HHS Public Access
Author manuscript
Scand Stat Theory Appl. Author manuscript; available in PMC 2019 June 01.

Published in final edited form as:
Scand Stat Theory Appl. 2018 June ; 45(2): 366–381. doi:10.1111/sjos.12296.

Efficient Robust Estimation for Linear Models with Missing Response at Random

Man-Lai Tang,
Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong

Niansheng Tang,
Department of Statistics, Yunnan University, P. R. of China

Puying Zhao, and
Department of Statistics, Yunnan University, P. R. of China

Hongtu Zhu
Department of Biostatistics, University of North Carolina at Chapel Hill, USA

Abstract

Coefficient estimation in linear regression models with missing data is routinely done in the mean regression framework. However, the mean regression theory breaks down if the error variance is infinite. In addition, correct specification of the likelihood function for existing imputation approach is often challenging in practice, especially for skewed data. In this paper, we develop a novel composite quantile regression and a weighted quantile average estimation procedure for parameter estimation in linear regression models when some responses are missing at random. Instead of imputing the missing response by randomly drawing from its conditional distribution, we propose to impute both missing and observed responses by their estimated conditional quantiles given the observed data and to use the parametrically estimated propensity scores to weigh check functions that define a regression parameter. Both estimation procedures are resistant to heavy-tailed errors or outliers in the response and can achieve nice robustness and efficiency. Moreover, we propose adaptive penalization methods to simultaneously select significant variables and estimate unknown parameters. Asymptotic properties of the proposed estimators are carefully investigated. An efficient algorithm is developed for fast implementation of the proposed methodologies. We also discuss a model selection criterion, which is based on an IC_Q-type statistic, to select the penalty parameters. The performance of the proposed methods is illustrated via simulated and real data sets.

Keywords

Conditional quantile; imputation; missing at random; model selection; quantile regression

---

Hongtu Zhu, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA. htzhu@email.unc.edu.

Supporting information
Additional information for this article is available online including technical assumptions, the proofs of Lemmas and Theorems, and some additional simulations.

1 Introduction

Missing data have long been a common problem in various settings, including surveys, clinical trials, and longitudinal studies, among many others. Ignoring the missing data will undermine study efficiency and sometimes introduce severe bias. There has been an enormous literature on different estimation and inference methods for various (conditional mean) regression analysis with missing response and/or covariates. See, for example, extensive reviews in Horton & Kleinman (2007), Horton & Laird (1999), and Ibrahim et al. (2005), and references therein. There are three streams of inference methods for mean regression with missing values: imputation method (Little & Rubin, 2002), inverse probability weighted method (Robins et al., 1994), and likelihood-based methods (Ibrahim et al., 2005). These existing approaches are expected to be sensitive to outliers and their efficiency may be significantly improved for many commonly used non-normal errors. However, the corrected specification of likelihood function used for imputation approach is often challenging in practice, especially for skewed data. Recently, quantile regression inference for missing data problems has received considerable attention due to the robustness of its regression coefficient estimates; see, for example, Yoon (2010), Wei et al. (2012), and Chen et al. (2015), among others.

Quantile regression as an important modeling tool (Koenker & Bassett, 1978) is a natural extension of classical least squares estimation of conditional mean models to the estimation of models for conditional quantile functions. See, for example, Koenker (2005) for an extensive review on quantile regression. It offers a systematic examination on the influence of covariates on the entire response distribution, while it is less sensitive to outliers. However, standard quantile regression focuses on estimator at each quantile and may be inefficient for many ‘global’ parameters of interest, which vary slowly across all quantiles. Since quantile regression exploits the whole conditional distribution, estimation efficiency could be potentially improved by combining quantile regression over multiple quantiles. For instance, in Zou & Yuan (2008), a composite quantile regression (CQR) estimator has been proposed for multiple quantile regression models. Furthermore, a weighted quantile average estimator (WQAE) proposed by Zhao & Xiao (2014) is a linear combination of quantile regression estimators across different quantiles. To the best of our knowledge, little work has been published on the CQR and/or WQAE-based inference under imputation for missing data.

The aim of this paper is to develop robust and efficient parameter estimation methods for regression models when some responses are missing at random. To increase both estimation robustness and efficiency, we propose a weighted CQR estimate (WCQR) and a WQAE estimate by borrowing information from multiple quantiles. We consider a novel parametric multiple imputation approach that imputes both missing and observed responses by their estimated conditional quantiles given the observed data. Specifically, following Robins et al. (1994), we use the parametrically estimated propensity scores to weigh check functions that define a regression parameter estimate. Moreover, the proposed multiple imputation approach yields more reliable results even in high dimensional scenario. We also use the smoothly clipped absolute deviation (SCAD) and adaptive least absolute shrinkage and selection operator (LASSO) regularization (Fan & Li, 2001; Zou, 2006) and a IC_Q statistic

(Ibrahim et al., 2008) to perform the selection of penalty parameters, variable selection, and estimation. We systematically investigate the asymptotic properties of the proposed estimators under the proposed multiple imputation approach. An efficient algorithm is developed for the fast implementation of the proposed methods. Unlike the existing mean regression procedure for missing data, our method is resistant to heavy-tailed errors or outliers in responses.

The rest of this paper is organized as follows. In Section 2, we describe the parametric imputation approaches and introduce WCQR and WQAE and their shrinkage estimators. We also systematically investigate the asymptotic properties of all parameter estimators. In Section 3, we present an efficient procedure of parameter estimation, variable selection, and tuning parameter selection. In Section 4, we evaluate the proposed methods in simulated datasets. In Section 5, we use a real data example to illustrate the strategy. In Section 6, we conclude the paper with some discussions. Technical proofs and additional numerical studies are given in the Supporting Information.

2 Methodology

2.1 Preliminaries

Let $\{(x_i, y_i), i = 1, \ldots, n\}$ be a random sample generated from a model given by

$$y_i = x_i^T \boldsymbol{\beta}^* + \varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \qquad (2.1)$$

where y_i and $x_i = (x_{i1}, \ldots, x_{id})^T$ are, respectively, a univariate response variable and a $d \times 1$ vector of covariates which contains no intercept term for the ith individual, $\boldsymbol{\beta}^*$ is a $d \times 1$ vector of regression coefficients, and ε_is are independent and identically distributed (i.i.d.) random errors that are independent of x_is. Throughout the paper, we assume that y_is may be missing, whereas x_is are always fully observed. That is, the data set consists of incomplete observations $\{(x_i, y_i, \delta_i) : i = 1, \ldots, n\}$, which are i.i.d. realizations from $(\mathbf{x}, \mathbf{y}, \delta)$, where $\delta = 0$ if y is missing and $\delta = 1$ otherwise. We also assume that $\mathbf{y}$ is missing at random, i.e., δ and $\mathbf{y}$ are conditionally independent given $\mathbf{x}$. That is, we have

$$\Pr(\delta = 1 | \mathbf{x}, \mathbf{y}) = \Pr(\delta = 1 | \mathbf{x}) =: p(\mathbf{x}),$$

where $p(\mathbf{x})$ is the propensity score that describes a pattern of selection bias in the missingness. In particular, we postulate a parametric model for the propensity score, that is, $p(\mathbf{x}) = p(\mathbf{x}, \boldsymbol{\gamma}^*)$, and further consider logistic model $\text{logit}\{p(\mathbf{x}, \boldsymbol{\gamma}^*)\} = \gamma_0^* + \gamma_1^{*T}\mathbf{x}$. An estimator $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\gamma}_1)^T$ is defined by maximizing the log-binomial likelihood given by

$$L(\boldsymbol{\gamma}) = \log\left\{\prod_{i=1}^{n} p(x_i, \boldsymbol{\gamma})^{\delta_i}(1 - p(x_i, \boldsymbol{\gamma}))^{1-\delta_i}\right\}. \qquad (2.2)$$

Under some regularity conditions, the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\gamma}^*$ even though the logistic model is misspecified (White, 1982). We denote $\hat{p}(\mathbf{x})$ as $p(\mathbf{x}, \hat{\boldsymbol{\gamma}})$.

By assuming that the regression coefficients are the same across different quantiles, Zou & Yuan (2008) proposed a CQR estimator, which minimizes a mixture of the objective functions from different quantiles. Let $\mathscr{T} = \{\tau_1, \dots, \tau_K\}$ be a set of quantiles. A simple estimator is a complete-case (CC) composite quantile regression estimator $\hat{\boldsymbol{\beta}}_{\mathrm{CC}}^{\mathrm{CQR}}$, which can be obtained by solving

$$(\hat{b}_1, \dots, \hat{b}_K, \hat{\boldsymbol{\beta}}_{\mathrm{CC}}^{\mathrm{CQR}}) = \underset{b_1, \dots, b_K, \boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \sum_{i=1}^{n} \delta_i \rho_{\tau_k}(y_i - x_i^T \boldsymbol{\beta} - b_k),$$

where $\rho_\tau(u) = u\{\tau - \mathbf{1}(u < 0)\}$ is the check function, in which $\mathbf{1}(\cdot)$ is an indicator function of an event. Let $f(\cdot)$ and $b_{\tau_k}^*$ be the density function and $100\tau_k\%$ quantile of residual $\boldsymbol{\varepsilon}$, respectively.

