} \left(\begin{array}{c} m \\ |i - j| \end{array} \right) & 0 \leq j - i \leq m \\ 0 & o.w. \end{array} \right. ,$$

where the notation $\binom{m}{k}$ is the choose function given by $(k!(m-k)!)^{-1} m!$ and m is the order of penalty. As discussed in Yoshida (2013), the difference penalty $\mathbf{b}^T(\tau)\mathbf{D}_m^T \mathbf{D}_m \mathbf{b}(\tau)$ is used to remove computational difficulty occurring when the penalty term is defined through an integral, and it controls the smoothness of the

estimated quantile regression function. When $m = 2$, $\mathbf{D}_m$ has an interpretation related to the integral of the square of the second derivative of the function defined by the B-spline. Because the second derivative of a straight line is zero, the use of $\mathbf{D}_m$ for $m = 2$ shrinks the estimated quantile regression function toward a straight line. The asymptotic property of $\hat{q}_\tau(x)$ is given in Section 3. Section 5 discusses about how we choose the tuning parameters (λ_n, m, K_n, p) in practice.

2.2. Parameter estimation through SQRI

Suppose the parameter of interest is $\boldsymbol{\theta}_0 = E\{\mathbf{r}(y_i, \mathbf{x}_i)\}$. Cheng (1994) showed that $n^{-1}\sum_{i=1}^n[\delta_i\mathbf{r}(y_i, \mathbf{x}_i) + (1-\delta_i)E_{y|x}\{\mathbf{r}(y_i, \mathbf{x}_i)\}]$ is a consistent estimator of $\boldsymbol{\theta}_0$ under MAR. By the probability integral transformation, $\tau = F_{y|x}(y|x)$, $E_{y|x}\{\mathbf{r}(y_i, \mathbf{x}_i)\} = \int_0^1 \mathbf{r}(q_\tau(\mathbf{x}_i), \mathbf{x}_i)d\tau$. Therefore an estimator of $E_{y|x}\{\mathbf{r}(y_i, \mathbf{x}_i)\}$ is $\int_0^1 \mathbf{r}(\hat{q}_\tau(\mathbf{x}_i), \mathbf{x}_i)d\tau$ where $\hat{q}_\tau(\mathbf{x}_i) = \mathbf{B}^T(x_i)\hat{\mathbf{b}}(\tau)$, and an estimator of $\boldsymbol{\theta}_0$ can be obtained as,

$$\widetilde{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^n[\delta_i\mathbf{r}(y_i, \mathbf{x}_i) + (1-\delta_i)\int_0^1 \mathbf{r}(\hat{q}_\tau(\mathbf{x}_i), \mathbf{x}_i)d\tau]. \tag{2.10}$$

The integral in (2.10) can be approximated by $J^{-1}\sum_{j=1}^J \mathbf{r}(\hat{q}_{\tau_j}(\mathbf{x}_i), \mathbf{x}_i)$, where τ_j ($j = 1, \cdots, J$) are J independent values generated from $Uniform[0, 1]$. This suggests the following SQRI imputation procedure. When y_i is missing, we generate J imputed values $\{y_{ij}^*\}_{j=1}^J$ by,

1. Simulate $\tau_j \sim$ Uniform(0,1) independently for $j = 1, 2, \cdots, J$;
2. For each $j = 1, 2, ..., J$, $\hat{\mathbf{b}}(\tau_j)$ is calculated as,

$$\hat{\mathbf{b}}(\tau_j) = \arg\min_{\mathbf{b}(\tau)}\sum_{i=1}^n \delta_i\rho_{\tau_j}[y_i - \mathbf{B}^T(x_i)\mathbf{b}(\tau)] + \frac{\lambda_n}{2}\mathbf{b}^T(\tau)\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}(\tau);$$

3. For the missing unit i, J independent values are generated as,

$$y_{ij}^*|x_i = \hat{q}_{\tau_j}(x_i) = \mathbf{B}^T(x_i)\hat{\mathbf{b}}(\tau_j), j = 1, 2, \cdots, J.$$

Repeat step 3 for every missing unit in the data set. Then an imputed version of $\hat{\boldsymbol{\theta}}_{full}$ in (2.4) is defined as,

$$\widehat{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^n[\delta_i\mathbf{r}(y_i, \mathbf{x}_i) + (1-\delta_i)\frac{1}{J}\sum_{j=1}^J \mathbf{r}(y_{ij}^*, x_i)]. \tag{2.11}$$

For some parametric imputation methods, imputation and estimation steps are entwined, in that updating parameters and re-imputing based on most recently updated parameters are iteratively done. This might require heavy computing time. In the SQRI described above, imputation and estimation steps are

totally separate, making general purpose parameter estimation possible. Also in SQRI, standard analytical tools can be directly applied to imputed data without re-imputation. The PFI by Kim (2011) avoids re-imputation by adjusting weights of imputed values based on iteratively updated parameters. However, any parametric imputation method, including PFI and MI, might suffer from model misspecification. Nonparametric imputation, such as HDI or the method proposed in Wang and Chen (2009) using kernel distance, assumes no parametric model, but the stability and accuracy of nonparametric estimators depend on sample size and dimensionality of the problem. The SQRI provides a useful compromise between a fully parametric approach and a purely nonparametric approach.

3. Large sample theories and variance estimation

In this section, we demonstrate the theoretical validity of $\widehat{\boldsymbol{\theta}}$ constructed with the imputed values generated from the SQRI procedure described in Section 2, provide a variance estimator for $\widehat{\boldsymbol{\theta}}$, and also introduce an alternative SQRI estimator using fixed τ_j's.

3.1. Asymptotic normality of $\widehat{\boldsymbol{\theta}}$

The derivation of the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ proceeds in two steps. In Lemma 1, we give the asymptotic distribution of the estimated quantile funtion $\hat{q}_\tau(x)$, which leads to the Bahadur representation of the estimated quantile regression coefficients $\hat{b}(\tau)$. In Theorem 1, we provide the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$. The Bahadur representation of $\hat{\mathbf{b}}(\tau)$ is crucial to the proof of Theorem 1.

Assuming the quantile function $q_\tau(x)$ is a function with $p + 1$ continuous derivatives, $K_n = O(n^{\frac{1}{2p+3}})$ and $\lambda_n = O(n^v)$ for $v \leq (2p+3)^{-1}(p+m+1)$, Barrow and Smith (1978) showed that there exists $\mathbf{b}^*(\tau)$ satisfying,

$$\sup_{x \in (0,1)} |q_\tau(x) + b_\tau^a(x) - \mathbf{B}^T(x)\mathbf{b}^*(\tau)| = o(K_n^{-(p+1)}), \qquad (3.1)$$

where $\mathbf{B}^T(x)\mathbf{b}^*(\tau)$ is the best L_∞ approximation for $q_\tau(x)$, and $b_\tau^a(x)$ is the approximation error for the true quantile function satisfying, $b_\tau^a(x) = O(K_n^{-(p+1)})$. For the detailed form of $b_\tau^a(x)$, see Appendix A. We now state Lemma 1.

Lemma 1. *Under the conditions given in Appendix, and assuming* $q_\tau(x) \in C^{p+1}$, $K_n = O(n^{\frac{1}{2p+3}})$, *and* $\lambda_n = O(n^v)$ *for* $v \leq (2p+3)^{-1}(p+m+1)$, *we have*

(i)

$$\sqrt{\frac{n}{K_n}}[\hat{q}_\tau(x) - \mathbf{B}^T(x)\mathbf{b}^*(\tau) + b_\tau^\lambda(x)] \to_d N(0, V_\tau(x)),$$

(ii)

$$\sqrt{\frac{n}{K_n}}[\hat{q}_\tau(x) - q_\tau(x) + b_\tau^a(x) + b_\tau^\lambda(x)] \to_d N(0, V_\tau(x)),$$

for a given $x \in (0,1)$ and $\tau \in (0,1)$, where

$$b_\tau^\lambda(x) = \frac{\lambda_n}{n} \mathbf{B}^T(x) \left(\mathbf{\Phi}(\tau) + \frac{\lambda_n}{n} \mathbf{D}_m^T \mathbf{D}_m \right)^{-1} \mathbf{D}_m^T \mathbf{D}_m \mathbf{b}^*(\tau),$$

$$V_\tau(x) = \lim_{n \to \infty} \frac{\tau(1-\tau)}{K_n} \mathbf{B}^T(x) \mathbf{H}_n^{-1}(\tau) \mathbf{\Phi} \mathbf{H}_n^{-1}(\tau) \mathbf{B}(x),$$

$$\mathbf{H}_n(\tau) = \mathbf{\Phi}(\tau) + \frac{\lambda_n}{n} \mathbf{D}_m^T \mathbf{D}_m,$$

$$\mathbf{\Phi} = \int_0^1 p(x) \mathbf{B}(x) \mathbf{B}^T(x) dF_x(x),$$

$$\mathbf{\Phi}(\tau) = \int_0^1 p(x) f_{y|x}(x, q_\tau(x)) \mathbf{B}(x) \mathbf{B}^T(x) dF_x(x).$$

Here $f_{y|x}(x,y)$ is the conditional density of Y given $X = x$. There exist two sources of asymptotic biases in $\hat{q}_\tau(x)$. One is $b_\tau^a(x)$ which is the model bias between the true function $q_\tau(x)$ and the spline model used, see equation (3.1). Another source of bias $b_\tau^\lambda(x)$ is introduced by adding the penalty term into the quantile regression. When there is no penalty term ($\lambda_n = 0$), this bias vanishes. Both of these two bias terms have an order of $O(n^{-\frac{p+1}{2p+3}})$. By the proof of Lemma 1, the estimator of the quantile regression coefficient $\hat{\mathbf{b}}(\tau)$ has the following Bahadur representation,

$$\sqrt{\tfrac{n}{K_n}}(\hat{\mathbf{b}}(\tau) - \mathbf{b}^*(\tau) + \tfrac{\lambda_n}{n} \mathbf{H}_n^{-1}(\tau) \mathbf{D}_m^T \mathbf{D}_m \mathbf{b}^*(\tau))$$

$$= \sqrt{\tfrac{n}{K_n}} \mathbf{H}_n^{-1}(\tau) \tfrac{1}{n} \sum_{i=1}^n \delta_i \mathbf{B}(x_i) \psi_\tau(e_i(\tau)) + o_p(1),$$

(3.2)

where $\psi_\tau(u) = \tau - 1_{u<0}$ and $e_i(\tau) = y_i - \mathbf{B}^T(x_i)\mathbf{b}^*(\tau)$. The property in (3.2) is extensively used in the derivation of Theorem 1 which regards the asymptotic normality of $\widehat{\boldsymbol{\theta}}$. We now state Theorem 1.

