Received: 4 June 2022 | Revised: 24 March 2023 | Accepted: 2 April 2023

DOI: 10.1002/sim.9745

RESEARCH ARTICLE

Statistics in Medicine WILEY

Quantile partially linear additive model for data with dropouts and an application to modeling cognitive decline

Adam Maidman[1] | **Lan Wang**[2] | **Xiao-Hua Zhou**[3] | **Ben Sherwood**[4]

[1]School of Statistics, University of Minnesota, Minneapolis, Minnesota,

[2]Miami Herbert Business School, University of Miami, Coral Gables, Florida,

[3]Department of Biostatistics and Beijing International Center for Mathematical Research, Peking University, Beijing, China

[4]School of Business, University of Kansas, Lawrence, Kansas,

Correspondence
Ben Sherwood, School of Business, University of Kansas, Lawrence, KS, USA.
Email: ben.sherwood@ku.edu

The National Alzheimer's Coordinating Center Uniform Data Set includes test results from a battery of cognitive exams. Motivated by the need to model the cognitive ability of low-performing patients we create a composite score from ten tests and propose to model this score using a partially linear quantile regression model for longitudinal studies with non-ignorable dropouts. Quantile regression allows for modeling non-central tendencies. The partially linear model accommodates nonlinear relationships between some of the covariates and cognitive ability. The data set includes patients that leave the study prior to the conclusion. Ignoring such dropouts will result in biased estimates if the probability of dropout depends on the response. To handle this challenge, we propose a weighted quantile regression estimator where the weights are inversely proportional to the estimated probability a subject remains in the study. We prove that this weighted estimator is a consistent and efficient estimator of both linear and nonlinear effects.

KEYWORDS
longitudinal data, missing data, quantile regression, semiparametric

1 | INTRODUCTION

The Uniform Data Set (UDS), maintained by the National Alzheimer's Coordinating Center, contains longitudinal data of patients' demographic information and other variables that may be related to Alzheimer's disease. Understanding factors associated with cognitive decline is a major goal that researchers hope to achieve with the UDS. Patients are expected to have appointments roughly once a year for ten years and the appointments include a battery of cognitive exams. These tests measure attention, processing speed, executive function, episodic memory, and language. Analysis of these test results on cognitively normal men and women have been used to model the average performance given a patient's age, years of education, and gender.[1] The residual of an individual's performance on an exam can be used to assign a percentile, using the common assumptions of normality of errors and constant variance. This can be used to provide a baseline of performance for an individual that implicitly incorporates variables that were not included in the model, such as intelligence.[2] Alternatively, fitting quantile regression models for multiple quantiles can provide an individual's percentile without requiring the assumptions of normality or constant variance, which are not plausible when modeling many of the test scores.[3] A patient diagnosed with Alzheimer's disease is often thought to have had the disease long before the diagnosis. Using longitudinal data to identify patients with slight changes in their cognitive ability could potentially be useful

Abbreviations: IPW, inverse probability weighting; UDS, Uniform Data Set.

in earlier identifying Alzheimer's patients.[4] The aforementioned models which control for age are particularly important because there is expected to be some cognitive decline with age.[5] However, the previous models only used data from the patients' first visits, ignoring the repeated measurements.[2,3] In addition, they assume that the relationships between the variables are linear, while this article provides some evidence to the contrary. In this article, we propose a model of the conditional quantile of a composite of test scores that can accommodate longitudinal observations, missing data due to dropout, and nonlinear relationships. This model provides a flexible estimate of a conditional percentile that does not depend on specifying the distribution of error terms.

There exist several challenges in the UDS that create significant difficulties for statistical analysis. Cognitive ability is highly skewed and heterogeneous. Some of the covariates have nonlinear effects on cognitive ability. The longitudinal structure of the data needs to be handled properly. Lastly, patients tended to drop out of the study before its conclusion. Of the 5350 patients in the UDS who attended the first appointment, about 75% dropped out after four follow-up appointments. Only about 10% of patients remained in the study for eight follow-up visits, and less than 2% of the original 5350 patients attended all nine follow-up appointments. Additionally, the probability that a patient drops out may depend on his or her cognitive ability at previous visits or covariates. Ignoring any of these features in the data can result in biased analysis. Many of these features are not unique to the UDS and have been studied previously.[6,7] However, this problem has not been studied when all these challenges are present.

One of the appeals of quantile regression is it does not require any assumption about the distribution of the error terms. Thus, quite naturally, there has been a lot of work in further relaxing the linearity assumption. A very popular approach is to use B-splines to estimate an unknown function. Early work demonstrated that for a single covariate, using B-splines to estimate an unknown nonlinear function can achieve the same, optimal, rate of convergence as least squares regression with splines.[8,9] Since then, using splines to provide more flexibility to quantile regression has been an active area of research. Applications include, but are not limited to, varying coefficients,[10] single index models,[11] partially linear models,[12,13] and variable selection for nonlinear models.[14,15]

Research on longitudinal models for quantile regression has been an active area. One approach is to consider a fixed effect for each individual and use a lasso penalty to avoid potentially over-fitting the model.[16] Mixed effects quantile models are also available.[17] In this article, we take a GEE approach to estimating a quantile regression model using longitudinal data. Similar approaches have been used in partially linear models,[18] single index models,[19] and using a smoothed version of quantile loss to efficiently handle an extremely large number of predictors.[20]

Estimating the conditional quantile in the presence of missing data has been studied in the literature. For linear quantile models with cross-sectional data, two popular approaches are imputation[21,22] and inverse probability weighting (IPW).[23-25] IPW has been proposed to estimate conditional quantiles using longitudinal data with dropouts,[26,27] where the inverse of the probability of dropping out at a given time is used as the weight. The asymptotic normality of the estimator has been derived for the special case of modeling the median.[27] This work generalizes the IPW approach for data with dropouts to the nonlinear model and develops the asymptotic statistical theory. Ignoring the longitudinal structure of the data can result in biased estimation. Missing data is another inherent challenge that can yield biased estimates if not handled properly. Partial linear quantile regression is a popular tool for analysis. Despite this, the literature lacks a theoretically justified partially linear quantile regression estimator for the longitudinal setting with missing data.

Section 2 formally defines the dropout model and provides intuition for the proposed estimator. Asymptotic properties are presented in Section 3. We demonstrate the performance of our estimator with Monte Carlo simulations in Section 4 and analyze the UDS in Section 5. We conclude with a discussion in Section 6.

2 | ESTIMATION

Let Y_{ij} be the univariate, continuous response of the ith subject at the jth time point, where $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$. Let $\mathbf{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^\top \in \mathbb{R}^p$ and $\mathbf{Z}_{ij} = (Z_{ij1}, \dots, Z_{ijq})^\top \in [0,1]^q$ be the linear and nonlinear predictors, respectively, for the ith subject at the jth time point. Assuming nonlinear covariates are bounded is a fairly common assumption[9,28] and without loss of generality we assume the support is $[0, 1]$ for all nonlinear covariates. Define $g_d^*(\cdot)$ as the nonlinear function for the dth nonlinear covariate and $E[g_d^*(Z_{ijd})] = 0$, for the identifiability of the model. We consider the model of

$$Y_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}^* + \alpha^* + \sum_{d=1}^{q} g_d^*\left(Z_{ijd}\right) + \epsilon_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta}^* + g^*(\mathbf{Z}_{ij}) + \epsilon_{ij}, \tag{1}$$

where ϵ_{ij} and ϵ_{kj} are independent when $i \neq k$ and $P(\epsilon_{ij} < 0 | X_{ij}, Z_{ij}) = \tau$. Note that $\boldsymbol{\beta}^*$, α^*, $g_d^*(\cdot)$ and ϵ_{ij} all depend on τ, but for ease of notation we are not indexing them by τ. The nonlinear functions are unknown and estimated using degree s B-splines with k_n quasi-uniform internal knots and $J_n = k_n + s$ corresponding functions. Define $w_{d,k}(\cdot)$ as the kth B-spline for the dth variable, $\mathbf{w}_d(Z_{ijd}) = \left[w_{d,1}(Z_{ijd}), \ldots, w_{d,J_n}(Z_{ijd})\right]^\top \in \mathbb{R}^{J_n}$ as the vector of B-splines for the dth variable and $\mathbf{w}(Z_{ij}) = \left[1, \mathbf{w}_1(Z_{ij1}), \ldots, \mathbf{w}_d(Z_{ijd})\right]^\top \in \mathbb{R}^{dJ_n+1}$ as the vector of all spline functions for the ith subject at the jth time point.