PROPOSITION 1—Under assumptions A1, A2 and A4 given in the Supporting Information, as $n \to \infty$, we have $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{CC}}^{\mathrm{CQR}} - \boldsymbol{\beta}^*) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma_0)$, where Σ_0 is given by

$$\Sigma_0 = [E\{p(\mathbf{x})\mathbf{x}\mathbf{x}^T\}]^{-1} \frac{\sum_{k, k'=1}^{K} \min(\tau_k, \tau_{k'})\{1 - \max(\tau_k, \tau_{k'})\}}{\{\sum_{k=1}^{K} f(b_{\tau_k}^*)\}^2}.$$

Proposition 1 shows that under the MAR assumption, the CC analysis also produces a consistent estimator. However, compared with the full sample analysis, CC analysis can lead to larger efficiency loss with the presence of a large amount of missing data. In what follows, we introduce a novel multiple imputation procedure that aims to improve the estimation robustness and efficiency of the existing methods and the CC estimator.

2.2 Augmented check functions

Instead of imputing the missing response $\mathbf{y}$ by randomly drawing from its conditional distribution $F(y \mid \mathbf{x})$ (Rubin, 1987), one can directly impute the missing response by its estimated conditional quantile given the observed data. This imputation procedure has advantages of being less sensitive to model misspecification and its ability to produce reliable estimates even under high-dimensional covariates, which have been previously studied in quantile regression with missing responses (Yoon, 2010) and M-regression with censored covariates (Wang & Feng, 2012). In this paper, we extend such imputation approach to the estimation of CQR models with ignorable missing response.

Define $Q_{\mathbf{y}}(\tau \mid \mathbf{x}) = \inf\{y : F(y \mid \mathbf{x}) \geq \tau\}$ to be the τ-th conditional quantile of $\mathbf{y}$ given $\mathbf{x}$, where $\tau \sim \mathrm{uniform}(\tau_L, \tau_U)$ with $0 < \tau_L < \tau_U < 1$. Under MAR assumption, we have $F(\mathbf{y} \mid \mathbf{x}) = F(\mathbf{y} \mid \mathbf{x}, \delta = 1) = F(\mathbf{y} \mid \mathbf{x}, \delta = 0)$, which leads to $Q_{\mathbf{y}}(\tau \mid \mathbf{x}) = Q_{\mathbf{y}}(\tau \mid \mathbf{x}, \delta = 1) = Q_{\mathbf{y}}(\tau \mid \mathbf{x}, \delta = 0)$.

Specifically, for $\tau \in [\tau_L, \tau_U]$, we assume the linear quantile regression model as $Q_{\mathbf{y}}(\tau|\mathbf{x}) = \boldsymbol{\theta}_0^*(\tau) + \mathbf{x}^T\boldsymbol{\theta}_1^*(\tau)$. It should be noted here that $\boldsymbol{\theta}_0^*(\tau) = b_\tau^*$ if such conditional quantile regression model under MAR assumption is correctly specified. Let $\boldsymbol{\theta}^*(\tau) = (\boldsymbol{\theta}_0^*(\tau), \boldsymbol{\theta}_1^*(\tau)^T)^T$, which is a $(d+1)\times 1$ vector of unknown quantile regression coefficients. Let $\mathbf{z} = (1, \mathbf{x}^T)$ and $z_i = (1, x_i^T)$ for $i = 1, \ldots, n$. Since we have $E\{\delta\mathbf{z}[\mathbf{1}(\mathbf{y} < \mathbf{z}^T\boldsymbol{\theta}^*(\tau)) - \tau]\} = 0$ under MAR assumption, an estimator $\hat{\boldsymbol{\theta}}(\tau)$ of $\boldsymbol{\theta}^*(\tau)$ can be calculated as

$$\hat{\boldsymbol{\theta}}(\tau) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \delta_i \rho_\tau(y_i - z_i^T\boldsymbol{\theta}), \quad (2.3)$$

and the τ-th conditional quantile is estimated by $\hat{Q}_{\mathbf{y}}(\tau|\mathbf{x}) = \hat{Q}_{y_i}(\tau|x_i, \delta_i = 1) = z_i^T\hat{\boldsymbol{\theta}}(\tau)$ for $i = 1, \ldots, n$. Although (2.3) assumes a linear form of the conditional quantiles of $\mathbf{x}$, our empirical studies in Section 4 suggest that such assumption is relatively robust to potential model misspecification. To achieve more flexible imputation, nonparametric or semiparametric quantile models may be employed in (2.3).

Let $\tilde{Y}_{i\nu}$ be the ν-th imputed value of the ith outcome for $\nu = 1, \ldots, m$, where m is the number of imputations. Given $0 < \tau_L < \tau_U < 1$, generate τ^ν from uniform(τ_L, τ_U) and set $\tilde{Y}_{i\nu} = z_i^T\hat{\boldsymbol{\theta}}(\tau^\nu)$, where $\hat{\boldsymbol{\theta}}(\tau^\nu)$ is obtained by solving (2.3) with τ replaced. It is not difficult to show that $\tilde{Y}_{i\nu} \sim \hat{F}_p(y \mid x_i)$, where $\hat{F}_p(y \mid x_i)$ is the inverse of the estimated conditional quantile function $\hat{Q}_{y_i}(\tau^\nu \mid x_i)$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)^T$ be a vector of known weights, $\boldsymbol{b} = (b_1, \ldots, b_K)^T$, and $0 < \tau_1 < \tau_2 < \cdots < \tau_K < 1$. Define an augmented inverse probability weighted (AIPW) composite quantile objective function with multiple imputation (MI) as

$$\mathcal{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}; \boldsymbol{\omega}, \mathcal{T}) = \sum_{k=1}^{K} \omega_k \mathcal{S}_{nm}(\boldsymbol{\beta}, b_k; \tau_k), \quad (2.4)$$

where $\mathcal{S}_{nm}(\boldsymbol{\beta}, b_k; \tau_k)$ is given by

$$\sum_{i=1}^{n} \left\{ \frac{\delta_i}{\hat{p}(x_i)} \rho_{\tau_k}(r_{ik}) + \left[1 - \frac{\delta_i}{\hat{p}(x_i)}\right] m^{-1} \sum_{\nu=1}^{m} \rho_{\tau_k}(r_{i\nu k}) \right\}, \quad (2.5)$$

in which $r_{ik} = y_i - x_i^T\boldsymbol{\beta} - b_k$ and $r_{i\nu k} = \tilde{Y}_{i\nu} - x_i^T\boldsymbol{\beta} - b_k$. Throughout the paper, we refer to the proposed imputation approach in (2.5) as “AIPW-MI procedure”.

REMARK 1—Compared with many existing approaches, the proposed AIPW-MI has several new features. First, it follows from (2.5) that the observed y_i is also imputed by its estimated conditional quantile given the observed data. See a similar idea used in Tang & Qin (2012), in which missing values were randomly drawn from a kernel estimator of the conditional distribution of the missing variables given the fully observable variables. However, Tang & Qin's kernel density estimation is impeded by the curse of dimensionality. Second, in the last term on the right hand side of (2.5), the check functions are averaged across multiple imputed values in order to control the variance of the imputed values.

Given $\hat{p}(\mathbf{x})$, an estimator of $\boldsymbol{\beta}^*$, say $\hat{\boldsymbol{\beta}}^{\text{PCQR}}$, can be estimated as follows

$$(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}^{\text{PCQR}}) = \arg\min_{\boldsymbol{b}, \boldsymbol{\beta}} \mathscr{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}; \boldsymbol{\omega}, \mathscr{T}).$$

We call it the *parametric-CQR* (PCQR) estimator of $\boldsymbol{\beta}^*$. Note that when $\delta_i = 1$ and $\hat{p}(x_i) = 1$ for all k, i, model (2.4) reduces to the weighted CQR model (see Jiang et al., 2012). When $\omega_k = 1/K$, $\delta_i = 1$ and $\hat{p}(x_i) = 1$ for all k, i, model (2.4) reduces to the CQR model (see, Zou & Yuan, 2008).

For single quantile regression (see, Koenker, 2005), $\boldsymbol{\beta}^*$ in model (2.1) can be estimated by solving

$$(\hat{b}_\tau, \hat{\boldsymbol{\beta}}_\tau^{QR}) = \arg\min_{b, \boldsymbol{\beta}} \mathscr{S}_{nm}(\boldsymbol{\beta}, b; \tau). \quad (2.6)$$

If ε_i follows a double-exponential distribution, the quantile regression estimator $\hat{\boldsymbol{\beta}}_{0.5}^{QR}$ is the most efficient estimator even compared with the least square estimator when there are no missing data. However, for other distributions, the relative efficiency of $\hat{\boldsymbol{\beta}}_\tau^{QR}$ over the least square estimator can be very small. By combining information based on estimators at different quantiles, we further define a weighted parametric quantile average estimate (PQAE) given by

$$\hat{\boldsymbol{\beta}}^{\text{PQAE}} = \sum_{k=1}^{K} \varpi_k \hat{\boldsymbol{\beta}}_{\tau_k}^{QR}, \quad \text{subject to} \quad \sum_{k=1}^{K} \varpi_k = 1,$$

where $\hat{\boldsymbol{\beta}}_\tau^{QR}$ is defined in (2.6). The PQAE is indeed motivated by a general weighting function $\int \hat{\boldsymbol{\beta}}_\tau^{QR} \varpi(\tau) d\tau$ (Portnoy & Koenker, 1989), where $\varpi(\tau)$ is a general weighting function.

2.3 Variable selection

Variable selection plays an important role in model building process. To achieve sparsity with high-dimensional covariates, we further develop adaptive penalization methods for variable selection based on the proposed AIPW-MI approach. Various penalization methods have been developed for variable selection for different models (Fan & Li, 2006; Tibshirani, 1996; Zou & Li, 2008; Zou, 2006).