Theorem 1. *Under the conditions given in the Appendix, and assuming $q_\tau(x) \in C^{p+1}$, $K_n = O(n^{\frac{1}{2p+3}})$, and $\lambda_n = O(n^v)$ for $v \leq (2p+3)^{-1}(p+m+1)$, as $n \to \infty$ and $J \to \infty$ we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)), \tag{3.3}$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = Var(\xi_{i,r}(\boldsymbol{\theta}_0)),$$

$$\xi_{i,r}(\boldsymbol{\theta}_0) = \mathbf{r}(y_i, x_i) - \boldsymbol{\theta}_0 + (1 - \delta_i) \left[\mu_{\mathbf{r}|x}(x_i) - \mathbf{r}(y_i, x_i) \right] + \delta_i C_p h_r(y_i, x_i) \mathbf{B}(x_i),$$

$$h_r(y_i, x_i) = \int_{-\infty}^{+\infty} \int_0^1 \dot{\mathbf{r}}_y(q_\tau(x), x) \mathbf{B}^T(x) \mathbf{H}_n^{-1}(\tau) \psi_\tau(e_i(\tau)) d\tau dF_x(x), \tag{3.4}$$

and $\dot{\mathbf{r}}_y(y,x) = \partial \mathbf{r}(y,x)/\partial y$, $\mu_{\mathbf{r}|x}(x) = E_{y|x}\{\mathbf{r}(y,x)\}$, and $C_p = E\{1 - p(x)\}$.

Remark 1: When there is no missing, $\xi_{i,r}(\boldsymbol{\theta}_0)$ in equation (3.4) coincides with $\mathbf{r}(y_i, x_i) - \boldsymbol{\theta}_0$.

The key step in the proof for Theorem 1 is to show that $n^{-1} \sum_{i=1}^{n} \xi_{i,r}(\boldsymbol{\theta}_0)$ is an asymptotic equivalence for $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$, i.e.

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \frac{1}{n} \sum_{i=1}^{n} \xi_{i,r}(\boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}}). \tag{3.5}$$

Appendix A gives the proof of Lemma 1, and Appendix B outlines the proof of Theorem 1.

3.2. Variance estimator for $\widehat{\boldsymbol{\theta}}$

We use the asymptotic matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = Var(\xi_{i,r}(\boldsymbol{\theta}_0))$ in (3.4) to estimate the variance of $\widehat{\boldsymbol{\theta}}$. Let an estimator of $\xi_{i,r}(\boldsymbol{\theta}_0)$ be $\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}})$, where

$$\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}}) = \mathbf{r}(y_i, x_i) - \widehat{\boldsymbol{\theta}} + (1 - \delta_i)\left\{\hat{\mu}_{r|x}(x_i) - \mathbf{r}(y_i, x_i)\right\} + \delta_i \hat{C}_p \hat{h}_r(y_i, x_i)\mathbf{B}(x_i),$$

$$\hat{h}_r(y_i, x_i) = \frac{1}{n}\frac{1}{J} \sum_{k=1}^{n} \sum_{j=1}^{J} \dot{\mathbf{r}}_y(\hat{q}_{\tau_j}(x_k), x_k)\mathbf{B}^T(x_k)\hat{\mathbf{H}}_n^{-1}(\tau_j)\psi_{\tau_j}(\hat{e}_i(\tau_j)),$$

$$\hat{\mathbf{H}}_n(\tau_j) = \hat{\boldsymbol{\Phi}}(\tau_j) + \frac{\lambda_n}{n}\mathbf{D}_m^T\mathbf{D}_m,$$

$$\hat{\boldsymbol{\Phi}}(\tau_j) = \frac{1}{n} \sum_{i=1}^{n} \delta_i \hat{f}_{y|x}(x_i, \hat{q}_{\tau_j}(x_i))\mathbf{B}(x_i)\mathbf{B}^T(x_i), \tag{3.6}$$

$\hat{\mu}_{r|x}(x_i) = J^{-1}\sum_{j=1}^{J}\mathbf{r}(\hat{q}_{\tau_j}(x_i), x_i)$, $\hat{e}_i(\tau_j) = y_i - \mathbf{B}^T(x_i)\hat{\mathbf{b}}(\tau_j)$ and $\hat{C}_p = n^{-1}\sum_{i=1}^{n}(1 - \delta_i)$. We use kernel density estimation to estimate $f_{y|x}(x, y)$ as,

$$\hat{f}_{y|x}(x, y) = \frac{(nab)^{-1}\sum_{i=1}^{n}\delta_i K(\frac{y-y_i}{a})K(\frac{x-x_i}{b})}{(na)^{-1}\sum_{i=1}^{n}\delta_i K(\frac{x-x_i}{a})}, \tag{3.7}$$

where $K(\cdot)$ is a Normal kernel and a (or b) is the bandwidth for x (or y).

Then the variance estimator of $\widehat{\boldsymbol{\theta}}$ is $\widehat{V}(\widehat{\boldsymbol{\theta}}) = n^{-1}\widehat{\Sigma}(\widehat{\boldsymbol{\theta}})$, where

$$\widehat{\Sigma}(\widehat{\boldsymbol{\theta}}) = \frac{1}{n-1} \sum_{i=1}^{n} \left\{\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}})\right\}\left\{\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_{i,r}(\widehat{\boldsymbol{\theta}})\right\}^T. \tag{3.8}$$

3.3. Alternative SQRI estimator using fixed τ_j's

When using random τ_j's generated from $Uniform[0, 1]$ to impute y_{ij}^*, it is possible to get some τ_j's that are very close to zero or one. This might cause instability in variance estimation due to extreme quantiles. To avoid this potential trouble, we propose an alternative SQRI estimator that uses fixed τ_j's.

The estimator $\widetilde{\boldsymbol{\theta}}$ defined in (2.10) is a consistent estimator for $\boldsymbol{\theta}_0 = E\{\mathbf{r}(y_i, x_i)\}$. Instead of approximating the integral $\int_0^1 \mathbf{r}(\hat{q}_\tau(\mathbf{x}_i), \mathbf{x}_i) d\tau$ in $\widetilde{\boldsymbol{\theta}}$ using J random τ's generated from $Uniform[0, 1]$, we can use a midpoint approximation approach (e.g. Nusser, Carriquiry, Dodd, and Fuller (1996)) which entails dividing the interval $[0, 1]$ into J sub-intervals and evaluating the function at the midpoints of the J sub-intervals. An estimator using the fixed midpoints can be defined as,

$$\widehat{\boldsymbol{\theta}}_{fixed} = \frac{1}{n} \sum_{i=1}^n [\delta_i \mathbf{r}(y_i, \mathbf{x}_i) + (1 - \delta_i) \frac{1}{J} \sum_{j=1}^J \mathbf{r}(\hat{q}_{\tau_j}(\mathbf{x}_i), x_i)], \qquad (3.9)$$

where τ_j $(j = 1, \cdots, J)$ are the fixed sequence of the midpoints of J evenly spaced sub-intervals of $[0, 1]$ with $0 < \tau_1 < \tau_2 < \cdots < \tau_J < 1$, and $\hat{q}_{\tau_j}(x_i) = \mathbf{B}^T(x_i)\hat{\mathbf{b}}(\tau_j)$.

By the proof given in Appendix C, the SQRI estimator $\widehat{\boldsymbol{\theta}}_{fixed}$ has the following asymptotic order,

$$\widehat{\boldsymbol{\theta}}_{fixed} - \boldsymbol{\theta}_0 = \frac{1}{n} \sum_{i=1}^n \xi_{i,r}(\boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{J}). \qquad (3.10)$$

Under the same conditions given in Theorem 1 and assuming further that $J = O(n^\alpha)$ for $\alpha > 1/2$, the two SQRI estimators $\widehat{\boldsymbol{\theta}}_{fixed}$ and $\widehat{\boldsymbol{\theta}}$ have the same asymptotic equivalence $n^{-1} \sum_{i=1}^n \xi_{i,r}(\boldsymbol{\theta}_0)$, i.e.

$$\widehat{\boldsymbol{\theta}}_{fixed} - \boldsymbol{\theta}_0 = \frac{1}{n} \sum_{i=1}^n \xi_{i,r}(\boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}}). \qquad (3.11)$$

Therefore the asymptotic normality and the variance estimator of $\widehat{\boldsymbol{\theta}}_{fixed}$ can be derived similarly as those for $\hat{\boldsymbol{\theta}}$ in Sections 3.1 and 3.2. To see whether there are any finite sample differences between the SQRI estimator using random τ_j's and the SQRI estimator using fixed τ_j's, both of the estimators are calculated and reported in the simulation studies of Section 5.