We consider longitudinal data with dropout. Define R_{ij} as an indicator if the ith subject is observed at the jth time point, it takes a value of one if this is true and zero otherwise. We assume a dropout structure. Each subject is assumed to make their first appoint. If a subject misses an appointment then we assume they fail to show up at all following appointments. That is, if $R_{ij} = 0$ then $R_{ij'} = 0$ for all $j' > j$ and $R_{i1} = 1$ for all $i \in \{1, \ldots, n\}$. A missing at random structure is assumed where the probability of dropout at time j depends on observations before time point j. Define $\mathbf{T}_{ij} = \left[1, Y_{i1}, \ldots, Y_{i(j-1)}, \mathbf{X}_{i1}^\top, \ldots, \mathbf{X}_{i(j-1)}, \mathbf{Z}_{i1}^\top, \ldots \mathbf{Z}_{i(j-1)}^\top\right]^\top \in \mathbb{R}^{j+(j-1)(p+q)}$ as the vector of all variables for subject i observed before time point j and let $\mathbf{Y}_i^c, \mathbf{X}_i^c$, and $\mathbf{Z}_i^c$ represent the responses and predictors for the ith subject that would be observed if they never dropped out of the study. Define $\pi_{ij}^* = P(R_{ij} = 1 | \mathbf{Y}_i^c, \mathbf{X}_i^c, \mathbf{Z}_i^c)$, we assume a monotone missing at random pattern,

$$\begin{aligned} \pi_{i1}^* &= 1 \ \text{ for all } \ i \in \{1, \ldots, n\}, \\ \pi_{ij}^* &= P\left(R_{ij} = 1 | \mathbf{T}_{ij}\right) \ \text{ for all } \ i \in \{1, \ldots, n\} \ \text{ and } \ j \in \{2, \ldots, m\}, \\ P[R_{ij} = 1 | R_{i(j-1)} = 0] &= 0. \end{aligned}$$

To motivate our proposed estimator and provide intuition about why ignoring the dropout structure is problematic we consider the ideal case where the functions $g_d^*(\cdot)$ are known. Define $\rho_\tau(u) = u[\tau - I(u < 0)]$, which is the quantile loss function.[29] In this setting, we only need to estimate the linear coefficients and if the dropout structure is ignored one approach is to minimize with respect to $\boldsymbol{\beta}$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij}\rho_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta} - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]. \tag{2}$$

This is the standard linear quantile regression estimator but it will be biased in this setting because of the dropouts. Define $\psi_\tau(u) = \tau - I(u < 0)$, which is the derivative of $\rho_\tau(u)$ where it is defined. The expected value of the estimating equation at the true value of $\boldsymbol{\beta}^*$ is

$$\sum_{i=1}^{n}\sum_{j=1}^{m} E\left\{R_{ij}\mathbf{X}_{ij}\psi_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta}^* - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]\right\}. \tag{3}$$

If there are no dropouts, $R_{ij} = 1$ for all i and j, then

$$E\left\{R_{ij}\mathbf{X}_{ij}\psi_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta}^* - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]\right\} = E\left(\mathbf{X}_{ij}E\left\{\psi_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta}^* - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]\Big|\mathbf{X}_{ij}\right\}\right) = 0. \tag{4}$$

However, with dropouts we have

$$E\left\{R_{ij}\mathbf{X}_{ij}\psi_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta}^* - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]\right\} = E\left(\mathbf{X}_{ij}E\left\{\pi_{ij}^*\psi_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta}^* - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]\Big|\mathbf{X}_{ij}\right\}\right). \tag{5}$$

Note, π_{ij}^* depends on $Y_{i(j-1)}$. In addition, $Y_{i(j-1)}$ and Y_{ij} are likely to be correlated in a longitudinal setting, so π_{ij}^* cannot be pulled out of the conditional expectation and the expected value of (5) is likely to not be zero and the estimator minimizing (2) is likely to be inconsistent. This can be solved by instead minimizing,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{R_{ij}}{\pi_{ij}^*}\rho_\tau\left[Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta} - \sum_{d=1}^{q} g_d^*(Z_{ijd})\right]. \tag{6}$$

However, minimizing (6) is not realistic because it depends on knowing the true value of π_{ij}^* and $g_d^*(\cdot)$. We propose replacing π_{ij}^* with estimates, $\hat{\pi}_{ij}$, and using a B-spline transformation of $\mathbf{Z}_i$ to estimate the unknown functions, $g_d^*(\cdot)$.

Let $\eta_{ij}^* = P[R_{ij} = 1 | T_{ij}, R_{i1} = \cdots = R_{i(j-1)} = 1]$ and assume

$$\eta_{ij}^* = \frac{\exp\left(\mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j^*\right)}{1 + \exp\left(\mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j^*\right)} \text{ for all } i \in \{1, \dots, n\} \text{ and } j \in \{2, \dots, m\}. \tag{7}$$

Let $\hat{\boldsymbol{\gamma}}_j$ be the maximum likelihood estimator of $\boldsymbol{\gamma}_j^*$. Then $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_2, \dots, \hat{\boldsymbol{\gamma}}_m)^\top$ is the MLE from the following log-likelihood

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \sum_{j=2}^{m} R_{i(j-1)} \left[R_{ij} \mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j - \log\left(1 + e^{\mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j}\right) \right]. \tag{8}$$

Given the monotone missing pattern structure, $\pi_{ij}^* = \prod_{j=2}^{j-1} \eta_{ij}^*$ and therefore $\hat{\pi}_{ij} = \prod_{j=2}^{j-1} \hat{\eta}_{ij}$ and $\hat{\pi}_{i1} = 1$ for all $i \in \{1, \dots, n\}$. The proposed estimator is

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}, \boldsymbol{\xi} \in \mathbb{R}^{qJ_n}}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{R_{ij}}{\hat{\pi}_{ij}} \rho_\tau \left[Y_{ij} - \left\{ \mathbf{X}_{ij}^\top \boldsymbol{\beta} + \mathbf{w}(\mathbf{Z}_{ij})^\top \boldsymbol{\xi} \right\} \right], \tag{9}$$

where $\hat{\boldsymbol{\xi}}_k^\top \in \mathbb{R}^{J_n}$, $\hat{g}_k(Z_{ijk}) = \mathbf{w}(Z_{ijk})^\top \hat{\boldsymbol{\xi}}_k$, $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\xi}}_1, \dots, \hat{\boldsymbol{\xi}}_q)^\top \in \mathbb{R}^{qJ_n}$ and $\hat{g}(\mathbf{Z}_{ij}) = \mathbf{w}(Z_{ij})^\top \hat{\boldsymbol{\xi}}$. Similar partial linear conditional quantile estimators have been considered for the case of longitudinal data with no missing data[18] and cross-sectional data with missing covariates.[30] Linear longitudinal data with dropouts have also been considered, but weights are based on the probability of dropout instead of the probability of being observed at a given time point.[26,27] However, none of the previously cited literature considers a partially linear quantile regression longitudinal model with dropouts. Note the construction of the weights relies on the monotone missing pattern structure. Therefore the proposed estimator only handles values that are missing due to monotone dropout and does not handle subjects missing an appointment at time j but appearing at $j+1$ or a subject being observed, but with some missing data. We assume the design is balanced, that is the number of potential subject observations, m, remains the same for all subjects.

3 | ASYMPTOTIC PROPERTIES

This section outlines the asymptotic properties of the estimators from (9). First, we provide some technical definitions and conditions.

Definition 1. Let $r \equiv d + v$, where m is a positive integer and $v \in (0, 1]$. Define $\mathcal{H}_r$ as the collection of functions $h(\cdot)$ on $[0, 1]$ whose mth derivative $h^{(m)}(\cdot)$ satisfies the Hölder condition of order v. That is, for any $h(\cdot) \in \mathcal{H}_r$, there exists some positive constant C such that

$$\left| h^{(d)}(z') - h^{(d)}(z) \right| \leq C |z' - z|^v, \quad \forall \ 0 \leq z', z \leq 1. \tag{10}$$

Define the set $\mathcal{H}_r^q = \left\{ \sum_{d=1}^{q} h_k(z) | h_k \in \mathcal{H}_r \right\}$.

Let $t_k(\mathbf{z}) = \mathrm{E}(x_{ijk} | \mathbf{z}_{ij} = \mathbf{z})$ and

$$h_k^*(\cdot) = \underset{h_k \in \mathcal{H}_r^q}{\arg\inf} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{E}\left[f_{ij}(0 | \mathbf{X}_{ij}, \mathbf{Z}_{ij}) \left\{ X_{ijk} - h_k(\mathbf{Z}_{ij}) \right\}^2 \right].$$

The function $h_k^*(\cdot)$ is the weighted projection of $t_k(\cdot)$ into $\mathcal{H}_r^q$ under the L_2 norm, where the weights $f_{ij}(0 | \mathbf{x}_{ij}, \mathbf{z}_{ij})$ are included to account for possibly heterogeneous errors. Define $\delta_{ijk} = X_{ijk} - h_k^*(\mathbf{Z}_{ij})$, $\boldsymbol{\delta}_{ij} = (\delta_{ij1}, \dots, \delta_{ijp})^\top \in \mathbb{R}^p$, $\boldsymbol{\delta}_i = (\boldsymbol{\delta}_{i1}^\top, \dots, \boldsymbol{\delta}_{im}^\top)^\top \in \mathbb{R}^{m \times p}$ and $\boldsymbol{\Delta}_n = (\boldsymbol{\delta}_{i1}, \dots, \boldsymbol{\delta}_{nm})^\top \in \mathbb{R}^{mn \times p}$. Define $\mathbf{H}$ as the $mn \times p$ matrix with $H_{m(i-1)+j,k} = h_k^*(\mathbf{Z}_{ij})$. Then $\mathbf{X} = \mathbf{H} + \boldsymbol{\Delta}_n$. Similar definitions have been used for partially linear quantile regression models to derive asymptotic

properties of the linear coefficients, while accounting for the nonlinear estimates.[31,32] For a vector $\mathbf{a}$ define $||\mathbf{a}||_2$ as the Euclidean norm. For a square matrix $\mathbf{A}$ define $\lambda_{\max}(\mathbf{A})$ as the maximum eigenvalue of $\mathbf{A}$ and for any matrix $\mathbf{A}$ define $||\mathbf{A}||_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\top \mathbf{A})}$. Below are the conditions used to prove the theoretical results.