We first consider the SCAD penalty (Fan & Li, 2001) with a tuning parameter λ_n, denoted by $\mathscr{P}_{\lambda_n}(\cdot)$, to be selected by a data-driven method. By filling in the missing y_i by $\widetilde{Y}_{i\nu} = z_i^T\widehat{\boldsymbol{\theta}}(\tau^{\nu})$, the SCAD estimator for penalized PCQR with missing data solves the following minimization problem

$$\underset{\boldsymbol{b},\boldsymbol{\beta}}{\arg\min}\left\{\mathcal{S}_{n\omega}(\boldsymbol{\beta},\boldsymbol{b};\boldsymbol{\omega},\mathcal{T}) + n\sum_{j=1}^{d}\mathscr{P}_{\lambda_n}(|\beta_j|)\right\}. \quad (2.7)$$

The penalty $\mathscr{P}_{\lambda_n}(\cdot)$ satisfies $\mathscr{P}_{\lambda_n}(0) = 0$, and its first-order derivative is given by

$$\mathscr{P}'_{\lambda_n}(\boldsymbol{\beta}) = \lambda_n\left\{\mathbf{1}(\boldsymbol{\beta} \le \lambda_n) + \frac{(a\lambda_n - \boldsymbol{\beta})_+}{(a-1)\lambda_n}\mathbf{1}(\boldsymbol{\beta} > \lambda_n)\right\},$$

where $a = 3.7$, and $(s)_+ = s$ for $s > 0$ and 0 otherwise.

It is very challenging to optimize (2.7) since $\mathscr{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}; \boldsymbol{\omega}, \mathscr{T})$ is nonconvex and both $\mathscr{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}; \boldsymbol{\omega}, \mathscr{T})$ and the penalty $\mathscr{P}_{\lambda_n}(|\beta_j|)$ are nondifferentiable. To overcome these difficulties, we adopt the one-step sparse estimate scheme in Zou & Li (2008). For any given initial value $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)}, \ldots, \beta_d^{(0)})^T$, we define a one-step penalized PCQR loss as

$$\mathscr{G}_n(\boldsymbol{b},\boldsymbol{\beta}) = \mathcal{S}_{n\omega}(\boldsymbol{\beta},\boldsymbol{b};\boldsymbol{\omega},\mathcal{T}) + n\sum_{j=1}^{d}\mathscr{P}'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|. \quad (2.8)$$

We define $(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}^{\mathrm{PSCAD}}) = \arg\min_{\boldsymbol{b},\boldsymbol{\beta}}\mathscr{G}_n(\boldsymbol{b}, \boldsymbol{\beta})$ as the one-step sparse PCQR-SCAD estimaror. By combining estimators corresponding to different quantiles, we can derive a one-step sparse PAQE-SCAD estimator as

$$\widehat{\boldsymbol{\beta}}^{\mathrm{PSCAD}} = \sum_{k=1}^{K}\varpi_k\widehat{\boldsymbol{\beta}}_{\tau_k}^{\mathrm{PSCAD}}, \quad \text{subject to} \quad \sum_{k=1}^{K}\varpi_k = 1,$$

where $(\widehat{b}_\tau, \widehat{\boldsymbol{\beta}}_\tau^{\mathrm{PSCAD}}) = \arg\min_{b,\boldsymbol{\beta}}\{\mathcal{S}_{nm}(\boldsymbol{\beta}, b; \tau) + n\sum_{j=1}^{d}\mathscr{P}'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|\}$ with $\tilde{Y}_{i\nu} \sim \hat{F}_p(y \mid x_i)$ in $\mathscr{S}_{nm}(\boldsymbol{\beta}, b; \boldsymbol{\tau})$. We also use the same notation $\hat{\boldsymbol{\beta}}^{\mathrm{PSCAD}}$ to denote the one-step sparse PAQE-SCAD estimator for simplicity.

Second, we consider the adaptive-LASSO penalty (Zou, 2006). Specifically, the adaptive LASSO PCQR loss is given by $\mathcal{S}_{n\omega}(\boldsymbol{\beta},\boldsymbol{b};\boldsymbol{\omega},\mathcal{T}) + \lambda_n\sum_{j=1}^{d}\widetilde{\omega}_j|\beta_j|$, where $\tilde{\omega_j} = 1/|\hat{\beta}_j|^2$ for $j = 1, \ldots, d$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_d)^T = \arg\min_{\boldsymbol{b},\boldsymbol{\beta}}\mathscr{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}; \boldsymbol{\omega}, \mathscr{T})$. By minimizing the above objective function with a proper penalty parameter λ_n, we can get a sparse estimator of $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}^{\mathrm{PLASSO}}$, which is called the adaptive PCQR-LASSO estimator. Similarly, we can derive the adaptive PAQE-LASSO estimator as

$$\hat{\beta}^{\mathrm{PLASSO}} = \sum_{k=1}^{K} \varpi_k \hat{\beta}_{\tau_k}^{\mathrm{PLASSO}}, \quad \text{subject to} \quad \sum_{k=1}^{K} \varpi_k = 1,$$

where $(\hat{b}_\tau, \hat{\boldsymbol{\beta}}_\tau^{\mathrm{PLASSO}}) = \arg\min_{b, \boldsymbol{\beta}} \{\mathcal{S}_{nm}(\boldsymbol{\beta}, b; \tau) + \lambda_n \sum_{j=1}^{d} \tilde{\omega}_j |\beta_j|\}$ with $\tilde{Y}_{i\nu} \sim \hat{F}_p(y \mid x_i)$ in $\mathcal{S}_{nm}(\boldsymbol{\beta}, b; \boldsymbol{\tau})$.

2.4 Asymptotic properties

We define $D = E\{p(\mathbf{x})^{-1}\mathbf{x}\mathbf{x}^T\}$, $\mathscr{W}(\tau, \tau') = \min(\tau, \tau')\{1 - \max(\tau, \tau')\}$, $\boldsymbol{\varpi} = (\varpi_1, \ldots, \varpi_K)^T$, $\mathrm{H} = \{\mathscr{W}(\tau_l, \tau_k)/f(b_{\tau_l}^*)f(b_{\tau_k}^*)\}_{1 \le l, k \le K}$, $\mathbb{V}_1 = D\sum_{k,k'=1}^{K} \omega_k \omega_{k'} \mathscr{W}(\tau_k, \tau_{k'})$ and $\mathbb{V}_2 = D\boldsymbol{\varpi}^T \mathrm{H} \boldsymbol{\varpi}$.

THEOREM 1—Under the assumptions A1–A6 given in the Supporting Information, as $n \to \infty$ and $m \to \infty$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{\mathrm{PCQR}} - \boldsymbol{\beta}^*) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \textstyle\sum_1(\boldsymbol{\omega})),$$

where $\sum_1(\boldsymbol{\omega}) = \mathbb{D}_1^{-1}\mathbb{V}_1\mathbb{D}_1^{-1}$ with $\mathbb{D}_1 = \sum_{k=1}^{K} \omega_k f(b_{\tau_k}^*) E\{\mathbf{x}\mathbf{x}^T\}$.

COROLLARY 1—Under the assumptions A1–A6 given in the Supporting Information, as $n \to \infty$ and m is fixed, we obtain $\sqrt{n}(\hat{\boldsymbol{\beta}}^{\mathrm{PCQR}} - \boldsymbol{\beta}^*) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \tilde{\sum}_1(\boldsymbol{\omega}))$, where $\tilde{\sum}_1(\boldsymbol{\omega}) = \mathbb{D}_1^{-1}\tilde{\mathbb{V}}_1\mathbb{D}_1^{-1}$ with

$$\tilde{\mathbb{V}}_1 = E\left\{[p(\mathbf{x})^{-1} + m^{-1}(p(\mathbf{x})^{-1} - 1)]\mathbf{x}\mathbf{x}^T\right\} \sum_{k,k'=1}^{K} \omega_k \omega_{k'} \mathscr{W}(\tau_k, \tau_{k'}).$$

Theorem 1 characterizes the asymptotic normality of $\hat{\boldsymbol{\beta}}^{\mathrm{PCQR}}$. For the fixed $\mathscr{T}$, the optimal $\boldsymbol{\omega}_* = \arg\min_{\boldsymbol{\omega}} \Sigma_1(\boldsymbol{\omega})$ can be calculated by solving

$$\boldsymbol{\omega}_* = \arg\min_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}^T W \boldsymbol{\omega}}{\boldsymbol{\omega}^T \mathbf{f}\mathbf{f}^T \boldsymbol{\omega}} \quad \text{or} \quad \boldsymbol{\omega}_* = \arg\max_{\boldsymbol{\omega}} \frac{\boldsymbol{\omega}^T \mathbf{f}\mathbf{f}^T \boldsymbol{\omega}}{\boldsymbol{\omega}^T W \boldsymbol{\omega}} = \frac{W^{-1}\mathbf{f}}{\|W^{-1/2}\mathbf{f}\|},$$

where $\|\cdot\|$ is the Euclidean norm of a vector, $W = (\mathscr{W}(\boldsymbol{\tau}_k, \boldsymbol{\tau}_{k'}))$, and $\mathbf{f} = (f(b_{\tau_1}^*), \ldots, f(b_{\tau_K}^*))^T$. Thus, the optimal $\Sigma_1(\boldsymbol{\omega}_*)$ is given by

$$\textstyle\sum_1(\boldsymbol{\omega}_*) = (\mathbf{f}^T W^{-1}\mathbf{f})^{-1} E\{\mathbf{x}\mathbf{x}^T\}^{-1} E\{p(\mathbf{x})^{-1}\mathbf{x}\mathbf{x}^T\} E\{\mathbf{x}\mathbf{x}^T\}^{-1}.$$