4. Estimating a parameter defined by an estimation equation through SQRI

In this section, we discuss how to use the SQRI to estimate more general parameters defined through an estimation equation. If the true parameter $\boldsymbol{\theta}_0$ is defined as an unique solution to,

$$E\{\mathbf{g}(Y, \mathbf{X}; \boldsymbol{\theta})\} = 0, \qquad (4.1)$$

where $\mathbf{g}(Y, \mathbf{X}; \boldsymbol{\theta})$ is a vector of d_g estimation functions for $d_g \geq d_\theta$. We consider the generalized method of moment (GMM) to estimate $\boldsymbol{\theta}_0$. The GMM estimator is obtained as,

$$\widehat{\boldsymbol{\theta}}_{gmm} = \arg\min_{\boldsymbol{\theta} \in \Theta} \mathbf{G}_n^T(\boldsymbol{\theta})\mathbf{G}_n(\boldsymbol{\theta}), \qquad (4.2)$$

where

$$\mathbf{G}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \{\delta_i \mathbf{g}(y_i, x_i; \boldsymbol{\theta}) + (1 - \delta_i) \frac{1}{J} \sum_{j=1}^{J} \mathbf{g}(y_{ij}^*, x_i; \boldsymbol{\theta})\}, \qquad (4.3)$$

and y_{ij}^* are J imputed values generated through the SQRI procedure either using random τ_j's from $Uniform[0,1]$ or using the fixed sequence of the midpoints of J equally spaced subintervals on $[0,1]$.

The asymptotic normality of $\widehat{\boldsymbol{\theta}}_{gmm}$ is very similar to Theorem 1, and is stated as follows,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \to_d N(0, \boldsymbol{\Sigma}_g(\boldsymbol{\theta}_0)), \qquad (4.4)$$

where

$$\boldsymbol{\Sigma}_g(\boldsymbol{\theta}_0) = \left\{\Gamma^T(\boldsymbol{\theta}_0)\Gamma(\boldsymbol{\theta}_0)\right\}^{-1} \Gamma^T(\boldsymbol{\theta}_0) \, Var\{\xi_{i,g}(\boldsymbol{\theta}_0)\}\Gamma(\boldsymbol{\theta}_0) \left\{\Gamma^T(\boldsymbol{\theta}_0)\Gamma(\boldsymbol{\theta}_0)\right\}^{-1},$$

$$\Gamma(\boldsymbol{\theta}_0) = E\{\frac{\partial \mathbf{g}(Y, X; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\},$$

$$\xi_{i,g}(\boldsymbol{\theta}_0) = \mathbf{g}(y_i, x_i; \boldsymbol{\theta}_0) + (1 - \delta_i) \left[\mu_{g|x}(x_i; \boldsymbol{\theta}_0) - \mathbf{g}(y_i, x_i; \boldsymbol{\theta}_0)\right]$$
$$+ \delta_i C_p h_g(y_i, x_i; \boldsymbol{\theta}_0)\mathbf{B}(x_i),$$

$$h_g(y_i, x_i; \boldsymbol{\theta}_0) = \int_{-\infty}^{+\infty} \int_0^1 \dot{\mathbf{g}}_y(q_\tau(x), x; \boldsymbol{\theta}_0)\mathbf{B}^T(x)\mathbf{H}_n^{-1}(\tau)\psi_\tau(e_i(\tau))d\tau dF_X(x),$$

$$(4.5)$$

$\dot{\mathbf{g}}_y(y, x; \boldsymbol{\theta}_0) = \partial \mathbf{g}(y, x; \boldsymbol{\theta}_0)/\partial y$, $\mu_{g|x}(x_i; \boldsymbol{\theta}_0) = E_{y|x}\{\mathbf{g}(y_i, x_i; \boldsymbol{\theta}_0)\}$, and the other notations are defined the same way as before.

The variance estimator for $Var\{\xi_{i,g}(\boldsymbol{\theta}_0)\}$ can be obtained by estimating $\xi_{i,g}(\boldsymbol{\theta}_0)$ similarly as estimating $\xi_{i,r}(\boldsymbol{\theta}_0)$ in Section 3.2. For example, see equation (3.6), where $\mathbf{r}(y_i, x_i)$ can be replaced by $\mathbf{g}(y_i, x_i; \widehat{\boldsymbol{\theta}}_{gmm})$ and $\dot{\mathbf{r}}_y(\hat{q}_{\tau_j}(x_k), x_k)$ can be replaced by $\dot{\mathbf{g}}_y(\hat{q}_{\tau_j}(x_k), x_k; \widehat{\boldsymbol{\theta}}_{gmm})$. The additional gradient matrix $\Gamma(\boldsymbol{\theta}_0)$ can be estimated by,

$$\hat{\Gamma}(\hat{\boldsymbol{\theta}}_{gmm}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i \frac{\partial \mathbf{g}(y_i, x_i; \hat{\boldsymbol{\theta}}_{gmm})}{\partial \boldsymbol{\theta}} + (1 - \delta_i) \frac{1}{J} \sum_{j=1}^{J} \frac{\partial \mathbf{g}(y_{ij}^*, x_i; \hat{\boldsymbol{\theta}}_{gmm})}{\partial \boldsymbol{\theta}} \right\}. \tag{4.6}$$

The proof of the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{gmm}$ requires additional regularity conditions regarding to the functional form of $\mathbf{g}(y_i, x_i; \boldsymbol{\theta})$, and is similar to the proof in Theorem 1. To conserve space, we defer the details of this proof to the supplement.

5. Simulation studies

The main objectives of the simulation studies are to compare SQRI to fully parametric and nonparametric alternatives, and to evaluate the performance of the variance estimator for SQRI. In addition, we also want to see if there are any differences in finite sample performances between the two SQRI estimators.

We specify the simulation set-up as follows. The response y_i is generated from a model $y_i = m(\mathbf{x}_i) + \epsilon_i$, where $m(\mathbf{x}_i)$ is the mean function and ϵ_i are *iid* $N(0, 0.1^2)$. We consider the following four different mean functions drawing from the design of simulation studies in Breidt, Opsomer, and Johnson (2007) to cover a range of correct and incorrect model specification.

$$\begin{aligned}
linear&: \quad m(x_i) = 1 + 2(x_i - 0.5), \\
bump&: \quad m(x_i) = 1 + 2(x_i - 0.5) + \exp\{-30(x_i - 0.5)^2\}, \\
cycle&: \quad m(x_i) = 0.5 + 2x_i + sin(3\pi x_i), \\
bivariate&: \quad m(x_{1i}, x_{2i}) = 1 + 2(x_{1i} - 0.5) + 2\exp\{-10(x_{2i} - 0.4)^2\}.
\end{aligned}$$

The covariate x_i for the first three univariate models (or x_{1i} and x_{2i} for the last *bivariate* model) are all independently and identically simulated from a truncated normal distribution $N(0.5, 0.3^2)$ on interval $[0, 1]$.

The response indicator, δ_i, is simulated from $Bernoulli(p_i)$, where

$$\begin{aligned}
logit(p_i) &= 1 + 0.5x_i & \text{for the } linear, bump, cycle \text{ models,} \\
\text{or} \quad logit(p_i) &= 0.2 + x_{1i} + 0.5x_{2i} & \text{for the } bivariate \text{ model.}
\end{aligned}$$
(5.1)

The missing rates in all scenarios are about 20%.

We are interested in estimating three parameters, the marginal mean of the response variable $\mu_y = E(Y)$, the marginal standard deviation of the response variable $\sigma_y = \sqrt{Var(Y)}$ and the correlation between the response and covariate variables $\rho = corr(X, Y)$. So $\boldsymbol{\theta} = (\mu_y, \sigma_y, \rho)$ and the corresponding estimation equation is defined as

$$\mathbf{g}(y_i, x_i; \mu_x, \mu_y, \sigma_x, \sigma_y, \rho) = \begin{pmatrix} x_i - \mu_x \\ y_i - \mu_y \\ (x_i - \mu_x)^2 - \sigma_x^2 \\ (y_i - \mu_y)^2 - \sigma_y^2 \\ (x_i - \mu_x)(y_i - \mu_y) - \rho\sigma_x\sigma_y \end{pmatrix}.$$

For *bivariate* model, $\boldsymbol{\theta} = (\mu_y, \sigma_y, \rho_1, \rho_2)$, where $\rho_1 = corr(X_1, Y)$ and $\rho_2 = corr(X_2, Y)$ and the estimation equation is defined in an analogous way. Note that μ_x and σ_x^2 are the mean and variance of the covariate and are treated as nuisance parameters.

For each model, 1000 Monte Carlo (MC) samples of size $n = 200$ are created. In order to focus on comparing different imputation methods, all of the following estimators are calculated using the GMM method once the missing values are filled in using the specified imputation approaches.

- **Full**: The GMM estimator using full observations.
- **Resp**: The GMM estimator using respondents only (where "Resp" comes from the word "respondents").
- **SQRI**: Our proposed estimator defined in (4.2), using random τ_j's from $Uniform[0, 1]$. Using the GMM estimator constructed with SQRI imputed values facilitates the joint estimation of μ_y, σ_y and ρ's, and their variance estimation.