(C1) (Conditions on the random error) The random error ϵ_{ij} has the conditional distribution function F_{ij} and continuous conditional density function f_{ij}. The conditional density function, f_{ij}, is uniformly bounded away from 0 and infinity in a neighborhood of zero and its first derivative f'_{ij} has a uniform upper bound in a neighborhood of zero, for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$. In addition $E[\epsilon_{ij}^4] < \infty$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$.

(C2) (Conditions on the covariates) There exist positive constants M_1 and M_2 such that, for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$, and $k \in \{1, \dots, p\}$, $|X_{ijk}| \le M_1$, $\mathrm{E}[\delta_{ij}^4] \le M_2$ and $E[X_{ijk}] = 0$. Let f_{z_k} be the pdf of the kth nonlinear covariate. The joint density of the nonlinear predictors is absolutely continuous and the density, $f_{\mathbf{z}}(\mathbf{z})$, is bounded away from zero and infinity by positive constants. There exists positive constants c_1 and c_2 such that $c_1 \le f_{z_k}(z_k) \le c_2$ for all $k \in \{1, \dots, q\}$ and $z_k \in [0, 1]$. There exist finite positive constants C_1 and C_2 such that with probability one

$$C_1 \le \lambda_{\max}\left(n^{-1}X^\top X\right) \le C_2, \quad C_1 \le \lambda_{\max}\left(n^{-1}\Delta^\top \Delta\right) \le C_2.$$

(C3) (Condition on the non-linear functions) For $r = m + v > 1.5$, $g_0 \in \mathcal{H}_r^q$.

(C4) (Condition on the B-spline basis) The number of internal knots, k_n, satisfies $k_n \approx n^{1/(2r+1)}$. In addition, the internal knots are quasi-uniform. That is, the ratio of the maximum and minimum distance between knots is bounded above by a positive constant.

(C5) (Condition on the dropout probability) There exist $0 < \alpha_\ell$ and $\alpha_u < 1$ such that $\alpha_\ell < \eta_{ij}^* < \alpha_u$ for all $(i, j) \in \{1, \dots, n\} \times \{2, \dots, m\}$.

(C6) (Condition on the estimator of weights) Define, $\eta_j(\mathbf{T}_{ij}, \boldsymbol{\gamma}_j) = \frac{\exp\left(\mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j\right)}{1+\exp\left(\mathbf{T}_{ij}^\top \boldsymbol{\gamma}_j\right)}$. Assume conditions for asymptotic normality of $\hat{\gamma}_j$ hold for all $j \in \{2, \dots, m\}$ and $||\partial \eta_j(\mathbf{T}_{ij}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}||_2$ and $||\partial^2 \eta_j(\mathbf{T}_{ij}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top||_2$ are bounded in a neighborhood of $\boldsymbol{\gamma}_j^*$ for all $j \in \{2, \dots, m\}$ and $i \in \{1, \dots, n\}$.

Conditions similar to C1 have been used for cross-sectional linear quantile regression with complete data.[33] The most prominent deviation is the assumption of a finite fourth moment for the error terms. This is used because the responses are used as predictors when modeling dropout probability. The proof of our central limit theorem relies on these predictors having a bounded fourth moment to apply Lindeberg's conditions. Similar assumptions have been made for this reason for quantile regression models that use IPW.[23] The assumptions for the nonlinear variables stated in Condition C2 are fairly common in work that models additive nonlinear functions with B-splines because under these assumptions one can establish lower bounds for the minimum eigenvalue of the B-spline transformed covariates Gram matrix.[28,34,35] The assumption that the linear terms are bounded is fairly common in quantile regression and the assumption that the variables have mean zero can be done by centering the predictors. Condition C3 provides that only reasonably smooth functions can be estimated by the proposed method. Condition C4 assumes that k_n increases at the optimal rate for estimating $g^*(\cdot)$.[28] Conditions C5 and C6 provide that the differences between the true and estimated probabilities have a well-defined Taylor approximation.

Let $\mathbf{I}(\boldsymbol{\gamma}^*)$ be the Fisher information matrix corresponding to (8), $\mathbf{F}_i$ be an $m \times m$ diagonal matrix with jth diagonal entry of $f_{ij}(0|\mathbf{X}_{ij}, \mathbf{Z}_{ij})$, $\mathbf{k}_i = \left[\psi_\tau(\epsilon_{i1}), \psi_\tau(\epsilon_{i2})R_{i2}/\pi_{i2}^*, \dots, \psi_\tau(\epsilon_{im})R_{im}/\pi_{im}^*\right]^\top \in \mathbb{R}^m$, and

$$\begin{aligned}
\boldsymbol{\Sigma}_1 &= \mathrm{E}\left[\boldsymbol{\delta}_i^\top \mathbf{F}_i \boldsymbol{\delta}_i\right], \\
\boldsymbol{\Sigma}_2 &= \mathrm{E}\left[\boldsymbol{\delta}_i^\top \mathbf{k}_i \mathbf{k}_i^\top \boldsymbol{\delta}_i\right], \\
\boldsymbol{\Sigma}_{3jk} &= I(k \le j)\mathrm{E}\left[\frac{\left(1-\eta_{ik}^*\right)}{\pi_{ij}^*}\psi_\tau(\epsilon_{ij})\boldsymbol{\delta}_{ij}\mathbf{T}_{ik}^\top\right] \qquad \text{for } k \in \{2, \dots, m\}, \\
\boldsymbol{\Sigma}_{3j} &= (\boldsymbol{\Sigma}_{3j2}^\top, \dots, \boldsymbol{\Sigma}_{3jm}^\top)^\top.
\end{aligned}$$

The following theorem presents asymptotic properties of the estimator for $n \to \infty$ and m fixed.

Theorem 1. *Let* $\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_2 - \sum_{j=2}^{m} \boldsymbol{\Sigma}_{3j} \mathbf{I}(\boldsymbol{\gamma}^*)^{-1} \boldsymbol{\Sigma}_{3j}^{\top}$. *Under conditions (C1)-(C6),*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \to N\left(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_m \boldsymbol{\Sigma}_1^{-1}\right), \tag{11}$$

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij} \left[\hat{g}_k(\mathbf{z}_{ij}) - g^*(\mathbf{Z}_{ij})\right]^2 = O_p\left[n^{-2/(2r+1)}\right]. \tag{12}$$

Theorem 1 proves that the estimator $\hat{g}(\cdot)$ achieves the optimal rate of convergence for nonlinear estimators of additive functions.[28] In addition, the linear terms have a central limit theorem. Note if the true weights were used the asymptotic variance would be $\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1}$, thus estimating the weights reduces the variance. A phenomenon that has previously been observed, and discussed, for IPW estimators.[36]

4 | MONTE CARLO STUDIES

We perform a simulation study to compare the proposed method for estimating effects with existing ones in the literature. We consider four covariates with linear effects, $X_1, \ldots, X_4$ and two covariates with nonlinear effects, Z_1 and Z_2, that do not change across time. In the simulations we keep the values of the predictors fixed but generate the initial values from a distribution. The variable X_1 has a Uniform(0,1) distribution and X_2, X_3, X_4 follow the standard normal distribution. The nonlinear variables Z_1 and Z_2 follow a Uniform(0,1) and Uniform$(-1, 1)$ distribution, respectively. The jth response for the ith subject is determined as follows,

$$Y_{ij} = X_{1i} + X_{2i} + X_{3i} + X_{4i} + \sin(2\pi Z_{1i}) + Z_{2i}^3 + \epsilon_{ij}. \tag{13}$$

Letting $(\xi_{i1}, \ldots, \xi_{im}) \sim N_m(0, \Sigma)$ where the (i,j)th element of Σ is $0.75^{|i-j|}$, we consider three settings for the errors: (1) $\epsilon_{ij} = 2\xi_{ij}$, (2) $\epsilon_{ij} = 2X_{1ij}\xi_{ij}$, and (3) $\epsilon_{ij} = \exp(2\xi_{ij}) - \exp(2)$. Settings 2 and 3 are interesting cases for quantile regression because due to heteroscedasticity or asymmetric errors, least squares regression with an assumption of normal errors would provide an incomplete description of the conditional distribution. For instance in setting 2, $\beta_1^* = 1 + \Phi_2^{-1}(\tau)$ where Φ_σ^{-1} is the inverse CDF of the normal distribution with mean 0 and variance σ^2. While β_0^* is not part of (13) depending on τ and the distribution of ϵ_{ij} there may be a non-zero intercept. For setting (1) $\beta_0^* = \Phi_2^{-1}(\tau)$; (2) $\beta_0^* = 0$ and (3) $\beta_0^* = \Gamma_2^{-1}(\tau)$, where Γ_σ^{-1} is the inverse CDF of a log-normal distribution that has been centered to have mean zero.