Theorem 1 also indicates that parametric estimation of propensity score has no asymptotic impact on the proposed *parametric-CQR* estimator. Moreover, from the proof of Theorem 1 in the Supporting Information, we may write

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{\text{PCQR}} - \boldsymbol{\beta}^*) = n^{-1/2}\sum_{i=1}^{n}\Lambda(x_i, y_i, \delta_i, \boldsymbol{\beta}^*) + o_p(1),$$

where $\Lambda(\mathbf{x}, \mathbf{y}, \delta, \boldsymbol{\beta}^*) = -\sum_{k=1}^{K}\omega_k\mathbb{D}_1^{-1}\delta p(\mathbf{x})^{-1}\mathbf{x}(\mathbf{1}(\mathbf{y} < \mathbf{x}^T\boldsymbol{\beta}^* + b_{\tau_k}^*) - \tau_k)$ is the influence function. Under assumption of missing response at random and model specification, we have $m_{\tau_k}(\mathbf{x}, \boldsymbol{\beta}^*, b_{\tau_k}^*) = E\{\mathbf{x}(\mathbf{1}(\mathbf{y} < \mathbf{x}^T\boldsymbol{\beta}^* + b_{\tau_k}^*) - \tau_k)|\mathbf{x}, \delta = 1\} = E\{\mathbf{x}(\mathbf{1}(\mathbf{y} < \mathbf{x}^T\boldsymbol{\beta}^* + b_{\tau_k}^*) - \tau_k)|\mathbf{x}\} = 0$ uniformly in k. This implies that $\Lambda(\mathbf{x}, \mathbf{y}, \delta, \boldsymbol{\beta}^*)$ equals

$$-\sum_{k=1}^{K}\omega_k\mathbb{D}_1^{-1}\left\{\frac{\delta}{p(\mathbf{x})}\mathbf{x}(\mathbf{1}(\mathbf{y} < \mathbf{x}^T\boldsymbol{\beta}^* + b_{\tau_k}^*) - \tau_k) + \left[1 - \frac{\delta}{p(\mathbf{x})}\right]m_{\tau_k}(\mathbf{x}, \boldsymbol{\beta}^*, b_{\tau_k}^*)\right\},$$

which is of a weighted composite AIPW form; thus, it is an efficient influence function.

THEOREM 2—Under the assumptions A1–A6 given in the Supporting Information, as $n \to \infty$ and $m \to \infty$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{\text{PQAE}} - \boldsymbol{\beta}^*) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \sum\nolimits_2(\boldsymbol{\varpi})),$$

where $\sum_2(\boldsymbol{\varpi}) = \mathbb{D}_2^{-1}\mathbb{V}_2\mathbb{D}_2^{-1}$ with $\mathbb{D}_2 = E\{\mathbf{x}\mathbf{x}^T\}$.

Theorem 2 characterizes the asymptotic normality of $\hat{\boldsymbol{\beta}}^{\text{PQAE}}$. For the fixed $\mathscr{T}$, the optimal $\boldsymbol{\varpi}_* = \arg\min_{\boldsymbol{\varpi}} \Sigma_2(\boldsymbol{\varpi})$ can be calculated by solving

$$\boldsymbol{\varpi}_* = \arg\min_{\boldsymbol{\varpi}} \boldsymbol{\varpi}^T \mathrm{H}\boldsymbol{\varpi} \text{ subject to } \mathbf{1}_K^T\boldsymbol{\varpi} = 1,$$

where $\mathbf{1}_K = (1, \ldots, 1)^T$ is a $K \times 1$ vector of ones. It is easy to show that the optimal $\Sigma_2(\boldsymbol{\varpi}_*)$ is given by

$$\sum\nolimits_2(\boldsymbol{\varpi}_*) = (\mathbf{1}_k^T\mathrm{H}^{-1}\mathbf{1}_K)^{-1}E\{\mathbf{x}\mathbf{x}^T\}^{-1}E\{p(\mathbf{x})^{-1}\mathbf{x}\mathbf{x}^T\}E\{\mathbf{x}\mathbf{x}^T\}^{-1}.$$

Using similar arguments as in Theorem 1, we can also show that the proposed weighted parametric quantile average estimator is efficient. From Theorems 1 and 2, we find that as $n \to \infty$ and $m \to \infty$, multiple imputation in (2.5) does not have asymptotic impact on the resultant estimators, and thus $\hat{\boldsymbol{\beta}}^{\text{PCQR}}$ and $\hat{\boldsymbol{\beta}}^{\text{PAQE}}$ are invariant to the model of the conditional quantile model of $\mathbf{y}$ given $\mathbf{x}$ and $\delta = 1$.

Since f is the density function of residual ε, the Fisher information of f is given by

$$\mathscr{I}_f = \int_{\mathbb{R}} \left[\frac{\partial \log f(u)}{\partial u}\right]^2 f(u)du = \int_0^1 \left[\frac{\partial f(b_\tau^*)}{\partial \tau}\right]^2 d\tau .$$

For equally spaced $\{\tau_k\}_{k=1}^K$, the following proposition shows that the optimal PQAE is asymptotically more efficient than the optimal PCQR. Only when the log density function log $f(u)$ of ε is concave, the optimal PCQR is asymptotically equivalent to the optimal PQAE, and both are nearly efficient as the oracle maximum likelihood estimator for various error distributions, a great advantage of the proposed methodology.

PROPOSITION 2—(Zhao & Xiao, 2014) Let $\tau_k = k/(K+1)$ for $k = 1, \ldots, K$, suppose that assumptions A1 and A2 given in the Supporting Information holds, we have the following results.

a. In general, $(\mathbf{1}_K^T \mathrm{H}^{-1}\mathbf{1}_K)^{-1} \le \min_{\boldsymbol{\omega}}\{\boldsymbol{\omega}^T W\boldsymbol{\omega}/\boldsymbol{\omega}^T \mathbf{f}\mathbf{f}^T\boldsymbol{\omega}\}$, with equality at optimal $\omega = \boldsymbol{\omega}_* = \{[2f(b_{\tau_k}^*) - f(b_{\tau_{k-1}}^*) - f(b_{\tau_{k+1}^*})]/[f(b_{\tau_1}^*) + f(b_{\tau_K}^*)]\}_{k=1}^K$ under the assumption that the log density function log $f(\boldsymbol{\tau})$ of $\boldsymbol{\varepsilon}$ is concave. Here, it is assumed that $f(b_{\tau_0}^*) = f(b_{\tau_{K+1}}^*) = 0$.

b. If $\lim_{\boldsymbol{\tau}\to 0}[g^2(\boldsymbol{\tau}) + g^2(1-\boldsymbol{\tau})]/\boldsymbol{\tau} = 0$ and $\lim_{\tau\to 0}\tau^2\int_\tau^{1-\tau}|g''(t)|^2 dt = 0$, where $g(\tau) = f(b_\tau^*)$ and $g''(t) = \partial^2 g(t)/\partial t^2$, then $\lim_{K\to\infty}\mathbf{1}_K^T\mathrm{H}^{-\mathbf{1}}\mathbf{1}_K = \mathscr{I}_f$.

However, compared with the optimal PCQR, the optimal PQAE has the disadvantage of being more sensitive to either near-zero or extreme values of the estimated propensity score because the proposed AIPW-MI procedure is independently repeated K times in PQAE.

In order to better understand the optimality of variable selection and parameter estimation, Fan & Li (2001) considered the oracle property under the true, but "unknown" subset $\mathscr{A}$, where $\mathscr{A} = \{j : \boldsymbol{\beta}_j \neq 0\}$. The oracle estimate would need to estimate $\boldsymbol{\beta}_{\mathscr{A}}^*$ and set $\boldsymbol{\beta}_{\mathscr{A}^c}^* = 0$. For any square matrix $\mathscr{C}$, denote $\mathscr{C}_{\mathscr{A}\mathscr{A}}$ as the sub-matrix of $\mathscr{C}$ with row and column indices in $\mathscr{A}$. According to Theorems 1 and 2, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^E_{(oracle)_{\mathscr{A}}} - \boldsymbol{\beta}^*_{\mathscr{A}}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \textstyle\sum_{oracle}),$$

where $\sum_{oracle} = \mathbb{D}^{-1}_{\imath\mathscr{A}\mathscr{A}}\mathbb{V}_{\imath\mathscr{A}\mathscr{A}}\mathbb{D}^{-1}_{\imath\mathscr{A}\mathscr{A}}$ for $\imath = 1$ and 2, and E denotes for PSCAD and PLASSO.

THEOREM 3—(Oracle properties) Suppose that assumptions A1–A6 given in the Supporting Information hold, as $n \to \infty$ and $m \to \infty$, we have the following results:

i. If $\sqrt{n}\lambda_n \to \infty$ and $\lambda_n \to 0$, then $\widehat{\boldsymbol{\beta}}^{\mathrm{PSCAD}} = (\widehat{\beta}_1^{\mathrm{PSCAD}}, \ldots, \widehat{\beta}_d^{\mathrm{PSCAD}})^T$ satisfies:

a. Consistent selection: $\Pr(\{j: \hat{\beta}_j^{\text{PSCAD}} \neq 0\} = \mathscr{A}) \to 1$;

b. Efficient estimation: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}^{\text{PSCAD}} - \boldsymbol{\beta}_{\mathscr{A}}^{*}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma_{oracle})$.

ii. If $\lambda_n/\sqrt{n} \to 0$ and $\lambda_n \to \infty$, then $\hat{\boldsymbol{\beta}}^{\text{PLASSO}} = (\hat{\beta}_1^{\text{PLASSO}}, \ldots, \hat{\beta}_d^{\text{PLASSO}})^T$ satisfies:

a. Consistent selection: $\Pr(\{j: \hat{\beta}_j^{\text{PLASSO}} \neq 0\} = \mathscr{A}) \to 1$;

b. Efficient estimation: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathscr{A}}^{\text{PLASSO}} - \boldsymbol{\beta}_{\mathscr{A}}^{*}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \Sigma_{oracle})$.