- **SQRI-fixed**: The proposed estimator defined in (4.2), but using the fixed midpoints of J evenly spaced sub-intervals of $[0, 1]$ as τ_j in step 1 of the SQRI procedure.
- **MI**: The multiple imputation method proposed in Rubin (1987). The R package 'mi' by Gelman, Hill, Su, Ya, and Pittau (2013) is employed to obtain J multiple imputed data sets. Then the GMM estimator is calculated for each imputed data set, and the MI estimator is obatined by taking an average of the GMM estimators across multiple imputed data sets.
- **PFI**: The parametric fractional imputation method proposed in Kim (2011). Under PFI, multiple imputed values $y_{ij}^*(j = 1, \cdots, J)$ are generated from a proposed conditional density $\tilde{f}(y|x)$ and their associated fractional weights w_{ij}^* are computed using $\tilde{f}(y|x)$ and the assumed conditional density $f(y|x; \hat{\eta}^0)$, where $\hat{\eta}^0$ is the given initial value for η in the conditional density formula. By maximizing the score function of the density $f(y_i|x_i; \eta)$ using the imputed values and their weights, $\hat{\eta}$ is updated, and the fractional weights w_{ij}^* are re-calculated iteratively until $\hat{\eta}$ converges. The PFI estimator is calculated using the GMM method, with the estimation equation for a missing unit i replaced by $\sum_{j=1}^J w_{ij}^* \mathbf{g}(y_{ij}^*, x_i; \boldsymbol{\theta})$.
- **NPI**: The non-parametric imputation method proposed in Wang and Chen (2009). In NPI, multiple imputed values $y_{ij}^*(j = 1, \cdots, J)$ are independently drawn from the respondent group ($\delta_i = 1$) with the probability of selecting y_s with $\delta_s = 1$ being

$$P(y_{ij}^* = y_s) = \frac{K\{(x_s - x_i)/h\}}{\sum_{m=1}^n \delta_m K\{(x_m - x_i)/h\}},$$

  where $K(\cdot)$ is a d_x-dimensional kernel function and h is a smoothing bandwidth. In our simulations, the Gaussian kernel is used with h prescribed by a cross-validation method. The NPI estimator is obtained using the GMM method with the estimation equation for a missing unit i replaced by $J^{-1} \sum_{j=1}^J \mathbf{g}(y_{ij}^*, x_i; \boldsymbol{\theta})$.
- **HDFI**: A hot-deck fractional imputation estimator. Under HDFI, multiple imputed values $y_{ij}^*(j = 1, \cdots, J)$ are independently drawn from a donor pool consisting of 20 nearest neighbors identified through the Euclidean distance. The HDFI estimator is calculated using the GMM method with the estimation equation for a missing unit i replaced by $J^{-1} \sum_{j=1}^J \mathbf{g}(y_{ij}^*, x_i; \boldsymbol{\theta})$.

The Full and the Resp estimators are included in order to help us gauge how far away our proposed estimators are from the ideal case and from the case of simply ignoring missing. Estimators NPI and HDFI are based on nonparametric imputation methods, while estimators MI and PFI are based on parametric imputation methods, where y_i is assumed to satisfy $Y|\mathbf{X} = \mathbf{x} \sim N(\boldsymbol{\beta}^T \mathbf{x}, \sigma^2)$ for some $\sigma > 0$. Our SQRI estimators are semiparametric as we use penalized B-spline to estimate conditional quantile regression. For penalized B-spline quantile estimators, typically the degree of B-spline p and the degree of the difference matrix m are fixed at low values, for example $p \leq 3$ and $m \leq 2$. We set

$p = 3$ and $m = 2$, a popular choice in practice as suggested in Yoshida (2013). For a given K_n (where $K_n =$ number of knots $+ 1$), the smoothing parameter λ_n is prescribed via the generalized approximation cross-validation (GACV) method discussed by Yuan (2006). We obtain results for a variety of choices of K_n and conclude $K_n = 5$ suffices in our examples. In the *bivariate* model, the same specifications are used to obtain bases $\mathbf{B}(x_1)$ and $\mathbf{B}(x_2)$ on x_1 and x_2 respectively, then $\mathbf{B}(\mathbf{x})$ is defined as, $\mathbf{B}(\mathbf{x}) = (\mathbf{B}^T(x_1), \mathbf{B}^T(x_2))^T$.

We discuss how to choose the number of imputed values, J, practically. Like all other imputation methods, the theoretical proofs of the SQRI estimators require J to go to infinity to ensure the consistency. But a finite number (J) imputation is necessary in practice. A midpoint approximation approach used for SQRI-fixed estimator sheds some light on our choice of J in practice. By the property of (3.10), we see that $\widehat{\boldsymbol{\theta}}_{fixed}$ and $\widehat{\boldsymbol{\theta}}$ have the same asymptotic equivalence when $J = O(n^\alpha)$ for $\alpha > 1/2$. We can use an integer which is close to $n^{1/2}$ as an initial choice of J. For example, in our simulation studies, $n^{1/2}$ with the sample size $n = 200$ equals 14.14. So $J = 10$ seems to be a reasonable choice and is used in the imputation. The simulation results with $J = 100$ are also provided in the paper for comparisons.

Tables 1–4 present the relative MC biases and the MC variances of all estimators with both $J = 10$ and $J = 100$ for the four models respectively. The relative MC bias is defined as the ratio of the MC bias to the true parameter value. The relative biases of the proposed SQRI estimators are less than 1% in all cases (Tables 1–4). In general, the proposed SQRI estimators have smaller relative biases and variances as compared with the Resp estimator because the SQRI estimators incorporate additional covariate information of the missing units while the Resp estimator totally ignores missing units.

We compare the SQRI estimators with the two parametric imputation alternatives, MI and PFI. When the *linear* model is correctly specified, the SQRI estimators have relative biases comparable to those in the MI and PFI estimators. When the model is misspecified (*bump, cycle, bivariate*), the SQRI estimators generally have significantly smaller biases than the MI and PFI estimators. The reduction in biases of the SQRI estimators in the *bump, cycle*, and *bivariate* models, relative to the MI and PFI, results because the SQRI procedure relies on fewer model assumptions. In terms of variances (Tables 1–4), the proposed SQRI estimators have variances comparable to those in the two parametric ones under the correct *linear* model, but have slightly smaller variances under the incorrect models.

We compare the SQRI estimators with the two nonparametric imputation alternatives, NPI and HDFI. The SQRI estimators have considerably smaller biases than NPI and HDFI estimators in most of the cases. The variances of the SQRI estimators are generally in line with those of the NPI and HDFI estimators (Table 1–4).

When comparing the SQRI estimator using random τ's with the SQRI-fixed estimator using fixed τ's, we don't see any consistent superiority of one over another in relative biases. And the variances of the two SQRI estimators are similar. The simulation studies show that $J = 10$ is sufficient for the proposed

TABLE 1

The relative MC biases and MC variances of all estimators for the linear model:
$m(x) = 1 + 2(x - 0.5)$. *The number of replicates in the Monte Carlo is 1000 and the sample*
size is 200. J is the number of imputed values. The table entries are the MC results
multiplied by 100.

		μ_y		σ_y		ρ	
		RBias ($\times 100$)	Var ($\times 100$)	RBias ($\times 100$)	Var ($\times 100$)	RBias ($\times 100$)	Var ($\times 100$)
	Full	-0.109	0.119	-0.137	0.038	-0.013	0.001
	Resp	2.961	0.166	-0.563	0.054	-0.032	0.001
	SQRI	-0.123	0.120	-0.556	0.041	0.005	0.001
	SQRI-fixed	-0.110	0.120	-0.383	0.041	-0.047	0.001
J=10	MI	-0.119	0.121	-0.414	0.040	-0.035	0.001
	PFI	-0.116	0.120	-0.426	0.040	-0.036	0.001
	NPI	-0.032	0.122	-0.758	0.041	-0.340	0.001
	HDFI	0.181	0.126	-2.044	0.045	-0.670	0.002
	SQRI	-0.123	0.121	-0.398	0.040	-0.043	0.001
	SQRI-fixed	-0.171	0.127	-0.370	0.041	-0.046	0.001
J=100	MI	-0.116	0.121	-0.423	0.040	-0.036	0.001
	PFI	-0.118	0.120	-0.422	0.040	-0.037	0.001
	NPI	-0.032	0.121	-0.738	0.041	-0.344	0.001
	HDFI	0.173	0.125	-2.044	0.045	-0.680	0.002

TABLE 2

The relative MC biases and MC variances of all estimators for the bump model:
$m(x) = 1 + 2(x - 0.5) + \exp\{-30(x - 0.5)^2\}$. *The number of replicates in the Monte Carlo is*
1000 and the sample size is 200. J is the number of imputed values. The table entries are
the MC results multiplied by 100.

		μ_y		σ_y		ρ	
		RBias ($\times 100$)	Var ($\times 100$)	RBias ($\times 100$)	Var ($\times 100$)	RBias ($\times 100$)	Var ($\times 100$)
	Full	0.101	0.184	-0.158	0.075	-0.128	0.054
	Resp	7.480	0.235	-3.338	0.117	-2.149	0.091
	SQRI	0.025	0.184	-0.365	0.082	-0.080	0.057
	SQRI-fixed	0.048	0.184	-0.212	0.082	-0.233	0.059
J=10	MI	1.072	0.215	-3.277	0.102	-1.980	0.080
	PFI	1.023	0.213	-3.255	0.102	-1.871	0.077
	NPI	0.535	0.187	-0.759	0.080	-0.705	0.058
	HDFI	0.557	0.188	-1.066	0.084	-0.533	0.058
	SQRI	0.017	0.184	-0.229	0.082	-0.212	0.058
	SQRI-fixed	0.016	0.192	-0.207	0.084	-0.219	0.059
J=100	MI	1.050	0.210	-3.237	0.100	-1.968	0.078
	PFI	1.008	0.212	-3.278	0.100	-1.885	0.077
	NPI	0.546	0.187	-0.748	0.080	-0.722	0.057
	HDFI	0.577	0.189	-1.089	0.083	-0.526	0.058

SQRI estimators to accurately estimate parameters in these simulation examples
when $n = 200$.