The probabilities of returning at time j are generated from the following model,

$$P(R_{ij} = 1 | R_{i,j-1} = 1) = \frac{\exp\left(\kappa - Y_{i,j-1}\right)}{1 + \exp\left(\kappa - Y_{i,j-1}\right)}. \tag{14}$$

Under this model, an individual with higher responses is more likely to drop out and the probability of dropping out will tend to increase with time. In the presented simulations $\kappa = 4$. The supplemental material includes results for the heteroscedastic setting, setting (2), where $\kappa \in \{3, 5\}$. All variables available before time point j are included as linear predictors when estimating (14). The supplemental material includes results where $Y_{i,j-1}$ in (14) is replaced with $Y_{i,j-1}^2$, but a misspecified model is fit that does not include the squared term.

We consider two different methods of estimating effects: the proposed inverse probability weighting method (IPW) and the method which ignores the dropout and only uses the complete observations (Naive). The Naive method would yield consistent estimates if the data were missing completely at random. Additionally, we compare results to the setting in which all data is observed (Oracle). This hypothetical situation is unattainable in practice but is useful for comparing the IPW and Naive methods to the gold standard.

When using B-splines the number of internal knots needs to be set. Many different approaches can be used to identify the number of internal knots, such as cross-validation and AIC. In the UDS analysis internal knots from zero to nine were evaluated using BIC, following suggestions from previous work using B-splines with quantile regression.[12,18,37,38] With one exception, zero internal knots provided the smallest BIC value for all quantile models. The one exception was for the IPW approach with $\tau = 0.9$, where one internal knot provided the minimum BIC value. In the simulations and the

UDS analysis cubic B-splines with no internal knots were used for all models and variables. We decided to use the same number of internal knots, zero, for all models and test the use of this selection in the simulations. However, there are two boundary knots and $J_n = 3$ because cubic B-splines are used.

We estimate the $\tau = 0.5$ and $\tau = 0.7$ quantiles for the settings where $m = 2$ and $m = 3$ and $n \in \{100,500,1000\}$. We report results for $\tau = 0.5$ for settings 1 and 3 and report $\tau = 0.7$ for setting 2, because non-central quantiles are more interesting when the errors are heteroscedastic. The supplemental material includes results for $\tau = 0.7$ for settings 1 and 3 and $\tau = 0.5$ for setting 2. One hundred replications are used for all settings. We report the bias of the estimator for each coefficient ($\hat{\beta}_j$), the mean squared error of the estimator for the linear coefficient vector (MSE), and the mean squared error of the estimator of the nonlinear functions (gMSE). Let $\hat{\beta}_{lk}$ be the estimate of kth coefficient in the lth simulation and similarly define $\hat{g}_{lk}$. Then,

$$\text{MSE}(\hat{\beta}) : \frac{1}{100}\sum_{l=1}^{100}\sum_{k=1}^{4}(\hat{\beta}_{lk} - \beta_k^*)^2;$$

$$\text{gMSE} : \frac{1}{100}\sum_{l=1}^{100}\left\{(\hat{\beta}_{l0} - \beta_0^*)^2 + \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{2}\hat{g}_{lk}(z_{lik}) - g_k^*(z_{lik})\right]^2\right\};$$

$$\text{Bias}(\hat{\beta}_k) : \left|\frac{1}{100}\sum_{l=1}^{100}\hat{\beta}_{lk} - \beta_k^*\right|.$$

The bias and MSE results are presented in Tables 1–3. In Tables 1 and 2, the results consistently show that MSE and gMSE tend to decrease with n for the IPW method. In addition, the Naive method's bias does not decrease with n and thus as n increases the MSE of the IPW method is smaller than the Naive method. For gMSE the two methods are performing similarly. Note for smaller sample sizes the MSE of the Naive method can be smaller than that of IPW. The motivation for IPW is to provide an unbiased estimator, but the weights increase the variance of the estimator. At smaller sample sizes

TABLE 1 MSE results for $\tau = 0.5$ from setting 1, normal errors.

Method	m	n	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}_2$)	Bias($\hat{\beta}_3$)	Bias($\hat{\beta}_4$)	MSE($\hat{\beta}$)	gMSE
Oracle	2	100	0.1689	0.0141	0.0226	0.0073	0.8341	0.3213
Naive	2	100	0.1074	0.0612	0.0620	0.0344	0.8253	0.3302
IPW	2	100	0.1167	0.0081	0.0366	0.0028	0.9431	0.3698
Oracle	3	100	0.1365	0.0144	0.0419	0.0032	0.6612	0.2672
Naive	3	100	0.0514	0.0785	0.0913	0.0719	0.7140	0.2825
IPW	3	100	0.1071	0.0248	0.0481	0.0120	0.8853	0.3355
Oracle	2	500	0.0172	0.0125	0.0208	0.0038	0.1297	0.0635
Naive	2	500	0.0479	0.0530	0.0609	0.0444	0.1342	0.0617
IPW	2	500	0.0123	0.0182	0.0219	0.0037	0.1364	0.0707
Oracle	3	500	0.0206	0.0151	0.0146	0.0047	0.1152	0.0538
Naive	3	500	0.0814	0.0762	0.0810	0.0696	0.1423	0.0545
IPW	3	500	0.0608	0.0373	0.0337	0.0164	0.1428	0.0650
Oracle	2	1000	0.0148	0.0079	0.0017	0.0069	0.0666	0.0313
Naive	2	1000	0.0656	0.0495	0.0424	0.0497	0.0759	0.0329
IPW	2	1000	0.0142	0.0144	0.0086	0.0115	0.0732	0.0354
Oracle	3	1000	0.0262	0.0075	0.0004	0.0059	0.0551	0.0279
Naive	3	1000	0.0947	0.0711	0.0651	0.0722	0.0777	0.0320
IPW	3	1000	0.0353	0.0280	0.0182	0.0242	0.0681	0.0366

TABLE 2 MSE results for $\tau = 0.7$ from setting 2, heteroscedastic errors.

Method	m	n	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}_2$)	Bias($\hat{\beta}_3$)	Bias($\hat{\beta}_4$)	MSE($\hat{\beta}$)	gMSE
Oracle	2	100	0.0755	0.0096	0.0205	0.0058	0.2273	0.0797
Naive	2	100	0.1727	0.0213	0.0252	0.0059	0.2408	0.0822
IPW	2	100	0.0976	0.0141	0.0171	0.0010	0.2607	0.0934
Oracle	3	100	0.0533	0.0048	0.0222	0.0028	0.1655	0.0708
Naive	3	100	0.2019	0.0298	0.0313	0.0124	0.1983	0.0656
IPW	3	100	0.1015	0.0221	0.0275	0.0021	0.1999	0.0814
Oracle	2	500	0.0437	0.0004	0.0104	0.0014	0.0274	0.0102
Naive	2	500	0.1497	0.0073	0.0171	0.0101	0.0489	0.0102
IPW	2	500	0.0463	0.0000	0.0105	0.0015	0.0352	0.0113
Oracle	3	500	0.0325	0.0003	0.0091	0.0007	0.0223	0.0089
Naive	3	500	0.1956	0.0115	0.0192	0.0142	0.0598	0.0090
IPW	3	500	0.0428	0.0033	0.0106	0.0013	0.0335	0.0104
Oracle	2	1000	0.0575	0.0061	0.0055	0.0054	0.0148	0.0066
Naive	2	1000	0.1635	0.0119	0.0114	0.0117	0.0367	0.0065
IPW	2	1000	0.0704	0.0080	0.0068	0.0063	0.0167	0.0069
Oracle	3	1000	0.0581	0.0061	0.0051	0.0056	0.0123	0.0071
Naive	3	1000	0.2194	0.0152	0.0147	0.0144	0.0562	0.0072
IPW	3	1000	0.0842	0.0098	0.0088	0.0077	0.0186	0.0076

TABLE 3 MSE results for $\tau = 0.5$ from setting 3, asymmetric errors.

Method	m	n	Bias($\hat{\beta}_1$)	Bias($\hat{\beta}_2$)	Bias($\hat{\beta}_3$)	Bias($\hat{\beta}_4$)	MSE($\hat{\beta}$)	gMSE
Oracle	2	100	0.1300	0.0114	0.0343	0.0019	0.8363	0.4263
Naive	2	100	0.0837	0.0148	0.0253	0.0186	0.5577	0.2670
IPW	2	100	0.0720	0.0216	0.0232	0.0108	0.6104	0.3012
Oracle	3	100	0.1479	0.0108	0.0465	0.0093	0.6446	0.3377
Naive	3	100	0.0789	0.0267	0.0360	0.0283	0.3885	0.1619
IPW	3	100	0.0592	0.0288	0.0314	0.0229	0.4082	0.1947
Oracle	2	500	0.0113	0.0138	0.0202	0.0015	0.1250	0.0618
Naive	2	500	0.0110	0.0190	0.0278	0.0091	0.0919	0.0449
IPW	2	500	0.0091	0.0166	0.0213	0.0060	0.1056	0.0518
Oracle	3	500	0.0160	0.0155	0.0172	0.0039	0.1091	0.0541
Naive	3	500	0.0254	0.0197	0.0223	0.0119	0.0704	0.0336
IPW	3	500	0.0222	0.0232	0.0172	0.0038	0.0831	0.0404
Oracle	2	1000	0.0143	0.0084	0.0024	0.0056	0.0639	0.0315
Naive	2	1000	0.0263	0.0188	0.0093	0.0138	0.0480	0.0238
IPW	2	1000	0.0273	0.0155	0.0035	0.0123	0.0521	0.0276
Oracle	3	1000	0.0299	0.0089	0.0002	0.0054	0.0530	0.0274
Naive	3	1000	0.0303	0.0173	0.0143	0.0166	0.0339	0.0180
IPW	3	1000	0.0423	0.0150	0.0092	0.0126	0.0418	0.0215

the bias-variance tradeoff may not be advantageous enough to support using IPW. Thus suggesting the proposed method is better suited to larger sample sizes. The results presented in Table 3, from the asymmetric setting are not as supportive of the IPW approach because the Naive method is consistently beating IPW with respect to MSE. Thus suggesting in data with a lot of noise that adding additional variance through the estimation and application of weights may not be helpful.