3 Computational issues

3.1 Estimation of the optimal weight

To calculate the optimal estimators, we need to compute the optimal weights $\boldsymbol{\omega}_*$ and $\boldsymbol{\varpi}_*$. It suffices to estimate $f(b_\tau)$ for the proposed AIPW-MI procedure. We can use some nonparametric density estimation methods based on estimated residuals $\hat{\varepsilon}_i$ to obtain a consistent estimator of $f(\cdot)$. For the sake of space, we focus on the calculation of the optimal weights $\boldsymbol{\omega}_*$, whereas extension to other cases is trivial.

I. For any y_i, draw τ^ν from uniform(τ_L, τ_U) with $0 < \tau_L < \tau_U < 1$, and substitute y_i by $\tilde{Y}_{i\nu} = z_i^T \hat{\boldsymbol{\theta}}(\tau^\nu)$ with $z_i = (1, x_i^T)$ for $i = 1, \ldots, n$. We can simply set $\tau_L = 0$ and $\tau_U = 1$. Repeating this imputation procedure m times yields the augmented data set $\{\delta_i \hat{p}(x_i)^{-1} y_i + [1 - \delta_i \hat{p}(x_i)^{-1}]\tilde{Y}_{i\nu}, x_i : i = 1, \ldots, n\}_{\nu=1}^{m}$. Here $\hat{p}(x_i)$ is the estimated propensity score.

II. Compute preliminary estimators $\hat{\boldsymbol{\beta}}$ and $\hat{b}_k$ based on the augmented data set and the uniform weight $\omega = (K^{-1}, \ldots, K^{-1})^T$. Then the estimated "residuals" $\hat{\varepsilon}_i$ can be obtained as follows:

$$\frac{1}{mK}\sum_{\nu=1}^{m}\sum_{k=1}^{K}\left\{\frac{\delta_i y_i}{\hat{p}(x_i)} + \left[1 - \frac{\delta_i}{\hat{p}(x_i)}\right]\tilde{Y}_{i\nu} - x_i^T\hat{\boldsymbol{\beta}} - \hat{b}_k\right\}.$$

III. Calculate $\hat{f}(v) = (nb)^{-1}\sum_{i=1}^{n}\mathscr{K}\{(v - \hat{\varepsilon}_i)/b\}$, where $\mathscr{K}(\cdot)$ is a univariate probability density function. Following Silverman (1986), we choose an optimal bandwidth b given by

$$b = 0.9n^{-1/5}\min\left\{\text{SD}(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n), \frac{\text{IQR}(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)}{1.34}\right\}, \quad (3.1)$$

where "SD" and "IQR" represent the sample standard deviation and interquartile range.

IV. Estimate $f(b_\tau)$ by $\hat{f}(\hat{b}_\tau)$, where $\hat{b}_\tau$ is the sample $100\tau\%$ quantile of $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$.

V. Substituting $\hat{f}(\hat{b}_\tau)$ into $\boldsymbol{\omega}_* = W^{-1}\mathbf{f}/\|W^{-1/2}\mathbf{f}\|$ to obtain an estimated optimal weight, denoted by $\hat{\boldsymbol{\omega}}_*$.

3.2 Majorize-Minimize algorithm

We develop an efficient computational procedure based on the Majorize-Minimize (MM) algorithm (Hunter & Lange, 2000) to optimize (2.8). Since extension to other cases is trivial, we omit them for the sake of space. It is difficult to minimize $\mathcal{G}_n(\boldsymbol{b}, \boldsymbol{\beta})$ because this function may possess multiple minima and $\rho_{\tau_k}(\cdot)$ and $|\beta_j|$ are not differentiable. The key idea of the MM algorithm is to approximate a complex optimization problem by a sequence of simple optimization problems, while the solutions of the new optimization problems converge to a solution of the original optimization problem. Here, we define $\boldsymbol{\eta} = (\boldsymbol{b}, \boldsymbol{\beta})$, $r_{ik} = y_i - x_i^T\boldsymbol{\beta} - b_k$ and $r_{i\nu k} = \widetilde{Y}_{i\nu} - x_i^T\boldsymbol{\beta} - b_k$. At the sth iteration, instead of directly minimizing $\mathcal{G}_n(\boldsymbol{b}, \boldsymbol{\beta})$, we minimize another objective function given by

$$Q_{\lambda_n}(\boldsymbol{\eta}|\boldsymbol{\eta}^{(s)}) = Q_\varepsilon(\boldsymbol{\eta}|\boldsymbol{\eta}^{(s)}) + n\sum_{j=1}^{d}\mathcal{P}'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|, \qquad (3.2)$$

where $\boldsymbol{\eta}^{(s)}$ denotes estimate of $\boldsymbol{\eta}$ at the sth iteration and $Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ is the surrogate function that majorizes $\mathcal{G}_n(\boldsymbol{b}, \boldsymbol{\beta})$ at $\boldsymbol{\eta}^{(s)}$. Let c_1 and c_2 be two constants. For any $\varepsilon > 0$, $Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ can be written as the sum of $Q_{1\varepsilon}(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ and $Q_{2\varepsilon}(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$, where

$$Q_{1\varepsilon}(\boldsymbol{\eta}|\boldsymbol{\eta}^{(s)}) = \frac{1}{4}\sum_{k=1}^{K}\sum_{i=1}^{n}\frac{\omega_k\delta_i}{\hat{p}(x_i)}\left[\frac{r_{ik}^2}{\varepsilon + |r_{ik}^{(s)}|} + (4\tau_k - 2)r_{ik} + c_1\right],$$

$$Q_{2\varepsilon}(\boldsymbol{\eta}|\boldsymbol{\eta}^{(s)}) = \frac{1}{4}\sum_{k=1}^{K}\sum_{i=1}^{n}\left\{1 - \frac{\delta_i}{\hat{p}(x_i)}\right\}\frac{\omega_k}{m}\sum_{\nu=1}^{m}\left[\frac{r_{i\nu k}^2}{\varepsilon + |r_{i\nu k}^{(s)}|} + (4\tau_k - 2)r_{i\nu k} + c_2\right].$$

Then, we calculate $\boldsymbol{\eta}^{(s+1)} = \arg\min_{\boldsymbol{\eta}} Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$. This process is repeated until the absolute difference between $\boldsymbol{\eta}^{(s)}$ and $\boldsymbol{\eta}^{(s+1)}$ is smaller than a prespecified small constant.

In our computational procedure, we also consider a quadratic approximation of $Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ in order to speed up the computation in (3.2). Specifically, we calculate the first-order and second-order derivatives of $Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ and denote them as $G = \partial_\eta Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)})$ and Hessian matrix $\mathcal{H} = \partial_\eta^2 Q_\varepsilon(\boldsymbol{\eta}|\boldsymbol{\eta}^{(s)})$, respectively. By following the Gauss-Newton approach (see, Kennedy & Gentle, 1980), we approximate $Q_\varepsilon(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(s)}) \approx 0.5(\mathbb{Y} - \mathbb{X}\,\boldsymbol{\eta})^T(\mathbb{Y} - \mathbb{X}\,\boldsymbol{\eta})$, where $\mathbb{X}^T\mathbb{X}$ is the Cholesky decomposition of $\mathcal{H}$ and $\mathbb{Y} = (\mathbb{X}^T)^{-1}(\mathcal{H}\boldsymbol{\eta} - G)$ is a pseudo response vector. By using the second-order Taylor expansion, at each iteration, we then minimize an objective function given by

$$\frac{1}{2}(\mathbb{Y} - \mathbb{X}\boldsymbol{\eta})^T(\mathbb{Y} - \mathbb{X}\boldsymbol{\eta}) + n\sum_{j=1}^{d}\mathscr{P}'_{\lambda_n}(|\beta_j^{(0)}|)|\beta_j|. \quad (3.3)$$

Since (3.3) is a simple penalized least square problem, it can be easily minimized by using many existing optimization algorithms. This algorithm leads to exact zeros for regression coefficients and converges quickly based on our empirical experience.

3.3 Penalty parameter selection procedure

To ensure good properties of $\hat{\boldsymbol{\eta}}_{\lambda_n}$, the penalty parameter λ_n has to be appropriately selected. Various techniques have been proposed in previous studies, such as the generalized cross-validation and BIC selectors. These criteria cannot be easily computed in the presence of missing data since $\mathscr{G}_n(\boldsymbol{b}, \boldsymbol{\beta})$ involves intractable multiple imputations. Instead, we develop an $\mathrm{IC_Q}$-type criterion (Ibrahim et al., 2008; Garcia et al., 2010) to select the penalty parameter. The $\mathrm{IC_Q}$ criterion selects the optimal λ_n by minimizing

$$\mathrm{IC_Q}(\lambda_n) = 2Q_\varepsilon(\hat{\boldsymbol{\eta}}_{\lambda_n}|\hat{\boldsymbol{\eta}}_0) + c_n(\hat{\boldsymbol{\eta}}_{\lambda_n}),$$

where $\hat{\boldsymbol{\eta}}_0 = \arg\min_{\boldsymbol{b},\boldsymbol{\beta}}\mathscr{S}_{n\omega}(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\omega}, \mathscr{T})$ with $\tilde{Y}_{i\nu} \sim \hat{F}_p(y \mid x_i)$ and $c_n(\boldsymbol{\eta})$ is a function of the data and the fitted model. For instance, if $c_n = 2d'$, where d' is the total number of parameters in the model, then we obtain an AIC-type criterion. Alternatively, we obtain a BIC-type criterion when $c_n(\boldsymbol{\eta}) = \log(n)df_{\lambda_n}/n$, where df_{λ_n} is the number of nonzero coefficients in the fitted model. Moreover, in the absence of missing data, $\mathrm{IC_Q}(\lambda_n)$ reduces to the usual AIC and BIC criteria. In our framework, we choose $c_n(\hat{\boldsymbol{\eta}}_{\lambda_n}) = df(\lambda_n)\log(nK)/nK$, where $df(\lambda_n)$ is the number of zero residuals based on the proposed AIPW-MI.