We evaluate the performance of variance estimation for both of the SQRI
estimators using random τ's and fixed τ's. Tables 5 and 6 contain the coverage
probabilities and the half interval widths of the 95% C.I.'s using the two SQRI
estimators for both $J = 10$ and $J = 100$. The 95% C.I.'s are calculated based on
the asymptotic normality and a bootstrapping method. For the SQRI estimator

TABLE 3

The relative MC biases and MC variances of all estimators for the cycle model:
$m(x) = 0.5 + 2x + \sin(3\pi x)$. *The number of replicates in the Monte Carlo is 1000 and the sample size is 200. J is the number of imputed values. The table entries are the MC results multiplied by 100.*

		μ_y		σ_y		ρ	
		RBias ($\times100$)	Var ($\times100$)	RBias ($\times100$)	Var ($\times100$)	RBias ($\times100$)	Var ($\times100$)
	Full	-0.026	0.370	-0.041	0.106	-0.156	0.206
	Resp	1.883	0.506	2.095	0.142	4.670	0.249
	SQRI	-0.018	0.369	-0.331	0.109	-0.160	0.210
	SQRI-fixed	-0.015	0.369	-0.275	0.109	-0.219	0.211
J=10	MI	-0.257	0.453	1.989	0.155	4.665	0.263
	PFI	-0.222	0.455	1.905	0.151	4.814	0.260
	NPI	-0.013	0.379	-0.312	0.115	-0.824	0.223
	HDFI	0.217	0.375	-0.322	0.114	-1.629	0.229
	SQRI	-0.017	0.368	-0.280	0.109	-0.212	0.211
	SQRI-fixed	-0.018	0.376	-0.282	0.110	-0.204	0.211
J=100	MI	-0.256	0.450	2.041	0.150	4.679	0.259
	PFI	-0.249	0.445	1.935	0.151	4.842	0.259
	NPI	-0.013	0.377	-0.317	0.115	-0.847	0.223
	HDFI	0.226	0.376	-0.328	0.114	-1.644	0.227

TABLE 4

The relative MC biases and MC variances of all estimators for the bivariate model:
$m(x) = 1 + 2(x_1 - 0.5) + 2\exp\{-10(x_2 - 0.4)^2\}$. *The number of replicates in the Monte Carlo is 1000 and the sample size is 200. J is the number of imputed values. The table entries are the MC results multiplied by 100.*

		μ_y		σ_y		ρ_1		ρ_2	
		RBias ($\times100$)	Var ($\times100$)	RBias ($\times100$)	Var ($\times100$)	RBias ($\times100$)	Var ($\times100$)	RBias ($\times100$)	Var ($\times100$)
	Full	0.067	0.295	-0.001	0.119	-0.063	0.173	0.682	0.455
	Resp	0.839	0.394	0.250	0.149	-0.532	0.217	4.910	0.565
	SQRI	0.064	0.297	-0.214	0.121	-0.038	0.175	0.653	0.456
	SQRI-fixed	0.069	0.296	-0.259	0.121	-0.007	0.174	0.719	0.457
J=10	MI	0.124	0.340	0.103	0.138	-0.705	0.206	4.393	0.520
	PFI	0.125	0.340	0.029	0.138	-0.511	0.209	4.629	0.518
	NPI	0.477	0.309	-1.010	0.131	-2.132	0.191	1.907	0.476
	HDFI	0.715	0.313	-1.538	0.134	-2.267	0.197	3.926	0.491
	SQRI	0.073	0.296	-0.215	0.121	-0.051	0.174	0.668	0.456
	SQRI-fixed	0.069	0.296	-0.218	0.121	-0.045	0.174	0.676	0.456
J=100	MI	0.122	0.333	0.151	0.136	-0.668	0.204	4.355	0.514
	PFI	0.126	0.336	0.009	0.137	-0.519	0.204	4.568	0.515
	NPI	0.479	0.308	-0.988	0.129	-2.129	0.189	1.899	0.470
	HDFI	0.718	0.312	-1.551	0.134	-2.269	0.197	3.975	0.488

using random τ's, the coverage probabilities based on normality are close to the nominal level of 0.95 except for ρ under the *linear* and *cycle* models, and (σ_y, ρ_1) under the *bivariate* model. The normality approximation of the SQRI estimator has undercoverage for ρ in the *linear* and *cycle* models and for σ_y in the *bivariate* model, but overestimates the coverage for ρ_1 in the *bivariate* model. When using the SQRI-fixed estimator, the coverage probabilities based on normality slightly improve in general, though the coverage probabilities for

TABLE 5

The coverage probabilities and the half interval widths of the 95% C.I. of the SQRI estimators under the three univariate models. The number of replicates in the Monte Carlo is 1000 and the sample size is 200. J is the number of imputed values.

(a). Model *linear*: $m(x) = 1 + 2(x - 0.5)$

C.I. Coverage	Probabilities	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.934	0.937	0.817	0.931	0.938	0.856
	Bootstrapping	0.928	0.953	0.933	0.932	0.953	0.967
SQRI-fixed	Normality	0.949	0.930	0.853	0.951	0.930	0.851
	Bootstrapping	0.951	0.947	0.916	0.949	0.947	0.925

C.I. Half	Interval Widths	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.067	0.038	0.005	0.067	0.038	0.005
	Bootstrapping	0.069	0.045	0.015	0.067	0.044	0.015
SQRI-fixed	Normality	0.067	0.038	0.005	0.067	0.038	0.005
	Bootstrapping	0.067	0.043	0.013	0.067	0.043	0.014

(b). Model *bump*: $m(x) = 1 + 2(x - 0.5) + \exp\{-30(x - 0.5)^2\}$

C.I. Coverage	Probabilities	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.930	0.940	0.940	0.947	0.942	0.942
	Bootstrapping	0.944	0.949	0.950	0.937	0.946	0.949
SQRI-fixed	Normality	0.953	0.938	0.939	0.953	0.938	0.941
	Bootstrapping	0.949	0.947	0.942	0.948	0.950	0.939

C.I. Half	Interval Widths	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.084	0.054	0.045	0.084	0.054	0.044
	Bootstrapping	0.085	0.062	0.050	0.084	0.062	0.049
SQRI-fixed	Normality	0.084	0.054	0.044	0.084	0.054	0.044
	Bootstrapping	0.084	0.062	0.049	0.084	0.063	0.049

(c). Model *cycle*: $m(x) = 0.5 + 2x + sin(3\pi x)$

C.I. Coverage	Probabilities	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.944	0.939	0.913	0.943	0.941	0.915
	Bootstrapping	0.943	0.947	0.941	0.932	0.947	0.943
SQRI-fixed	Normality	0.943	0.939	0.925	0.944	0.938	0.925
	Bootstrapping	0.940	0.933	0.941	0.939	0.930	0.943

C.I. Half	Interval Widths	J=10			J=100		
		μ_y	σ_y	ρ	μ_y	σ_y	ρ
SQRI	Normality	0.119	0.063	0.081	0.119	0.063	0.081
	Bootstrapping	0.118	0.065	0.088	0.117	0.065	0.088
SQRI-fixed	Normality	0.119	0.063	0.080	0.119	0.063	0.080
	Bootstrapping	0.118	0.065	0.088	0.118	0.065	0.088

TABLE 6
The coverage probabilities and the half interval widths of the 95% C.I. of the SQRI estimators under the bivariate model: $m(x) = 1 + 2(x_1 - 0.5) + 2\exp\{-10(x_2 - 0.4)^2\}$. *The number of replicates in the Monte Carlo is 1000 and the sample size is 200. J is the number of imputed values.*

C.I. Coverage	Probabilities	J=10				J=100			
		μ_y	σ_y	ρ_1	ρ_2	μ_y	σ_y	ρ_1	ρ_2
SQRI	Normality	0.953	0.923	0.972	0.939	0.953	0.928	0.977	0.942
	Bootstrapping	0.953	0.945	0.963	0.950	0.948	0.944	0.958	0.947
SQRI-fixed	Normality	0.942	0.939	0.945	0.950	0.943	0.940	0.952	0.950
	Bootstrapping	0.936	0.950	0.950	0.947	0.938	0.951	0.948	0.946

C.I. Half	Interval Widths	J=10				J=100			
		μ_y	σ_y	ρ_1	ρ_2	μ_y	σ_y	ρ_1	ρ_2
SQRI	Normality	0.071	0.041	0.028	0.149	0.071	0.041	0.027	0.148
	Bootstrapping	0.071	0.048	0.024	0.138	0.071	0.047	0.024	0.138
SQRI-fixed	Normality	0.071	0.041	0.027	0.148	0.071	0.041	0.027	0.148
	Bootstrapping	0.070	0.047	0.024	0.138	0.070	0.047	0.024	0.138

ρ in the *linear* and *cycle* models are still low (about 85.3% and 92.5%). The confidence interval lengths of the two SQRI estimators based on normality are comparable. A bootstrapping method then is conducted as a remedy to obtain the confidence intervals. The bootstrapping algorithm is described as follows.

1. Draw a simple random sample χ_n^* with replacement from the original sample $\chi_n = (\mathbf{x}_i, y_i, \delta_i)_{i=1}^n$;
2. Implement semiparametric quantile regression to impute values for the missing cells in χ_n^*;
3. Estimate $\hat{\boldsymbol{\theta}}$ using the SQRI and SQRI-fixed estimators in equation (4.2).
4. Repeat step $1 \sim 3$ for B times, then we have $\hat{\boldsymbol{\theta}}^{[1]}, \hat{\boldsymbol{\theta}}^{[2]}, \cdots, \hat{\boldsymbol{\theta}}^{[B]}$.

The 2.5-th and 97.5-th percentiles of $\{\hat{\boldsymbol{\theta}}^{[b]}\}_{b=1}^B$ give the lower and upper bounds of the 95% confidence interval. We use $B = 400$ in our simulation. In general, the bootstrapping method has a slightly better performance over normal approximation method, offering satisfactory coverage probabilities close to 0.95 even when J is small.

In summary, the simulation study results confirm that the SQRI estimators can be used as effective alternatives in imputation. Both SQRI and SQRI-fixed estimators perform similarly in point estimates and variance estimation. However, if one has a concern about having extreme imputed values, the SQRI-fixed estimator might be preferable because using fixed τ's can prevent it from happening.