In addition, to estimation accuracy, we also provided coverage results. To derive confidence intervals we use a bootstrap approach. Many different bootstrap approaches have been proposed that apply to quantile regression, but much of the research has been focused on cross-sectional data.[39-43] We re-sample the subjects with replacement, re-estimate the weights and then fit a new model, repeating what has been done for similar methods.[26,44] For the Naive and Oracle methods a similar approach is used but without weights. In the simulations, we use 100 bootstrap iterations. Then 90% confidence intervals are calculated by using the 0.05 and 0.95 quantiles of the bootstrap coefficients across the 100 iterations.

In Tables 4–6 we report coverage of the confidence intervals and their average length. Coverage is the proportion of times the true parameter was in a calculated confidence interval and thus the correct coverage would be 0.9. For the case of normally distributed or heteroscedastic errors, Tables 4 and 5 respectively, both methods tend to have coverage close to 0.9 for smaller sample sizes. However, for larger samples sizes the coverage for the Naive method can be noticeably worse than IPW due to the non-diminishing bias of the Naive estimator. This is particularly true for β_1^* in the heteroscedastic case, recall that in this setting X_1 influenced both the location and scale of the response. In that setting with $n = 1000$ the coverage of the IPW approach for $m = 2$ and $m = 3$ is 0.82 and 0.77 respectively. While the coverage for the Naive approach is 0.49 and 0.16. The coverages using IPW do not meet the benchmarks of 0.9 but the IPW coverage is much closer to 0.9 than the Naive approach. Similar, but less drastic patterns can be seen for the other coefficients. Again, Table 6 demonstrates that in this setting the IPW method does not improve much upon the Naive method. The coverages are similar, but the lengths of the IPW intervals are larger. Note, in all settings the IPW lengths are longer, but this is expected because the weights increase the variance of the estimator.

TABLE 4 Coverage (and average length) for bootstrap confidence intervals of each coefficient from simulation setting 1, normal errors with $\tau = 0.5$.

Method	m	n	Coverage (length) β_1^* C.I.	Coverage (length) β_2^* C.I.	Coverage (length) β_3^* C.I.	Coverage (length) β_4^* C.I.
Oracle	2	100	0.92 (2.67)	0.91 (0.84)	0.93 (0.86)	0.96 (0.84)
Naive	2	100	0.93 (2.72)	0.92 (0.85)	0.91 (0.86)	0.96 (0.86)
IPW	2	100	0.91 (2.82)	0.93 (0.88)	0.92 (0.88)	0.99 (0.88)
Oracle	3	100	0.91 (2.49)	0.9 (0.78)	0.91 (0.78)	0.9 (0.76)
Naive	3	100	0.91 (2.49)	0.93 (0.8)	0.92 (0.78)	0.91 (0.78)
IPW	3	100	0.9 (2.62)	0.95 (0.85)	0.92 (0.83)	0.93 (0.82)
Oracle	2	500	0.94 (1.13)	0.9 (0.34)	0.85 (0.33)	0.92 (0.33)
Naive	2	500	0.93 (1.13)	0.88 (0.34)	0.84 (0.34)	0.9 (0.33)
IPW	2	500	0.94 (1.21)	0.91 (0.36)	0.86 (0.37)	0.95 (0.36)
Oracle	3	500	0.91 (1.02)	0.95 (0.31)	0.89 (0.31)	0.95 (0.31)
Naive	3	500	0.88 (1.03)	0.82 (0.32)	0.81 (0.32)	0.82 (0.31)
IPW	3	500	0.91 (1.15)	0.89 (0.35)	0.86 (0.35)	0.9 (0.34)
Oracle	2	1000	0.94 (0.8)	0.94 (0.23)	0.92 (0.23)	0.9 (0.24)
Naive	2	1000	0.92 (0.79)	0.84 (0.23)	0.84 (0.23)	0.84 (0.24)
IPW	2	1000	0.94 (0.85)	0.94 (0.25)	0.93 (0.25)	0.88 (0.26)
Oracle	3	1000	0.92 (0.72)	0.93 (0.21)	0.92 (0.21)	0.93 (0.22)
Naive	3	1000	0.91 (0.73)	0.81 (0.22)	0.75 (0.21)	0.78 (0.23)
IPW	3	1000	0.97 (0.83)	0.91 (0.25)	0.9 (0.24)	0.91 (0.25)

TABLE 5 Coverage (and average length) for bootstrap confidence intervals of each coefficient from simulation setting 2, heteroscedastic errors with $\tau = 0.7$.

Method	m	n	Coverage (length) β_1^* C.I.	Coverage (length) β_2^* C.I.	Coverage (length) β_3^* C.I.	Coverage (length) β_4^* C.I.
Oracle	2	100	0.9 (1.43)	0.92 (0.42)	0.95 (0.41)	0.93 (0.38)
Naive	2	100	0.82 (1.42)	0.92 (0.42)	0.91 (0.41)	0.94 (0.38)
IPW	2	100	0.84 (1.46)	0.92 (0.44)	0.91 (0.43)	0.93 (0.39)
Oracle	3	100	0.9 (1.29)	0.89 (0.38)	0.96 (0.37)	0.92 (0.34)
Naive	3	100	0.8 (1.29)	0.91 (0.38)	0.92 (0.37)	0.88 (0.34)
IPW	3	100	0.89 (1.38)	0.88 (0.42)	0.92 (0.39)	0.87 (0.36)
Oracle	2	500	0.86 (0.52)	0.91 (0.11)	0.9 (0.11)	0.94 (0.11)
Naive	2	500	0.7 (0.52)	0.92 (0.11)	0.88 (0.11)	0.94 (0.12)
IPW	2	500	0.83 (0.57)	0.94 (0.12)	0.91 (0.13)	0.93 (0.12)
Oracle	3	500	0.91 (0.47)	0.94 (0.1)	0.9 (0.1)	0.94 (0.1)
Naive	3	500	0.55 (0.47)	0.88 (0.1)	0.85 (0.1)	0.88 (0.1)
IPW	3	500	0.84 (0.55)	0.92(0.11)	0.88 (0.12)	0.88 (0.12)
Oracle	2	1000	0.82 (0.34)	0.96 (0.06)	0.94 (0.06)	0.92 (0.06)
Naive	2	1000	0.49 (0.33)	0.91 (0.07)	0.86 (0.06)	0.88 (0.06)
IPW	2	1000	0.82 (0.37)	0.93 (0.07)	0.89 (0.07)	0.91 (0.07)
Oracle	3	1000	0.81 (0.31)	0.92 (0.06)	0.93 (0.05)	0.89 (0.06)
Naive	3	1000	0.16 (0.31)	0.85 (0.06)	0.74 (0.06)	0.8 (0.06)
IPW	3	1000	0.77 (0.36)	0.9 (0.07)	0.88 (0.06)	0.91 (0.07)

TABLE 6 Coverage (and average length) for bootstrap confidence intervals of each coefficient from simulation setting 3, asymmetric errors with $\tau = 0.5$.