4 Simulations

We conducted a set of simulations to examine the finite sample performance of the proposed estimators when responses are missing at random. The simulation data was simulated from model (2.1), in which we set $\boldsymbol{\beta}^* = (1, 1.5, 0.5)^T$ and generated $x_i = (x_{i1}, x_{i2}, x_{i3})^T$ from $\mathscr{N}(0, \Sigma_\mathbf{x})$ with $\Sigma_\mathbf{x} = (0.5^{|i-j|})$ for $1 \le i, j \le 3$. The missing y_is' were simulated according to the following four scenarios: M1. $p(x_i) = 0.6$; M2. $\mathrm{logit}\{p(x_i)\} = \gamma_0 + \gamma_1\sin(x_{i1}) + \gamma_2 x_{i2}^2 + \gamma_3 x_{i1}x_{i3}$ with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (0.5, 0.50, 0.20, 0.25)$; M3. $\mathrm{logit}\{p(x_i)\} = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3}$ with $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (1.5, 0.50, 0.50, 0.05)$; and M4. $p(x_i) = 0.5 + 0.5\{0.1(x_{i1} - 1)^2 + 0.1(x_{i2} - 1)^2 + |x_{i1} - 1|\}$ if $0.1(x_{i1} - 1)^2 + 0.1(x_{i2} - 1)^2 + |x_{i1} - 1| < 1$, and 0.5 elsewhere. The average proportions of missing data corresponding to M1–M4 are about 40%, 32.49%, 21.34%, and 39.41%, respectively. We set the working model of the missing data mechanism as $\mathrm{logit}\{p(x_i)\} = \gamma_0^* + \gamma_1^{*T}x_i$. In this case, models M1 and M3 are correctly specified, whereas M2 and M4 are misspecified. We use M2 and M4 to investigate the robustness of our PCQR and PAQE estimators. Moreover, we consider four different error distributions included *C1.* $\mathscr{N}(0, 3)$; *C2.* $t(3)$; *C3.* standard Cauchy distribution; and *C4.* $\varepsilon = (1 + \mathbf{z})\mathbf{w}$, in which $\mathbf{z} \sim$ Bernoulli(0.5) and $\mathbf{w} \sim \mathscr{N}(0, 1)$.

For each case, we simulated 500 incomplete data sets with $n = 200$. For each simulated data set, we used the reweighted least squares iterative algorithm to compute $\hat{p}(x_i)$ and compared five estimators of $\boldsymbol{\beta}^*$ as follows. The first one, denoted as S0, is the least square estimator using regression imputation as the solution to $\sum_{i=1}^{n} x_i\{\delta_i \hat{p}(x_i)^{-1} y_i + [1 - \delta_i \hat{p}(x_i)^{-1}] x_i^T \hat{\boldsymbol{\beta}}_{cc}^{ls} - x_i^T \boldsymbol{\beta}\} = 0$, where $\hat{\boldsymbol{\beta}}_{cc}^{ls}$ is the least square estimator of $\boldsymbol{\beta}^*$ using the complete-case analysis. The second one, denoted as S1, is the proposed PCQR estimator. The third one, denoted as S2, is the proposed PQAE estimator. The fourth one, denoted as S3, is based on Wang & Feng's (2012) approach, given by $\hat{\boldsymbol{\beta}}^{\mathrm{MI}} = m^{-1} \sum_{\nu=1}^{m} \hat{\boldsymbol{\beta}}_{\nu}^{\mathrm{MI}}$, where

$$\hat{\boldsymbol{\beta}}_{\nu}^{\mathrm{MI}} = \underset{b_1, \dots b_K, \boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \sum_{i=1}^{n} \left\{ \delta_i \rho_{\tau_k}(r_{ik}) + (1 - \delta_i) \rho_{\tau_k}(r_{i\nu k}) \right\},$$

in which $r_{ik} = y_i - x_i^T \boldsymbol{\beta} - b_k$, $r_{i\nu k} = \tilde{Y}_{i\nu} - x_i^T \boldsymbol{\beta} - b_k$, and $\tilde{Y}_{i\nu}$ is the imputed value for missing Y_i in the ν-th imputation for $\nu = 1, \dots, m$. Although Wang & Feng's multiple imputation is developed for M-regression models with censored covariates, it can be extended to the analysis of linear models with missing response at random. The fifth one, denoted as S4, is the CQR estimator based on complete-case analysis.

For the estimation of the optimal weights for S1 and S2, we selected the optimal bandwidth b according to (3.1). Following Bradic et al. (2011), we set the number of quantiles to be $K = 9$ and the quantile vector $\mathscr{T} = (0.1, 0.2, \dots, 0.9)$. To implement our proposed two approaches, and Wang & Feng's approach, we imputed y from the parametrically estimated conditional quantile of $\mathbf{y}$ given $\mathbf{x}$, and set $m = 20$. Specifically, in the νth imputation for $\nu = 1, \dots, m$, for missing scenarios M1 and M2, we used conditional quantile $Q_{\mathbf{y}}(\tau^{\nu} | \mathbf{x}) = \boldsymbol{\theta}_0^*(\tau^{\nu}) + \mathbf{x}^T \boldsymbol{\theta}_1^*(\tau^{\nu})$ to impute the missing or observed y_i by using $\tilde{Y}_{i\nu} = \hat{\boldsymbol{\theta}}_0(\tau^{\nu}) + x_i^T \hat{\boldsymbol{\theta}}_1(\tau^{\nu})$; for M3 and M4, we used conditional quantile $Q_{\mathbf{y}}(\tau^{\nu} | \mathbf{x}) = \boldsymbol{\theta}_0^*(\tau^{\nu}) + \exp(\mathbf{x}^T \boldsymbol{\theta}_1^*(\tau^{\nu}))$ to impute the missing or observed y_i by using $\tilde{Y}_{i\nu} = \hat{\boldsymbol{\theta}}_0(\tau^{\nu}) + \exp(x^T \hat{\boldsymbol{\theta}}_1(\tau^{\nu}))$ in order to investigate the robustness of our proposed method under the violation of linearity assumption in the parametric multiple imputation. In the ν-th imputation procedure, $\boldsymbol{\tau}^{\nu}$ was randomly generated from the uniform distribution on (0, 1) and estimators $\hat{\boldsymbol{\theta}}_0(\boldsymbol{\tau}^{\nu})$ and $\hat{\boldsymbol{\theta}}_1(\boldsymbol{\tau}^{\nu})$ for both conditional quantile regression models are computed by using Majorize-Minimize algorithm for single quantile regression.

Tables 1 and 2 report the bias, the root mean square (RMS) and the standard deviation estimate (SD) of the five estimators S0–S4 under different scenarios. We have the following observations. (i) Under the correctly specified model for imputation, Wang & Feng's method outperforms all other estimation methods in terms of RMS and SD. However, under model misspecification for imputation, estimator computed by Wang & Feng's method suffers from severe bias resulting in much larger RMS and SD compared with other estimation methods. (ii) Under all scenarios, our PCQR and PAQE estimators are quite close to the true values, and the RMS is close to its corresponding SD. Therefore, our proposed methods are not only efficient, but also robust to the misspecified linearity assumption and missing mechanism. Such robustness may be due to the use of the parametrically estimated propensity scores. (iii) The proposed PCQR and PAQE estimates outperform the least square estimate for all error distributions and such efficiency gain can be substantial in most cases. As expected, under the Cauchy distribution, the least square estimator based on the regression imputation suffers from severe bias. It is also noteworthy that least square method has the worst performance among all methods even under normal error. This phenomena can be ascribed to the sensitivity of a mean regression procedure to either near-zero or extreme values of the estimated propensity score.

5 AIDS Clinical Trials Group 175 study

To illustrate our proposed methods, we consider a data set taken from 2139 patients from the AIDS Clinical Trials Group (ACTG) protocol 175 (Hammer et al., 1996), a study that randomized patients to four antiretroviral regimes in equal proportions: zidovudine (ZDV) monotherapy, ZDV +didanosine (ddI), ZDV+zalcitabine, and ddI monotherapy. Let $\mathbf{y}$ be the CD4 count at 96 ± 5 weeks and let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_6)^T$ be the six baseline characteristics: Age, Weight, CD4 counts at baseline (CD40), CD4 counts at 20 ± 5 weeks (CD420), CD8 counts at baseline (CD80) and CD8 counts at 20 ± 5 weeks (CD820). Since death and dropout, data on $\mathbf{y}$ have been missing, but $\mathbf{x}$ were fully observed. The response rate of $\mathbf{y}$ is about 62.74%. We are interested in understanding whether the CD4 count at 96 ± 5 weeks ($\mathbf{y}$) depends on the possible influential factors ($\mathbf{x}$). To do this, we fit the linear regression model $\mathbf{y} = \mathbf{x}^T\boldsymbol{\beta}^* + \varepsilon$, where ε is random error with an unknown distribution function. Following Hu et al. (2010), it may be reasonable to assume that $\mathbf{y}$ is missing at random in that the baseline characteristics can faithfully predict the missingness in $\mathbf{y}$. All of the covariates are standardized. In addition, we consider the following model for missing data mechanism: $\mathrm{logit}\{p(\mathbf{x})\} = \gamma_0 + \gamma_1^T\mathbf{x}$. The response model is estimated by using the reweighted least squares iterative algorithm.