6. Future work

There will be some future work along this line. In this paper, we propose quantile regression imputation method for missing response cases. Wei, Ma, and Caroll

(2012) used multiple imputation with the goal of estimating the quantile regression coefficients of y on $\mathbf{x}$ when covariates are subject to missing. Combining their method with ours, we can propose a quantile regression imputation method to handle data with both missing covariates and missing responses. Another limitation of the current approach is that we focus on low dimensional covariates. It is important to extend the proposed method to high dimensional covariates. Model selection for high dimensional data has been intensively studied in the area of quantile regression. One of the popular model selection methods is the Least Absolute Shrinkage and Selection Operator (LASSO) method. We can use LASSO quantile regression to reduce the dimension first, then a conditional quantile regression can be estimated based on the selected model. We will investigate on these topics in our future work.

Appendix

The notation of $|\cdot|$ represents the norm of a matrix, defined as $|A| = \sqrt{trace(A'A)}$. We first give the assumptions.

Conditions:

(a) There exists $\gamma > 0$ such that $E[|y|^{2+\gamma}] < \infty$.

(b) The order of the difference matrix $\mathbf{D}_m$ is $m < p$.

(c) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} p(x_i) B(x_i) B^T(x_i)$ exists and converges to $\mathbf{\Phi}$ where $\mathbf{\Phi} = \int_0^1 \mathbf{B}(x)\mathbf{B}^T(x)p(x)dF_x(x)$.

(d) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} p(x_i) f_{y|x}(x, q_\tau(x)) \ \mathbf{B}(x_i)\mathbf{B}^T(x_i)$ exists and converges to $\mathbf{\Phi}(\tau)$, where $\mathbf{\Phi}(\tau) = \int_0^1 \mathbf{B}(x)\mathbf{B}^T(x)p(x) \ f_{y|x}(x, q_\tau(x))dF_x(x)$.

(f) $\mathbf{r}(y, x)$ has a bounded second derivative function with respect to y.

A. Proof of Lemma 1

Assume the number of knots $K_n - 1$ and the smoothing parameter λ_n depend on n. By Barrow and Smith (1978), there exists $\mathbf{b}^*(\tau)$ that satisfies

$$\sup_{x\in(0,1)} |q_\tau(x) + b_\tau^a(x) - \mathbf{B}^T(x)\mathbf{b}^*(\tau)| = o(K_n^{-(p+1)}), \tag{A.1}$$

where $b_\tau^a(x)$ is an approximation error defined as

$$b_\tau^a(x) = \frac{q_\tau^{(p+1)}(x)}{(p+1)!K_n^{p+1}} Br_p(\frac{x - \kappa_{k-1}}{K_n^{-1}}) \text{ if } \kappa_{k-1} \le x < \kappa_k,$$

and $q_\tau^{(p+1)}(x)$ is the $(p+1)$-th derivative of $q_\tau(x)$ with respect to x. Here $Br_p(\cdot)$ is the p-th Bernoulli polynomial inductively defined as $Br_0(x) = 1$, and $Br_p(x) = \int_0^x pB_{p-1}(z)dz + b_p$, where $b_p = -p\int_0^1\int_0^x Br_{p-1}(z)dzdx$ is the p-th Bernoulli number (Barrow and Smith (1978) and Yoshida (2013)).

Proof: The following proof needs the Knight's identity,

$$\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v I(u \leq s) - I(u \leq 0)ds, \tag{A.2}$$

where $\psi_\tau(x) = \tau - I(x < 0)$.

Define

$$U_n(\boldsymbol{\Delta}) = \sum_{i=1}^n \delta_i[\rho_\tau(e_i(\tau) - \sqrt{\frac{K_n}{n}}\mathbf{B}^T(x_i)\boldsymbol{\Delta}) - \rho_\tau(e_i(\tau))]$$

$$+ \frac{\lambda_n}{2}\left[\mathbf{b}^*(\tau) + \sqrt{\frac{K_n}{n}}\boldsymbol{\Delta}\right]^T \mathbf{D}_m^T\mathbf{D}_m \left[\mathbf{b}^*(\tau) + \sqrt{\frac{K_n}{n}}\boldsymbol{\Delta}\right]$$

$$- \frac{\lambda_n}{2}\mathbf{b}^*(\tau)^T\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}^*(\tau), \tag{A.3}$$

where $e_i(\tau) = y_i - \mathbf{B}^T(x_i)\mathbf{b}^*(\tau)$. It is easy to see that $\hat{\boldsymbol{\Delta}}_n(\tau) = \sqrt{\frac{n}{K_n}}\left[\hat{\mathbf{b}}(\tau) - \mathbf{b}^*(\tau)\right]$ is the minimizer of $U_n(\boldsymbol{\Delta})$. $U_n(\boldsymbol{\Delta})$ can be decomposed as

$$U_n(\boldsymbol{\Delta}) = \left\{-\sum_{i=1}^n \delta_i\sqrt{\frac{K_n}{n}}\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau)) + \lambda_n\sqrt{\frac{K_n}{n}}\mathbf{b}^*(\tau)^T\mathbf{D}_m^T\mathbf{D}_m\right\}\boldsymbol{\Delta}$$

$$+ \sum_{i=1}^n \delta_i \int_0^{\sqrt{\frac{K_n}{n}}\mathbf{B}^T(x_i)\boldsymbol{\Delta}} [I(e_i(\tau) \leq s) - I(e_i(\tau) < 0)]ds$$

$$+ \frac{\lambda_n}{2}\frac{K_n}{n}\boldsymbol{\Delta}^T\mathbf{D}_m^T\mathbf{D}_m\boldsymbol{\Delta}. \tag{A.4}$$

Similar to Lemma 3 in Yoshida (2013), we can show that

$$\sum_{i=1}^n \delta_i \int_0^{\sqrt{\frac{K_n}{n}}\mathbf{B}^T(x_i)\boldsymbol{\Delta}} [I(e_i(\tau) \leq s) - I(e_i(\tau) < 0)]ds$$

$$= \frac{1}{2}K_n\boldsymbol{\Delta}^T\boldsymbol{\Phi}(\tau)\boldsymbol{\Delta}[1 + O_p(K_n^{-(p+1)})], \tag{A.5}$$

where $\boldsymbol{\Phi}(\tau) = \int_0^1 p(x_i)\mathbf{B}^T(x_i)\mathbf{B}(x_i)f_{Y|X}(q_\tau(x_i)|x_i)dF_x(x_i)$. Thus, by equation (A.4) and (A.5), we have

$$U_n(\boldsymbol{\Delta}) = \left\{-\sum_{i=1}^n \delta_i\sqrt{\frac{K_n}{n}}\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau)) + \lambda_n\sqrt{\frac{K_n}{n}}\mathbf{b}^*(\tau)^T\mathbf{D}_m^T\mathbf{D}_m\right\}\boldsymbol{\Delta}$$

$$+ \frac{1}{2}\boldsymbol{\Delta}^T K_n\left[\frac{\lambda_n}{n}\mathbf{D}_m^T\mathbf{D}_m + \boldsymbol{\Phi}(\tau)(1 + O_p(K_n^{-(p+1)}))\right]\boldsymbol{\Delta}. \tag{A.6}$$

We can see that $U_n(\boldsymbol{\Delta})$ is convex with respect to $\boldsymbol{\Delta}$, so it has a unique minimizer $\hat{\boldsymbol{\Delta}}_n(\tau)$ as

$$\hat{\boldsymbol{\Delta}}_n(\tau) = (\mathbf{H}_n(\tau) + \boldsymbol{\Phi}(\tau)O_p(K_n^{-(p+1)}))^{-1}\left\{\sum_{i=1}^n \delta_i \sqrt{\frac{1}{nK_n}}\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau))\right.$$

$$\left. - \lambda_n\sqrt{\frac{1}{nK_n}}\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}^*(\tau)\right\}, \tag{A.7}$$

where $\mathbf{H}_n(\tau) = \frac{\lambda_n}{n}\mathbf{D}_m^T\mathbf{D}_m + \boldsymbol{\Phi}(\tau)$. Therefore, we have

$$\sqrt{\frac{n}{K_n}}\left\{\hat{\mathbf{b}}(\tau) - \mathbf{b}^*(\tau) + \mathbf{H}_n^{-1}(\tau)\left[\frac{\lambda_n}{n}\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}^*(\tau)\right]\right\}$$
$$= \mathbf{H}_n^{-1}(\tau)\left\{\sqrt{\frac{1}{K_n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau))\right]\right\}$$
$$\quad - \sqrt{\frac{1}{K_n}}H_n^{-1}(\tau)\boldsymbol{\Phi}(\tau)\{I + \frac{\lambda_n}{\sqrt{n}}\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}^*(\tau)\}H_n^{-1}(\tau)O_p(K_n^{-(p+1)})$$
$$= \mathbf{H}_n^{-1}(\tau)\left\{\sqrt{\frac{1}{K_n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau))\right]\right\} + o_p(1). \tag{A.8}$$

The last equality holds in (A.8) because of Fact 1 and 2 in the Appendix of the supplement. We can show that

$$W_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\mathbf{B}^T(x_i)\psi_\tau(e_i(\tau)) \sim_d N(0, V_w),$$

where $V_w = \tau(1-\tau)\boldsymbol{\Phi}$ and $\boldsymbol{\Phi} = \int p(x_i)\mathbf{B}^T(x_i)\mathbf{B}(x_i)dF_X(x_i)$.