Method	m	n	Coverage (length) β_1^* C.I.	Coverage (length) β_2^* C.I.	Coverage (length) β_3^* C.I.	Coverage (length) β_4^* C.I.
Oracle	2	100	0.91 (3.02)	0.9 (0.96)	0.95 (0.96)	0.95 (0.96)
Naive	2	100	0.92 (2.51)	0.93 (0.78)	0.92 (0.8)	0.93 (0.8)
IPW	2	100	0.92 (2.61)	0.92 (0.81)	0.92(0.83)	0.94 (0.83)
Oracle	3	100	0.9 (2.62)	0.9 (0.86)	0.94 (0.86)	0.85 (0.84)
Naive	3	100	0.96 (2.08)	0.93 (0.64)	0.93 (0.66)	0.91 (0.64)
IPW	3	100	0.96 (2.14)	0.93 (0.66)	0.94 (0.68)	0.89 (0.67)
Oracle	2	500	0.91 (1.06)	0.9 (0.31)	0.85 (0.31)	0.9 (0.3)
Naive	2	500	0.92 (0.88)	0.91 (0.26)	0.86 (0.26)	0.94 (0.26)
IPW	2	500	0.93 (0.94)	0.87 (0.28)	0.86 (0.28)	0.92 (0.28)
Oracle	3	500	0.91 (0.98)	0.94 (0.29)	0.88 (0.29)	0.94 (0.28)
Naive	3	500	0.89 (0.74)	0.9 (0.23)	0.86 (0.22)	0.95 (0.22)
IPW	3	500	0.92 (0.8)	0.93 (0.24)	0.89 (0.24)	0.96 (0.24)
Oracle	2	1000	0.92 (0.76)	0.95 (0.22)	0.92 (0.22)	0.94 (0.23)
Naive	2	1000	0.92 (0.64)	0.95 (0.18)	0.91 (0.18)	0.92 (0.19)
IPW	2	1000	0.93 (0.69)	0.93 (0.2)	0.92 (0.2)	0.91 (0.2)
Oracle	3	1000	0.89 (0.7)	0.93 (0.21)	0.9 (0.2)	0.93 (0.21)
Naive	3	1000	0.94 (0.54)	0.94 (0.16)	0.89 (0.15)	0.92 (0.16)
IPW	3	1000	0.91 (0.6)	0.93 (0.17)	0.91 (0.17)	0.93 (0.17)

The supplemental material includes results for $\tau = 0.7$ for settings 1 and 3, and for $\tau = 0.5$ for setting 2. Those results are similar to what was discussed above, in settings 1 and 2 the advantage of the IPW approach becomes clear for larger sample sizes while in setting 3 IPW does not offer much of a benefit. The supplemental material also includes results for the heteroscedastic error case but with different rates of dropout, specifically with $\kappa \in \{3, 5\}$. Recall, κ is the intercept of the dropout model and $\kappa = 4$ in the presented results. In both settings, the IPW method outperforms the Naive method with respect to MSE when the sample size is larger. When there are fewer dropouts both approaches provide reasonable coverage but with more dropouts the Naive method provides very poor coverage for larger sample sizes, while the IPW approach provides close to the expected coverage of 0.9. Finally, the supplemental material presents results for heteroscedastic errors with a misspecified model for the probability of a subject not dropping out. As mentioned earlier the model relies on a squared term of the previous response but the fitted model only includes linear variables. In that setting, the IPW approach does worse than the Naive approach both in terms of coverage and MSE. Suggesting that a reasonable model for the weights is needed, an issue we discuss more in Section 6.

5 | ANALYSIS OF THE UNIFORM DATA SET

In this section, we analyze data from the National Alzheimer's Coordinating Center's UDS. In particular, we are interested in the effects of covariates on cognitive ability. To measure cognitive ability, we create a composite cognitive (CC) score which is the sum of standardized scores from the logical memory IA and IIA tests, digit span backwards test, animals, vegetables, and Boston naming tests, digit symbol, Trail A and B tests, and the mini-mental state exam.[45-47] The scores for the Trail A and B tests are times until completion, where a shorter time is interpreted as indicative of higher cognitive ability, unlike the other tests. To handle this discrepancy, we instead record the time to complete as a percentage of the maximum score observed, 150 for Trail A and 300 for Trail B. Under this transformation the quicker the completion time the lower the score. All binary predictors were treated as linear variables. Specifically, female (sex), belonging to any group besides non-Hispanic white (race2), have hypertension (hyperten), diabetic (diabetes), have had a stroke (stroke), have been depressed in the last 2 years (depression), history of alcohol abuse (alcohol), and current smoker (smoke). In addition, years of education (education) were treated as linear due to a limited number of values. While the nonlinear predictors are the continuous variables, age, cognitive score at the initial appointment (CC1), and days from the initial visit (dfiv).

In the NACC dataset, data was collected on the first visit and patients were asked to return for nine follow-up visits spaced about 1 year apart. Though many patients followed-up for 1 or 2 years, very few patients remained for all ten visits and we restrict the analysis to the first five follow-up visits. The data does not specify whether an appointment was missed or not but includes the total number of appointments and the time between appointments. Each visit number is recorded as the total number of attended appointments the patient had up to that point, and thus the data follows a monotone missing structure. The length between visits can vary with 96% of follow-up time being between 245 and 694 days. No patients were removed for having too long, or short, a time between visits. The proposed method works only for monotone missing data patterns and thus any individuals that made appointments, but had missing values, were removed from the analysis. To study the covariates' effects on cognitive decline, we define the response variable as the difference in CC scores between the initial visit and each follow-up visit (CCdiff). We consider patients aged 65 or older at the initial visit. A patient was removed if he or she increased their CC score from the initial CC score by more than 1. Such a large increase suggests that the score is not reliable. About 7% of patients were removed because of this restriction.

The number of patients at each follow-up appointment and the means and standard deviations of the variables are provided in Table 7. Notable changes are the number of patients dropping from 3581 to 2381, the percentage with hypertension increasing from 0.58 to 0.67, and the average decrease in CC score increasing from 0.03 to 0.14.

Table 8 contains the estimates of the linear covariates in the UDS for the proposed IPW method and the Naive method at $\tau = 0.9$ and also includes results for a Naive and IPW approach for a mean model, where squared error loss is used. The 90% confidence intervals are created using the same bootstrap approach used for the simulations in Section 4. We focus on the 0.9 conditional quantile because that corresponds to larger declines in CC scores, which would be patients that are more likely to be at risk for non-normal cognitive decline. Note, the coefficients for the IPW and Naive approaches are different. For the models of $\tau = 0.9$, there are noticeable differences between the coefficient estimates of race2, stroke,

TABLE 7 Means (and standard deviations for non-binary variables) of patients at each follow-up appointment.

Follow-up appointment	1	2	3	4	5
Patients observed	3581	3424	3254	2381	1728
Sex	0.67	0.67	0.67	0.67	0.66
race2	0.21	0.21	0.21	0.21	0.2
Hyperten	0.58	0.61	0.64	0.66	0.67
Diabetes	0.12	0.13	0.13	0.13	0.14
Stroke	0.02	0.03	0.03	0.03	0.04
Depression	0.17	0.17	0.18	0.18	0.2
Alcohol	0.03	0.03	0.03	0.04	0.04
Smoke	0.03	0.02	0.03	0.02	0.02
Education	15.59 (2.94)	15.6 (2.93)	15.6 (2.93)	15.62 (2.92)	15.71 (2.87)
dfiv	435.44 (143.73)	849.24 (196.92)	1248.53 (222.95)	1641.45 (241.53)	2012.09 (239.84)
Age	76.69 (6.88)	77.74 (6.83)	78.69 (6.72)	79.64 (6.62)	80.52 (6.54)
CC1	6.46 (0.88)	6.48 (0.88)	6.49 (0.88)	6.52 (0.87)	6.58 (0.86)
CCdiff	0.03 (0.43)	0.05 (0.48)	0.07 (0.52)	0.11 (0.57)	0.14 (0.64)

Note: The means for variables from sex to smoke are the proportion of patients that belong to that group. Patients observed is the number of patients observed at a given visit.

TABLE 8 Linear coefficient estimates for quantile regression, with $\tau = 0.9$ and mean regression (with 90% bootstrapped confidence intervals).

Coefficient	QR IPW	QR Naive	LS IPW	LS Naive
(Intercept)	0.3718 (−0.0616, 0.7706)	0.2317 (0.0012, 0.4852)	−0.1934 (−0.4966, 0.0354)	−0.2349 (−0.4705, −0.0226)
Sex	−0.0257 (−0.0575, 0.0126)	−0.0256 (−0.0488, 0.0016)	−0.0403 (−0.0701, −0.0105)	−0.0433 (−0.0663, −0.0160)
race2	0.0167 (−0.0219, 0.0583)	0.0489 (0.0191, 0.0752)	0.0951 (0.0499, 0.1395)	0.1206 (0.0845, 0.1642)
Hyperten	0.0240 (0.0031, 0.0511)	0.0155 (−0.0020, 0.0380)	−0.0008 (−0.0246, 0.0269)	−0.0082 (−0.0333, 0.0170)
Diabetes	0.0638 (0.0085, 0.1141)	0.0373 (0.0049, 0.0695)	0.0674 (0.0256, 0.1126)	0.0464 (0.0104, 0.0849)
Stroke	0.2653 (0.1018, 0.3767)	0.1467 (0.0794, 0.2204)	0.0732 (0.0024, 0.1538)	0.0489 (−0.0142, 0.1276)
Depression	0.1464 (0.1003, 0.2012)	0.1078 (0.0753, 0.1360)	0.1167 (0.0771, 0.1487)	0.0852 (0.0581, 0.1116)
Alcohol	−0.1039 (−0.1679, −0.0205)	−0.0747 (−0.1308, 0.0020)	−0.1057 (−0.2173, −0.029)	−0.0906 (−0.1878, −0.0196)
Smoke	0.0504 (−0.0097, 0.1199)	0.0335 (−0.0146, 0.1086)	0.0648 (−0.0088, 0.1231)	0.0765 (0.0102, 0.1266)
Education	−0.0055 (−0.0098, 0.0017)	−0.0075 (−0.0106, −0.0027)	−0.0111 (−0.0160, −0.0045)	−0.0135 (−0.0174, −0.0069)

diabetes, and depression. For the quantile models the signs of the coefficients are the same but if we account for the confidence intervals there are differences in significance. The IPW approach identifies hypertension and alcohol as significant variables, but the Naive approach does not, while the Naive approach identifies race2 and education as significant, while the IPW does not. Which of these results is correct is difficult to say because previous studies have found all of these variables related to cognitive differences.[48-51]