Similar to the simulation studies, we also analyzed the real data by using the five estimation methods S0–S4. We used the same number of quantiles K and quantiles τ as those used in simulation study. For all imputation procedures, we imputed $\mathbf{y}$ from the estimated linear conditional quantiles of $\mathbf{y}$ given all covariates, and set imputation time $m = 10$. We also applied the variable selection method based on the SCAD and adaptive LASSO regularization to the five estimation methods. The proposed $\mathrm{IC_Q}$ criterion was used to select the penalty parameter λ_n for all methods except that the penalized least square estimation was based on the cross validation.

Table 3 presents the estimation results. For the unpenalized estimates, their standard errors (SE) were computed by using the bootstrap method. The parameter estimates of each parameter are quite close to each other for all estimation methods, but our PCQR and PAQE methods yield smaller standard errors compared with the least square method and Wang & Feng's (2012) method. All estimators indicate that the CD8 counts at 20 ± 5 weeks have negative effects on the CD4 count at 96 ± 5 weeks. Adaptive LASSO and SCAD penalties perform similarly in variable selection. The penalized least square method yields the smallest models, and our proposed penalized PCQR and penalized PAQE methods yield sparser models than Wang & Feng's method. All proposed penalized estimators identify CD4 counts at baseline (CD40) and CD4 counts at 20 ± 5 weeks (CD420) as significant predictors.

6 Conclusions

In this paper, we have proposed to impute both missing and observed responses by using their estimated conditional quantiles given the observed data. The augmented inverse probability weighted approach is used to reconstruct the check function that defines a regression parameter. Efficient and robust parametric estimation for the regression coefficients is based on the weighted composite quantile regression estimation and weighted quantile average estimation procedures. Although we assume a linear form of the conditional quantiles of $\mathbf{x}$ in the imputation step, our empirical studies suggest that the proposed estimators are less sensitive to potential model misspecification than those obtained from misspecified parametric likelihood. For more flexible imputation, nonparametric or semiparametric quantile models can also be employed.

In this paper, we propose using complete-case estimator obtained from (2.3) to impute the missing response. Such imputation procedure is valid because we assume that only theresponse is missing at random. When some covariates are missing at random and the propensity depends on the response, our methods do not work any more because the proposed complete-case estimator will be biased and inconsistent. In this case, to mitigate the effects of missing covariates data and improve the the performance of the proposed methods, we can develop a conditional quantiles imputation procedure based on inverse probability weighted and/or AIPW approaches. Extension of the method to missing covariates is interesting and challenging, which is a topic of our future research. The proposed imputation procedure also has some important implications. It can be easily extended to more complicated models, such as partial linear models, single index models, varying-coefficient models and even survival data analysis, among others. An important extension is to consider the estimating equation with missing data. Further study is needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to the Editor, an Associate Editor and referees for constructive suggestions that greatly improved the paper. All authors made equal contributions to this work and the order of authorship carries only alphabetical significance. Tang's work was supported by grants from the National Science Fund for Distinguished Young Scholars of China (11225103). The research of Dr. Zhu was supported by NSF grants SES-1357666 and DMS-1407655, NIH grants MH086633, and a grant from Cancer Prevention Research Institute of Texas.

References

Bradic J, Fan J, Wang W. Penalized composite quasi-likelihood for ultrahigh-dimensional variable selection. J. Roy. Statist. Soc. Ser. B. 2011; 73:325–349.

Chen X, Wan A, Zhou Y. Efficient quantile regression analysis with missing observations. J. Amer. Statist. Assoc. 2015; 110:723–741.

FanJ, , LiR. Proceedings of the Madrid International Congress of Mathematicians 2006 III 595–622EMS; Zurich: 2006Statistical challenges with high dimensionality: Feature selection in knowledge discovery.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 2001; 96:1348–1360.

Garcia RI, Ibrahim JG, Zhu H. Variable selection for the cox regression model with covariates missing at random. Biometrics. 2010; 66:97–104. [PubMed: 19459831]

Hammer SM, et al. A trial comparing nucleotide monotherapy with combined therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine. 1996; 335:1081–1090. [PubMed: 8813038]

Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models (plus appendix). The American Statistician. 2007; 61:79–90. [PubMed: 17401454]

Horton NJ, Laird NM, Zahner GEP. Use of multiple informant data as a predictor in psychiatric epidemiology. International Journal of Methods in Psychiatric Research. 1999; 8:6–18.

Hu Z, Follmann DA, Qin J. Semiparametric dimension reduction estimation for mean response with missing data. Biometrika. 2010; 97:305–319. [PubMed: 23049121]

Hunter DR, Lange K. Quantile regression via an MM algorithm. J. Comput. Graph. Statist. 2000; 9:60–77.

Ibrahim JG, Chen MH, Lipsitz SR, Herring AH. Missing data methods for generalized linear models: a comparative review. J. Amer. Statist. Assoc. 2005; 100:332–346.

Ibrahim JG, Zhu HT, Tang NS. Model selection criteria for missing data problems via the EM algorithm. J. Amer. Statist. Assoc. 2008; 103:1648–1658.

Jiang X, Jiang J, Song X. Oracle model selection for nonlinear models based on weighted composite quantile regression. Statist. Sinica. 2012; 22:1479–1506.

KennedyWJ, , GentleJE. Statistical computingNew York: Marcel Dekker; 1980

KoenkerR. Quantile regressionNew York: Cambridge University Press; 2005

Koenker R, Bassett G. Regression quantiles. Econometrica. 1978; 46:33–50.

LittleRJA, , RubinDB. Statistical analysis with missing data2. New York: Wiley; 2002

Portnoy S, Koenker R. Adaptive L-estimation of linear models. Ann. Statist. 1989; 17:362–381.

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J. Amer. Statist. Assoc. 1994; 89:846–866.

RubinDB. Multiple imputation for nonresponse in surveysNew York: Wiley & Sons; 1987

SilvermanBW. Density estimationChapman and Hall; London: 1986

Tang CY, Qin Y. An efficient empirical likelihood approach for estimating equations with missing data. Biometrika. 2012; 99:1001–1007.

Tibshirani R. Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B. 1996; 58:267–288.

Wang H, Feng X. Multiple imputation for M-regression with censored covariates. J. Amer. Statist. Assoc. 2012; 107:194–204.

Wei Y, Ma Y, Carroll RJ. Multiple imputation in quantile regression. Biometrika. 2012; 99:423–438. [PubMed: 24944347]

White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50:1–25.

Yoon J. Quantile regression analysis with missing response, with application to inequality measure and data combination. Working paper. 2010

Zhao Z, Xiao Z. Efficient regressions via optimally combining quantile information. Econometric Theory. 2014; 30:1272–1314. [PubMed: 25484481]

Zou H, Li R. One-step estimation in nonconcave penalized likelihood models. Ann. Statist. 2008; 36:1509–1533.

Zou H. The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. 2006; 101:1418–1429.

Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. Ann. Statist. 2008; 36:1108–1126.

Table 1

Simulation results under missing scenarios M1 and M2.

		$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
c.d.f.	**Methods**	**Bias**	**RMS**	**SD**	**Bias**	**RMS**	**SD**	**Bias**	**RMS**	**SD**
Missing Pattern M1										
C1	S0	−0.005	0.188	0.188	0.004	0.196	0.196	−0.012	0.177	0.177
	S1	−0.010	0.189	0.189	0.001	0.187	0.187	−0.014	0.171	0.170
	S2	−0.010	0.184	0.184	−0.011	0.187	0.187	−0.017	0.163	0.162
	S3	−0.018	0.156	0.155	−0.014	0.158	0.157	−0.019	0.138	0.136
	S4	−0.010	0.184	0.183	−0.001	0.183	0.183	−0.013	0.166	0.166
C2	S0	0.007	0.177	0.177	−0.005	0.205	0.205	−0.010	0.186	0.185
	S1	0.000	0.125	0.125	−0.007	0.139	0.139	−0.013	0.128	0.128
	S2	−0.006	0.115	0.115	−0.009	0.127	0.127	−0.019	0.116	0.115
	S3	−0.009	0.105	0.104	−0.012	0.116	0.116	−0.020	0.107	0.106
	S4	0.003	0.127	0.127	−0.003	0.140	0.140	−0.015	0.133	0.133
C3	S0	3042	68012	68012	−71	1581	1581	1444	32307	32307
	S1	−0.021	0.209	0.208	−0.006	0.225	0.225	−0.013	0.202	0.202
	S2	−0.020	0.155	0.154	−0.015	0.173	0.173	−0.013	0.146	0.146
	S3	−0.024	0.139	0.137	−0.012	0.153	0.153	−0.018	0.134	0.133
	S4	−0.020	0.182	0.181	−0.008	0.197	0.197	−0.014	0.177	0.176
C4	S0	−0.006	0.163	0.163	0.010	0.191	0.191	0.001	0.164	0.164
	S1	−0.004	0.161	0.161	0.007	0.183	0.183	0.003	0.161	0.161
	S2	−0.004	0.151	0.151	0.001	0.167	0.167	−0.006	0.150	0.151
	S3	−0.013	0.122	0.121	−0.010	0.138	0.138	−0.011	0.121	0.120
	S4	−0.004	0.157	0.157	0.006	0.176	0.176	0.002	0.155	0.155
Missing Pattern M2										
C1	S0	−0.002	0.178	0.178	0.006	0.193	0.193	−0.006	0.183	0.183
	S1	−0.008	0.172	0.172	0.000	0.187	0.188	−0.002	0.177	0.177
	S2	−0.012	0.166	0.166	−0.009	0.182	0.182	−0.011	0.171	0.171
	S3	−0.034	0.145	0.141	−0.002	0.161	0.161	−0.011	0.147	0.147
	S4	−0.004	0.164	0.164	−0.002	0.181	0.181	−0.009	0.170	0.170