Let $\hat{q}_\tau(x) = \mathbf{B}^T(x)\hat{\mathbf{b}}(\tau)$ and $b_\tau^\lambda(x) = \frac{\lambda_n}{n}\mathbf{H}_n(\tau)^{-1}\mathbf{B}^T(x)\mathbf{D}_m^T\mathbf{D}_m\mathbf{b}^*(\tau)$, so

$$\sqrt{\frac{n}{K_n}}\left[\hat{q}_\tau(x) - \mathbf{B}^T(x)\mathbf{b}^*(\tau) + b_\tau^\lambda(x)\right] = \mathbf{B}^T(x)\mathbf{H}_n^{-1}(\tau)\sqrt{\frac{1}{K_n}}W_n^T. \tag{A.9}$$

By equation (A.1), we have

$$\sqrt{\frac{n}{K_n}}[\hat{q}_\tau(x) - q_\tau(x) + b_\tau^\lambda(x) + b_\tau^a(x)] \to_p \mathbf{B}^T(x)\mathbf{H}_n^{-1}(\tau)\sqrt{\frac{1}{K_n}}W_n^T. \tag{A.10}$$

The results in (i) and (ii) of Lemma 1 follow immediately.

B. Proof of Theorem 1

To save space, we sketch the proof of Theorem 1. For details, readers are referred to the supplement which provides the detailed derivation for the asymptotic normality of the GMM estimator $\widehat{\boldsymbol{\theta}}_{gmm}$ defined in Section 4. Note that the estimator $\widehat{\boldsymbol{\theta}}$ in Theorem 1 is a special case of the GMM estimator when defining $\mathbf{g}(y_i, x_i; \boldsymbol{\theta}) = \mathbf{r}(y_i, x_i) - \boldsymbol{\theta}$. So the proofs are very similar.

Proof: We can decompose $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ into three terms as follows,

$$
\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad = \quad & \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{r}(y_i, x_i) - \boldsymbol{\theta}_0]}_{:=B_1} \\
& + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [(1 - \delta_i)(\mu_{r|x}(x_i) - \mathbf{r}(y_i, x_i)]}_{:=B_2} \\
& + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [(1 - \delta_i)(\hat{\mu}_{r|x}(x_i) - \mu_{r|x}(x_i)]}_{:=B_3}.
\end{aligned}
\tag{B.1}
$$

Terms of B_1 and B_2 are easy since they are sums of i.i.d. random variables. The key idea in our proof is to replace B_3 by $\tilde{B}_3 = E(B_3 | A_R)$ where $A_R = \{\delta_i, (y_i, x_i) | \delta_i = 1; i = 1, \cdots, n\}$, and to show the following two results:

(1) $\tilde{B}_3 = n^{-1/2} \sum_{i=1}^{n} \delta_i C_p h_r(x_i, y_i) \mathbf{B}(x_i) + o_p(1)$, and

(2) $\tilde{B}_3 - B_3 = o_p(1)$.

(1) To show $\tilde{B}_3 = n^{-1/2} \sum_{i=1}^{n} \delta_i C_p h_r(x_i, y_i) \mathbf{B}(x_i) + o_p(1)$: We further decompose $\tilde{B}_3$ into two terms,

$$
\begin{aligned}
\tilde{B}_3 \quad = \quad & \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left\{ (1 - \delta_i) \frac{1}{J} \sum_{j=1}^{J} [\mathbf{r}(\hat{q}_{\tau_j}(x_i), x_i) - \mathbf{r}(q_{\tau_j}(x_i), x_i)] | A_R \right\}}_{:=\tilde{B}_{31}} \\
& + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left\{ (1 - \delta_i) \frac{1}{J} \sum_{j=1}^{J} [\mathbf{r}(q_{\tau_j}(x_i), x_i) - \mu_{r|x}(x_i)] | A_R \right\}}_{:=\tilde{B}_{32}}.
\end{aligned}
\tag{B.2}
$$

It is obvious that $\tilde{B}_{32} = 0$ because $E_{\tau|x}[\mathbf{r}(q_\tau(x), x)] = E_{y|x}[\mathbf{r}(y, x)] = \mu_{r|x}(x)$ for any x. For $\tilde{B}_{31}$, assuming that $\mathbf{r}(x_i, y_i)$ is twice differentiable with respect to y, then we have

$$
\begin{aligned}
& \mathbf{r}(\hat{q}_\tau(x_i), x_i) - \mathbf{r}(q_\tau(x_i), x_i) \\
= \quad & \dot{\mathbf{r}}_y(q_\tau(x_i), x_i)[\hat{q}_\tau(x_i) - q_\tau(x_i)] + \ddot{\mathbf{r}}_y(\tilde{q}_\tau(x_i), x_i)[\hat{q}_\tau(x_i) - q_\tau(x_i)]^2,
\end{aligned}
\tag{B.3}
$$

for $\tilde{q}_{\tau_j}(x_i)$ lying between $q_{\tau_j}(x_i)$ and $\hat{q}_{\tau_j}(x_i)$. By equation (B.3), we have

$$\tilde{B}_{31} \quad = \quad \frac{n_m}{\sqrt{n}} E\left\{ \frac{1}{J} \sum_{j=1}^{J} \dot{\mathbf{r}}_y(q_{\tau_j}(x), x)[\hat{q}_{\tau_j}(x) - q_{\tau_j}(x)]|A_R \right\}$$

$$+ \frac{n_m}{\sqrt{n}} E\left\{ \frac{1}{J} \sum_{j=1}^{J} \ddot{\mathbf{r}}_y(\tilde{q}_{\tau_j}(x), x)[\hat{q}_{\tau_j}(x) - q_{\tau_j}(x)]^2|A_R \right\},$$

where $n_m = n - \sum_{i=1}^{n} \delta_i$ and $x \perp A_R$. By Fact 3 of the supplement, we have

$$E\left\{ \frac{1}{J} \sum_{j=1}^{J} \ddot{\mathbf{r}}_y(\tilde{q}_{\tau_j}(x), x)[\hat{q}_{\tau_j}(x) - q_{\tau_j}(x)]^2|A_R \right\} = O(\frac{K_n}{n}). \qquad \text{(B.4)}$$

By Lemma 1, we have

$$\sqrt{n}\,(\hat{q}_\tau(x) - q_\tau(x)) \quad = \quad \frac{1}{\sqrt{n}} \mathbf{B}^T(x) \mathbf{H}_n^{-1}(\tau) \sum_{i=1}^{n} \delta_i \mathbf{B}(x_i) \psi_\tau(e_i(\tau))$$
$$- \frac{\lambda_n}{\sqrt{n}} \mathbf{B}^{(}x) \mathbf{H}_n^{-1}(\tau) C_n(\tau) - \sqrt{n} b_\tau^a(x) + o_p(1), \qquad \text{(B.5)}$$

where $C_n(\tau) = \mathbf{D}_m^T \mathbf{D}_m \mathbf{b}^*(\tau)$. Then $\tilde{B}_{31}$ can be written as

$$\tilde{B}_{31} = \frac{C_p}{\sqrt{n}} \sum_{i=1}^{n} \delta_i h_r(x_i, y_i) \mathbf{B}(x_i) - \sqrt{n} C_p C_{1n} - \sqrt{n} C_p C_{2n} + o_p(1), \qquad \text{(B.6)}$$

where $h_r(x_i, y_i)$ and C_p are defined in Theorem 1. The asymptotic orders of C_{1n} and C_{2n} are

$$C_{1n} = \frac{\lambda_n}{n} E_{x,\tau} \left\{ \dot{\mathbf{r}}_y(q_\tau(x), x) \mathbf{B}^T(x) \mathbf{H}_n^{-1}(\tau) C_n(\tau) \right\} = O(K_n^{-(p+2)}),$$

by Fact 2 of the supplement, and

$$C_{2n} = E[\dot{\mathbf{r}}_y(q_\tau(x), x) \mathbf{b}_\tau^a(x)] = O(K_n^{-(p+2)}).$$

Thus we have $\tilde{B}_{31} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i C_p h_r(x_i, y_i) \mathbf{B}(x_i) + o_p(1)$.

(2) To show $\tilde{B}_3 - B_3 = o_p(1)$: By Chebychev's inequality, we only need to show that $E[\tilde{B}_3 - B_3]^{\otimes 2} \leq E\{[\hat{\mu}_{r|x}(x_i) - \mu_{r|x}(x_i)]^{\otimes 2}\} = o(1)$. First of all, we can decompose $\hat{\mu}_{r|x}(x_i) - \mu_{r|x}(x_i)$ into two terms,

$$\hat{\mu}_{r|x}(x_i) - \mu_{r|x}(x_i)$$

$$= \quad \underbrace{\frac{1}{J} \sum_{j=1}^{J} [\mathbf{r}(\hat{q}_{\tau_j}(x), x) - \mathbf{r}(q_{\tau_j}(x), x)]}_{:=S_n} + \underbrace{\frac{1}{J} \sum_{j=1}^{J} [\mathbf{r}(q_{\tau_j}(x), x) - \mu_{r|x}(x_i)]}_{:=Q_n}.$$

It is equivalent to show that $E[Q_n^{\otimes 2}] = o(1)$, $E[Q_n S_n^T] = o(1)$ and $E[S_n^{\otimes 2}] = o(1)$. Details to show these orders can be found in the supplement.

Combining step (1) & (2), together with equation (B.1), we can write

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{r}(y_i, x_i) - \boldsymbol{\theta}_0] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [(1 - \delta_i)(\mu_{r|x}(x_i) - \mathbf{r}(y_i, x_i)]$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i C_p h_r(x_i, y_i) \mathbf{B}(x_i) + o_p(1),$$

$$= \frac{1}{\sqrt{n}} \sum_{i} \xi_{i,r}(\boldsymbol{\theta}_0) + o_p(1),$$

where $\xi_{i,r}(\boldsymbol{\theta}_0)$ is defined in (3.4). Then by the central limit theorem, we have

$$[V\{\xi_{i,r}(\boldsymbol{\theta}_0)\}]^{-1/2} n^{-1/2} \sum_{i=1}^{n} \xi_{i,r}(\boldsymbol{\theta}_0) \sim_d N(\mathbf{0}, \mathbf{I}_{\mathbf{d}_\theta \times \mathbf{d}_\theta}),$$

where

$$V(\xi_{i,r}(\boldsymbol{\theta}_0)) \doteq \sigma^2_{r|x}(x) - E[(1 - p(x))\sigma^2_{r|x}(x)]$$

$$+ C_p^2 E\{p(x_i) h_r(x_i, y_i) \mathbf{B}(x_i) \mathbf{B}^T(x_i) h_r^T(x_i, y_i)\}$$

$$+ 2C_p E\{\delta_i h_r(x_i, y_i) \mathbf{B}(x_i) \mathbf{r}^T(y_i, x_i)\},$$

and $\sigma^2_{r|x}(x) = V_{y|x}\{\mathbf{r}(y, x)\}$.