A comparison of the mean and quantile models presented in Table 8 highlights that a model of the conditional quantile is providing information that would be lost by only modeling the conditional mean. The results show that the differences in the conditional 0.9 quantile between the two categories for race are smaller than the differences observed for the conditional mean. While the 0.9 quantile coefficient estimates for stroke and hypertension are noticeably larger than the mean model estimates. Figure 1 plots how the quantile regression coefficients for race2 change with τ, and provides a comparison to the mean regression estimate. Similar plots are provided for the other variables in the supplemental material,

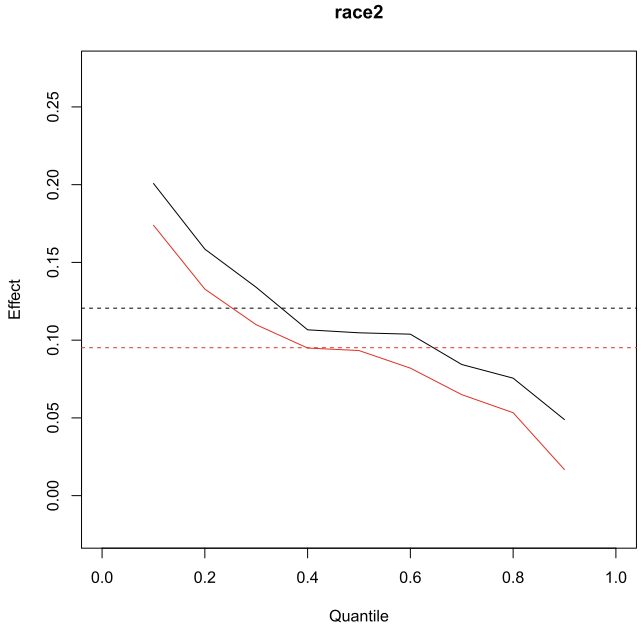

FIGURE 1 Naive (black) and IPW (red) coefficient estimates for the race2 covariate. The solid lines are estimates from quantile regression at the deciles and the dashed lines are estimates from least squares.

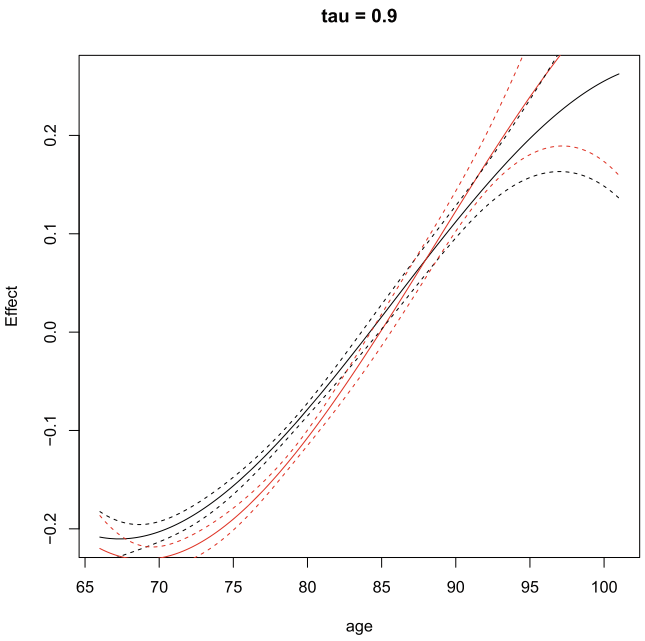

FIGURE 2 Naive (black) and IPW (red) nonlinear fits for age. Dashed lines are one bootstrap standard error above and below the 0.9 quantile estimate.

with education and stroke coefficients showing noticeable differences by the quantiles. These estimates changing by the quantile are evidence that quantile regression is providing insight about differences at the conditional quantiles that would be lost by only considering mean regression.

Figures 2–4 present the nonlinear fits for age, CC1, and dfiv. Differences between the IPW and Naive methods are most noticeable for CC1 and dfiv, seen in Figures 3 and 4, respectively. In addition, these figures present noticeably nonlinear effects. The confidence bands tend to be wider at the extremes, where there are fewer observations with those values. Similar plots for the other quantiles and mean models are available in the supplemental material.

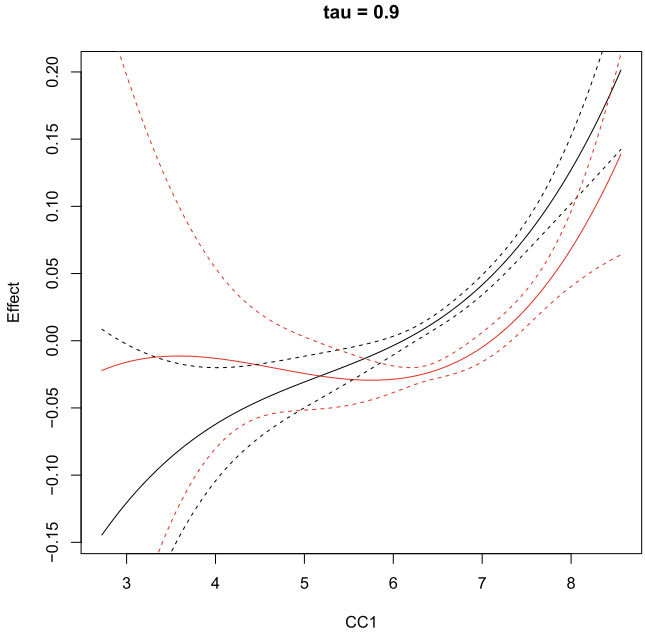

FIGURE 3 Naive (black) and IPW (red) nonlinear fits for CC1, cognitive score at the initial appointment. Dashed lines are one bootstrap standard error above and below the 0.9 quantile estimate.

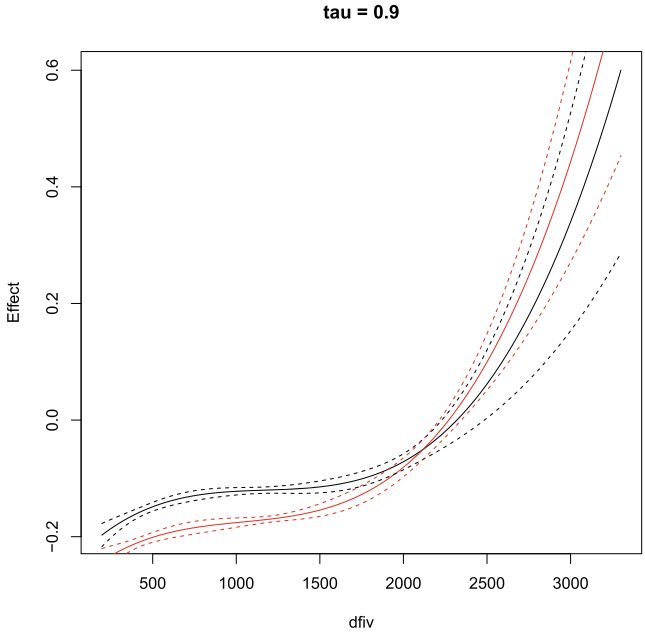

FIGURE 4 Naive (black) and IPW (red) nonlinear fits for dfiv, days from the initial visit. Dashed lines are one bootstrap standard error above and below the 0.9 quantile estimate.

6 | DISCUSSION

This article proposes a GEE approach with IPW to estimate a partial linear quantile regression model using longitudinal data with monotone missing values. There are alternative approaches to estimating a quantile regression model, as well as different types of missing data. For instance, a mixed effects approach is a popular alternative to GEE models that this article does not cover. In addition, observations can have a non-monotone missing pattern. For instance, a subject appears at time 1, is missing at time 2 but then appears again at time 3. In addition, a subject can be observed but have

missing values. The missing model of (9) could be updated to account for this, where the weights would be the inverse of the probability a subject had complete data. However, how to account for both missing values and a dropout structure using IPW is not immediately clear to us. One could consider imputing values instead of using the inverse weighting approach.

An alternative approach in monotone missing models is to use IPW where the weights are inverses of the probability a subject would drop out at their observed dropout time. Say subject i drops out at time 4 then $\hat{\pi}_i = \hat{\eta}_2\hat{\eta}_3(1-\hat{\eta}_4)$. In this framework the objective function of (9) is replaced by

$$\sum_{i=1}^{n}\hat{\pi}_i^{-1}\sum_{j=1}^{m}R_{ij}\rho_\tau\left[Y_{ij}-\left\{\mathbf{X}_{ij}^\top\boldsymbol{\beta}+\mathbf{w}(\mathbf{Z}_{ij})^\top\boldsymbol{\xi}\right\}\right]. \tag{15}$$

In a weighted model when an observation is deemed unlikely then that observation receives more weight. The intuition is that less likely observations should be weighted more heavily to account for similar observations that are likely missing. However, if the weights are very large this creates unstable estimates. A potential advantage of modeling each observation, instead of modeling the overall monotone pattern, is the weights may tend to be smaller because for any individual that dropped out of the study the overall probability of dropout will be smaller than the probability they would be observed at any specific time point. In addition, the structure of (9) can more easily accommodate different missing structures discussed earlier because the weights can vary with subject and time, whereas the weights in (15) can only vary by subject.