c.d.f.	Methods	$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
		Bias	**RMS**	**SD**	**Bias**	**RMS**	**SD**	**Bias**	**RMS**	**SD**
C2	S0	−0.006	0.178	0.178	0.004	0.184	0.184	−0.007	0.165	0.165
	S1	−0.006	0.124	0.124	−0.002	0.136	0.137	−0.002	0.120	0.120
	S2	−0.006	0.111	0.111	−0.009	0.131	0.131	−0.008	0.117	0.117
	S3	−0.036	0.105	0.099	−0.008	0.114	0.114	−0.007	0.099	0.099
	S4	−0.006	0.123	0.123	−0.003	0.136	0.136	−0.005	0.121	0.121
C3	S0	−1.093	28.262	28.269	−0.693	29.664	29.685	0.420	23.377	23.396
	S1	−0.005	0.175	0.175	−0.017	0.180	0.179	−0.001	0.181	0.181
	S2	−0.013	0.142	0.141	−0.019	0.155	0.154	−0.018	0.136	0.135
	S3	−0.034	0.137	0.133	−0.014	0.149	0.148	−0.013	0.134	0.134
	S4	−0.008	0.178	0.178	−0.016	0.185	0.184	−0.008	0.188	0.188
C4	S0	−0.003	0.160	0.160	0.002	0.177	0.178	−0.008	0.161	0.161
	S1	−0.004	0.152	0.152	0.002	0.170	0.171	0.000	0.153	0.153
	S2	−0.006	0.145	0.145	−0.008	0.168	0.168	−0.005	0.145	0.145
	S3	−0.036	0.125	0.120	−0.008	0.143	0.143	−0.008	0.120	0.120
	S4	−0.002	0.152	0.152	−0.002	0.169	0.169	−0.006	0.149	0.149

Table 2

Simulation results under missing scenarios M3 and M4.

c.d.f.	Methods	$\hat{\beta}_1$ Bias	$\hat{\beta}_1$ RMS	$\hat{\beta}_1$ SD	$\hat{\beta}_2$ Bias	$\hat{\beta}_2$ RMS	$\hat{\beta}_2$ SD	$\hat{\beta}_3$ Bias	$\hat{\beta}_3$ RMS	$\hat{\beta}_3$ SD
					Missing Pattern M3					
C1	S0	0.011	0.165	0.164	0.016	0.190	0.190	−0.002	0.176	0.176
	S1	0.005	0.178	0.178	−0.002	0.202	0.202	−0.011	0.181	0.180
	S2	0.001	0.183	0.183	−0.001	0.206	0.206	−0.012	0.181	0.181
	S3	−0.219	0.272	0.162	−0.247	0.319	0.202	−0.052	0.188	0.181
	S4	0.008	0.145	0.145	0.007	0.171	0.171	−0.010	0.165	0.165
C2	S0	0.005	0.166	0.166	−0.017	0.186	0.186	0.012	0.159	0.159
	S1	−0.004	0.136	0.136	−0.018	0.143	0.142	0.001	0.132	0.132
	S2	−0.006	0.128	0.127	−0.024	0.141	0.139	−0.002	0.124	0.124
	S3	−0.189	0.232	0.134	−0.227	0.280	0.163	−0.033	0.142	0.138
	S4	0.002	0.114	0.114	−0.011	0.130	0.129	0.004	0.114	0.114
C3	S0	−2.004	30.869	30.835	2.553	24.040	23.928	0.871	31.701	31.721
	S1	−0.029	0.188	0.186	−0.023	0.215	0.214	0.000	0.185	0.186
	S2	−0.030	0.164	0.162	−0.041	0.203	0.199	−0.006	0.202	0.202
	S3	−0.206	0.273	0.179	−0.225	0.303	0.202	−0.033	0.180	0.177
	S4	−0.020	0.156	0.155	−0.011	0.177	0.177	0.001	0.152	0.152
C4	S0	−0.007	0.152	0.152	0.004	0.178	0.178	−0.003	0.149	0.149
	S1	−0.012	0.163	0.162	0.001	0.189	0.189	−0.010	0.153	0.153
	S2	−0.014	0.163	0.163	−0.001	0.188	0.189	−0.012	0.156	0.155
	S3	−0.221	0.272	0.158	−0.237	0.299	0.183	−0.044	0.162	0.156
	S4	−0.010	0.141	0.141	0.007	0.163	0.163	−0.001	0.137	0.138
					Missing Pattern M4					
C1	S0	0.012	0.201	0.201	−0.004	0.215	0.215	−0.007	0.186	0.186
	S1	0.029	0.187	0.185	−0.036	0.188	0.185	−0.030	0.161	0.158
	S2	0.033	0.211	0.209	−0.041	0.207	0.203	−0.030	0.185	0.183
	S3	−0.128	0.264	0.231	−0.016	0.266	0.266	0.002	0.207	0.207
	S4	0.011	0.187	0.187	−0.004	0.200	0.200	−0.011	0.166	0.166

c.d.f.	Methods	$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
		Bias	RMS	SD	Bias	RMS	SD	Bias	RMS	SD
C2	S0	−0.010	0.194	0.194	−0.010	0.215	0.215	0.004	0.201	0.201
	S1	0.028	0.154	0.152	−0.039	0.158	0.154	−0.019	0.131	0.130
	S2	0.009	0.137	0.136	−0.041	0.153	0.147	−0.018	0.125	0.124
	S3	−0.131	0.227	0.185	−0.022	0.219	0.219	0.004	0.176	0.176
	S4	0.000	0.134	0.134	−0.016	0.153	0.152	−0.004	0.129	0.129
C3	S0	4.305	114.741	114.775	−3.204	92.973	93.011	2.193	41.506	41.489
	S1	0.079	0.287	0.276	−0.035	0.274	0.272	−0.031	0.264	0.262
	S2	−0.005	0.230	0.230	−0.052	0.234	0.229	−0.032	0.205	0.203
	S3	−0.118	0.267	0.239	0.021	0.282	0.281	0.020	0.247	0.247
	S4	0.000	0.179	0.179	−0.010	0.202	0.202	−0.003	0.182	0.182
C4	S0	−0.003	0.180	0.181	0.001	0.207	0.207	0.006	0.170	0.170
	S1	0.044	0.191	0.186	−0.026	0.201	0.200	−0.007	0.170	0.170
	S2	0.006	0.168	0.168	−0.026	0.183	0.181	−0.007	0.156	0.156
	S3	−0.147	0.255	0.209	−0.011	0.244	0.244	0.010	0.197	0.197
	S4	−0.004	0.167	0.167	−0.001	0.193	0.194	0.009	0.161	0.161

Table 3

Estimates of ACTG 175 Data.

	Un-penalized Estimator									
	S0		S1		S2		S3		S4	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
$\mathbf{x}_1$	2.923	4.420	1.761	1.971	1.045	2.952	4.727	4.853	1.854	1.839
$\mathbf{x}_2$	2.073	4.429	3.178	1.982	3.009	3.027	−3.726	6.554	3.407	1.793
$\mathbf{x}_3$	16.147	5.977	14.796	2.882	12.294	3.620	12.795	6.622	15.790	2.643
$\mathbf{x}_4$	48.265	6.008	46.530	2.633	43.551	6.077	62.944	6.083	46.847	2.557
$\mathbf{x}_5$	4.671	7.513	2.497	3.133	2.022	4.555	16.865	7.911	3.894	3.083
$\mathbf{x}_6$	−11.652	7.501	−10.817	3.490	−8.309	8.395	−23.328	8.827	−12.330	3.529
	Penalized Estimator									
	S0		S1		S2		S3		S4	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
$\mathbf{x}_1$	0.000	0.000	0.000	0.000	0.000	0.000	2.1951	0.000	0.000	0.000
$\mathbf{x}_2$	0.000	0.000	0.000	0.000	0.000	0.000	3.111	0.000	0.000	0.558
$\mathbf{x}_3$	14.416	11.641	12.374	15.899	5.874	7.815	16.225	17.478	17.178	16.538
$\mathbf{x}_4$	46.554	37.845	46.116	45.437	47.799	47.147	47.453	46.278	45.147	17.758
$\mathbf{x}_5$	0.000	0.000	−0.000	0.000	0.000	0.000	3.500	0.000	0.000	0.000
$\mathbf{x}_6$	0.000	0.000	−3.340	−8.787	−0.532	−0.742	−11.374	−8.244	7.7797	−8.833

NOTE: In the column labeled “P1” presents variable selection using the adaptive LASSO, “P2” presents variable selection using the SCAD.