C. Proof of property (3.10)

We can decompose $\widehat{\boldsymbol{\theta}}_{fixed} - \boldsymbol{\theta}_0$ into three terms as follows,

$$\widehat{\boldsymbol{\theta}}_{fixed} - \boldsymbol{\theta}_0 = \underbrace{\frac{1}{n} \sum_{i=1}^{n} [\delta_i \mathbf{r}(y_i, x_i) + (1 - \delta_i) \int_0^1 \mathbf{r}(q_\tau(x_i), x_i) d\tau] - \boldsymbol{\theta}_0}_{:=C_1}$$

$$+ \underbrace{\frac{1}{n} \sum_{i=1}^{n} (1 - \delta_i) [\frac{1}{J} \sum_{j=1}^{J} \mathbf{r}(q_\tau(x_i), x_i) - \int_0^1 \mathbf{r}(q_\tau(x_i), x_i) d\tau]}_{:=C_2}$$

$$+ \underbrace{\frac{1}{n} \sum_{i=1}^{n} (1 - \delta_i) \frac{1}{J} \sum_{j=1}^{J} [\mathbf{r}(\hat{q}_\tau(x_i), x_i) - \mathbf{r}(q_\tau(x_i), x_i)]}_{:=C_3}.$$

$$(C.1)$$

It is easy to see

$$C_1 = \frac{1}{n} \sum_{i=1}^{n} [\mathbf{r}(y_i, x_i) - \boldsymbol{\theta}_0 + (1 - \delta_i) \mu_{r|x}(x_i)]. \qquad (C.2)$$

For term C_2, we show that

$$\frac{1}{J}\sum_{j=1}^{J}\mathbf{r}(q_\tau(x_i), x_i) - \int_0^1 \mathbf{r}(q_\tau(x_i), x_i)d\tau$$

$$= \sum_{j=1}^{J}[\frac{1}{J}\mathbf{r}(q_\tau(x_i), x_i) - \int_{\frac{j-1}{J}}^{\tau_j}\mathbf{r}(q_\tau(x_i), x_i)d\tau - \int_{\tau_j}^{\frac{j}{J}}\mathbf{r}(q_\tau(x_i), x_i)d\tau]$$

$$= \sum_{j=1}^{J}[\frac{1}{J}\mathbf{r}(q_\tau(x_i), x_i)$$

$$- \int_{\frac{j-1}{J}}^{\tau_j}\{\mathbf{r}(q_{\tau_j}(x_i), x_i) + \dot{\mathbf{r}}_y(q_{\tau_{1j}}(x_i), x_i)\frac{\partial q_{\tau_{1j}}(x_i)}{\partial\tau}(\tau - \tau_j)\}d\tau$$

$$- \int_{\tau_j}^{\frac{j}{J}}\{\mathbf{r}(q_{\tau_j}(x_i), x_i) + \dot{\mathbf{r}}_y(q_{\tau_{2j}}(x_i), x_i)\frac{\partial q_{\tau_{2j}}(x_i)}{\partial\tau}(\tau - \tau_j)\}d\tau]$$

$$= \sum_{j=1}^{J}[\dot{\mathbf{r}}_y(q_{\tau_{1j}}(x_i), x_i)\frac{\partial q_{\tau_{1j}}(x_i)}{\partial\tau}\frac{-8}{J^2} + \mathbf{r}_y(q_{\tau_{2j}}(x_i), x_i)\frac{\partial q_{\tau_{2j}}(x_i)}{\partial\tau}\frac{8}{J^2}]$$

$$= O(\frac{1}{J}), \tag{C.3}$$

where τ_j's are the fixed sequence of the midpoints of J evenly spaced subintervals on $[0,1]$, and τ_{1j} and τ_{2j} are some values between $[\frac{j-1}{J}, \tau_j]$ and $[\tau_j, \frac{j}{J}]$ respectively. The last equality in (C.3) holds because of condition (f) and that $q_\tau(x) \in C^{p+1}$.

By the proof of Theorem 1 in Appendix B, similary we can show that

$$C_3 = \frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i)\frac{1}{J}\sum_{j=1}^{J}[\mathbf{r}(\hat{q}_\tau(x_i), x_i) - \mathbf{r}(q_\tau(x_i), x_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i)\frac{1}{J}\sum_{j=1}^{J}[\dot{\mathbf{r}}_y(q_\tau(x_i), x_i)[\hat{q}_\tau(x_i) - q_\tau(x_i)]$$

$$+ \ddot{\mathbf{r}}_y(\tilde{q}_\tau(x_i), x_i)[\hat{q}_\tau(x_i) - q_\tau(x_i)]^2]$$

$$= \frac{1}{n}\sum_{i=1}^{n}C_p\delta_i h_r(x_i, y_i)\mathbf{B}(x_i) + o_p(\frac{1}{\sqrt{n}}). \tag{C.4}$$

Combining equations (C.1) to (C.4), we have showed that

$$\widehat{\boldsymbol{\theta}}_{fixed} - \boldsymbol{\theta}_0 = \frac{1}{n}\sum_{i=1}^{n}\xi_{i,r}(\boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{J}).$$

Acknowledgements

Supplementary Material

Supplement to "Parameter estimation through semiparametric quantile regression imputation"
(doi: [10.1214/16-EJS1208SUPP](); .pdf). In this supplement, we provide very detailed proofs needed for the theories.

References

ANDRIDGE, R. R. and LITTLE, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review* **78** 40–64.

BARROW, D. L. and SMITH, P. W. (1978). Asymptotic properties of the best $L_2[0,1]$ approximation by splines with variable knots. *Quart. Appl. Math.* **36** 293–304. [MR0508773]()

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D., and FERNANDEZ-VAL, I. (2016). Conditional quantile processes based on series or many regressors. *ArXiv preprint [arXiv:1105.6154]()*.

BREIDT, F. J., OPSOMER, J. D., JOHNSON, A. A., and RANALLI, M. G. (2007). Semiparametric model-assisted estimation for natural resource surveys. *Statistics Canada* **33** 35–44.

BRICK, J. M. and KALTON, G. (1996). Handling missing data in survey research. *Stat. Methods Med. Res.* **5** 215–238.

CHAO, S., VOLGUSHEV, S., and CHENG, G. (2016). Quantile processes for semi and nonparametric regression. *ArXiv preprint [arXiv:1604.02130]().*

CHEN, X., HONG, H., and TAROZZI, A. (2008). Semiparametric efficiency in GMM models with auxiliary data. *Ann. Statist.* **36** 808–843. [MR2396816]()

CHENG, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* **89** 81–87.

DE BOOR, C. (2001). *A practical guide to splines.* Springer, New York. [MR1900298]()

GELMAN, A., HILL, J., SU, Y. S., YA, M., and PITTAU, M. G. (2013). Missing data imputation and model checking. [http://www.R-project.org/]()

KIM, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98** 119–132. [MR2804214]()

KIM, J. K., BRICK, J. M., FULLER, W. A., and KALTON, G. (2006). On the bias of the multiple imputation variance estimator in survey sampling. *J. R. Statist. Soc. B* **68**, 509–521. [MR2278338]()

KIM, J. K. and FULLER, W. (2004). Fractional hot deck imputation. *Biometrika* **91** 559–578. [MR2090622]()

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644]()

KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. [MR1326417]()

KOENKER, R. (2005). *Quantile regression.* Cambridge Univ. Press, Cambridge. [MR2268657]()

LITTLE, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics* **37** 23–38.

MENG, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statist. Sci.* **9** 538–558.

MUNOZ, J. F. and RUEDA, M. (2009). New imputation methods for missing data using quantiles. *J. Comput. Appl. Math.* **232** 305–317. MR2555402

NUSSER, S. M., CARRIQUIRY, A. L., DODD, K. W., and FULLER, W. A. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *J. Amer. Statist. Assoc.* **91** 1440–1449.

PAKES, A. and POLLARD, D. (1989). Simulation and the asymptotic of optimization estimators. *Econometrica* **57** 1027–1057. MR1014540

RAO, J. N. K. and SHAO, J. (1992). Jackknife variance estimation with survey data under hot-deck imputation. *Biometrika* **79** 811–822. MR1209480

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

RUBIN, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley, New York. MR0899519

WANG, D. and CHEN, S. X. (2009). Empirical likelihood for estimating equations with missing values. *Ann. Statist.* **37** 490–517. MR2488360

WANG, Q. and RAO, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30** 896–924. MR1922545

WANG, N. and ROBINS, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika.* **85** 935–948. MR1666715

WEI, Y., MA, Y., and CAROLL, R. J. (2012). Multiple imputation in quantile regression *Biometrika* **99** 423–438. MR2931263

YOON, J. (2013). Quantile regression analysis with missing response, with application to inequality measure and data combination. to be appear.

YOSHIDA, T. (2013). Asymptotics for penalized spline estimators in quantile regression. *Communications in Statistics – Theory and Methods*, to appear.

YU, K. and JONES, M. C. (1994). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

YUAN, M. (2006). GACV for quantile smoothing splines. *Comput. Statist. Data Anal.* **50** 813–829. MR2207010