If the proposed model and missing data framework are relevant there remains work that can be done on this problem. One major challenge is the optimal estimation of the weights. In our simulations and applied analysis we propose using all the available variables to fit a glm model, but alternative approaches exist. For instance, one could consider fitting a more flexible, nonlinear model such as using splines to provide a more flexible estimate of the missing probabilities. This would be one potential way to guard against a misspecified model, though these approaches often require a simplifying assumption of an additive model to avoid the curse of dimensionality. If the dimension of the predictors for the missing probability model is a concern then one could consider model selection. Model selection for glm models is well studied and all the available tools such as information criteria or penalized estimators would apply to this setting. What is not as clear, is how the results or conditions needed for Theorem 1 may change with how the weights were estimated. An in-depth exploration of the different approaches for estimating the weights would be an interesting extension of this work.

DATA AVAILABILITY STATEMENT

The data analyzed in this article came from the NACC Uniform Data Set, https://naccdata.org/data-collection/forms-documentation/uds-3. The code used in the simulations is available at https://github.com/bssherwood/quantileDropout.

ORCID

Xiao-Hua Zhou https://orcid.org/0000-0001-7935-1222
Ben Sherwood https://orcid.org/0000-0002-4419-8288

REFERENCES

1. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' uniform data set (UDS): The neuropsychologic test battery. *Alzheimer Dis Assoc Disord*. 2009;23:91-101.
2. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzhemier Res Ther*. 2011;3:32.
3. Sherwood B, Zhou AXH, Weintraub S, Wang L. Using quantile regression to create baseline norms for neuropsychological tests. *Alzheimers Dement*. 2016;2:12-18.
4. Sperling R, Aisen P, Beckett L, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-292.
5. Hedden T, Gabrieli JD. Insights into the ageing mind: A view from cognitive neuroscience. *Nat Rev Neurosci*. 2004;5:87-96.
6. Zhou XH, Stroupe KT, Tierney WM. Regression analysis of health care charges with heteroscedasticity. *J R Stat Soc Ser C Appl Stat*. 2001;50(3):303-312.
7. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. *Stat Med*. 2004;23(9):1455-1497.

8. He X, Shi P. Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J Nonparametr Stat*. 1994;3:299-308.
9. Stone CJ. Optimal global rates of convergence for nonparametric regression. *Ann Stat*. 1982;10(4):1040-1053.
10. Kim MO. Quantile regression with varying coefficients. *Ann Stat*. 2007;35(1):92-108.
11. Wu TZ, Yu K, Yu Y. Single-index quantile regression. *J Multivar Anal*. 2010;101(7):1607-1621.
12. He X, Shi P. Bivariate tensor-product B-splines in a partly linear model. *J Multivar Anal*. 1996;58(2):162-181.
13. Maidman A, Wang L. New semiparametric method for predicting high-cost patients. *Biometrics*. 2017;74:1104-1111.
14. Zhang Y, Lian H, Yu Y. Ultra-high dimensional single-index quantile regression. *J Mach Learn Res*. 2020;21(224):1-25.
15. Sherwood B, Maidman A. Additive nonlinear quantile regression in ultra-high dimension. *J Mach Learn Res*. 2022;23(63):1-47.
16. Koenker R. Quantile regression for longitudinal data. *J Multivar Anal*. 2004;91(1):74-89.
17. Geraci M, Bottai M. Linear quantile mixed model. *Stat Comput*. 2014;24:461-479.
18. He X, Zhu ZY, Fung WK. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*. 2002;89(3):579-590.
19. Zhao W, Lian H, Liang H. GEE analysis for longitudinal single-index quantile regression. *J Stat Plan Inference*. 2017;187:78-102.
20. Zu T, Lian H, Green B, Yu Y. Ultra-high dimensional quantile regression for longitudinal data: An application to blood pressure analysis. *J Am Stat Assoc*. 2022;118(541):97-108.
21. Wei Y, Ma Y, Carroll RJ. Multiple imputation in quantile regression. *Biometrika*. 2012;99(2):423.
22. Wei Y, Yang Y. Quantile regression with covariates missing at random. *Stat Sin*. 2014;24:1277-1299.
23. Sherwood B, Wang L, Zhou XH. Weighted quantile regression for analyzing health care cost data with missing covariates. *Stat Med*. 2013;32(28):4967-4979.
24. Liu T, Yuan X. Weighted quantile regression with missing covariates using empirical likelihood. *Statistics*. 2016;50(1):89-113.
25. Chen X, Wan AT, Zhou Y. Efficient quantile regression analysis with missing observations. *J Am Stat Assoc*. 2015;110(510):723-741.
26. Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *J R Stat Soc Ser C Appl Stat*. 1997;46(4):463-476.
27. Yi GY, He W. Median regression models for longitudinal data with dropouts. *Biometrics*. 2009;65(2):618-625.
28. Stone CJ. Additive regression and other nonparametric models. *Ann Stat*. 1985;13(2):689-705.
29. Koenker R, Bassett G. Regression quantiles. *Econometrica*. 1978;46(1):33-50.
30. Sherwood B. Variable selection for additive partial linear quantile regression with missing covariates. *J Multivar Anal*. 2016;152:206-223.
31. Wang HJ, Zhu Z, Zhou J. Quantile regression in partially linear varying coefficient models. *Ann Stat*. 2009;37:3841-3866.
32. Sherwood B, Wang L. Partially linear additive quantile regression in ultra-high dimension. *Ann Stat*. 2016;44(1):288-317.
33. Koenker R. *Quantile Regression*. Cambridge: Cambridge University Press; 2005.
34. Zhou S, Shen X, Wolfe D. Local asymptotics for regression splines and confidence regions. *Ann Stat*. 1998;26(5):1760-1782.
35. Chen Z, Fan J, Li R. Supplemental material of error variance estimation in ultrahigh-dimensional additive models. *J Am Stat Assoc*. 2018;113(521):315-327.
36. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc*. 1994;89(427):846-866.
37. Doksum K, Koo JY. On spline estimators and prediction intervals in nonparametric regression. *Comput Stat Data Anal*. 2000;35(1):67-82.
38. Horowitz JL, Lee S. Nonparametric estimation of an additive quantile regression model. *J Am Stat Assoc*. 2005;100(472):1238-1249.
39. Efron B. Bootstrap methods: Another look at the jackknife. *Ann Stat*. 1979;7(1):1-26.
40. Feng X, He X, Hu J. Wild bootstrap for quantile regression. *Biometrika*. 2011;98(4):995.
41. Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994;81(2):341-350.
42. Bose A, Chatterjee S. Generalized bootstrap for estimators of minimizers of convex functions. *J Stat Plan Inference*. 2003;117(2):225-239. doi:10.1016/S0378-3758(02)00386-5
43. He X, Hu F. Markov chain marginal bootstrap. *J Am Stat Assoc*. 2002;97(459):783-795.
44. Jacqmin-Gadda H, Rouanet A, Mba RD, Philipps V, Dartigues JF. Quantile regression for incomplete longitudinal data with selection by death. *Stat Methods Med Res*. 2020;29(9):2697-2716.
45. Sano M, Zhu CW, Grossman H, Schimming C. Longitudinal cognitive profiles in diabetes: results from the National Alzheimer's Coordinating Center's uniform data. *J Am Geriatr Soc*. 2017;65(10):2198-2204.
46. Nandipati S, Luo X, Schimming C, Grossman HT, Sano M. Cognition in non-demented diabetic older adults. *Curr Aging Sci*. 2012;5(2):131-135.
47. Cosentino SA, Stern Y, Sokolov E, et al. Plasma β-amyloid and cognitive decline. *Arch Neurol*. 2010;67(12):1485-1490.
48. Zsembik BA, Peek MK. Race differences in cognitive functioning among older adults. *J Gerontol B Psychol Sci Soc Sci*. 2001;56(5):S266-S274. doi:10.1093/geronb/56.5.S266
49. Topiwala A, Wang C, Ebmeier KP, et al. Associations between moderate alcohol consumption, brain iron, and cognition in UK Biobank participants: Observational and mendelian randomization analyses. *PLoS Med*. 2022;19(7):1-26. doi:10.1371/journal.pmed.1004039

50. Clouston SA, Smith DM, Mukherjee S, et al. Education and cognitive decline: An integrative analysis of global longitudinal studies of cognitive aging. *J Gerontol B Psychol Sci Soc Sci*. 2020;75(7):151-160.
51. Iadecola C, Yaffe K, Biller J, et al. Impact of hypertension on cognitive function: A scientific statement from the American Heart Association. *Hypertension*. 2016;68(6):e67-e94.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Maidman A, Wang L, Zhou X-H, Sherwood B. Quantile partially linear additive model for data with dropouts and an application to modeling cognitive decline. *Statistics in Medicine*. 2023;42(16):2729-2745. doi: 10.1002/sim.9745